Géométrie et concentration de la mesure
Présentation du groupe de travail

Yann Ollivier

2 mars 2004

Introduction

La concentration de la mesure est un phénomène qui intervient lorsqu'un espace inconnu est observé au moyen de fonctions définies sur cet espace : on parle de concentration lorsque vu au travers de ces fonctions, l'espace a apparemment une taille beaucoup plus faible que ce qu'il n'est réellement. Cette notion permet en particulier de donner un sens à l'affirmation intuitive « si une fonction dépend de N variables indépendantes, chacune intervenant pour au plus $1/N$ dans le résultat, alors la fonction est constante à $1/\sqrt{N}$ près ».

L'exemple le plus simple est la loi des grands nombres : si l'on observe la fonction « proportion de pile » sur l'espace $\{\mathrm{pile}, \mathrm{face}\}^N$, cette fonction est concentrée (à environ $1/\sqrt{N}$ près) autour de $1/2$, alors qu'en principe elle pourrait varier entre 0 et 1. En fait cette fonction n'est pas du tout la seule à manifester ce comportement, c'est pourquoi on peut parler de concentration de l'espace $\{\mathrm{pile}, \mathrm{face}\}^N$.

Un exemple apparemment bien différent est la sphère S^N de grande dimension (de rayon 1 dans $\mathbb{R}^{N+1}$) : bien que cette sphère soit bien sûr de diamètre π pour sa distance intrinsèque, toute fonction 1-lipschitzienne sur cette sphère varie d'environ $1/\sqrt{N}$ seulement autour de sa valeur moyenne. Appliqué par exemple à la fonction « projection sur un axe vertical », ceci montre que presque toute la masse de la sphère se concentre autour de l'équateur. En fait, en grande dimension, toute variété à courbure (de Ricci) positive présente ce phénomène.

Comme l'examen de la table ci-dessous le suggère, les moyens d'approche de la concentration de la mesure sont très divers, allant bien sûr des probabilités (généralisations de la loi des grands nombres, méthodes de martingales...) à la géométrie riemannienne (courbure de Ricci, isopérimétrie...) voire algébrique (concentration des variétés algébriques complexes, cf [Gro99], section $3\frac{1}{2}.29$), en passant par l'analyse (inégalités de Sobolev logarithmiques, noyaux de la chaleur généralisés).

Table

1 Concentration de la mesure : quelques définitions **3**

2 Concentration et isopérimétrie **4**
2.1 Le cas de la sphère 4
2.2 Plus formellement 6
2.3 Variantes isopérimétriques 6

3 Méthodes de martingales **7**
3.1 Inégalités de Laplace 7
3.2 La loi des grands nombres généralisée — concentration sur le cube 9
3.3 Généralisations 11
3.4 Produits et concentration 12

4 Inégalités fonctionnelles **14**
4.1 Spectre du laplacien, inégalité de Poincaré, variétés produits . . . 14
4.2 Inégalité de Sobolev logarithmique 16

5 Courbure et concentration **18**
5.1 Inégalités courbure-dimension 18
5.2 Courbure et transport de mesure 20

6 Le cadre métrique mesuré **21**

Le premier objectif de ce groupe de travail est donc d'exposer des méthodes de base pour la concentration de la mesure, dans des cadres aussi variés que possible (géométrie, analyse, probabilités, statistiques), avec des exemples d'applications.

Un second objectif est de développer un cadre géométrique général dans lequel ces méthodes prendraient place. En effet, la concentration de la mesure s'exprime uniquement en termes d'une mesure et d'une métrique. Or les approches habituelles de la concentration sont beaucoup plus gourmandes en hypothèses (utilisation du tenseur de courbure d'une variété, ou bien d'un laplacien continu ou discret), et même, dans un certain nombre de cas, on a des démonstrations conceptuellement identiques appartenant à des domaines différents, qui ne sont à l'heure actuelle pas techniquement unifiées.

Par exemple, on sait que la courbure positive implique la concentration, pour des variétés riemanniennes. On aimerait un résultat plus robuste dans le cadre « métrique mesuré » ; il faudrait pour cela généraliser la notion géométrique de courbure positive. Une généralisation possible utiliserait les inégalités de courbure-dimension (Bakry-Émery, voir par exemple l'introduction [Led00]) pour les noyaux d'opérateurs de Markov. Une autre piste ([SvR]) consiste à trouver des conditions équivalentes, sur une variété riemannienne, à la courbure de Ricci positive : de nature analytique (convexité de la fonction entropie sur les mesures avec la distance de transport L^2) ou probabiliste (comportement de la marche aléatoire sur les variétés).

Par ailleurs, isopérimétrie et concentration sont des notions équivalentes, et des progrès récents en isopérimétrie ([Gro03] qui travaille sur la sphère), utilisant de la topologie algébrique, pourraient avoir des conséquences intéressantes. Ce texte de Gromov, qui étudie de la concentration à plusieurs dimensions, est encore très mal compris, et mériterait d'être étudié.

À long terme, l'objectif principal est d'arriver à exprimer des théorèmes de concentration généralisant les critères connus dans chaque domaine (courbure, laplacien, entropies...), avec des hypothèse se fondant uniquement sur ce qui est nécessaire à la formulation de la concentration : une métrique et une mesure. Le cadre correct a été fixé par Gromov dans [Gro99] (chapitre $3\frac{1}{2}$), c'est celui de tous les espaces métriques mesurés muni de sa topologie "naturelle" (qui ne l'est pas tant que cela et que nous rappelons donc plus bas). Dans cette topologie, un espace proche d'un espace concentré est encore concentré. Il paraît donc raisonnable de demander qu'un théorème de concentration ait des hypothèses telles qu'un espace proche (dans cette topologie) d'un espace les vérifiant les vérifie encore ; pour les critères géométriques (courbure) ou analytique (laplacien, entropie) c'est actuellement très loin d'être le cas.

1 Concentration de la mesure : quelques définitions

La concentration de la mesure précise ces notions. Soit donc (X, d, μ) un espace métrique mesuré de masse totale 1, que l'on supposera toujours légèrement régulier[1].

L'idée est que l'on ne connaît pas X mais qu'on l'observe au travers d'une fonction numérique f (mettons 1-lipschitzienne pour la normalisation), c'est-à-dire que l'on tire un point de X au hasard suivant la mesure de probabilité μ, et qu'on observe $f(x)$. Si l'on répète cette observation un grand nombre de fois, on parle de concentration lorsque les variations typiques de f sont très petites devant le diamètre de X.

On s'intéresse plus particulièrement au cas gaussien et au cas exponentiel ; cela évite de donner des définitions trop générales.

Définition 1 (Concentration gaussienne, exponentielle) –

On dit que l'espace métrique mesuré X est de diamètre observable gaussien $D > 0$ si, pour toute fonction 1-lipschitzienne $f : X \to \mathbb{R}$, il existe une valeur $m_f \in \mathbb{R}$ telle que pour tout $t > 0$ on ait

$$\mu\left(\{x \in X, |f(x) - m_f| > t\}\right) \leqslant 2e^{-t^2/2D^2}$$

[1] Par cela on entend par exemple que X est polonais et que la mesure μ est définie pour la tribu borélienne provenant de la topologie subordonnée à la distance d.

On dit que l'espace métrique mesuré X est de diamètre observable exponentiel $D > 0$ si, pour toute fonction 1-lipschitzienne $f : X \to \mathbb{R}$, il existe une valeur $m_f \in \mathbb{R}$ telle que pour tout $t > 0$ on ait

$$\mu\left(\{x \in X, |f(x) - m_f| > t\}\right) \leqslant 2e^{-t/D}$$

L'idée fondamentale, généralisant la loi des grands nombres, est que, très souvent, si l'on a un phénomène dépendant d'un grand nombre N de paramètres, alors ce phénomène présentera de la concentration avec un diamètre observable d'environ $1/\sqrt{N}$. Ceci même si les paramètres ne sont pas très indépendants, comme le montre l'exemple des variétés de dimension N à courbure positive.

Exemple trivial : $\mathbb{R}$ muni de sa mesure gaussienne normalisée est de diamètre observable gaussien 1. (Les constantes numériques de la définition sont bien sûr modelées sur ce cas ; elles importent peu en général.)

Beaucoup moins évident est le fait que, sur $\mathbb{R}^N$ euclidien, la mesure gaussienne canonique de densité $\frac{1}{(2\pi)^{N/2}} e^{-\|x\|^2/2}$ est aussi de diamètre observable gaussien 1, bien qu'un élément tiré selon cette mesure soit de norme environ $\sqrt{N}$. On retrouve le rapport $1/\sqrt{N}$ entre le diamètre observable et le « diamètre » réel.

Bien sûr, l'inégalité de concentration ne dit rien sur la valeur de m_f, qui peut être (par translation) arbitrairement grande. En général, on prouve la concentration en prenant pour m_f la moyenne de f, ou sa médiane, ou des constructions approchantes.

2 Concentration et isopérimétrie

2.1 Le cas de la sphère

À titre d'archétype on donne la démonstration isopérimétrique de la concentration de la mesure sur la sphère. On peut faire remonter ce théorème à Lévy ([Lév51]).

Théorème 2 – *Soit S^N la sphère de rayon 1 dans $\mathbb{R}^{N+1}$, munie de sa mesure canonique normalisée. Alors S^N est de diamètre observable gaussien $1/\sqrt{N-1}$.*

Démonstration – La démonstration utilise comme élément crucial le résultat naturel suivant. On note

$$V_\varepsilon(A) = \{x \in S^N, d(x, A) \leqslant \varepsilon\}$$

le ε-voisinage d'une partie $A \subset S^N$.

Proposition bien connue – *Les parties $A \subset S^N$ minimisant le volume de $V_\varepsilon(A)$ à volume de A fixé sont les calottes.*

Cette proposition est bien sûr analogue au fait que, dans le plan, la figure minimisant le périmètre à aire fixée est le disque.

En particulier, si A est de mesure $1/2$, alors $\mu(V_\varepsilon(A))$ est au moins égal à la mesure du ε-voisinage d'une demi-sphère. Or on peut facilement estimer cette mesure :

Proposition 3 – *Soit S_+ une demi-sphère de S^N. Alors pour tout $\varepsilon > 0$, on a*

$$\mu(V_\varepsilon(S_+)) \geqslant 1 - e^{-(N-1)\varepsilon^2/2}$$

Cela se montre par un calcul explicite pas si compliqué. Une manière d'interpréter ce résultat est de dire que presque toute la masse de la sphère est concentrée dans un voisinage de taille environ $1/\sqrt{N-1}$ autour d'un équateur. (Bien sûr, tous les équateurs sont équivalents ! le terme « concentration de la mesure » ne s'applique pas à l'espace de départ, mais à l'espace d'arrivée $\mathbb{R}$ des fonctions 1-lipschitziennes par lesquelles on fait les observations).

Maintenant apparaît le lien entre isopérimétrie et concentration. Soit donc $f : S^N \to \mathbb{R}$ une fonction 1-lipschitzienne. On doit montrer qu'il existe un nombre m_f tel que, pour tout ε, la mesure des points où $|f(x) - m_f| \geqslant \varepsilon$ est inférieure à $2e^{-(N-1)\varepsilon^2/2}$.

Soit m_f une médiane de f, c'est-à-dire un nombre tel que

$$\mu\left(\{x \in S^N, f(x) \geqslant m_f\}\right) \geqslant 1/2 \text{ et } \mu\left(\{x \in S^N, f(x) \leqslant m_f\}\right) \geqslant 1/2$$

Une telle quantité existe toujours. On peut avoir deux inégalités strictes lorsque f vaut exactement m_f sur une partie de mesure non nulle. Ici on négligera ce dernier problème : on peut supposer (par un découpage adéquat) que les deux mesures en question sont exactement $1/2$.

Soient donc $S_{f+} = \{x \in S^N, f(x) \geqslant m_f\}$ et $S_{f-} = \{x \in S^N, f(x) \leqslant m_f\}$.

Par définition, S_{f-} est de mesure $1/2$. Donc, par isopérimétrie, la mesure d'un ε-voisinage de S_{f-} est supérieure à celle du ε-voisinage d'une calotte de mesure $1/2$, c'est-à-dire d'une demi-sphère. Or on vient de voir que cette dernière mesure est supérieure à $1 - e^{-(N-1)\varepsilon^2/2}$. Donc

$$\mu(V_\varepsilon(S_{f-})) \geqslant 1 - e^{-(N-1)\varepsilon^2/2}$$

Mais par définition, sur S_{f-}, la fonction f vaut au plus m_f. Comme f est 1-lipschitzienne, si un point x est à distance au plus ε de S_{f-}, alors $f(x)$ vaut au plus $m_f + \varepsilon$. Donc sur $V_\varepsilon(S_{f-})$, la fonction f est inférieure à $m_f + \varepsilon$ et en particulier

$$\mu(\{x \in S^N, f(x) > m_f + \varepsilon\}) \leqslant e^{-(N-1)\varepsilon^2/2}$$

Bien sûr, en répétant l'argument avec S_{f+} on obtient le résultat. $\square$

2.2 Plus formellement

Ce que l'on vient de démontrer à coups de médianes est en fait la proposition suivante :

Proposition 4 – *Soit (X, d, μ) un espace métrique mesuré de masse 1. Supposons que pour toute partie mesurable $A \subset X$ de mesure $1/2$, pour tout $\varepsilon > 0$, la mesure du ε-voisinage de A vérifie*

$$\mu(V_\varepsilon(A)) \geqslant 1 - e^{-\varepsilon^2/2D^2}$$

pour un certain nombre D indépendant de A.

Alors X est de diamètre observable gaussien D.

Inversement, si l'on sait qu'un espace présente de la concentration gaussienne, en appliquant cette concentration à la fonction « distance à une partie de mesure $1/2$ » on obtient la réciproque, avec une perte dans les constantes :

Proposition 5 – *Soit (X, d, μ) un espace de diamètre observable gaussien D. Alors, pour toute partie mesurable $A \subset X$ de mesure $1/2$, pour tout $\varepsilon > 0$, la mesure du ε-voisinage de A vérifie*

$$\mu(V_\varepsilon(A)) \geqslant 1 - 8e^{-\varepsilon^2/4D^2}$$

Ceci montre que profil isopérimétrique et concentration de la mesure sont essentiellement le même phénomène.

2.3 Variantes isopérimétriques

La démonstration de la concentration de la sphère par la méthode isopérimétrique se généralise bien. En particulier, toute variété de dimension N à courbure (de Ricci) plus grande que celle de la sphère S^N sera aussi concentrée que cette sphère (théorème de Lévy-Gromov, voir l'appendice par Gromov dans [MS86]). *Ce très joli théorème devrait faire l'objet d'un exposé ultérieur.*

En partant de l'inégalité isopérimétrique sur la sphère, Gromov a réussi à démontrer la concentration pour les variétés algébriques complexes ([Gro99], section $3\frac{1}{2}.29$, voir aussi [Oll99]) : si $X \subset \mathbb{C}\mathrm{P}^N$ est une variété algébrique complexe de degré d et codimension k, plongée dans $\mathbb{C}\mathrm{P}^N$, munie de sa métrique riemannienne comme sous-variété de $\mathbb{C}\mathrm{P}^N$ et de la mesure riemannienne normalisée associée, il existe une constante $C_{k,d}$ telle que X est de diamètre observable gaussien $C_{k,d}\,(\log N/N)^{1/2d}$. La démonstration mêle allègrement géométries riemannienne et algébrique. On ne sait pas du tout dans quelle mesure ce résultat est précis : le $\log N$ est probablement de trop, mais on ne sait pas si la puissance $1/d$ par rapport au cas standard $1/\sqrt{N}$ est nécessaire ou non.

Enfin, Gromov a récemment amélioré (sans doute définitivement) les résultats d'isopérimétrie sur la sphère en les étendant à plusieurs dimensions. L'une des étapes de la démonstration ci-dessus du théorème 2 peut être énoncée de la manière suivante. Soit $f : S^N \to \mathbb{R}$ une fonction 1-lipschitzienne. Il existe une valeur $m_f \in \mathbb{R}$ (sa médiane) telle que pour tout $\varepsilon > 0$ on a

$$\mu(V_\varepsilon(f^{-1}(\{m_f\}))) \geqslant \mu(V_\varepsilon(S^{N-1}))$$

où $S^{N-1} \subset S^N$ est un équateur. De là on concluait en observant que les mesures des voisinages des équateurs étaient grandes, et que donc f était très proche de m_f. Gromov généralise cet énoncé à plusieurs dimensions de la manière suivante.

Pour $k \leqslant n$, soit $f : S^N \to \mathbb{R}^k$ une fonction *continue* (dans l'argument ci-dessus, la lipschitzianité n'intervient qu'au moment de la conclusion, pas pour l'existence de la médiane). On s'attend à ce que les fibres $f^{-1}(\{z\})$, pour $z \in \mathbb{R}^k$, soient de dimension $N - k$. Le résultat affirme qu'il existe une de ces fibres qui est aussi « grosse » que S^{N-k}, au sens suivant : il existe un point $z \in \mathbb{R}^k$ tel que, pour tout $\varepsilon > 0$, on a

$$\mu(V_\varepsilon(f^{-1}(\{z\}))) \geqslant \mu(V_\varepsilon(S^{N-k}))$$

où $S^{N-k} \subset S^N$ est une $N - k$-sphère équatoriale (i.e. de rayon 1).

Ce résultat est bien sûr optimal (prendre pour f une projection). Il permet de faire de la concentration de la mesure avec un espace d'arrivée de dimension grande sans perdre sur le diamètre observable. La démonstration est longue et utilise apparemment de la topologie algébrique et de la concavité de mesures. *Ce théorème devrait faire l'objet d'un exposé ultérieur.*

3 Méthodes de martingales

Revenons à des préoccupations plus proches des probabilités. La concentration de la mesure généralise des inégalités classiques comme l'inégalité de Hoeffding ou (asymptotiquement) le théorème central limite, qui affirment que les variations typiques de combinaisons linéaires de N variables indépendantes sont de l'ordre de $\sqrt{N}$. La concentration permet d'affaiblir considérablement l'hypothèse de linéarité, pour ne garder que le fait qu'individuellement, chaque variable influe peu.

Cette section est fortement inspirée de [Led01], chapitre 1.

3.1 Inégalités de Laplace

Un des outils les plus simples, très classique, est la fonction de partition (ou transformée de Laplace). Si f est une variable aléatoire réelle, on note pour $\lambda \in \mathbb{R}$

$$Z_f(\lambda) = \mathbb{E}e^{\lambda f} = \int e^{\lambda f(x)} d\mu(x)$$

Ici c'est la moyenne de f plutôt que sa médiane qui est utilisée. La remarque de base est la suivante :

Proposition 6 – *Supposons que pour un certain $D \geqslant 0$, pour tout $\lambda \geqslant 0$ on a*

$$\mathbb{E}e^{\lambda(f-\mathbb{E}f)} \leqslant e^{D^2\lambda^2/2}$$

Alors pour tout $\varepsilon > 0$, on a

$$\Pr(f \geqslant \mathbb{E}f + \varepsilon) \leqslant e^{-\varepsilon^2/2D^2}$$

Ceci s'obtient simplement en observant que $\Pr(f \geqslant \mathbb{E}f + \varepsilon) = \Pr(e^{\lambda f - \mathbb{E}f} \geqslant e^{\lambda t})$, en prenant $\lambda = t/D^2$ et en appliquant l'inégalité de Markov.

On a bien sûr par translation $Z_f(\lambda) = e^{\lambda \mathbb{E}f} Z_{f-\mathbb{E}f}(\lambda)$. Posons donc

$$Z_X(\lambda) = \sup Z_f(\lambda)$$

le sup portant sur toutes les fonctions 1-lipschitziennes de moyenne nulle sur notre espace X.

Corollaire 7 – *Soit (X, d, μ) un espace métrique probabilisé et supposons qu'il existe $D \geqslant 0$ tel que*

$$Z_X(\lambda) \leqslant e^{D^2\lambda^2/2}$$

pour tout $\lambda \geqslant 0$.

Alors X est de diamètre observable gaussien D. De plus, la concentration d'une fonction a lieu autour de son espérance.

Ce critère permet de démontrer un résultat d'apparence anodine :

Corollaire 8 – *Un espace métrique probabilisé de diamètre D est de diamètre observable gaussien D.*

Démonstration – Appliquons le critère. Soit f une fonction 1-lipschitzienne de moyenne nulle. On a par convexité

$$\int e^{\lambda f} = \int_x e^{\lambda\left(f(x) - \int_y f(y)\right)} \leqslant \int_x \int_y e^{\lambda(f(x)-f(y))}$$

Développons l'exponentielle en série, les termes d'ordre impair disparaissent par symétrie entre x et y. En utilisant que $|f(x) - f(y)| \leqslant D$ on obtient que $\int e^{\lambda f} \leqslant \sum (D\lambda)^{2i}/(2i)! \leqslant e^{D^2\lambda^2/2}$, d'où la conclusion. $\square$

Avant de donner des applications plus importantes de ce critère, mentionnons une réciproque qui prouve que c'est seulement une manière de réécrire la concentration (on perd un peu sur les constantes).

Proposition 9 – *Soit (X, d, μ) un espace métrique probabilisé, de diamètre observable gaussien D. Alors $Z_X(\lambda) \leqslant 7e^{D^2\lambda^2}$.*

Démonstration – Soit f une fonction 1-lipschitzienne de moyenne nulle. Alors

$$\begin{aligned}
\mathbb{E}e^{\lambda f} &= \int_{t\geqslant 0} \Pr\left(e^{\lambda f} \geqslant t\right) dt = \int_{t\geqslant 0} \Pr(f \geqslant \ln t\,/\lambda)\, dt \\
&\leqslant 1 + \int_{t\geqslant 1} \Pr(f \geqslant \ln t\,/\lambda)\, dt \leqslant 1 + \int_{t\geqslant 1} 2e^{-(\ln t)^2/2D^2\lambda^2}\, dt \\
&= 1 + \int_{u\geqslant 0} 2e^{-u^2/2D^2\lambda^2} e^u du = 1 + \int_{u\geqslant 0} 2e^{-\frac{1}{2}\left(\frac{u}{D\lambda}-D\lambda\right)^2+\frac{D^2\lambda^2}{2}}\, du \\
&\leqslant 1 + 2e^{D^2\lambda^2/2}\sqrt{2\pi}\, D\lambda \leqslant 7e^{D^2\lambda^2}
\end{aligned}$$

□

En augmentant la valeur de 7, on peut même ramener un coefficient arbitrairement proche de 1/2 dans l'exposant.

3.2 La loi des grands nombres généralisée — concentration sur le cube

Un énorme avantage du critère de la fonction de partition est son excellent comportement vis-a-vis des produits ℓ^1 :

Proposition 10 – *Soient (X_1, d_1, μ_1) et (X_2, d_2, μ_2) deux espaces métriques probabilisés tels que $Z_{X_1}(\lambda) \leqslant e^{D_1^2\lambda^2/2}$ et $Z_{X_2}(\lambda) \leqslant e^{D_2^2\lambda^2/2}$.*

Alors l'espace métrique probabilisé $(X, d, \mu) = (X_1 \times X_2, d_1 + d_2, \mu_1 \otimes \mu_2)$ vérifie

$$Z_X(\lambda) \leqslant e^{\left(D_1^2+D_2^2\right)\lambda}$$

et est donc de diamètre observable gaussien $\sqrt{D_1^2 + D_2^2}$.

Démonstration – Si f est une fonction 1-lipschitzienne de moyenne nulle sur $X_1 \times X_2$ on a

$$\int_{x_1}\int_{x_2} e^{\lambda f(x_1,x_2)} = \int_{x_1} e^{\lambda \mathbb{E} f(x_1,\cdot)} \int_{x_2} e^{\lambda(f(x_1,x_2)-\mathbb{E}f(x_1,\cdot))}$$

où $\mathbb{E}f(x_1,\cdot)$ est la moyenne de f en x_2, pour x_1 fixé. Mais à x_1 fixé, $f(x_1,x_2) - \mathbb{E}f(x_1,\cdot)$ est une fonction 1-lipschitzienne de x_2, de moyenne nulle par construction, et donc

$$\int_{x_2} e^{\lambda(f(x_1,x_2)-\mathbb{E}f(x_1,\cdot))} \leqslant e^{D_2^2\lambda^2/2}$$

par hypothèse. Donc

$$\int_{x_1}\int_{x_2} e^{\lambda f(x_1,x_2)} \leqslant e^{D_2^2\lambda^2/2} \int_{x_1} e^{\lambda \mathbb{E} f(x_1,\cdot)}$$

Mais $\mathbb{E}f(x_1,\cdot)$ est une fonction 1-lipschitzienne de x_1, en tant que moyenne de fonctions 1-lipschitziennes. En outre, elle est de moyenne nulle car f l'est. Donc le dernier terme est inférieur à $e^{D_1^2\lambda^2/2}$ par hypothèse, d'où le résultat. □

Remarque 11 – Dans ce qui précède, on n'a utilisé de la distance sur $X_1 \times X_2$ que la propriété

$$d((x_1,x_2),(x_1',x_2)) \leqslant d_1(x_1,x_1') \quad \text{et} \quad d((x_1,x_2),(x_1,x_2')) \leqslant d_2(x_2,x_2')$$

En particulier, la conclusion est vraie pour toutes les distances $d_{\ell^p} = (d_1^p + d_2^p)^{1/p}$ sur $X_1 \times X_2$, pour $p \geqslant 1$. Mais c'est bien pour la distance ℓ^1 que l'affirmation est la plus intéressante, puisque c'est pour celle-là que le diamètre observable augmentera le moins vite par rapport au diamètre réel. Par mauvais exemple, appliquée au produit X^N d'une variété riemanienne compacte X, ce qui revient à prendre un produit ℓ^2, cette proposition donne un diamètre observable de l'ordre du diamètre, alors qu'on verra plus bas par des méthodes spectrales que le diamètre observable exponentiel de X^N est constant, donc bien en $1/\sqrt{N}$ du diamètre.

À titre de corollaire, on obtient le théorème important suivant, que l'on peut attribuer à Talagrand (voir l'introduction [Tal96]) et qui énonce une loi des grands nombres (plus exactement une inégalité de Hoeffding) généralisée.

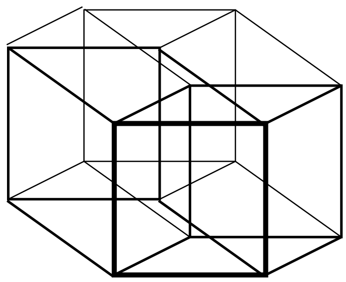
Théorème 12 – *Soit $X = \{0,1\}^N$ le cube discret de dimension N, muni de la mesure uniforme, et de la distance ℓ^1 (ou de Hamming) qui attribue à chaque arête une longueur $1/N$, de sorte que le diamètre de X soit 1. Alors le diamètre observable gaussien de X est $1/\sqrt{N}$.*

Ce théorème généralise fortement les propriétés de type loi des grands nombres ou théorème central limite (dont il est aussi une formulation géométrique) : en effet, ces derniers ne portent que sur des fonctions très particulières (la fonction « proportion de 0 », ou des combinaisons linéaires de fonctions de chaque composante), alors qu'ici le diamètre observable porte sur les fonctions *lipschitziennes*, qui sont évidemment beaucoup plus nombreuses. Être 1-lipschitzienne pour la métrique des arêtes signifie que la modification de l'une des composantes ne modifie jamais le résultat de plus de $1/N$, d'où la formulation intuitive du théorème : une fonction dépendant de N variables indépendantes, telle que chaque variable influe d'au plus $1/N$, est constante à $1/\sqrt{N}$ près. (On a bien sûr des raffinements techniques de l'hypothèse d'indépendance.) Par contre, les constantes données ici ne sont pas optimales.

Démonstration – Démontrons le résultat avec une autre normalisation des distances, dans laquelle chaque arête est de longueur 1 (le diamètre est N) de sorte qu'on veut démontrer que le diamètre observable est $\sqrt{N}$. (Cette normalisation, moins naturelle si l'on pense à la loi des grands nombres ou bien à la comparaison avec le théorème 2 sur la sphère, est plus pratique techniquement.)

D'après le corollaire anodin 8 ci-dessus, comme $\{0,1\}$ est de diamètre 1, on a $Z_{\{0,1\}}(\lambda) \leqslant e^{\lambda^2/2}$. Il suffit alors d'appliquer la proposition 10 pour passer au produit de N facteurs. □

Géométriquement, on a envie de dire que la contrainte d'être 1-lipschitzienne est rendue très forte dans le cube discret à cause des très nombreux chemins distincts entre deux points fixés (à l'opposé, dans un arbre par exemple, être lipschitzienne n'impose presque aucune contrainte, on peut choisir la fonctions arbitrairement dans les différentes directions). On se rapproche de la notion de courbure positive évoquée plus haut, qui, pour les variétés, impliquait la concentration ; en effet, intuitivement la courbure positive dit que l'espace « se referme sur lui-même » (et techniquement cela se traduit par des théorèmes de type Bonnet ou Myers), et cela semble bien être le cas du cube aussi.



Un cube en dimension 4, presque une sphère

La comparaison des théorèmes 2 et 12 sur la sphère et le cube tend à montrer que la grande dimension efface les différences : si on a un espace inconnu de dimension N et que l'on fait des mesures dessus, on n'observera que des gaussiennes et on ne pourra pas différencier un cube d'une sphère !

Ceci est bien entendu à rapprocher du théorème de Dvoretzky ([Dvo61], voir aussi [MS86] pour une démonstration simple utilisant la concentration) sur les sections sphériques des corps convexes, théorème qui est le premier exemple d'application de la concentration de la mesure en-dehors des probabilités.

3.3 Généralisations

Le résultat ci-dessus admet de nombreuses généralisations. Dans la démonstration ci-dessus on n'a pas utilisé le fait que toutes les composantes du produit $\{0,1\}^N$ étaient identiques. On peut aussi attribuer des longueurs différentes aux arêtes ayant des directions différentes. En recopiant directement la démonstration on obtient alors

Théorème 13 – *Soient $X_1, \ldots, X_N$ des espaces de probabilité et soit $X = X_1 \times \ldots \times X_N$ leur produit. Soient $a_1, \ldots, a_N$ des nombres positifs et soit $D = \sqrt{\sum a_i^2}$.*

Soit f une fonction de X vers $\mathbb{R}$ telle que, pour tout i, on ait

$$\left| f(x_1, \ldots, x_i, \ldots, x_N) - f(x_1, \ldots, x_i', \ldots, x_N) \right| \leqslant a_i$$

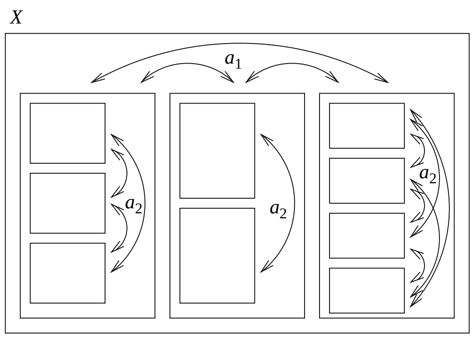
pour tous x_i. Alors pour tout $\varepsilon > 0$ on a

$$\Pr(|f - \mathbb{E}f| \geqslant \varepsilon) \leqslant 2e^{-\varepsilon^2/2D^2}$$

Encore une fois, l'inégalité de Hoeffding classique, de conclusion semblable, ne traite que le cas où la fonction f est une combinaison *linéaire*.

Bien sûr, il arrive qu'on fasse face à des situations où, par exemple, borner l'influence de la variable i par a_i n'est pas possible ou pas optimal (si pour certains points l'influence de certaines variables est très grande tandis que l'influence d'autres variables diminue), ou encore, des cas où l'indépendance n'est pas tout à fait vérifiée. Il existe des variantes des théorèmes ci-dessus permettant de traiter différentes situations (cf. [Led01], chapitre 4).

Donnons à titre de curiosité l'une de ces variantes, dans un vocabulaire géométrique ou combinatoire. Soit X un espace métrique fini, et on suppose que l'on a sur X une suite de partitions de plus en plus fines $Y_0, Y_1, \ldots, Y_N$, où Y_0 est la partition grossière $Y_0 = \{X\}$, où Y_N est la partition discrète $Y_N = \{\{x\}_{x \in X}\}$, et où Y_i est un raffinement de Y_{i-1}. On se donne en outre une suite $a_1, \ldots, a_N$ de nombres positifs et on suppose que, dès que A_1^i, A_2^i sont deux éléments de la partition Y_i tous deux inclus dans un même élément A^{i-1} de la partition Y^{i-1}, alors il existe une bijection $\varphi : A_1^i \to A_2^i$ telle que $d(x, \varphi(x)) \leqslant a_i$, c'est-à-dire que φ déplace les points d'une distance au plus a_i. Alors, X muni de sa mesure uniforme est de diamètre observable gaussien $\sqrt{\sum a_i^2}$.



Cela s'applique par exemple à un groupe discret (compact suffit) G muni d'une métrique invariante : si on a $G = G_0 \supset G_1 \supset \ldots \supset G_n = \{e\}$ une suite de sous-groupes (fermés) imbriqués, tels que G_{i-1}/G_i est de diamètre a_i, alors G muni de sa mesure de Haar est de diamètre observable gaussien $\sqrt{\sum a_i^2}$. (Ceci a été en particulier appliqué aux groupes de permutations.)

3.4 Produits et concentration

On vient de voir qu'un espace produit est un espace concentré. On peut se demander dans quelle mesure la concentration reflète toujours l'action d'un grand nombre de paramètres indépendants. Par exemple sur la sphère $S^N \subset \mathbb{R}^{N+1}$, qui est concentrée, deux fonctions coordonnées sont essentiellement indépendantes ; en fait S^N est très proche dans la topologie de Gromov définie plus bas, de l'espace $\mathbb{R}^{N+1}$ muni de la mesure gaussienne, espace qui est un espace produit. On peut donc se demander dans quelle mesure tout espace concentré est proche d'un espace produit.

Mais la concentration est stable par les « contractions » (applications diminuant les distances et préservant la mesure). On peut donc, au mieux, espérer démontrer qu'un espace concentré est proche d'une contraction d'un espace produit.

Le nombre de paramètres indépendants peut être évalué de la manière suivante : dans les cas rencontrés jusqu'ici, le diamètre observable était toujours de l'ordre de $1/\sqrt{N}$ fois le diamètre. Pour pouvoir aussi traiter des espaces non compacts (comme par exemple $\mathbb{R}^N$ gaussien), il vaut mieux remplacer le diamètre par le rayon moyen $R = \mathbb{E}d(\cdot, \cdot)$. On a vu sur les exemples que le diamètre observable D était de l'ordre de $R/\sqrt{N}$, ce qui conduit, étant donné un espace concentré de diamètre observable gaussien D et de rayon moyen R, à définir sa "dimension statistique" $N = (R/D)^2$.

On a un début de résultat dans ce sens : sur un espace concentré, il existe des fonctions jouant le rôle de fonctions coordonnées, telles que deux d'entre elles sont relativement indépendantes au sens où elles vérifient une inégalité du type

$$\Pr(f_1 \geqslant a, f_2 \geqslant b) \leqslant \Pr(f_1 \geqslant a) \Pr(f_2 \geqslant 2)$$

(comme la concentration est stable par les contractions, ces fonctions pourraient prendre des valeurs anormalement petites, et on ne demande donc le contrôle que du côté des grandes valeurs, d'où les inégalités). Plus exactement ([Oll03]) :

Théorème 14 – *Soit $m \geqslant 1$ un entier. Soit $(X, d, \Pr)$ un espace métrique probabilisé, de rayon moyen $R = \mathbb{E}d(\cdot, \cdot)$. On suppose que X est de diamètre observable gaussien D et on pose $N = (R/D)^2$.*

Alors il existe une famille de fonctions $f_i : X \to \mathbb{R}$ pour $1 \leqslant i \leqslant m$ telles que, asymptotiquement quand $N \to \infty$, on a

- *Chaque f_i est 2-lipschitzienne, satisfait $|\mathbb{E}f_i| = O(D)$ et prend ses valeurs dans un intervalle de taille au moins $R(1 - O(1/\sqrt{N}))$.*
- *Pour tout entier $1 \leqslant k \leqslant m$, pour toute sous-famille $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant m$, pour tout k-uplet $t_j \in \mathbb{R}$ vérifiant $D\sqrt{\log \log N} \ll t_j \leqslant 2R$, on a*

$$\Pr\left(\left|f_{i_j}\right| \geqslant t_j \ \forall j\right) \leqslant e^{-(\sum t_j^2)/(2+o(1))D^2}$$

Autrement dit, l'évaluation de la probabilité que plusieurs fonctions coordonnées f_i dépassent simultanément certaines valeurs, est égale au produit des évaluations naturelles pour chaque f_i.

Bien sûr, sans la contrainte sur la taille de l'intervalle des valeurs (qui n'est pas si faible dans un espace concentré), le théorème serait trivial car $f_i = 0$ conviendrait. Mais on aimerait plutôt obtenir une borne inférieure sur la variance des f_i, qui devrait être de l'ordre de D^2 avec D la valeur optimale du diamètre observable gaussien.

Si X est concentré, alors bien sûr les fonctions lipschitziennes de X vers $\mathbb{R}^m$, pour m fixé, sont aussi concentrées, mais ce qu'on obtient par cette dernière observation est bien moins bon que le théorème ci-dessus, puisqu'on obtient alors en exposant une moyenne plutôt qu'une somme des t_i^2. La concentration des fonctions lipschitziennes vers $\mathbb{R}^m$ ne suffit pas à obtenir le résultat ci-dessus, puisque par exemple si $f_1 = f_2 = \ldots = f_m$ l'estimation du théorème est fausse.

À noter que dans le cas du cube $\{0,1\}^n$, les fonctions que l'on construit ainsi ne sont pas les fonctions coordonnées standard (qui ont peu de signification statistique : la 372e coordoonnée est-elle un 0 ou un 1 ?), mais plutôt des fonctions globales telles que la différence entre le total des 0 et le total des 1, cette même différence dans la moitié de droite, etc., qui sont statistiquement plus sensées. Géométriquement, ceci est encore relié au fait qu'un cube en grande dimension n'apparaît cubique qu'observé dans certaines directions très particulières, alors que dans la plupart des autres directions il est plutôt sphérique.

Ce résultat peut certainement être amélioré : en particulier, on s'attendrait à ce que le nombre de fonctions que l'on peut extraire soit de l'ordre de la dimension statistique N.

4 Inégalités fonctionnelles

Supposer des inégalités fonctionnelles sur l'espace de départ (spectre du laplacien, log-Sobolev...) permet de démontrer de la concentration de la mesure. Ces inégalités s'appliquent à une grande variété d'objets (variétés, graphes...). Elles ont en outre souvent un comportement très agréable par produit.

4.1 Spectre du laplacien, inégalité de Poincaré, variétés produits

On commence par donner un énoncé un peu informel, que l'on peut attribuer à Gromov et Milman (voir [GM83]).

Théorème 15 – *Soit (X, d, μ) un espace métrique probabilisé sur lequel on dispose d'une notion de laplacien et telle que la plus petite valeur propre non nulle du laplacien soit λ_1. Alors X est de diamètre observable exponentiel $1/\sqrt{\lambda_1}$.*

Les deux cas les plus importants sont bien sûr les variétés riemanniennes compactes et les graphes non orientés (plus généralement, toutes les chaînes de Markov réversibles sur des espaces discrets), pour lequel (si X est connexe) on a toujours $\lambda_1 > 0$.

Remarquons d'abord l'homogénéité du résultat : les valeurs propres du laplacien ont l'homogénéité de l'inverse du carré d'une longueur, et donc $1/\sqrt{\lambda_1}$ est bien une longueur.

Il y a deux religions pour le choix du signe du laplacien. On prend celle qui en fait un opérateur positif sur $L^2(X)$, soit, pour une variété, $-\sum \frac{\partial^2}{\partial x_i^2}$, et, pour un graphe symétrique, $\Delta f(x) = f(x) - \frac{1}{\deg x}\sum_{y\sim x} f(y)$, qui vaut aussi 1 moins l'opérateur de marche aléatoire. Sur une variété riemannienne compacte on prend comme mesure la mesure riemannienne normalisée, et sur un graphe non orienté fini, la mesure normalisée qui à un sommet donne un poids proportionnel à son degré, de sorte que Δ soit un opérateur symétrique.

On va démontrer le théorème ci-dessus dans le cas d'une variété riemannienne compacte. La démonstration pour un graphe ou une chaîne de Markov réversible est entièrement similaire, bien que, curieusement, les deux démonstrations ne soient pas techniquement unifiées dans un cadre commun.

Démonstration – On va utiliser un critère de type Laplace, mais exponentiel plutôt que quadratique : si, pour toute fonction 1-lipschitzienne de moyenne nulle sur X, pour un certain λ positif on a $\mathbb{E}e^{\lambda f} \leqslant 2$, alors X est de diamètre exponentiel D par l'inégalité de Markov

$$\Pr(f \geqslant t) = \Pr\left(e^{\lambda f} \geqslant e^{\lambda t}\right) \leqslant 2/e^{\lambda t}$$

(notons qu'ici on n'a besoin que d'un seul λ, alors que dans le cas quadratique le choix de λ optimal dépend de t).

L'hypothèse de valeur propre du laplacien s'écrit, pour toute fonction différentiable g :

$$\langle \Delta g, g\rangle \geqslant \lambda_1 \langle g - \mathbb{E}g, g - \mathbb{E}g\rangle$$

ou encore

$$\|\nabla g\|_2^2 \geqslant \lambda_1 \operatorname{Var} g$$

qui a l'avantage d'être une formulation ayant un sens pour toute fonction lipschitzienne (∇g n'étant pas défini, mais $|\nabla g|$ au point x étant le module de continuité de g en x), et d'être valable pour ces fonctions avec cette interprétation de $\|\nabla g\|_2^2 = \int |\nabla g|^2$.

Soit donc f une fonction 1-lipschitzienne de moyenne nulle sur X ; $e^{\lambda f}$ est une fonction lipschitzienne. L'idée est d'estimer $\mathbb{E}e^{\lambda f}$ en fonction de $\mathbb{E}e^{\lambda f/2}$, puis d'itérer.

$$\begin{aligned}\mathbb{E}e^{\lambda f} &= \left\|e^{\lambda f/2}\right\|_2^2 = \operatorname{Var} e^{\lambda f/2} + \left(\mathbb{E}e^{\lambda f/2}\right)^2 \\ &\leqslant \frac{1}{\lambda_1}\left\|\nabla e^{\lambda f/2}\right\|_2^2 + \left(\mathbb{E}e^{\lambda f/2}\right)^2\end{aligned}$$

or on a pour les modules de continuité $\left|\nabla e^{\lambda f/2}\right| \leqslant \frac{\lambda}{2}|\nabla f|\, e^{\lambda f/2}$ et donc, puisque $|\nabla f| \leqslant 1$:

$$\mathbb{E}e^{\lambda f} \leqslant \frac{1}{\lambda_1}\int \frac{\lambda^2}{4} e^{\lambda f} + \left(\mathbb{E}e^{\lambda f/2}\right)^2$$

d'où enfin

$$\mathbb{E}e^{\lambda f}\left(1 - \frac{\lambda^2}{4\lambda_1}\right) \leqslant \left(\mathbb{E}e^{\lambda f/2}\right)^2$$

Par récurrence on obtient

$$\mathbb{E}e^{\lambda f}\prod_{i=1}^{k}\left(1-\frac{\lambda^2}{2^{2i}\lambda_1}\right)^{2^{i-1}} \leqslant \left(\mathbb{E}e^{\lambda f/2^k}\right)^{2^k}$$

Or il est très facile de voir que $\lim_{k\to\infty}\left(\mathbb{E}e^{\lambda f/2^k}\right)^{2^k} = e^{\mathbb{E}f} = 1$. Prenons $\lambda = \sqrt{\lambda_1}$. Alors le produit infini $\prod_{i\geqslant 1}\left(1-\frac{\lambda^2}{2^{2i}\lambda_1}\right)^{2^{i-1}}$ est supérieur à $1-\sum_{i\geqslant 1}2^{i-1}/2^{2i} = 1/2$. La conclusion est que

$$\mathbb{E}e^{f\sqrt{\lambda_1}} \leqslant 2$$

ce qui achève la démonstration (au détail près qu'on obtient 4 plutôt que 2 comme constante multiplicative dans la définition de la concentration gaussienne).

Dans cette démonstration, nous n'avons pas mentionné le problème de l'existence de $\mathbb{E}e^{\lambda f}$ (qui pourrait être a priori infini). Un simple raisonnement de limite à partir des fonctions bornées lève l'obstacle. □

Ce critère a l'énorme avantage de très bien passer aux produits de variétés riemanniennes. En effet, si X_1 et X_2 sont des variétés riemanniennes compactes on a $\lambda_1(X_1\times X_2) = \min(\lambda_1(X_1), \lambda_1(X_2))$. En particulier $\lambda_1(X^N) = \lambda_1(X)$. On obtient immédiatement :

Proposition 16 – *Soit X une variété riemannienne compacte connexe. Le diamètre observable exponentiel de X^N est borné lorsque $N\to\infty$.*

(Le diamètre, lui, augmente comme $\sqrt{N}$: on retrouve toujours le même facteur $1/\sqrt{N}$.)

4.2 Inégalité de Sobolev logarithmique

L'inégalité de Sobolev logarithmique (vérifiée, notamment, sur $\mathbb{R}^N$ gaussien, avec des constantes indépendantes de N) permet d'obtenir de la concentration gaussienne. Pour une excellente introduction à cette inégalité, on pourra consulter [ABCFGMRS00], qui donne de nombreuses applications y compris à la concentration.

Soit (X, d, μ) un espace métrique probabilisé et f une fonction positive sur X. On définit l'*entropie* de f par

$$\operatorname{Ent} f = \int f\log f - \int f\,\log\int f$$

qui est une quantité positive, homogène de degré 1 en f, et convexe, mesurant l'information relative de la mesure $f\mu$ par rapport à la mesure μ (c'est-à-dire la quantité moyenne de $f\mu$-information qu'on obtient en tirant un élément de X au hasard selon μ).

On suppose que X est muni d'une notion de norme du gradient $|\nabla f|$ comme ci-dessus (par exemple, si X est une variété). On dit que (X, d, μ) satisfait l'inégalité de Sobolev logarithmique de constante C si, pour toute fonction sur X, on a

$$\operatorname{Ent}(f^2) \leqslant 2C \left\| \nabla f \right\|_2^2$$

où l'on prend un carré sous l'entropie, premièrement pour avoir une fonction positive, mais surtout pour que cette inégalité soit homogène. Noter que C est homogène au carré d'une longueur.

L'inégalité de Sobolev logarithmique, appliquée à la fonction $1 + \varepsilon f$ avec $\varepsilon \to 0$, implique en particulier l'inégalité de Poincaré

$$\operatorname{Var} f \leqslant C \left\| \nabla f \right\|_2^2$$

d'où le choix du facteur 2 devant C. En effet $\operatorname{Ent}(1 + \varepsilon f) \sim \frac{\varepsilon^2}{2} \operatorname{Var} f$.

Le théorème suivant peut être attribué à Herbst (voir [ABCFGMRS00] ou les notes du chapitre 5 de [Led01]).

Théorème 17 – *Soit (X, d, μ) un espace métrique probabilisé vérifiant l'inégalité de Sobolev logarithmique de constante C, pour une bonne notion de norme du gradient. Alors X est de diamètre observable gaussien $\sqrt{C}$.*

Démonstration – On va encore utiliser le critère de Laplace (Proposition 7). L'inégalité de Poincaré permettait de comparer $Z_f(\lambda)$ et $Z_f(\lambda/2)$. L'inégalité de Sobolev logarithmique permet de contrôler la dérivée logarithmique de $Z_f(\lambda)$ (dont l'expression fait apparaître l'entropie de $e^{\lambda f}$). Plus exactement, comme pour les petits λ on a $\ln Z_f(\lambda) \approx \lambda \mathbb{E} f$, posons

$$h(\lambda) = \frac{\ln Z_f(\lambda)}{\lambda}$$

et $h(0) = \mathbb{E} f$ par continuité. On vérifie alors immédiatement que

$$\frac{dh(\lambda)}{d\lambda} = \frac{1}{\lambda^2} \frac{\operatorname{Ent} e^{\lambda f}}{\mathbb{E} e^{\lambda f}}$$

or par hypothèse, $\operatorname{Ent} e^{\lambda f} \leqslant 2C \left\| \nabla e^{\lambda f/2} \right\|_2^2 \leqslant 2C \frac{\lambda^2}{4} \left\| f \right\|_{\mathrm{Lip}}^2 \left\| e^{\lambda f/2} \right\|_2^2$. On obtient donc, si f est 1-lipschitzienne, que $h'(\lambda) \leqslant C/2$, d'où $h(\lambda) \leqslant \mathbb{E} f + \lambda C/2$, ou enfin

$$Z_f(\lambda) \leqslant e^{\lambda \mathbb{E} f} e^{\lambda^2 C/2}$$

d'où la conclusion.

Là encore, un argument de limite permet de s'assurer qu'il suffit de raisonner sur les fonctions f bornées, pour lesquelles l'existence de l'entropie de $e^{\lambda f}$ ne pose pas de problème. $\square$

L'inégalité de Sobolev logarithmique a elle aussi un comportement très agréable sous les produits ℓ^2, grâce à la convexité de l'entropie, qui implique

que l'entropie d'une fonction sur un espace produit est inférieure à la somme de ses entropies partielles moyennes. Ceci prouve que la constante $C(X_1 \times X_2)$ vaut $\max(C(X_1), C(X_2))$.

L'approche entropique de la concentration de la mesure donne lieu à des développements importants en statistiques (voir par exemple [BLM00]), qui dans certains cas affaiblissent beaucoup les hypothèses nécessaires à la concentration de la mesure et permettent en particulier de passer de l'hypothèse lipschitzienne à l'hypothèse « lipschitzienne en moyenne ». *Ceci devrait faire l'objet d'un exposé ultérieur.* Pour d'autres applications, voir [ABCFGMRS00] ou par exemple [Led01], chapitre 5.

5 Courbure et concentration

On rappelle le théorème (Lévy-Gromov) énoncé plus haut : une variété de dimension N, dont la courbure de Ricci est en tout point supérieure à celle de la sphère S^N, a un diamètre observable gaussien inférieur à celui de S^N (soit $1/\sqrt{N-1}$).

L'hypothèse sur la courbure de Ricci est extrêmement fragile puisqu'elle nécessite de se donner une variété riemannienne, tandis que la conclusion s'exprime uniquement en termes d'une métrique et d'une mesure. On veut donc étudier quelques voies permettant peut-être d'adopter une définition plus générale de la courbure de Ricci, sur des espaces moins réguliers, qui entraînerait la concentration.

Déjà, on peut se demander si ce théorème reste valable pour des espaces à courbure positive au sens des triangles de comparaison (ce qui, pour une variété, est analogue à suppose que la courbure sectionnelle est positive, ce qui est plus fort que la courbure de Ricci ; voir [BBI01]) ; ce n'est pas évident et mériterait d'être étudié.

Ensuite, on donne ci-dessous deux approches qui font ressortir des caractérisations alternatives équivalentes à « courbure de Ricci minorée » : les inégalités courbure-dimension, et le transport de mesure.

5.1 Inégalités courbure-dimension

Les inégalités de courbure-dimension, ou de Bakry-Émery (voir par exemple le survol [Led00]), sont des inégalités apparaissant dans les semi-groupes de diffusion sur les variétés (l'exemple typique étant l'équation de la chaleur, éventuellement dans un potentiel). *Cette théorie devrait faire l'objet d'un exposé ultérieur.*

Soit donc $(P_t) = \left(e^{tL}\right)$ un semi-groupe engendré par un opérateur L d'ordre 2, tel que P_t est positif, et tel que, pour tout t, on ait $P_t 1 = 1$. On suppose L défini sur une certaine algèbre de fonctions qu'on ne précisera pas. On suppose en outre qu'il existe une mesure μ stationnaire pour P_t (c'est-à-dire que pour toute fonction f on a $\int P_t f \, d\mu = \int f \, d\mu$), et que le processus est réversible par rapport à μ (c'est-à-dire que pour toutes fonctions f et g, on a $\int f\, P_t g\, d\mu = \int g\, P_t f\, d\mu$; intuitivement cela signifie que non seulement μ est une mesure d'équilibre, mais qu'en outre à l'équilibre, pour tout couple de points, la masse qui passe de l'un vers l'autre est la même que celle passant du second vers le premier).

On définit l'opérateur « carré du champ » Γ par

$$\Gamma(f,g) = \frac{1}{2}\left(L(fg) - g\, Lf - f\, Lg\right)$$

qui est une forme bilinéaire positive. On notera aussi $\Gamma f = \Gamma(f,f)$. Cette définition est motivée par le fait que, dans le cas standard où L est le laplacien, alors $\Gamma f = |\nabla f|^2$. On a de plus $\int f\, Lf\, d\mu = -\int \Gamma f\, d\mu$.

Si on veut faire apparaître de la courbure il faut dériver une fois de plus. L'opérateur « carré du champ itéré » est défini par

$$\Gamma_2(f,g) = \frac{1}{2}\left(L\Gamma(f,g) - \Gamma(f, Lg) - \Gamma(g, Lf)\right)$$

et on notera encore $\Gamma_2 f = \Gamma_2(f,f)$. Si L est le laplacien sur $\mathbb{R}^n$, on a simplement $\Gamma_2 f = \|\mathrm{Hess}\, f\|_{HS}^2$ (norme de Hilbert-Schmidt). De manière plus intéressante, si L est le laplacien sur une variété riemannienne de dimension n, on a

$$\Gamma_2 f = \|\mathrm{Hess}\, f\|_{HS}^2 + \mathrm{Ric}\,(\nabla f, \nabla f)$$

où Ric est le tenseur de Ricci de la variété riemannienne (formule de Bochner, voir par exemple [GHL87]).

On s'intéresse à des caractérisations alternatives de « la courbure de Ricci est supérieure à k ». On peut en outre remarquer que $(\Delta f)^2 \leqslant n\, \|\mathrm{Hess}\, f\|_{HS}^2$ par Cauchy-Schwarz (car Δf est la trace de $\mathrm{Hess}\, f$). On obtient donc que si la courbure de Ricci est supérieure à k on a pour toute fonction f l'inégalité

$$\Gamma_2 f \geqslant \frac{1}{n}\,(\Delta f)^2 + k\, |\nabla f|^2$$

en tout point.

Ceci amène à poser qu'un opérateur de diffusion L satisfait l'inégalité de courbure-dimension (k,n) en un point si, pour toute fonction f, on a

$$\Gamma_2 f \geqslant \frac{1}{n}(Lf)^2 + k\,\Gamma f$$

en cohérence avec le cas du laplacien sur les variétés riemanniennes. À noter que cette inégalité a un sens ponctuellement.

On a alors la généralisation suivante du théorème de Lévy-Gromov :

Théorème 18 – *Soit $(P_t) = \left(e^{tL}\right)$ un semi-groupe de diffusion admettant une mesure de probabilité μ réversible et stationnaire. On suppose de plus que (P_t) est ergodique (c'est-à-dire que $P_t f \to \int f\, d\mu$: intuitivement, P_t atteint tout l'espace).*

Si ce semi-groupe satisfait l'inégalité de courbure-dimension (k, ∞) *en (presque) tout point, alors toute fonction* f *satisfait l'inégalité de Sobolev logarithmique généralisée*

$$\operatorname{Ent}(f^2) \leqslant \frac{2}{k} \int \Gamma f \, d\mu$$

En particulier, une variété riemannienne de dimension n et de courbure de Ricci plus grande que k, avec pour L le laplacien, satisfait l'inégalité de courbure-dimension (k, n) et donc a fortiori (k, ∞). Elle vérifie donc l'inégalité de Sobolev logarithmique $\operatorname{Ent}(f^2) \leqslant \frac{2}{k} \int |\nabla f|^2$, car dans ce cas $\Gamma = |\nabla \ |^2$; par conséquent, d'après les résultats des sections précédentes, cette variété est de diamètre observable gaussien $\sqrt{2/k}$, ce qui est, à une constante près et par une voie plus détournée, le théorème de Lévy-Gromov.

5.2 Courbure et transport de mesure

Une autre caractérisation de la courbure de Ricci positive pour les variétés riemanniennes, ne faisant intervenir que la métrique et la mesure sans structure supplémentaire, a été récemment démontrée par Sturm et von Renesse ([SvR]). Il s'agit d'une propriété globale de l'espace des mesures de probabilité sur la variété.

On rappelle d'abord la notion de K-convexité d'une fonction, qui est un léger renforcement de la notion de convexité. Supposons qu'on a un espace A muni d'une notion de plus courts chemins entre deux points (i.e. un espace géodésique, cf. [BBI01]). Si x, y sont deux points de A, et qu'on s'est donné un plus court chemin $[xy]$ de x à y, on note $[xy]_t$, pour $0 \leqslant t \leqslant 1$, le point de $[xy]$ situé à distance $t \operatorname{dist}(x, y)$ de x (ainsi $[xy]_0 = x$ et $[xy]_1 = y$). On dit qu'une fonction $f : A \to \mathbb{R}$ est K-convexe si pour tous x, y dans A reliés par un plus court chemin $[xy]$, on a

$$f([xy]_t) \leqslant (1-t)f(x) + tf(y) - \frac{K}{2} t(1-t) \operatorname{dist}(x, y)^2$$

autrement dit, le graphe de la fonction se situe non seulement sous le segment reliant $f(x)$ à $f(y)$, mais, mieux, sous une parabole de courbure K située sous ce segment. Bien sûr K est homogène à l'inverse du carré d'une longueur.

On rappelle aussi la notion de distance de transport L^2 entre deux mesures sur un même espace (ou distance de Wasserstein ; voir [Vil03] pour une introduction au transport de mesure) : si μ et ν sont deux mesures de probabilité sur un espace métrique X, cette distance consiste à imaginer que μ représente une répartition de masse qu'on veut transformer en ν en transportant la masse, et en supposant que transporter une unité de masse sur une distance d coûte d^2. Plus précisément :

$$d_2^W(\mu, \nu) = \inf_\pi \sqrt{\int_{X \times X} \operatorname{dist}(x_1, x_2)^2 \, d\pi(x_1, x_2)}$$

où l'inf est pris sur toutes les mesures de probabilité π sur $X \times X$ dont les deux projections sur X sont μ et ν (intuitivement $d\pi(x_1, x_2)$ représente la quantité de masse qu'on fait passer de x_1 à x_2).

Sur une variété riemannienne (connexe complète) X, la distance de Wasserstein fait de l'espace $\mathcal{P}(X)$ des mesures de probabilité (boréliennes) sur X un espace métrique dans lequel on a une notion de plus courts chemins.

On a alors la caractérisation suivante :

Théorème 19 ([SvR]) – *Soit X une variété riemannienne complète connexe, dont on note* vol *la mesure riemannienne. Soit $K \in \mathbb{R}$. Alors sont équivalents :*

1. $\mathrm{Ric}(X) \geqslant K$ *;*
2. *la fonction entropie*

$$\mathrm{Ent}_{\mathrm{vol}}(\mu) = \int_X \frac{d\mu}{d\,\mathrm{vol}} \log \frac{d\mu}{d\,\mathrm{vol}}\, d\,\mathrm{vol}$$

   est K-convexe sur l'espace $\mathcal{P}(X)$ muni de la structure géodésique issue de la distance de transport L^2.

Un des avantages de cet énoncé est que la caractérisation obtenue a un sens sur des espaces beaucoup plus généraux que des variétés riemanniennes. Noter aussi que le théorème est valable pour les K négatifs, ainsi que pour des variétés non compactes. *Ce théorème devrait faire l'objet d'un exposé ultérieur.*

La question qui se pose très naturellement est alors : est-ce que cette caractérisation, sur un espace autre qu'une variété riemannienne, implique de la concentration de la mesure ?

6 Le cadre métrique mesuré

Le cadre naturel pour la concentration de la mesure consiste à dire que si on modifie un peu la distance, alors la concentration ne va pas être affectée. De même, si on modifie beaucoup la distance mais sur une partie de très faible mesure, la concentration ne va pas non plus être très affectée. Ceci permet de définir, suivant Gromov ([Gro99], chapitre $3\frac{1}{2}$), une topologie naturelle sur l'espace des espaces métriques mesurés, topologie pour laquelle un espace proche d'un espace concentré sera encore concentré.

Soient d'abord (X, d, μ) et (X, d', μ) deux métriques sur le même espace mesuré (X, μ), mettons de mesure 1. On dit qu'elles sont (ε, κ)-proches (ε est une petite distance et κ est une petite mesure) s'il existe $Y \subset X$ de mesure au moins $1 - \kappa$, telle que pour tous $x_1, x_2 \in Y$ on a

$$\left| d(x_1, x_2) - d'(x_1, x_2) \right| \leqslant \varepsilon$$

Ensuite, pour comparer abstraitement deux espaces mesurés (X, d, μ) et (X', d', μ'), il faut essayer de « recoller » au mieux X et X' (comparer avec la définition de la topologie de Hausdorff-Gromov sur la classe des espaces compacts). Supposons pour simplifier que X et X' sont tous deux de masse 1. Sous nos hypothèses de « légère régularité » des espaces métriques mesurés, tout espace de masse 1 peut être paramétré par le segment $[0;1]$ de manière à préserver la mesure ; plus précisément, il existe une fonction $\varphi : [0;1] \to X$, surjective (à un ensemble de mesure nulle près), envoyant la mesure de Lebesgue sur la mesure de X (de plus, si X n'a pas d'atomes, alors φ est bijective à des ensembles de mesure nulle près). De même, il existe $\varphi' : [0;1] \to X'$ jouissant des mêmes propriétés.

On dit que (X, d, μ) et (X', d', μ') sont (ε, κ)-proches pour les paramétrisations φ et φ', si les deux distances $d \circ \varphi$ et $d' \circ \varphi'$ ramenées sur $[0;1]$ sont (ε, κ)-proches. On dit que (X, d, μ) et (X', d', μ') sont (ε, κ)-proches tout court s'il existe des paramétrisations φ et φ' pour lesquelles ils sont (ε, κ)-proches. Autrement dit, ces paramétrisations permettent d'identifier X et X' à une petite déformation ($\leqslant \varepsilon$) de la métrique près, éventuellement en jetant des parties de petite mesure ($\leqslant \kappa$).

Cette topologie est métrisable, par exemple en disant que deux espaces sont à distance inférieure à δ s'ils sont (δ, δ)-proches (noter que cette définition n'est pas homogène : on a choisi un facteur de proportionnalité entre une longueur et une masse). Il est vrai mais non évident qu'il s'agit vraiment d'une métrique, c'est-à-dire que deux espaces (δ, δ)-proches pour tout δ sont bien isomorphes (à des parties de mesure nulle près).

Les propriétés de cette topologie devraient faire l'objet d'un exposé ultérieur. L'une des plus pratiques, qui évite de se poser des problèmes de mesurabilité, est que tout espace métrique probabilisé peut être approché par des espaces de probabilité finis. Un inconvénient corollaire est que dans cet espace, localement à très petite échelle il peut se passer n'importe quoi.

Il existe un ordre naturel sur cet espace : X domine Y s'il existe une application 1-lipschitzienne de X vers Y préservant la mesure (autrement dit, Y est une contraction de X). Bien sûr, la concentration de la mesure est stable par contractions, ce que nous avons déjà évoqué.

L'un des objectifs du groupe de travail serait donc de formuler des théorèmes de concentration dont les hypothèses se placeraient naturellement dans cette topologie, c'est-à-dire qu'un espace proche d'un espace les satisfaisant les satisferait encore. C'est déjà le cas pour certains critères (inégalité de Laplace, ce qui est normal puisqu'on a vu qu'elle était équivalente à la concentration), mais pas du tout, par exemple, des critères faisant intervenir de la courbure riemannienne ou des gradients ([Gro99] donne des pistes pour ce dernier cas). Satisfaire cette exigence naturelle permettra certainement de développer de nouveaux concepts sur les espaces métriques mesurés.

Références

[ABCFGMRS00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*, Panoramas et synthèses **10**, SMF (2000).

[BBI01] D. Burago, Yu. Burago, S. Ivanov, *A course in metric geometry*, AMS (2001).

[BLM00] S. Boucheron, G. Lugosi, P. Massart, *A sharp concentration inequality with applications*, Random Structures and Algorithms **16** (2000), No. 3, 277–292.

[Dvo61] A. Dvoretzky, *Some results on convex bodies and Banach spaces*, Proc. Symp. on Linear Spaces (Jerusalem, 1961), 123–160.

[GHL87] S. Gallot, D. Hulin, J. Lafontaine, *Riemannian geometry*, Springer (1987).

[GM83] M. Gromov, V. Milman, *A topological application of the isoperimetric inequality*, Amer. J. Math. **105** (1983), 843–854.

[Gro99] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*, Progress in Math. **152**, Birkhäuser (1999).

[Gro03] M. Gromov, *Isoperimetry of waists and concentration of maps*, Geom. Funct. Anal. **13** (2003), No. 1, 178–215.

[Led00] M. Ledoux, *The geometry of Markov diffusion generators*, Ann. Fac. Sci. Toulouse **IX**, 305–366 (2000).

[Led01] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs **89**, AMS (2001).

[Lév51] P. Lévy, *Leçons d'analyse fonctionnelle*, Gauthier-Villars, Paris (1922), réédité en 1951 sous le titre *Problèmes concrets d'analyse fonctionnelle*.

[MS86] V. Milman, G. Schechtman, *Asymptotic theory of finite dimensional normed spaces*, Lecture Notes in Mathematics 1200, Springer-Verlag, Berlin (1986).

[Oll99] Y. Ollivier, *Diamètre observable des sous-variétés de S^n et $\mathbb{C}\mathrm{P}^n$*, mémoire de DEA, université d'Orsay (1999).

[Oll03] Y. Ollivier, *Concentrated spaces seen as product spaces*, in *Probabilités sur les espaces de configuration d'origine géométrique*, thèse de doctorat, université Paris-Sud, 2003.

[SvR] K.-T. Sturm, M.-K. von Renesse, *Transport Inequalities, Gradient Estimates, Entropy and Ricci Curvature*, preprint.

[Tal96] M. Talagrand, *A new look at independence*, Ann. Prob. **24** (1996), No. 1, 1–37.

[Vil03] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics **58**, AMS (2003).