Processus stochastiques, Convexité et Inégalités fonctionnelles

Joseph Lehec

Université Paris-Dauphine
Mémoire d'habilitation à diriger des recherches

Spécialité : Mathématiques

Processus stochastiques, Convexité, et Inégalités fonctionnelles

Joseph Lehec

Habilitation soutenue le 15 novembre 2016 devant un jury composé de :

- Sergey Bobkov
- Djalil Chafaï
- Arnak Dalalyan
- Jean Dolbeault
- Michel Ledoux
- Bernard Maurey

Au vu des rapports de :

- Fabrice Baudoin
- Sergey Bobkov
- Michel Ledoux

Remerciements

Je suis très honoré que Fabrice Baudoin, Sergey Bobkov et Michel Ledoux aient accepté d'être rapporteurs pour ce travail. Je suis également très reconnaissant envers Sergey Bobkov et Michel Ledoux de s'être déplacés pour la soutenance, de très loin en ce qui concerne le premier nommé.

Je remercie chaleureusement Djalil Chafaï d'avoir accepté de coordonner cette habilitation, notamment d'avoir relu le mémoire, et de m'avoir permis de l'améliorer par ses commentaires toujours constructifs.

Je remercie Arnak Dalalyan et Jean Dolbeault d'avoir accepté de faire partie du jury, ainsi que Bernard Maurey. J'en profite pour remercier encore une fois ce dernier pour tout ce que j'ai appris à son contact, notamment à l'époque où j'étais en thèse.

Les travaux présentés dans ce mémoire ont tous été conduits depuis mon arrivée à Dauphine, en septembre 2009. Je remercie tous mes collègues du Ceremade, qui contribuent à faire de ce laboratoire un lieu de travail particulièrement agréable et stimulant. À cet égard je voudrais souligner l'importance du travail accompli par Olivier Glass en tant que directeur. Je remercie aussi tout particulièrement César Faivre pour sa disponibilité, et je m'excuse auprès de lui pour mes demandes d'ordre de mission toujours déposées au dernier moment. Enfin je voudrais aussi remercier toute l'équipe du département MIDO pour son efficacité.

Contents

Publication list

Journal articles

1. The symmetric property (τ) for the Gaussian measure. *Ann. Fac. Sci. Toulouse Math. (6)* 17 (2008), no. 2, 357–370.

2. A direct proof of the functional Santaló inequality. *C. R. Math. Acad. Sci. Paris* 347 (2009), no. 1–2, 55–58.

3. Partitions and functional Santaló inequalities. *Arch. Math. (Basel)* 92 (2009), no. 1, 89–94.

4. On the Yao-Yao partition theorem. *Arch. Math. (Basel)* 92 (2009), no. 4, 366–376.

5. Moments of the Gaussian chaos. *Séminaire de Probabilités XLIII*, 327–340, Lecture Notes in Math., 2006, *Springer, Berlin*, 2011.

6. Representation formula for the entropy and functional inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.* 49 (2013), no. 3, 885–899.

7. Cover times and generic chaining. *J. Appl. Probab.* 51 (2014), no. 1, 247–261.

8. A short probabilistic proof of the Brascamp-Lieb and Barthe theorems. *Canad. Math. Bull.* 57 (2014), no. 3, 585–597.

9. with R. Eldan. Bounding the norm of a log-concave vector via thin-shell estimate. *Geometric Aspects of Functional Analysis*, 107–122, Lecture Notes in Math. 2116, *Springer, Cham*, 2014.

10. with U. Caglar, M. Fradelizi, O. Guédon, C. Schütt, and E. Werner. Functional Versions of L_p-Affine Surface Area and Entropy Inequalities. *Int. Math. Res. Not. IMRN* 2016 (2016), no. 4, 1223–1250.

11. Regularization in L_1 for the Ornstein-Uhlenbeck semigroup. *Ann. Fac. Sci. Toulouse Math. (6)* 25 (2016), no. 2, 191–204.

Preprints

12. with S. Bubeck and R. Eldan. Sampling from a log-concave distribution with Projected Langevin Monte Carlo, submitted.

13. Borell's formula for a Riemannian manifold and applications, submitted.

Proceedings

- with E. Boissard, N. Gozlan, C. Léonard, G. Menz, and A. Schlichting, Some recent developments in functional inequalities. Journées MAS 2012, 338–354, ESAIM Proc., 44, *EDP Sci., Les Ulis*, 2014.

- with S. Bubeck, and R. Eldan. Finite-Time Analysis of Projected Langevin Monte Carlo. Advances in Neural Information Processing Systems 28 (NIPS 2015).

Foreword

This memoir presents an overview of the research I conducted after I received my Ph.D. at the end of 2008. The first 5 items of the above list, which were part of this Ph.D. thesis are not addressed here. All the other articles are discussed, although in different degrees of detail. The material is organized as follows.

- Chapter 1: articles 6, 8 and 13.
- Chapter 2: article 11.
- Chapter 3: article 12.
- Chapter 4: article 9.
- Chapter 5: article 7.
- Chapter 6: article 10.

Version française abrégée

Dans ce mémoire on s'intéresse aux inégalités fonctionnelles, c'est-à-dire aux inégalités impliquant des intégrales de fonctions, et à leur interaction avec les processus stochastiques. Cette interaction s'exprime de deux manières. On peut tirer des informations sur le comportement d'un processus stochastique donné à partir d'une inégalité fonctionnelle. À l'inverse les processus et le calcul stochastique s'avèrent être de puissants outils pour démontrer des inégalités. Donnons tout de suite un exemple.

On note γ_n la mesure gaussienne standard sur $\mathbb{R}^n$. L'inégalité de Sobolev logarithmique joue un rôle central tout au long de ce mémoire. Elle affirme que toute mesure de probabilité μ sur $\mathbb{R}^n$ admettant une densité ρ par rapport à γ_n suffisamment régulière vérifie

$$\mathrm{H}(\mu \mid \gamma_n) \leq \frac{1}{2}\,\mathrm{I}(\mu \mid \gamma_n)$$

où $\mathrm{H}(\mu \mid \gamma_n) = \int_{\mathbb{R}^n} \log \rho \, d\mu$ est l'entropie relative de μ et

$$\mathrm{I}(\mu \mid \gamma_n) = \int_{\mathbb{R}^n} |\nabla \log \rho|^2 \, d\mu$$

est son information de Fisher. En appliquant cette inégalité à ρ de la forme $1 + \epsilon f$ et en faisant tendre ϵ vers 0 on obtient l'inégalité de Poincaré :

$$\mathrm{Var}_{\gamma_n}(f) \leq \int_{\mathbb{R}^n} |\nabla f|^2 \, d\gamma_n.$$

L'inégalité de log-Sobolev peut donc se voir comme un renforcement de l'inégalité de Poincaré. Elle a énormément de conséquences, à commencer par le phénomène de concentration suivant. Soit f une fonction 1-Lipschitzienne et soit X un vecteur gaussien standard. Alors $f(X)$ se concentre autour de sa moyenne. Plus précisément on a

$$\mathsf{P}\left(|f(X) - \mathsf{E}[f(X)]| \geq t\right) \leq 2\,\mathrm{e}^{-t^2/2}, \quad \forall t \geq 0.$$

On obtient cette inégalité en appliquant l'inégalité de Sobolev logarithmique à la mesure de probabilité dont la densité est proportionnelle à $\mathrm{e}^{\lambda f}$, puis en utilisant l'inégalité de Markov et en optimisant en λ. C'est une propriété extrêmement importante, depuis son introduction par V. Milman dans sa célèbre démonstration du théorème de Dvoretzky, le concept de concentration de la mesure est devenu un sujet de recherche à part entière. Le livre de Ledoux [42] offre une excellente vue d'ensemble de ce domaine. Ceci étant dit, le phénomène de concentration de la mesure n'est

pas la seule conséquence de l'inégalité de log-Sobolev, qui admet bien d'autres aspects. En effet, considérons l'équation différentielle stochastique suivante

$$dX_t = \sqrt{2}\, dB_t - X_t\, dt,$$

où (B_t) est un mouvement Brownien standard sur $\mathbb{R}^n$. La solution (X_t) est un processus de Markov appelé processus d'Ornstein-Uhlenbeck. On note (P_t) le semigroupe associé, défini par $P_t f(x) = \mathsf{E}_x[f(X_t)]$ pour toute fonction test f. Comme d'habitude, l'indice x à côté de l'espérance désigne le point de départ du processus (X_t). Il est facile de voir que la mesure gaussienne est stationnaire et que le processus est ergodique, au sens où X_t converge vers γ_n en loi quand t tend vers $+\infty$. Par ailleurs, en dérivant l'entropie relative le long du semigroupe (P_t) on montre aisément que l'inégalité de log-Sobolev est complètement équivalente à la propriété suivante. Pour toute mesure de probabilité μ on a

$$\mathrm{H}(\mu P_t \mid \gamma_n) \leq \mathrm{e}^{-2t}\, \mathrm{H}(\mu \mid \gamma_n), \quad \forall t \geq 0,$$

où μP_t est la loi au temps t du processus d'Ornstein-Uhlenbeck ayant loi μ au temps 0. Par conséquent, l'inégalité de log-Sobolev montre que le processus converge en entropie vers γ_n exponentiellement vite. Une autre conséquence de log-Sobolev s'exprime de la manière suivante. Soit $p > 1$ et soit $t > 0$, pour toute fonction $f \in L^p(\gamma_n)$, la fonction $P_t f$ appartient à $L^q(\gamma_n)$ où $q = 1 + \mathrm{e}^{2t}(p-1)$, remarquer que $q > p$. L'opérateur P_t est même une contraction de L^p dans L^q, on dit que le semigroupe (P_t) est *hypercontractif*. En utilisant la dualité pour les espaces L^p, on peut reformuler cette propriété de la manière suivante. Soient $\rho, \alpha, \beta \in [0,1]$ vérifiant l'inégalité

$$\rho^2 \leq (\alpha^{-1} - 1)(\beta^{-1} - 1)$$

et soit (X, Y) un vecteur Gaussien sur $\mathbb{R}^n \times \mathbb{R}^n$ de moyenne nulle et de matrice de covariance

$$\begin{pmatrix} I_n & \rho\, I_n \\ \rho I_n & I_n \end{pmatrix},$$

où I_n désigne la matrice identité de $\mathbb{R}^n$. Alors pour toutes fonctions positives f et g on a

$$\mathsf{E}[f(X)^\alpha g(Y)^\beta] \leq \mathsf{E}[f(X)]^\alpha \mathsf{E}[g(Y)]^\beta.$$

Donnons maintenant une démonstration simple de ce résultat basée sur le calcul stochastique. Cet argument est dû à Neveu [55]. Soit (X_t, Y_t) un mouvement Brownien à valeurs dans $\mathbb{R}^{2n}$, partant de 0 et de variation quadratique donnée par

$$[(X,Y)]_t = t \begin{pmatrix} I_n & \rho\, I_n \\ \rho I_n & I_n \end{pmatrix}$$

pour tout t. On considère ensuite les martingales (M_t) et (N_t) données par

$$M_t = \mathsf{E}[f(X_1) \mid \mathcal{F}_t], \quad N_t = \mathsf{E}[g(Y_1) \mid \mathcal{F}_t],$$

$(\mathcal{F}_t)$ désignant la filtration naturelle du processus (X_t, Y_t). Comme les martingales Browniennes peuvent se représenter comme des intégrales stochastiques, il existe deux processus adaptés (u_t) et (v_t) tels que

$$dM_t = \langle u_t, dX_t \rangle, \quad dN_t = \langle v_t, dY_t \rangle.$$

En utilisant la structure de covariance du processus (X_t, Y_t) et la formule d'Itô, on montre facilement que

$$\begin{aligned} d\left(M_t^\alpha N_t^\alpha\right) &= M_t^\alpha N_t^\alpha \left(\alpha \langle \widetilde{u}_t, dX_t\rangle + \beta \langle \widetilde{v}_t, dY_t\rangle\right) \\ &\quad + \frac{1}{2} M_t^\alpha N_t^\alpha \left(\alpha(\alpha-1)\,|\widetilde{u}_t|^2 + \beta(\beta-1)\,|\widetilde{v}_t|^2 + 2\alpha\beta\rho\,\langle \widetilde{u}_t, \widetilde{v}_t\rangle\right)\, dt, \end{aligned}$$

où $\widetilde{u}_t = u_t/M_t$ et $\widetilde{v}_t = v_t/N_t$. Les hypothèses faites sur α, β, ρ impliquent que la matrice

$$\begin{pmatrix} \alpha(\alpha-1) & \alpha\beta\rho \\ \alpha\beta\rho & \beta(\beta-1) \end{pmatrix}$$

est négative. Par conséquent la partie absolument continue de l'équation précédente est négative. Le processus $(M_t^\alpha N_t^\beta)$ est donc une sur-martingale. En particulier

$$\mathsf{E}[M_1^\alpha N_1^\beta] \leq M_0^\alpha N_0^\beta,$$

ce qui est l'inégalité cherchée. On sait depuis l'article fondateur de Gross [34] que l'hypercontractivité est en fait équivalente à l'inégalité de log-Sobolev. L'argument précédent fournit donc une démonstration simple de cette dernière par calcul stochastique. On verra dans le premier chapitre du mémoire une autre démonstration de log-Sobolev par calcul stochastique.

Comme le titre du mémoire le laisse entendre la convexité joue aussi un rôle prépondérant tout au long de ce travail. Illustrons ceci par l'inégalité de Prékopa-Leindler, qui s'écrit de la manière suivante. Soit $\lambda \in [0,1]$ et soient f, g, h trois fonctions positives vérifiant

$$f(x)^{1-\lambda} g(y)^\lambda \leq h\left((1-\lambda)x + \lambda y\right)$$

pour tous $x, y \in \mathbb{R}^n$. Alors

$$\left(\int_{\mathbb{R}^n} f(x)\,dx\right)^{1-\lambda} \left(\int_{\mathbb{R}^n} g(y)\,dy\right)^{\lambda} \leq \int_{\mathbb{R}^n} h(x)\,dx.$$

Si on applique ceci à des indicatrices on obtient tout de suite la version multiplicative de l'inégalité de Brunn-Minkowski : pour tous sous-ensembles compacts (disons) A, B de $\mathbb{R}^n$ on a

$$|A|^{1-\lambda}\,|B|^\lambda \leq |(1-\lambda)A + \lambda B|, \tag{1}$$

où $|\cdot|$ désigne la mesure de Lebesgue et

$$(1-\lambda)A + \lambda B = \{(1-\lambda)x + \lambda y,\ x \in A,\ y \in B\}$$

est la combinaison de Minkowski des ensembles A et B. Plus généralement, l'inégalité de Prékopa-Leindler implique que toute mesure sur $\mathbb{R}^n$ de densité log-concave par rapport à la mesure de Lebesgue est log-concave, au sens où elle vérifie (1). Par exemple la mesure gaussienne est log-concave, de même que la mesure uniforme sur un ensemble convexe. Dans le premier chapitre on donne un argument stochastique qui permet de réduire l'inégalité de Prékopa-Leindler à la

convexité de la norme euclidienne. La question des propriétés de concentration des mesures log-concaves est un problème important encore très largement ouvert. Rappelons que la constante de Poincaré d'un vecteur aléatoire X la meilleure constante dans l'inégalité

$$\mathrm{Var}(f(X)) \leq C\,\mathsf{E}[|\nabla f(X)|^2].$$

Une conjecture de Kannan, Lovasz et Simonovits affirme que si X est log-concave alors, à un facteur universel près, la constante de Poincaré de X s'obtient en testant l'inégalité sur des fonctions f linéaires. Pour comprendre le rôle de la log-concavité dans cette conjecture, considérons le cas où le vecteur X est uniforme sur un ensemble A. Il est bien connu que si l'ensemble A possède un goulet d'étranglement alors le vecteur X aura une constante de Poincaré très grande. En effet, supposons par exemple que A est constitué de deux boules de rayon 1 disjointes reliée par un pont très fin, et soit f une fonction valant 0 et 1 sur chacune des deux boules et qui soit disons affine sur le pont. Alors $\mathrm{Var}(f(X))$ est de l'ordre de 1, tandis que $\mathsf{E}[|\nabla f(X)|^2]$ est négligeable, ce qui montre que la constante de Poincaré de X explose. De manière informelle la conjecture de Kannan, Lovasz et Simonovits affirme que c'est l'unique façon pour le vecteur X d'avoir une constante de Poincaré très grande. Le rôle de la log-concavité, qui dans ce cas se ramène à la convexité de l'ensemble A, est donc de proscrire les goulets d'étranglement.

Résumons maintenant brièvement le contenu de chaque chapitre.

Le **Chapitre 1**, qui est basé sur les articles [44], [45] et [48], tourne autour d'une formule stochastique due à Borell et qui s'énonce ainsi. Étant donné un mouvement Brownien standard sur $\mathbb{R}^n$ et une fonction f on a

$$\log\left(\int_{\mathbb{R}^n} \mathrm{e}^f \, d\gamma_n\right) = \sup\left\{\mathsf{E}\left[f\left(B_1 + \int_0^1 u_s\,ds\right) - \frac{1}{2}\int_0^1 |u_s|^2\,ds\right]\right\}$$

le supremum étant pris sur tous les processus adaptés (u_t). En réalité cette formule est un cas particulier d'un résultat antérieur de Boué et Dupuis qui autorise la fonction f à dépendre de toute la trajectoire et pas seulement du point final. Ceci étant dit le réel apport de Borell est une preuve très simple de l'inégalité de Prékopa-Leindler basée sur sa formule, ce qui montre la puissance de cet outil pour démontrer des inégalités fonctionnelles. Le résultat principal du chapitre 1 est une version duale de cette formule qui s'exprime de la manière suivante.

Théorème. *Soit μ une mesure de probabilité sur $\mathbb{R}^n$. On a*

$$\mathrm{H}(\mu \mid \gamma_n) = \frac{1}{2}\inf\left\{\mathsf{E}\left[\int_0^1 |u_s|^2\,ds\right]\right\},$$

l'infimum étant pris sur tous les processus adaptés (u_t) vérifiant la contrainte $B_1 + \int_0^1 u_s\,ds = \mu$ en loi.

En d'autres termes, l'énergie minimale nécessaire pour contraindre (de manière adaptée) le mouvement Brownien à avoir une loi donnée au temps 1 coïncide avec l'entropie relative de cette loi. On montre aussi que dans les bons cas il existe une dérive (u_t) optimale qui est par ailleurs solution d'une équation différentielle stochastique faisant intervenir la densité de μ. On obtient également une formule analogue pour les lois de trajectoires, la mesure gaussienne γ_n étant

remplacée par la mesure de Wiener. L'ingrédient essentiel de la démonstration est le théorème de Girsanov. De manière informelle, celui-ci affirme que les processus (X_t) de la forme $X_t = B_t + \int_0^t u_s\,ds$ pour un certain processus adapté (u_s) sont précisément ceux dont la loi est absolument continue par rapport à la mesure de Wiener et donne une formule pour leur densité. Suivant les traces de Borell, on donne un certain nombre d'applications de cette formule aux inégalités fonctionnelles. Prenons l'exemple de l'inégalité de Talagrand. Si μ et ν sont deux mesures de probabilités sur $\mathbb{R}^n$ on note $T_2(\mu,\nu)$ le coût de transport associé à la fonction de coût donné par la distance euclidienne au carré :

$$T_2(\mu,\nu) = \inf\left\{\mathsf{E}\left[|X-Y|^2\right]\right\}$$

l'infimum étant pris sur tous les couplages (X,Y) de (μ,ν). Soit μ une mesure de probabilité sur $\mathbb{R}^n$ et soit (u_t) un processus adapté vérifiant $B_1 + \int_0^1 u_s\,ds = \mu$ en loi. Comme $B_1 = \gamma_n$ en loi, on a par définition de T_2

$$T_2(\mu,\gamma_n) \le \mathsf{E}\left[\left|\int_0^1 u_s\,ds\right|^2\right] \le \mathsf{E}\left[\int_0^1 |u_s|^2\,ds\right].$$

En prenant maintenant l'infimum sur les (u_s) et en utilisant le théorème on obtient l'inégalité de Talagrand

$$T_2(\mu,\gamma_n) \le 2\,\mathrm{H}(\mu \mid \gamma_n).$$

On donne aussi une démonstration simple de l'inégalité de log-Sobolev ainsi que des inégalités de Brascamp-Lieb et de Brascamp-Lieb inverse. On établit ensuite des généralisations au cas où le processus de référence est une diffusion ou un mouvement Brownien sur une variété. On donne aussi quelques applications de ces généralisations. Le chapitre se conclut par une version de la formule de Borell pour un processus de Poisson qui est encore à l'état de conjecture.

Le **Chapitre 2** est basé sur l'article [47] et contient une extension du théorème d'hypercontractivité mentionné plus haut. Notons en effet que celui ci ne dit rien pour les fonctions qui sont seulement dans $L^1(\gamma_n)$, et on peut d'ailleurs montrer que si $p > 1$, l'opérateur d'Ornstein-Uhlenbeck P_t n'est pas borné de L^1 dans L^p. Notons que si f est une fonction positive vérifiant

$$\int_{\mathbb{R}^n} f\,d\gamma_n = 1,$$

alors $P_t f$ vérifie les mêmes propriétés. On a donc d'après l'inégalité Markov

$$\gamma_n\left(\{P_t f \ge r\}\right) \le \frac{1}{r},$$

pour tout $r \ge 1$. Le résultat principal du chapitre 2 s'énonce ainsi.

Théorème. *Soit f une fonction positive d'intégrale 1. $\mathbb{R}^n$ On a pour tout $t > 0$ et pour tout $r > 1$*

$$\gamma_n\left(\{P_t f > r\}\right) \le C\,\frac{\max(1,t^{-1})}{r\sqrt{\log r}},$$

où C est un constante universelle.

De manière informelle, le théorème affirme que l'opérateur P_t envoie L_1 dans un espace $L_1\sqrt{\log L_1}$ faible. En testant l'inégalité sur des fonctions f dont le logarithme est affine on peut de plus montrer que la dépendance en $1/\sqrt{\log r}$ est optimale. Un tel phénomène avait été conjecturé par Talagrand dans un contexte légèrement différent, la mesure γ_n étant remplacée par la mesure uniforme sur le cube discret $\{-1,1\}^n$ et le semigroupe d'Ornstein-Uhlenbeck par le semigroupe de la marche aléatoire sur le cube. Le résultat précédent est en fait essentiellement dû à Eldan et Lee. L'apport de l'article [47] est surtout d'avoir simplifié leur argument mais aussi d'avoir un tout petit peu amélioré leur résultat, lequel contenait un facteur $(\log\log r)^4$ superflu. Dans les deux articles, l'ingrédient principal de la démonstration est la formule de représentation de l'entropie énoncée plus haut.

Le **Chapitre 3** est basé sur l'article [14], qui est un travail en commun avec S. Bubeck et R. Eldan. Dans ce chapitre on se donne un corps convexe K de $\mathbb{R}^n$ et on étudie un algorithme permettant de simuler la mesure uniforme sur K. C'est une question de nature pratique qui a beaucoup d'applications en informatique théorique, en particulier en apprentissage, même si ces applications ne sont pas abordées dans ce mémoire. Pour que le problème ait une chance d'être soluble il faut faire une hypothsèe de normalisation. On suppose donc que le convexe K contient la boule euclidienne de rayon 1 et est contenu dans la boule de rayon R, où R est un paramètre plus grand que 1. Dans ces conditions, l'algorithme de rejet suivant produit un point uniforme sur K.

- Produire X uniforme sur RB_2^n ;
- Si $X \in K$, retourner X ;
- Sinon recommencer.

Le problème est que le temps moyen de calcul de cet algorithme vaut $|RB_2^n|/|K|$ ce qui est typiquement exponentiel en la dimension. Or dans la plupart des applications la dimension est gigantesque et donc cet algorithme ne fonctionne pas en pratique. L'objectif est donc de produire un algorithme qui soit polynomial en la dimension. On se donne un paramètre $\eta > 0$ et une suite i.i.d. $\xi_1, \xi_2, \dots$ de gaussiennes standard dans $\mathbb{R}^n$. L'algorithme proposé dans ce chapitre est la chaine de Markov suivante, qu'on appellera Monte Carlo projeté par la suite. On fixe $X_0 = 0$ arbitrairement et on pose pour tout $k \geq 0$

$$X_{k+1} = \mathcal{P}_K\left(X_k + \sqrt{\eta}\,\xi_{k+1}\right),$$

où $\mathcal{P}_K$ désigne la projection euclidienne sur K. Autrement dit $\mathcal{P}_K(x)$ est le point de K qui minimise la distance à x. Une étape de l'algorithme consiste donc à faire un pas gaussien et à projeter sur K. Le résultat principal s'énonce ainsi

Théorème. *Soit $\epsilon > 0$. Si le paramètre η et le nombre d'étapes N vérifient*

$$\eta \approx \frac{R^2}{N}, \quad N \approx \frac{R^6 n^7}{\epsilon^8}$$

alors $\mathrm{TV}(X_N, \mu) \leq \epsilon$*, où μ désigne la mesure uniforme sur K, et* TV *la distance en variation totale.*

Le symbole $\approx$ cache des constantes universelles ainsi de possibles dépendances logarithmiques. Quoiqu'il en soit, ce que le théorème dit réellement est que le Monte Carlo projeté permet d'approcher la mesure uniforme sur K en un temps polynomial en tous les paramètres du problème, y compris la dimension. Notons que la question du calcul de la projection $\mathcal{P}_K$ n'est pas abordée ici. Disons toutefois qu'il s'agit de minimiser une distance euclidienne sur un convexe ce qui est en général faisable.

Ce résultat s'inscrit dans une longue lignée de travaux. Le premier algorithme polynomial pour simuler la mesure uniforme sur un convexe est dû à Dyer, Frieze et Kannan [22]. Essentiellement leur méthode consistait à approximer le convexe par son intersection avec un réseau et à réaliser une marche aléatoire sur ce graphe. Le résultat le plus précis à ce jour, dû à Lovasz et Vempala [49] utilise une marche appelée *hit-and-run*. Lovasz et Vempala montrent qu'essentiellement n^4 étapes de cette marche suffisent à approcher la mesure uniforme sur K. Du coup notre résultat est nettement moins bon que le leur. Ceci étant dit, notre algorithme est la discrétisation naturelle du mouvement Brownien réfléchi à la frontière de K, et le simple fait de savoir qu'il converge en temps polynomial, ce qui semble-t-il n'était pas connu jusque là, nous parait intéressant en soit. Par ailleurs, il est probable que la borne obtenue soit assez loin de la vérité et donc possible que le Monte Carlo projeté soit en fait aussi efficace que le *hit-and-run*.

L'algorithme et la démonstration du résultat s'inspirent de l'article [20] dans lequel Dalalyan cherche à simuler une variable dont la densité est proportionnelle $\mathrm{e}^{-V(x)}$ où V est un potentiel strictement convexe et régulier. Il considère la chaine de Markov donnée par

$$X_{k+1} = X_k - \frac{\eta}{2}\nabla V(X_k) + \sqrt{\eta}\,\xi_{k+1},$$

et il observe que cette chaine est la discrétisation naturelle du processus de Langevin, solution de

$$dY_t = dB_t - \frac{1}{2}\nabla V(Y_t)\,dt,$$

où (B_t) est un mouvement Brownien standard. L'analyse de la convergence de l'algorithme consiste alors à estimer d'une part la vitesse à laquelle (Y_t) converge vers sa mesure stationnaire et d'autre part ce qui est perdu dans la discrétisation. De manière analogue, l'algorithme de Monte Carlo projeté décrit plus haut est la discrétisation du mouvement Brownien réfléchi à la frontière de K tel que le définit Tanaka dans [63]. La démonstration suit alors le même plan que celle de Dalalyan même si les deux étapes sont compliquées par la réflexion à la frontière de K, qui est un processus très singulier. Pour conclure disons que la méthode permet en fait d'étudier un mélange de ces deux situations. On montre en effet que pour V convexe et régulier, l'algorithme

$$X_{k+1} = \mathcal{P}_K\left(X_k - \frac{\eta}{2}\nabla V(X_k) + \sqrt{\eta}\,\xi_{k+1}\right),$$

permet d'approcher la mesure de probabilité dont la densité est proportionnelle à $\mathrm{e}^{-V}1_K$ en temps polynomial.

Le **Chapitre 4** est basé sur l'article [26] qui est un travail en commun avec Eldan. On y décrit d'abord un processus stochastique à valeur mesure introduit par Eldan et qu'il appelle *localisation stochastique*. Une version simplifiée de ce processus consiste, étant donnée une mesure de probabilité μ sur $\mathbb{R}^n$, à regarder la fonction aléatoire

$$\rho_t(x) = \exp\left(\langle x, B_t\rangle - \frac{t}{2}|x|^2\right)$$

où (B_t) est un Brownien standard et à appeler μ_t la mesure aléatoire ayant pour densité ρ_t par rapport à la mesure μ. Noter que pour x fixé, le processus $(\rho_t(x))_{t\geq 0}$ est une martingale. En particulier $\mathsf{E}[\rho_t(x)] = 1$ pour tout t ce qui montre que la mesure μ_t est en moyenne égale à μ. D'un autre côté, à cause du facteur $\mathrm{e}^{-t|x|^2/2}$ la mesure μ_t se concentre de plus en plus à mesure que t augmente, d'où le nom de localisation stochastique. Le processus d'Eldan est nettement plus compliqué que ce processus simplifié mais l'idée reste la même.

Rappelons qu'un vecteur aléatoire X est dit isotrope s'il vérifie $\mathsf{E}[X] = 0$ et $\mathsf{E}[XX^T] = I_n$. Posons

$$D_n = \sup\left\{\mathrm{Var}(|X|);\ X \text{ log-concave et isotrope sur } \mathbb{R}^n\right\}.$$

La conjecture de la variance affirme que la suite (D_n) est bornée. Il est évident que $D_n \leq n$ pour tout n et l'estimation la plus précise à ce jour, due à Guédon et E. Milman [35] s'écrit $D_n = O(n^{2/3})$. La conjecture de la variance consiste essentiellement à restreindre la conjecture de Kannan, Lovasz et Simonovits dont on a parlé plus haut à la seule fonction $f(x) = |x|$. Elle est donc a priori beaucoup plus faible. En utilisant le processus de localisation stochastique, Eldan a montré que les deux conjectures étaient en fait essentiellement équivalentes. Le résultat principal de ce chapitre s'écrit ainsi.

Théorème. *Soit X un vecteur isotrope et log-concave de $\mathbb{R}^n$, soit G un vecteur gaussien standard sur $\mathbb{R}^n$ et soit $\|\cdot\|$ une norme. On a*

$$\mathsf{E}[\|X\|] \lesssim \sqrt{D_n}\,\mathsf{E}[\|G\|].$$

Notons qu'en utilisant des techniques complètement différente, Bourgain à montré que

$$\mathsf{E}[\|X\|] \lesssim n^{1/4}\mathsf{E}[\|G\|],$$

ce qui est mieux que ce qu'on obtient en combinant notre résultat avec l'estimation de Guédon-Milman $D_n = O(n^{2/3})$. Notre résultat n'est donc intéressant que conditionnellement à la conjecture de la variance, ou du moins à une amélioration conséquente de la borne.

La démonstration utilise la localisation stochastique. Plus précisément, soit μ la loi du vecteur X et (μ_t) le processus de localisation stochastique partant de μ. On note ensuite a_t le barycentre de μ_t. On montre alors que le processus (a_t) est une martingale qui converge vers μ en loi. En utilisant le travail d'Eldan, On montre aussi que sa variation quadratique est contrôlée par la constante D_n. L'inégalité s'obtient ensuite en utilisant une inégalité de corrélation due à Maurey pour de telles martingales. La méthode permet d'ailleurs d'attraper aussi le théorème suivant, dû à Hargé.

Théorème. *Soit μ une mesure de probabilité ayant son barycentre en 0 et possédant une densité log-concave par rapport à la mesure gaussienne standard γ_n. Alors pour toute fonction ϕ convexe on a*

$$\int_{\mathbb{R}^n} \phi\, d\mu \leq \int_{\mathbb{R}^n} \phi\, d\gamma_n,$$

Le **Chapitre 4** est un peu à part, au sens où il a peu de rapport avec les inégalités fonctionnelles. Dans ce chapitre on se donne une chaine de Markov (X_n) sur un espace d'état fini M et

on s'intéresse au temps moyen que la chaine met à visiter tous les points de M. Plus précisément, pour tout $x \in M$ on appelle $T_x = \inf\{n \geq 0 \colon X_n = x\}$ le temps d'atteinte de x et on note

$$\mathrm{cov}(M) = \sup_{x \in M} \left\{ \mathsf{E}_x \left[\sup_{y \in M} \{T_y\} \right] \right\}.$$

Cette quantité est appelée temps de recouvrement de M. L'objectif est d'estimer $\mathrm{cov}(M)$ en fonction des propriétés de l'espace métrique (M, d) où d est la distance donnée par

$$d(x, y) = \mathsf{E}_x[T(y)] + \mathsf{E}_y[T(x)].$$

C'est le temps moyen qu'il faut, partant de x, pour toucher y et revenir en x. L'origine de cette question vient d'une analogie avec les processus gaussiens. On sait en effet que si (X_t) est un processus Gaussien, l'étude de la quantité $\mathsf{E}\left[\sup_{t \in T}\{X_t\}\right]$ se ramène à celle de l'espace métrique (T, d_X) où d_X est la distance donnée par

$$d_X(s, t)^2 = \mathsf{E}[(X_s - X_t)^2].$$

Un résultat fondamental de Talagrand affirme en effet que

$$\mathsf{E}\left[\sup_{t \in T} \{X_t\} \right] \approx \gamma_2(T, d_X),$$

où $\gamma_2(T, d_X)$ est une quantité qu'on ne définira pas précisément ici, mais qui ne dépend que de la distance d_X et qui mesure en un sens la complexité de l'espace métrique (T, d_X). L'ensemble des techniques développées par Talagrand pour démontrer ce résultat est appelé *chainage générique*. La question analogue pour les chaines de Markov a été résolue récemment par Ding, Lee et Peres [21] qui ont montré que pour toute chaine réversible on a

$$\mathrm{cov}(M) \approx \gamma_2(T, \sqrt{d})^2.$$

Leur démonstration utilise le théorème d'isomorphisme de Ray-Knight pour ramener le problème à l'étude du champs libre gaussien associé à la chaine. L'argument est donc très sophistiqué et peu naturel. L'objet de l'article [46] est de donner une démonstration directe de ce résultat, en appliquant directement les techniques de chainage à l'espace métrique (M, d). Malheureusement nous ne sommes pas parvenu à retrouver complètement le résultat de Ding, Lee et Peres. En effet, si nous obtenons bien la borne supérieure $\mathrm{cov}(M) \lesssim \gamma_2(M, \sqrt{d})^2$, qui est la plus facile, nous ne retrouvons la borne inférieure qu'à un facteur $\log\log \mathrm{card}(M)$ près.

Le **Chapitre 6** est basé sur l'article [15], qui est un travail en commun avec U. Caglar, M. Fradelizi, O. Guédon, C. Schütt, et E. Werner. Le point de départ de ce chapitre est la version euclidienne de l'inégalité de Sobolev logarithmique. Soit μ est une mesure de probabilité sur $\mathbb{R}^n$ possédant une densité ρ par rapport à la mesure de Lebesgue. On note

$$\mathrm{S}(\mu) = -\int_{\mathbb{R}^n} \log \rho \, d\mu \quad \text{et} \quad \mathrm{J}(\mu) = \int_{\mathbb{R}^n} |\nabla \log \rho|^2 \, d\mu$$

son entropie de Shannon et son information de Fisher, respectivement. Remarquer que la convention de signe diffère entre S et H. L'inégalité de log-Sobolev euclidienne s'écrit alors

$$\mathrm{S}(\gamma_n) - \mathrm{S}(\mu) \leq \frac{n}{2} \log\left(\frac{\mathrm{J}(\mu)}{n}\right).$$

Cette inégalité s'obtient à partir de l'inégalité de log-Sobolev gaussienne en faisant des dilatations et en optimisant sur le coefficient de dilatation. Dans ce chapitre on établit une minoration du saut entropique $\mathrm{S}(\gamma_n) - \mathrm{S}(\mu)$ pour les mesures log-concaves. Plus précisément on démontre le théorème suivant.

Théorème. *Soit μ une mesure de probabilité log-concave sur $\mathbb{R}^n$ possédant une densité ρ par rapport à la mesure de Lebesgue. On a*

$$\frac{1}{2} \int_{\mathbb{R}^n} \log\det(\nabla^2 \phi)\, d\mu \leq \mathrm{S}(\gamma_n) - \mathrm{S}(\mu),$$

où $\phi = -\log\rho$ est le potentiel de μ.

Remarquons qu'une intégration par partie montre que $\mathrm{J}(\mu) = \int_{\mathbb{R}^n} \Delta\phi\, d\mu$. Par suite notre minoration du saut entropique s'obtient à partir de la majoration en passant le logarithme sous le signe intégrale et en remplaçant la moyenne arithmétique des valeurs propres de $\nabla^2\phi$ par leur moyenne géométrique. Ce théorème est en fait dû à Artstein, Klartag, Schütt et Werner, mais leur argument était très compliqué. L'apport de l'article [15] est d'avoir remarqué que ce théorème se déduit facilement de l'inégalité fonctionnelle de Santaló : si ϕ une fonction convexe vérifiant $\int_{\mathbb{R}^n} x\, \mathrm{e}^{-\phi(x)}\, dx = 0$ alors

$$\int_{\mathbb{R}^n} \mathrm{e}^{-\phi(x)}\, dx \int_{\mathbb{R}^n} \mathrm{e}^{-\phi^*(x)}\, dx \leq \left(\int_{\mathbb{R}^n} \mathrm{e}^{-|x|^2/2}\, dx\right)^2 = (2\pi)^n,$$

où ϕ^* la transformée de Legendre de ϕ, donnée par

$$\phi^*(y) = \sup_{x \in \mathbb{R}^n} \left\{\langle x, y\rangle - \phi(x)\right\}.$$

Donc si μ est log-concave et centrée, et si ϕ est le potentiel de μ, l'inégalité de Santaló fonctionnelle montre que

$$\log\left(\int_{\mathbb{R}^n} \mathrm{e}^{-\phi^*(y)}\, dy\right) \leq \frac{n}{2} \log(2\pi).$$

Ensuite on fait le changement de variable $y = \nabla\phi(x)$ dans l'intégrale précédente, et le résultat s'obtient au bout de quelques lignes de calcul relativement élémentaires.

Introduction

This memoir deals with functional inequalities, which is a fancy way of saying inequalities involving integrals, and their interplay with stochastic processes. The interaction goes both ways, we sometimes infer properties of a given stochastic process from a functional inequality, and we sometimes use stochastic calculus to prove inequalities. Let us start with an example.

Throughout these notes γ_n denotes the standard Gaussian measure in dimension n. The logarithmic Sobolev inequality for the Gaussian measure plays a central part in this memoir. It asserts that every probability measure μ on $\mathbb{R}^n$ having density ρ with respect to γ_n satisfies

$$\mathrm{H}(\mu \mid \gamma_n) \leq \frac{1}{2}\,\mathrm{I}(\mu \mid \gamma_n)$$

where $\mathrm{H}(\mu \mid \gamma_n) = \int_{\mathbb{R}^n} \log \rho \, d\mu$ is the relative entropy of μ and

$$\mathrm{I}(\mu \mid \gamma_n) = \int_{\mathbb{R}^n} |\nabla \log \rho|^2 \, d\mu$$

is its Fisher information. Applying this to $\rho = 1 + \epsilon f$ and letting ϵ go to 0 yields the more familiar Poincaré inequality:

$$\mathrm{Var}_{\gamma_n}(f) \leq \int_{\mathbb{R}^n} |\nabla f|^2 \, d\gamma_n.$$

So the log-Sobolev inequality can be seen as a reinforcement of the Poincaré inequality. It has many consequences, including the following concentration property. If f is a 1-Lipschitz function and X is a standard Gaussian vector then $f(X)$ is close to its mean with very high probability. More precisely we have

$$\mathsf{P}\left(|f(X) - \mathsf{E}[f(X)]| \geq t\right) \leq 2\,\mathrm{e}^{-t^2/2}, \quad \forall t \geq 0.$$

This is obtained by applying the log-Sobolev inequality to the probability measure whose density is proportional to $\mathrm{e}^{\lambda f}$, using Markov's inequality and optimizing in λ. This is certainly an important inequality, since its introduction by V. Milman in his famous proof of Dvoretzky's theorem in 1971, the concept of concentration of measure has become a field of research in itself, see for instance Ledoux's book [42] for an excellent account on this topic. Nevertheless, concentration is only one part of the story, and the log-Sobolev inequality has many other counterparts. To describe these, we need to introduce the stochastic differential equation

$$dX_t = \sqrt{2}\, dB_t - X_t \, dt,$$

where (B_t) is a standard Brownian motion in $\mathbb{R}^n$. The solution (X_t) is called the Ornstein-Uhlenbeck process. This is a Markov process and we let (P_t) be the associated semigroup, defined by $P_t f(x) = \mathsf{E}_x[f(X_t)]$ for every test function f. As usual the subscript x denotes the starting point of (X_t). It is easily seen that the Gaussian measure γ_n is stationary and ergodic, in the sense that X_t converges to γ_n in law as t tends to $+\infty$, for every initial distribution. Standard techniques show that the log-Sobolev inequality is equivalent to the following property. For every distribution μ we have

$$\mathrm{H}(\mu P_t \mid \gamma_n) \leq \mathrm{e}^{-2t}\, \mathrm{H}(\mu \mid \gamma_n), \quad \forall t \geq 0$$

where μP_t denotes the law of X_t when X_0 has law μ. Therefore, the log-Sobolev inequality implies that the Ornstein-Uhlenbeck process converges in entropy towards γ_n exponentially fast. It possesses yet another important consequence, discovered by Gross, and which reads as follows. If a function f belongs to $L^p(\gamma_n)$ for some $p > 1$, then $P_t f$ belongs to $L^q(\gamma_n)$ for some $q > p$, namely for $q = 1 + \mathrm{e}^{2t}(p-1)$. Moreover the operator P_t is actually a contraction from L^p to L^q. This property is called *hypercontractivity*. Using duality, this can be reformulated as follows. Let $\rho, \alpha, \beta \in [0,1]$ be such that

$$\rho^2 \leq (\alpha^{-1} - 1)(\beta^{-1} - 1)$$

and let (X, Y) be a Gaussian vector on $\mathbb{R}^n \times \mathbb{R}^n$ centered at 0 and having covariance matrix

$$\begin{pmatrix} I_n & \rho\, I_n \\ \rho I_n & I_n \end{pmatrix},$$

where I_n denotes the $n \times n$ identity matrix. Then for every non-negative functions f, g we have

$$\mathsf{E}[f(X)^\alpha g(Y)^\beta] \leq \mathsf{E}[f(X)]^\alpha \mathsf{E}[g(Y)]^\beta.$$

Let us give a short proof of this fact based on stochastic calculus. This argument is due to Neveu, see [55]. Let (X_t, Y_t) be a Brownian motion taking values in $\mathbb{R}^{2n}$, starting from 0 and having covariation given by

$$[(X,Y)]_t = t \begin{pmatrix} I_n & \rho\, I_n \\ \rho I_n & I_n \end{pmatrix}$$

for every t. Now consider the martingales (M_t) and (N_t) given by

$$M_t = \mathsf{E}[f(X_1) \mid \mathcal{F}_t], \quad N_t = \mathsf{E}[g(Y_1) \mid \mathcal{F}_t]$$

where $(\mathcal{F}_t)$ is the natural filtration of the process (X_t, Y_t). Brownian martingales can be represented as stochastic integrals, so there exist two adapted processes (u_t) and (v_t) such that

$$dM_t = \langle u_t, dX_t \rangle, \quad dN_t = \langle v_t, dY_t \rangle.$$

From the covariation structure of the Brownian motion (X_t, Y_t) and using Itô's formula we easily get

$$\begin{aligned} d\left(M_t^\alpha N_t^\alpha\right) &= M_t^\alpha N_t^\alpha \left(\alpha \langle \widetilde{u}_t, dX_t \rangle + \beta \langle \widetilde{v}_t, dY_t \rangle\right) \\ &\quad + \frac{1}{2} M_t^\alpha N_t^\alpha \left(\alpha(\alpha-1)\,|\widetilde{u}_t|^2 + \beta(\beta-1)\,|\widetilde{v}_t|^2 + 2\alpha\beta\rho\, \langle \widetilde{u}_t, \widetilde{v}_t \rangle\right) dt, \end{aligned}$$

where $\widetilde{u}_t = u_t/M_t$ and $\widetilde{v}_t = v_t/N_t$. Now the hypothesis made on α, β, ρ insure that the matrix

$$\begin{pmatrix} \alpha(\alpha-1) & \alpha\beta\rho \\ \alpha\beta\rho & \beta(\beta-1) \end{pmatrix}$$

is negative. As a result the absolutely continuous part of the previous equation is non-positive. Therefore $(M_t^\alpha N_t^\beta)$ is a super-martingale, in particular

$$\mathsf{E}[M_1^\alpha N_1^\beta] \leq M_0^\alpha N_0^\beta,$$

which is the result. It can actually be shown that hypercontractivity is equivalent to the log-Sobolev inequality, so the above argument provides a simple proof of the log-Sobolev inequality based on stochastic calculus. We shall see another stochastic approach in the first chapter.

As the title suggests convexity also plays a prominent role throughout these notes. This is well illustrated by the case of the Prékopa-Leindler inequality, which we will encounter time and again, and which states as follows. Let $\lambda \in [0,1]$ and let f, g, h be non-negative functions satisfying

$$f(x)^{1-\lambda} g(y)^\lambda \leq h\left((1-\lambda)x + \lambda y\right)$$

for every $x, y \in \mathbb{R}^n$. Then

$$\left(\int_{\mathbb{R}^n} f(x)\,dx\right)^{1-\lambda} \left(\int_{\mathbb{R}^n} g(y)\,dy\right)^{\lambda} \leq \int_{\mathbb{R}^n} h(x)\,dx.$$

Applied to indicators of sets, this immediately gives the multiplicative form of the Brunn-Minkowski inequality: for every compact (say) subsets A, B of $\mathbb{R}^n$ we have

$$|A|^{1-\lambda}\,|B|^\lambda \leq |(1-\lambda)A + \lambda B|, \tag{2}$$

where $|\cdot|$ denotes the Lebesgue measure, and

$$(1-\lambda)A + \lambda B = \{(1-\lambda)x + \lambda y,\ x \in A,\ y \in B\}$$

is the Minkowski combination of the sets A and B. More generally, the Prékopa-Leindler inequality shows that any measure having a density whose logarithm is concave is *log-concave*, in the sense that it satisfies (2). For instance the Gaussian measure is log-concave, as well as the uniform measure on a convex set. In the first chapter we shall see a stochastic argument which reduces the Prékopa-leindler inequality to the convexity of the Euclidean norm. One of the main challenges in the field of functional inequalities is to understand the concentration properties of log-concave measures. Recall that the Poincaré constant of a random vector X is the best constant in the inequality

$$\mathrm{Var}(f(X)) \leq C\,\mathsf{E}[|\nabla f(X)|^2].$$

A conjecture of Kannan, Lovasz and Simonovits asserts that if X is log-concave, then, up to a universal factor, the Poincaré constant can be estimated by plugging in linear functions f. To understand the role of log-concavity in this conjecture, consider a random vector X being uniform on a set A. It is well known that if A has a bottleneck, then X has a bad Poincaré constant.

Indeed, think of A being the union of two disjoint Euclidean balls linked by a tiny bridge. If f is equal to 0 and 1 respectively on each ball, and, say, affine through the bridge, then $\mathrm{Var}(f(X))$ has order 1, whereas $\mathsf{E}[|\nabla f(X)|^2]$ is negligible, so the Poincaré constant is huge. Loosely speaking the Kannan-Lovasz-Simonovits conjecture asserts that this is the only way the Poincaré constant of X can be large. So the role of log-concavity, which in this case amounts to convexity of the set A, is to rule out bottlenecks.

Let us close this introduction with a short overview of the manuscript. In Chapter 1 we give a stochastic formula for the Gaussian relative entropy and we derive many functional inequalities from it. We also discuss extensions of the formula beyond the Gaussian case. In Chapter 2 we give an extension of the hypercontractivity property mentioned above. The proof is based on the stochastic formula of Chapter 1. In Chapter 3 we study an algorithm allowing to sample from a convex body. Functional inequalities and stochastic calculus are combined together to estimate the running time of the algorithm. Chapter 4 is devoted to a stochastic process invented by Ronen Eldan and called stochastic localization. We use it to give partial answers to some conjectures about log-concave measures such as the one mentioned above. Chapter 5 is slightly off topic, in this chapter we use Talagrand's generic chaining to estimate the cover time of a graph. Lastly, in Chapter 6 we show that a functional form of the famous Blaschke-Santaló inequality yields a reversed form of the log-Sobolev inequality for log-concave measures.

Chapter 1

Around Borell's formula

This chapter presents results obtained in the papers [44], [45] and [48].

1.1 Introduction

Let (B_t) be a standard n-dimensional Brownian motion and let γ_n be the standard Gaussian measure on $\mathbb{R}^n$. Borell's formula [10] asserts that for any function f bounded from below we have

$$\log\left(\int_{\mathbb{R}^n} \mathrm{e}^f \, d\gamma_n\right) = \sup\left\{\mathsf{E}\left[f\left(B_1 + \int_0^1 u_s \, ds\right) - \frac{1}{2}\int_0^1 |u_s|^2 \, ds\right]\right\}$$

where the supremum is taken on every progressively measurable process (u_t) taking values in $\mathbb{R}^n$ and satisfying

$$\int_0^1 |u_s|^2 \, ds < +\infty, \quad \text{almost surely.}$$

The term *progressively measurable* essentially means that for every time t, the random vector u_t only depends on the past values $\{B_s, s \leq t\}$ of the process B. The name *Borell's formula* is probably unfair to Boué and Dupuis, who proved in an earlier article [11] a more general formula, allowing the function f to depend on the whole trajectory of the process, rather than just the terminal point (see below for a precise formulation). Anyways, both articles agree that the formula already appears in Fleming and Soner's work on stochastic optimal control, although in a rather cryptic way, see [30, section VI.9]. In any case Borell should definitely be credited for bringing this formula into the context of functional inequalities. Before stating the more general Boué-Dupuis formula, let us specify the setting of this chapter and introduce a couple of notations. Let $\mathbb{W}$ be the space of continuous paths $w\colon [0,1] \to \mathbb{R}^n$ satisfying $w(0) = 0$, equipped with the topology of uniform convergence, let $\mathcal{B}$ be the associated Borel σ-field, and let γ be the Wiener measure. So γ is the law of a standard Brownian motion. The space $(\mathbb{W}, \mathcal{B}, \gamma)$ is called the *Wiener space*. We let $\mathbb{H}$ be the Cameron-Martin space, namely the space of absolutely continuous paths $u\colon [0,1] \to \mathbb{R}^n$ satisfying $u(0) = 0$ and

$$\|u\|_{\mathbb{H}}^2 = \int_0^1 |\dot{u}_s|^2 \, ds < +\infty.$$

The Cameron-Martin space $\mathbb{H}$ is a Hilbert space that embeds continuously in $\mathbb{W}$. We are also given a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ equipped with a filtration $(\mathcal{F}_t)$ and carrying a standard n-dimensional Brownian motion (B_t). Throughout this chapter, we call *drift* any process U which is progressively measurable and belongs to $\mathbb{H}$ almost surely. With this formalism Borell's formula can be restated as follows. For every $f\colon \mathbb{R}^n \to \mathbb{R}$, measurable and bounded from below we have

$$\log\left(\int_{\mathbb{R}^n} \mathrm{e}^f \, d\gamma_n\right) = \sup\left\{\mathsf{E}\left[f(B_1 + U_1) - \frac{1}{2}\|U\|_{\mathbb{H}}^2\right]\right\},$$

where the supremum is taken over all drifts U. Now the Boué-Dupuis formula reads as follows. For every functional $F\colon \mathbb{W} \to \mathbb{R}$ which is measurable and bounded from below we have

$$\log\left(\int_{\mathbb{W}} \mathrm{e}^F \, d\gamma\right) = \sup\left\{\mathsf{E}\left[F(B + U) - \frac{1}{2}\|U\|_{\mathbb{H}}^2\right]\right\},$$

where again the supremum is again taken on all drifts U. Of course Borell's formula is easily retrieved applying Boué-Dupuis to the function $F(w) = f(w(1))$.

1.2 Representation formula for the entropy

The setting of this section is the same as before. Let us mention that we are working with the Wiener space on the interval $[0, 1]$ for simplicity, but that everything we shall say remains valid on $[0, +\infty[$. In [44] we establish a dual version of the Boué-Dupuis formula that involves relative entropy. Recall its definition: if μ is a probability measure on $(\mathbb{W}, \mathcal{B})$, the relative entropy of μ with respect to γ is defined as

$$\mathrm{H}(\mu \mid \gamma) = \int_{\mathbb{W}} \log\left(\frac{d\mu}{d\gamma}\right) d\mu$$

if μ is absolutely continuous with respect to γ and $\mathrm{H}(\mu \mid \gamma) = +\infty$ otherwise. Here is the main result of [44].

Theorem 1.1. *Let μ be a probability measure on $\mathbb{W}$ absolutely continuous with respect to γ and whose density satisfies certain technical assumptions. Then*

$$\mathrm{H}(\mu \mid \gamma) = \inf\left\{\frac{1}{2}\mathsf{E}\left[\|U\|_{\mathbb{H}}^2\right]\right\}$$

where the infimum is taken on every drift U such that $B + U$ has law μ.

In the sequel we call *energy* of the drift U the quantity $\frac{1}{2}\mathsf{E}[\|U\|_{\mathbb{H}}^2]$. The theorem thus asserts that the minimal energy needed to constrain in a progressively measurable way the Brownian motion to have law μ is the relative entropy of μ. Let us not specify the *technical assumptions* for now and let us explain why the Boué-Dupuis formula follows immediately from this theorem. It is well known that there is a convex duality between relative entropy and the log-Laplace transform. Namely, for every function F we have

$$\log\left(\int_{\mathbb{W}} \mathrm{e}^F \, d\gamma\right) = \sup_{\mu}\left\{\int_{\mathbb{W}} F \, d\mu - \mathrm{H}(\mu \mid \gamma)\right\},$$

where the supremum is taken on every probability measure μ. Now applying the previous theorem we get

$$\begin{aligned}\log\left(\int_{\mathbb{W}} \mathrm{e}^F \, d\gamma\right) &= \sup_{\mu}\left\{\int_{\mathbb{W}} F\, d\mu - \mathrm{H}(\mu \mid \gamma)\right\} \\ &= \sup_{\mu}\left\{\int_{\mathbb{W}} F\, d\mu - \inf_{B+U\sim\mu}\left\{\frac{1}{2}\mathsf{E}\|U\|_{\mathbb{H}}^2\right\}\right\} \\ &= \sup_{\mu}\sup_{B+U\sim\mu}\left\{\int_{\mathbb{W}} F\, d\mu - \frac{1}{2}\mathsf{E}\|U\|_{\mathbb{H}}^2\right\} \\ &= \sup_{U}\left\{\mathsf{E}\left[F(B+U) - \frac{1}{2}\|U\|_{\mathbb{H}}^2\right]\right\}\end{aligned}$$

which is exactly the Boué-Dupuis formula.

The main ingredient of the proof of Theorem 1.1 is the Girsanov transform. If U is a drift, Itô's formula shows that the process

$$D_t = \exp\left(-\int_0^t \langle \dot{U}_s, dB_s\rangle - \frac{1}{2}\int_0^t |\dot{U}_s|^2\, ds\right)$$

is a non-negative local martingale. Under a suitable integrability condition, for instance when $\|U\|_{\mathbb{H}}$ is bounded, the process (D_t) becomes a genuine positive martingale of expectation 1 and we can define a new probability measure Q by setting $d\mathsf{Q} = D_1\, d\mathsf{P}$. Girsanov's formula then asserts that under Q the process $X = B + U$ is a standard Brownian motion. In other words X has law γ under Q. Letting μ be the law of X under P, we thus get

$$\mathrm{H}(\mu \mid \gamma) \le \mathrm{H}(\mathsf{P} \mid \mathsf{Q}).$$

On the other hand

$$\mathrm{H}(\mathsf{P} \mid \mathsf{Q}) = \mathsf{E}^{\mathsf{P}}[-\log(D_1)] = \frac{1}{2}\mathsf{E}^{\mathsf{P}}[\|U\|_{\mathbb{H}}^2]$$

since the stochastic integral $\int \langle \dot{U}_s, dB_s\rangle$ has expectation 0. This essentially proves the following proposition.

Proposition 1.2. *Let U be a drift and let μ be the law of $B + U$. Then*

$$\mathrm{H}(\mu \mid \gamma) \le \frac{1}{2}\mathsf{E}[\|U\|_{\mathbb{H}}^2].$$

In order to show that equality can be achieved in the latter proposition we use an observation of Föllmer [32] which is also based on Girsanov's theorem. Given a probability measure μ on $\mathbb{W}$ absolutely continuous with respect to γ, Föllmer considers the probability space $(\mathbb{W}, \mathcal{B}, \mu)$ equipped with the natural filtration of the coordinate process (w_t), and he shows that there exists a progressively measurable map $u\colon \mathbb{W} \to \mathbb{H}$ such that the process $y = w - u(w)$ is a standard Brownian motion and such that

$$\mathrm{H}(\mu \mid \gamma) = \frac{1}{2}\int_{\mathbb{W}} \|u\|^2\, d\mu.$$

Furthermore, the Föllmer drift u can be expressed in terms of the Malliavin derivative of the density of μ with respect to γ, see below for an explicit expression in a particular case. So, on the probability space $(\mathbb{W}, \mathcal{B}, \mu)$ the coordinate process (which by definition has law μ) is a standard Brownian motion plus a drift whose energy equals the relative entropy of μ. This exactly gives equality in Proposition 1.2, but only a weak sense, up to a change of the probability space. To finish the proof of Theorem 1.1, we need to do the same thing in a strong sense, namely when the probability space and the Brownian motion are fixed from the start. It is easily seen that this task amounts to showing that if u is the Föllmer map associated to the probability measure μ, then the stochastic differential equation

$$X = B + u(X), \tag{1.1}$$

has a unique strong solution, in the sense of Ikeda and Watanabe, see [37, chapter 4]. We will not give anymore details in these notes, let us just say that in [44] we give reasonable sufficient conditions for this to be the case.

Let us now emphasize the particular case where the measure μ takes the form

$$\mu(dw) = \rho(w(1))\,\gamma(dw),$$

where ρ is a density on $\mathbb{R}^n$ with respect to the standard Gaussian measure. Some aspects of what follows are also treated in Baudoin's article [8]. Let ν be the probability measure on $\mathbb{R}^n$ having density ρ with respect to γ_n. An alternative description of μ reads as follows: a process X has law μ if X_1 has law ν and if the bridges of X coincide with those of the Brownian motion, where *bridge* means the law of the trajectory given the endpoint. Note that we have the equality

$$\mathrm{H}(\mu \mid \gamma) = \mathrm{H}(\nu \mid \gamma_n)$$

and that if $\widetilde{\mu}$ is any other measure on $\mathbb{W}$ whose marginal at time 1 coincide with that of μ, the entropy of $\widetilde{\mu}$ can only be larger. Moreover one can show that in that case the Föllmer drift of μ is given by

$$\dot{u}_t(w) = \nabla \log P_{1-t}\rho(w_t), \quad t \in [0,1]$$

where (P_t) is the heat semigroup, given by $P_t f(x) = \mathsf{E}[f(x + B_t)]$ for every test function f. The stochastic differential equation (1.1) thus becomes

$$\begin{cases} X_0 = 0 \\ dX_t = dB_t + \nabla \log P_{1-t}\rho(X_t)\,dt, \quad t \in [0,1]. \end{cases} \tag{1.2}$$

Note that it is well-posed if, say, ρ is Lipschitz and bounded away from 0. Putting everything together, we obtain the following dual version of Borell's formula.

Theorem 1.3. *Let ν and ρ be as above. We have*

$$\mathrm{H}(\nu \mid \gamma_n) = \frac{1}{2} \inf \left\{ \mathsf{E}[\|U\|_{\mathbb{H}}^2] \right\},$$

where the infimum is taken on all drifts U such that $B_1 + U_1 = \nu$ in law. The infimum is attained by the drift U given by

$$\dot{U}_t = \nabla \log P_{1-t}\rho(X_t), \quad t \in [0,1],$$

where X is the solution of (1.2).

In words, the minimal energy needed to constrain the Brownian motion to have a prescribed law at time 1 coincides with the relative entropy of that particular measure with respect to γ_n. Again this yields easily Borell's formula by duality. We close this section with a key property of the optimal drift which seems to have been missed by Borell.

Lemma 1.4. *The derivative of the optimal drift is a martingale.*

There are several ways to see this. In [44] we derive it directly from the explicit expression of the law of X. Alternatively one can apply Itô's formula to $\nabla \log P_{1-t}\rho(X_t)$ and check that the absolutely continuous term vanishes. Let us present briefly here a third option, which was pointed out to us by Rémi Lassalle. The optimality of U implies that

$$\mathsf{E}[\|U\|_{\mathbb{H}}^2] \leq \mathsf{E}[\|U+V\|_{\mathbb{H}}^2],$$

for every drift V satisfying $V_1 = 0$. This clearly implies $\mathsf{E}[\langle U, V\rangle_{\mathbb{H}}] = 0$ for any such V. Now a simple duality trick shows that this property is equivalent to the derivative of U being a square integrable martingale (see [27]).

1.3 Applications

Following Borell, we will concentrate on applications to functional inequalities. Let us just mention that the Boué-Dupuis formula also has applications to large deviation theory. For instance, it is almost immediate to derive Schilder's theorem and in [11], Boué and Dupuis actually recover the more general Freidlin-Wentzell principle.

Prékopa-Leindler. Let us explain first Borell's proof of the Prékopa-Leindler inequality based on his formula. Let f, g, h be functions satisfying

$$f(x)^{1-\lambda} g(y)^{\lambda} \leq h\left((1-\lambda)x + \lambda y\right).$$

for every $x, y \in \mathbb{R}^n$ and for some fixed $\lambda \in [0,1]$. Assume without loss of generality that f, g, h are bounded away from 0, let $F = \log f$ and define G and H similarly. Let (B_t) be a standard Brownian motion and let U and V be two drifts. Using the Prékopa-Leindler hypothesis, the convexity of the Cameron-Martin norm and applying Borell's formula to the function H and to the drift $(1-\lambda)U + \lambda V$ yields

$$\begin{aligned}
(1-\lambda)\mathsf{E}\left[F(B_1+U_1) - \frac{1}{2}\|U\|_{\mathbb{H}}^2\right] + \lambda\,\mathsf{E}\left[G(B_1+V_1) - \frac{1}{2}\|V\|_{\mathbb{H}}^2\right] \\
\leq \mathsf{E}\left[H\left(B_1 + (1-\lambda)U_1 + \lambda V_1\right) - \frac{1}{2}\|(1-\lambda)U + \lambda V\|_{\mathbb{H}}^2\right] \\
\leq \log\left(\int_{\mathbb{R}^n} \mathrm{e}^H \, d\gamma_n\right).
\end{aligned}$$

Now taking the supremum over all drifts U and V and applying Borell's formula to F and G we obtain

$$\left(\int_{\mathbb{R}^n} f\, d\gamma_n\right)^{1-\lambda} \left(\int_{\mathbb{R}^n} g\, d\gamma_n\right)^{\lambda} \leq \int_{\mathbb{R}^n} h\, d\gamma_n.$$

So we obtain the conclusion, but for the Gaussian measure rather than the Lebesgue measure. On the other hand, it is actually easy to see that the Prékopa-Leindler inequality for the Lebesgue measure follows from this: just replace f by $f(x/t)$, similarly for g and h, and let t tend to $+\infty$.

Brascamp-Lieb. A Brascamp-Lieb datum on $\mathbb{R}^n$ is a finite sequence

$$(c_1, B_1), \ldots, (c_m, B_m)$$

where c_i is a positive number and $B_i \colon \mathbb{R}^n \to \mathbb{R}^{n_i}$ is linear and onto. The Brascamp-Lieb inequality [13] asserts that the best constant in the inequality

$$\int_{\mathbb{R}^n} \prod_{i=1}^m (f_i \circ B_i)^{c_i} \, dx \leq C \prod_{i=1}^m \left(\int_{\mathbb{R}^{n_i}} f_i \, dx \right)^{c_i} \tag{1.3}$$

can be computed by restricting it to Gaussian functions. In other words if (1.3) holds for every functions $f_1, \ldots, f_m$ of the form

$$f_i(x) = \mathrm{e}^{-\langle A_i x, x \rangle / 2}$$

where A_i is a symmetric positive definite matrix on $\mathbb{R}^{n_i}$ then it holds for every set of functions $f_1, \ldots, f_m$. The reversed Brascamp-Lieb inequality, first proved by Barthe [7], asserts that the same holds true for the inequality

$$\prod_{i=1}^m \left(\int_{\mathbb{R}^{n_i}} f_i \, dx \right)^{c_i} \leq C_r \int_{\mathbb{R}^n} f \, dx$$

where f is the function

$$f(x) = \sup \left\{ \prod_{i=1}^m f_i(x_i)^{c_i}; \ \sum_{i=1}^m c_i B_i^* x_i = x \right\}.$$

In [45] we derive both the Brascamp-Lieb inequality and its reversed version from Borell's formula. For the direct version, we prove first the following Gaussian inequality. Let X be a centered Gaussian vector on $\mathbb{R}^n$ and let A be its covariance matrix. If A is invertible and satisfies

$$\sum_{i=1}^m c_i B_i^* (B_i A B_i^*)^{-1} B_i \leq A^{-1}, \tag{1.4}$$

then for every set of non negative functions $f_1, \ldots, f_m$ we have

$$\mathsf{E}\left[\prod_{i=1}^m f_i(B_i X)^{c_i} \right] \leq \prod_{i=1}^m \mathsf{E}\left[f_i(B_i X) \right]^{c_i}.$$

This is an easy consequence of Borell's formula. Indeed, let (X_t) be a Brownian motion on $\mathbb{R}^n$ having quadratic covariation given by $[X]_t = tA$ for all t. The process X is not a standard Brownian motion but it is easy to see that Borell's formula holds just the same, provided the

Cameron-Martin space $\mathbb{H}$ is modified in order to take A into account. Namely, for every absolutely continuous path u taking values in $\mathbb{R}^n$ we have

$$\|u\|_{\mathbb{H}}^2 = \int_0^1 \langle A^{-1}\dot{u}_t, \dot{u}_t\rangle \, dt.$$

Similarly, for every $i \leq m$ we let $\mathbb{H}_i$ be the Cameron-Martin space associated to the Brownian motion $B_i X$, whose covariation matrix is $B_i A B_i^*$. Then it is easily seen that the hypothesis (1.4) implies that

$$\sum_{i=1}^m c_i \|B_i u\|_{\mathbb{H}_i}^2 \leq \|u\|_{\mathbb{H}}^2,$$

for every $u \in \mathbb{H}$. Assume (without loss of generality) that the functions $f_1, \ldots, f_m$ are bounded away from 0, set $g_i = \log f_i$ and $g(x) = \sum c_i g_i(B_i x)$. Let U be an $\mathbb{R}^n$-valued drift, using the previous inequality and Borell's formula for the Brownian motions $(B_i X)_{i\leq m}$ we get

$$\begin{aligned}\mathsf{E}\left[g(X_1+U_1) - \frac{1}{2}\|U\|_{\mathbb{H}}^2\right] &\leq \sum_{i=1}^m c_i \mathsf{E}\left[g_i(B_i X_1 + B_i U_1) - \frac{1}{2}\|B_i U\|_{\mathbb{H}_i}^2\right] \\ &\leq \sum_{i=1}^m c_i \log \mathsf{E}\left[\mathrm{e}^{g_i(B_i X_1)}\right].\end{aligned}$$

Now taking the supremum in U and using Borell's formula for X we get

$$\log \mathsf{E}\left[\mathrm{e}^{g(X_1)}\right] \leq \sum_{i=1}^m c_i \log \mathsf{E}\left[\mathrm{e}^{g_i(B_i X_1)}\right],$$

which is the result. Deriving Brascamp and Lieb's theorem from this Gaussian inequality is pretty standard, we omit this part and refer to [45]. The argument for Barthe's theorem is very similar and is omitted as well. To conclude, let us note that the method is very robust and leaves plenty of room for variations. For instance, the recent inequality of Chen, Dafnis and Paouris [17] can also be recovered this way. Let us also mention the article [18], in which Cordero and Maurey obtain variants of the Barthe-Brascamp-Lieb inequalities using Borell's formula.

Talagrand and log-Sobolev inequalities. We now give two applications of the stochastic formula for the entropy. Talagrand's inequality [61] asserts that for every probability measure ν on $\mathbb{R}^n$ we have

$$T_2(\nu, \gamma_n) \leq 2\,\mathrm{H}(\nu \mid \gamma_n),$$

where T_2 is the transportation cost between ν and γ_n associated to the cost function given by the Euclidean distance squared:

$$T_2(\nu, \gamma_n) = \inf\left\{\mathsf{E}[|X-Y|^2]\right\}$$

where the infimum is taken over all pairs of random vector (X, Y) such that X and Y have law ν and γ_n respectively. The results of the previous section yield very easily a Wiener space version of this. Given two probability measures $\mu, \widetilde{\mu}$ on $\mathbb{W}$, we let $T_2(\mu, \widetilde{\mu})$ be their transportation cost

associated to the cost function given by the Cameron-Martin norm squared. As we have seen in the previous section one can find a Brownian motion B and a drift U such that $B+U$ has law μ and such that

$$\mathrm{H}(\mu \mid \gamma) = \frac{1}{2}\mathsf{E}[\|U\|_{\mathbb{H}}^2].$$

On the other hand, since B has law γ and by definition of T_2 we have

$$T_2(\mu,\gamma) \leq \mathsf{E}[\|B+U-B\|_{\mathbb{H}}^2] = \mathsf{E}[\|U\|_{\mathbb{H}}^2].$$

This gives the following result.

Theorem 1.5. *For every absolutely continuous measure μ on $(\mathbb{W},\mathcal{B})$ we have*

$$T_2(\mu,\gamma) \leq 2\,\mathrm{H}(\mu \mid \gamma).$$

Of course, this infinite dimensional version of Talagrand's inequality is not new, it is proved in Gentil's PhD thesis [33], and in Feyel and Üstunel's article [29].

We now move on to the logarithmic Sobolev inequality. Let ν be a probability measure having smooth density ρ with respect to γ_n. Let (B_t) be a Brownian motion, let (X_t) be the solution of (1.2) and let $v_t = \nabla \log P_{1-t}\rho(X_t)$. According to Theorem 1.3 the random vector X_1 has law ν and we have

$$\mathrm{H}(\nu \mid \gamma_n) = \frac{1}{2}\mathsf{E}\left[\int_0^1 |v_t|^2\,dt\right]$$

Besides, Lemma 1.4 asserts that the process (v_t) is a martingale. So $(|v_t|^2)$ is a sub-martingale, in particular $\mathsf{E}[|v_t|^2] \leq \mathsf{E}[|v_1|^2]$ for every $t \leq 1$. But since X_1 has law μ

$$\mathsf{E}[|v_1|^2] = \mathsf{E}[|\nabla \log \rho(X_1)|^2] = \int_{\mathbb{R}^n} |\nabla \log \rho|^2\,d\nu,$$

which is the Fisher information of μ. We thus have obtained the logarithmic Sobolev inequality for the Gaussian measure

$$\mathrm{H}(\nu \mid \gamma_n) \leq \frac{1}{2}\,\mathrm{I}(\nu \mid \gamma_n).$$

We now describe an open problem related to this proof, which was brought to our attention by Max Fathi. Using Itô's formula, one can push the above argument one step further to get

$$\mathrm{I}(\nu \mid \gamma_n) - 2\,\mathrm{H}(\nu \mid \gamma_n) \geq \mathsf{E}\left[\int_0^1 t\|\nabla^2 \log P_{1-t}\rho(X_t)\|_{\mathrm{HS}}^2\,dt\right],$$

where ∇^2 denotes the Hessian matrix and $\|\cdot\|_{\mathrm{HS}}$ denotes the Hilbert-Schmidt norm. While it is pretty clear that this implies that equality holds in the log-Sobolev inequality if and only $\log\rho$ is an affine function, which correspond to ν being a translate of γ_n, we were not able to derive from this inequality a stability estimate such as the one obtained by Fathi, Indrei and Ledoux in [28].

To close this section, let us mention that [44] contains a couple more applications of the entropy formula, including a short proof of the Shannon-Stam inequality, that we shall not spell out here. Lastly, another application is described in details in the next chapter.

1.4 Extensions

We consider a stochastic differential equation of the form

$$dX_t = \sigma(X_t)\, dB_t + b(X_t)\, dt \tag{1.5}$$

where σ is matrix valued and b takes values in $\mathbb{R}^n$. Let us assume that (1.5) has a unique strong solution defined for all time. This hypothesis is satisfied in particular if σ and b are locally Lipschitz and grow at most linearly. The process (X_t) is then a diffusion with generator L given by

$$Lf = \frac{1}{2}\sum_{ij} a_{ij}\, \partial^2_{ij} f + \sum_i b_i\, \partial_i f,$$

for every $\mathcal{C}^2$-smooth function f, and where a is the matrix $a = \sigma\sigma^T$. We denote the associated semigroup by (P_t): for any test function f

$$P_t f(x) = \mathsf{E}_x\left[f(X_t)\right],$$

where the subscript x denotes the starting point of (X_t). Fix a finite time horizon T and fix $x \in \mathbb{R}^n$. In this section the Wiener space is considered on the interval $[0, T]$ and the Cameron-Martin space $\mathbb{H}$ is modified accordingly. In [48] we prove that in this context, Borell's formula reads as follows.

Theorem 1.6. *For any function $f\colon \mathbb{R}^n \to \mathbb{R}$ bounded from below we have*

$$\log P_T(\mathrm{e}^f)(x) = \sup_U \left\{ \mathsf{E}\left[f(X^U_T) - \frac{1}{2}\|U\|^2_{\mathbb{H}} \right] \right\},$$

where the supremum is taken over all drifts U and the process X^U is the solution of

$$dX^U_t = \sigma(X^U_t)\,(dB_t + dU_t) + b(X^U_t)\, dt, \tag{1.6}$$

starting from x.

We also give a dual formulation in terms of entropy. In what follows $\delta_x P_T$ is the law at time T of the process (X_t) initiated from x.

Theorem 1.7. *Let ν have smooth and positive density ρ with respect to $\delta_x P_T$. Then*

$$\mathrm{H}(\nu \mid \delta_x P_T) = \frac{1}{2}\inf\left\{\mathsf{E}[\|U\|^2_{\mathbb{H}}]\right\}$$

where the infimum is taken over every drift U such that X^U_T has law ν, where again (X^U_t) is the solution of (1.6) starting from x. Moreover the infimum is attained by the drift (V_t) given by

$$\dot{V}_t = \nabla \log P_{T-t}\rho(Y_t), \quad t \le T,$$

where (Y_t) is the solution of

$$dY_t = \sigma(Y_t)\,(dB_t + \nabla \log P_{T-t}\rho(Y_t)\, dt) + b(Y_t)\, dt$$

starting from x.

In general the derivative of the optimal drift is no longer a martingale, this depends on the diffusion coefficients σ and b. Let us spell out the case of the Langevin diffusion: assume that $\sigma = I_n$ (the identity matrix) and that $b = -\frac{1}{2}\nabla V$ for a certain potential V. In that case the measure $\mu(dx) = \mathrm{e}^{-V(x)}dx$ is stationary. If e^{-V} is integrable we normalize it to be a probability measure. Assume additionally that there exists $\lambda \in \mathbb{R}$ such that

$$\nabla^2 V \geq \lambda\, I_n$$

pointwise. In this case, one can establish using Itô's formula that the process

$$\mathrm{e}^{-\lambda t}|\nabla \log P_{T-t}\rho(Y_t)|^2$$

is a submartingale. In particular

$$\mathsf{E}[|\nabla \log P_{T-t}\rho(Y_t)|^2] \leq \mathrm{e}^{\lambda(t-T)}\mathsf{E}[|\nabla \log \rho(Y_T)|^2].$$

Since Y_T has law ν the latter expectation equals $\mathrm{I}(\nu \mid \delta_x P_T)$, the Fisher information of ν with respect to the measure $\delta_x P_T$. So that integrating the inequality between 0 and T yields the following log-Sobolev inequality for $\delta_x P_T$:

$$\mathrm{H}(\nu \mid \delta_x P_T) \leq \frac{1-\mathrm{e}^{-\lambda T}}{2\lambda}\mathrm{I}(\nu \mid \delta_x P_T).$$

If λ is positive, which implies that μ is a probability measure, we can let T tend to $+\infty$ and obtain

$$\mathrm{H}(\nu \mid \mu) \leq \frac{1}{2\lambda}\mathrm{I}(\nu \mid \mu).$$

We thus have shown that if $\nabla^2 V \geq \lambda I_n$ for some $\lambda > 0$ then the probability measure proportional to $\mathrm{e}^{-V}dx$ satisfies the log-Sobolev inequality with constant $1/(2\lambda)$. This is of course well-known, as a particular case of the celebrated Bakry-Émery criterion, see for instance [4, section 5.7].

We also extend the results of the first section to the Riemannian setting. This extension is based on the intrinsic construction of the Brownian motion on a Riemannian manifold, sometimes called *rolling without slipping*. Let us give a very informal description of this. Given a Riemannian manifold (M, g) of dimension n and a standard Brownian motion on $\mathbb{R}^n$ we construct a process (X_t) taking values in M such that $dX_t = \Phi_t(dB_t)$ where the process (Φ_t) is above (X_t) in the orthonormal frame bundle $\mathcal{O}(M)$ and moves by parallel transport along (X_t). Although (X_t) is not smooth there is a way to define this properly. We say that (X_t) is the *stochastic development* of (B_t). We now formulate Borell's formula for a Riemannian manifold. In the next theorem, we fix a time horizon T, a point $x \in M$, and we denote the heat semigroup on M by (P_t).

Theorem 1.8. *Let $f\colon M \to \mathbb{R}$ be measurable and bounded from below. Then*

$$\log P_T(\mathrm{e}^f)(x) = \sup_U \left\{ \mathsf{E}\left[f(X_T^U) - \frac{1}{2}\|U\|_{\mathbb{H}}^2 \right] \right\}$$

where the supremum is taken on every drift U and the process (X_t^U) denotes the stochastic development of $B + U$ starting from x.

Based on this formula, we give in [48] a short proof of the spherical Brascamp-Lieb inequality of Carlen, Lieb and Loss [16]. It would be very nice to recover also with this method the Riemannian version of the Prékopa-Leindler inequality of Cordero, McCann and Schmuckenschläger [19] but so far, we were not able to do so. Let us conclude this section by saying that [48] also contains a dual version for the entropy from which we can recover the log-Sobolev inequality under a curvature condition. Namely we prove that if $\mathrm{Ric} \geq \kappa\, g$ pointwise for some positive κ then for every probability measure μ on M we have

$$\mathrm{H}(\mu \mid m) \leq \frac{1}{2\kappa} \mathrm{H}(\mu \mid m), \tag{1.7}$$

where m denotes the volume measure normalized to be a probability measure. Roughly speaking, the argument consists in computing the Itô derivative of the optimal drift and applying Bochner's formula. Actually, this can be pushed one step further to get the following improvement of (1.7):

$$\mathrm{H}(\mu \mid m) \leq \frac{n}{2} \log\left(1 + \frac{\mathrm{I}(\mu \mid m)}{n\,\kappa}\right).$$

Let us note that improvement in not new. Indeed, we are in a situation where the so-called curvature dimension condition $\mathrm{CD}(\rho, n)$ holds true, under which two different dimensional improvements of (1.7) are known to be true, the one we already mentioned and

$$\mathrm{H}(\mu \mid m) \leq \frac{n-1}{\kappa\, n}\, \mathrm{I}(\mu \mid m),$$

see [4, section 5.7]. Unfortunately, we were not able to recover the latter with our method.

1.5 Further developments

We close this chapter by presenting a Borell type formula for the Poisson process. This is a work in progress and the formula should be considered a conjecture at this stage.

A map $x\colon \mathbb{R}_+ \to \mathbb{N}$ is called a counting path if it starts form 0, if it is right-continuous, non-decreasing and if

$$\Delta x(t) = x(t) - x(t-) \in \{0, 1\} \quad \text{for every } t.$$

The set of all counting paths is denoted by $\mathcal{N}$. Let $(\Omega, \mathcal{A}, \mathsf{P})$ be a probability space and let $(\mathcal{F}_t)$ be a filtration. A process X defined on Ω is called a counting process if it is adapted and if the path (X_t) belongs to $\mathcal{N}$ almost surely. A counting process X is said to have stochastic intensity (λ_t) if the process (λ_t) is progressively measurable, non-negative and satisfies

$$\lambda(t)\, dt = \mathsf{E}[\Delta X(t) \mid \mathcal{F}_{t-}]$$

for every t. There is a way to give a precise meaning to this informal definition but we shall not do it here. With this definition at hand, a Poisson process is a counting process whose stochastic intensity is deterministic. In the sequel we let π be the law of the Poisson process having intensity 1. It turns out that the law of a counting process is absolutely continuous with respect to π if and only if the process has a stochastic intensity. In some sense stochastic intensities have a similar

role as drifts in the previous sections. Before stating our Borell type formula we need to couple non trivially processes having different intensities. To do so, wet N be a Poisson point process on $(\mathbb{R}_+)^2$ having intensity the Lebesgue measure. Next let $(\mathcal{F}_t)$ be the filtration given by

$$\mathcal{F}_t = \sigma\left(\{N(A),\ A \subset [0,t] \times \mathbb{R}_+\}\right)$$

for every t. Given a non-negative progressively measurable process (λ_t), we let X_λ be the process given by

$$X_\lambda(t) = N(\{(s,u)\colon s \leq t,\ u \leq \lambda_s\}).$$

Then X_λ is a counting process having stochastic intensity (λ_t). We are now in a position to state our Borell formula. For every $f\colon \mathcal{N} \to \mathbb{R}$ we have

$$\log\left(\int_{\mathcal{N}} \mathrm{e}^f \, d\pi\right) = \sup_\lambda \left\{ \mathsf{E}\left[f(X_\lambda) - \int_{\mathbb{R}_+} (\lambda_t \log \lambda_t - \lambda_t + 1)\ dt \right] \right\},$$

where the supremum is taken on every stochastic intensity (λ_t). We believe that this should have several interesting consequences, in terms of transport inequalities or discrete versions of the Prékopa-Leindler inequality.

Chapter 2

Regularization in L^1 for the Ornstein-Uhlenbeck semigroup

This chapter is based on the paper [47].

2.1 Introduction and statement of the result

Let (B_t) be a standard Brownian motion. Recall that the Ornstein-Uhlenbeck process is the solution of

$$dX_t = \sqrt{2}\, dB_t - X_t\, dt, \tag{2.1}$$

and let (Q_t) be the corresponding semigroup. From (2.1) we easily get

$$X_t = \mathrm{e}^{-t} X_0 + \sqrt{2} \int_0^t \mathrm{e}^{s-t}\, dB_s$$

hence the following explicit expression for $Q_t f$, called Mehler's formula:

$$Q_t f(x) = \int_{\mathbb{R}^n} f\left(\mathrm{e}^{-t} x + \sqrt{1 - \mathrm{e}^{-2t}}\, y\right)\, \gamma_n(dy),$$

where γ_n is still the standard Gaussian measure on $\mathbb{R}^n$. Note that this implies in particular that γ_n is stationary for (Q_t) and that $Q_t f \to \int_{\mathbb{R}^n} f\, d\gamma_n$ as t tends to $+\infty$. Recall the hypercontractive property of (Q_t), first established by Nelson in [54]. If $p > 1$ and $t > 0$ then Q_t is a contraction from $L_p(\gamma_n)$ to $L_q(\gamma_n)$ where

$$q = 1 + \mathrm{e}^{2t}(p-1).$$

As we already mentioned in the introduction, this turns out to be equivalent to the logarithmic Sobolev inequality, see the classical article by Gross [34]. Observe that Nelson's gives no information for $p = 1$. Besides, it can be shown that Nelson's result is sharp in a very strong sense: if $r > 1 + \mathrm{e}^{2t}(p-1)$ then Q_t is not even bounded from L^p to L^r. So there is no hope to get hyperboundedness for functions in L^1. In this chapter we address the question of a weaker regularizing effect of (Q_t) for such functions. If f is non-negative and satisfies

$$\int_{\mathbb{R}^n} f\, d\gamma_n = 1,$$

then so does $Q_t f$. So by Markov inequality

$$\gamma_n\left(\{Q_t f \geq r\}\right) \leq \frac{1}{r},$$

for all $r \geq 1$. Now Markov inequality is only sharp for indicator functions and $Q_t f$ cannot be an indicator function, so it may be the case that this inequality can be improved. More precisely one might conjecture that for any fixed $t > 0$ (or at least for t large enough) there exists a function $\alpha(r)$ tending to 0 as r tends to $+\infty$ such that

$$\gamma_n\left(\{Q_t f \geq r\}\right) \leq \frac{\alpha(r)}{r}, \tag{2.2}$$

for every $r \geq 1$ and for every non-negative function f of integral 1. The function α should be independent of the dimension n, just as the hypercontractivity result stated above. Such a phenomenon was actually conjectured by Talagrand in [60] in a slightly different context. He conjectured that the same inequality holds true when γ_n is replaced by the uniform measure on the discrete cube $\{-1, 1\}^n$ and the Ornstein-Uhlenbeck semigroup is replaced by the semigroup associated to the random walk on the discrete cube. The Gaussian version of the conjecture would follow from Talagrand's discrete version by the central limit theorem.

In [5], Ball, Barthe, Bednorz, Oleszkiewicz and Wolff showed that in dimension 1 the inequality (2.2) holds with decay

$$\alpha(r) = \frac{C}{\sqrt{\log r}},$$

where the constant C depends on the time parameter t. Moreover the authors provide an example showing that the $1/\sqrt{\log r}$ decay is sharp. They also have a result in higher dimension but they loose a factor $\log\log r$ and, more importantly, their constant C then tends to $+\infty$ (actually exponentially fast) with the dimension. The deadlock was broken by Eldan and Lee who showed in [25] that (2.2) holds with function

$$\alpha(r) = C\,\frac{(\log\log r)^4}{\sqrt{\log r}},$$

with a constant C that is independent of the dimension. Again up to the $\log\log$ factor the result is optimal. The purpose of the article [47] is to simplify the argument of Eldan and Lee, and to remove the extra $\log\log$ factor. More precisely the main result reads as follows.

Theorem 2.1. *Let f be a non-negative function on $\mathbb{R}^n$ satisfying $\int_{\mathbb{R}^n} f\, d\gamma_n = 1$ and let $t > 0$. Then for every $r > 1$*

$$\gamma_n\left(\{Q_t f > r\}\right) \leq C\,\frac{\max(1, t^{-1})}{r\sqrt{\log r}},$$

where C is a universal constant.

As in Eldan and Lee's paper, the Ornstein-Uhlenbeck semigroup only plays a role through the following inequality. For every $f\colon \mathbb{R}^n \to \mathbb{R}_+$ and for every $t > 0$, we have

$$\nabla^2 \log(Q_t f) \geq -\frac{1}{2t}\, I_n, \tag{2.3}$$

pointwise, where ∇^2 denotes the Hessian matrix. Let us mention that this has nothing to do with the following implication

$$\nabla^2 \log f \leq 0 \quad \Rightarrow \quad \nabla^2 \log Q_t f \leq 0,$$

valid for every $t > 0$. The latter is a deep fact that follows from the Prékopa-Leindler inequality, whereas (2.3) is a straighforward consequence of Mehler's formula. What we actually prove is the following.

Theorem 2.2. *Let f be a positive function on $\mathbb{R}^n$ satisfying $\int_{\mathbb{R}^n} f \, d\gamma_n = 1$. Assume that f is smooth and satisfies*

$$\nabla^2 \log f \geq -\beta \, I_n \tag{2.4}$$

pointwise, for some $\beta \geq 0$. Then for every $r > 1$

$$\gamma_n \left(\{ f > r \} \right) \leq \frac{C \, \max(\beta, 1)}{r \sqrt{\log r}},$$

where C is a universal constant.

Note that this is an inequality about the Gaussian measure, and that the Ornstein-Uhlenbeck semigroup does not appear anymore. As we already mentioned the dependence in r in both theorems is optimal. This can be seen by plugging in functions of the form $f(x) = \mathrm{e}^{\langle u, x \rangle + c}$ and noting that this class of functions is preserved by the operator Q_t. We refer to [47] for the details.

2.2 Sketch of proof

The argument, which is a mere refinement of that of Eldan and Lee, rely heavily on Theorem 1.3 from the previous chapter. Let f be a probability density with respect to γ_n. For simplicity let us consider only the case $\beta = 0$ of Theorem 2.2. We need to show that if $\log f$ is convex then

$$\gamma_n(\{ f \geq r \}) \leq \frac{C}{r \sqrt{\log r}}$$

for every $r \geq 1$. An easy computation shows that it is enough to prove that

$$\mu(\{ f \in [r, 2r] \}) \leq \frac{C}{\sqrt{\log r}}$$

for every $r \geq 1$, where μ is the probability measure having density f with respect to γ_n. Let (B_t) be a standard Brownian motion, let (P_t) be the heat semigroup and let (X_t) be the solution of

$$dX_t = dB_t + \nabla \log P_{1-t} f(X_t) \, dt$$

starting from 0. By Theorem 1.3, the vector X_1 has law μ. The key idea of Eldan and Lee is to introduce a perturbed version of (X_t): fix $\delta > 0$ and let

$$X_t^\delta = X_t + \delta \int_0^1 v_s \, ds = B_t + \int_0^t (1 + \delta) v_s \, ds,$$

where $v_t = \nabla \log P_{1-t} f(X_t)$. Now write

$$\begin{aligned}\mu(\{f \in [r, 2r]\}) &= \mathsf{P}(f(X_1) \in [r, 2r]) \\ &\leq \mathsf{P}(f(X_1) \geq r,\ f(X_1^\delta) \leq 2r) + \mathrm{TV}(X_1, X_1^\delta),\end{aligned}$$

where TV denotes the total variation. We then show that if δ is chosen appropriately then both terms are smaller than $1/\sqrt{\log r}$. Let us handle the total variation term first. Let μ^δ be the law of X_1^δ, and observe that

$$\mathrm{H}(\mu^\delta \mid \mu) = \mathrm{H}(\mu^\delta \mid \gamma_n) - \int_{\mathbb{R}^n} \log(f)\ d\mu^\delta.$$

Since $\log f$ is convex we have

$$\begin{aligned}\log(f)(X_1^\delta) &\geq \log f(X_1) + \langle \nabla \log f(X_1), X_1^\delta - X_1 \rangle \\ &= \log f(X_1) + \delta \int_0^1 \langle v_1, v_t \rangle\, dt.\end{aligned} \tag{2.5}$$

Hence

$$\mathrm{H}(\mu^\delta \mid \mu) \leq \mathrm{H}(\mu^\delta \mid \gamma_n) - \mathrm{H}(\mu \mid \gamma_n) - \delta\, \mathsf{E}\left[\int_0^1 \langle v_1, v_t \rangle\, dt\right].$$

Now we apply Theorem 1.3 to (v_t), which is the optimal drift for μ, and to $((1+\delta)v_t)$, which has no reason to be optimal for μ^δ. We obtain

$$\begin{aligned}\mathrm{H}(\mu \mid \gamma_n) &= \frac{1}{2}\mathsf{E}\left[\int_0^1 |v_t|^2\, dt\right] \\ \mathrm{H}(\mu^\delta \mid \gamma_n) &\leq \frac{(1+\delta)^2}{2}\mathsf{E}\left[\int_0^1 |v_t|^2\, dt\right].\end{aligned}$$

Lastly since (v_t) is a martingale $\mathsf{E}[\langle v_1, v_t \rangle] = \mathsf{E}[|v_t|^2]$ for every $t \leq 1$. Putting everything together we obtain

$$\mathrm{H}(\mu^\delta \mid \mu) \leq \frac{\delta^2}{2}\mathsf{E}\left[\int_0^1 |v_t|^2\, dt\right].$$

Combining this with Pinsker's inequality we obtain a bound for $\mathrm{TV}(\mu, \mu^\delta)$. We now give a heuristic argument for the end of the proof. We have seen that $\log f(X_1)$, $\frac{1}{2}\int |v_t|^2$, and $\frac{1}{2}\int \langle v_1, v_t \rangle$ coincide in average. Let us assume that all three of them are concentrated around $\log r$. Then (2.5) shows that

$$\log f(X_1^\delta) \geq (1 + 2\delta) \log r.$$

So that choosing $\delta \approx 1/\log r$ kills the term $\mathsf{P}(f(X_1^\delta) \leq 2r;\ f(X_1) \geq r)$. On the other hand, for this value of δ, the above argument gives

$$\mathrm{TV}(X_1, X_1^\delta) \lesssim \frac{1}{\sqrt{\log r}},$$

hence the result. Of course the precise proof is a little more intricate, in particular we need to modify slightly the definition of X_t^δ, but all the ideas are presented here, and the rest consists of mere technicalities.

2.3 Perspectives

It is natural to ask whether other models satisfy this L_1-regularization property. A related question is the relationship of L_1-regularization to functional inequalities such as Poincaré or log-Sobolev. These questions are completely open, nothing beyond the Gaussian case is known. In particular, Talagrand's conjecture about the discrete cube is still open. In this section we just give an example of a semigroup not satisfying the L_1-regularization property. The model, which was suggested to us by Itai Benjamini, is a sort of Cauchy-Ornstein-Uhlenbeck semigroup. Recall that the Cauchy distribution on $\mathbb{R}$ is the probability measure μ given by

$$\mu(dx) = \frac{1}{\pi(1+x^2)}.$$

This is a 1-stable distribution: if X and Y are i.i.d. Cauchy distributed then for every $s, t > 0$

$$sX + tY = (s+t)X \quad \text{in law.} \tag{2.6}$$

We thus define a semigroup (P_t) by setting

$$P_t f(x) = \mathsf{E}[f(\mathrm{e}^{-t}x + (1-\mathrm{e}^{-t})X)]$$

for every test function f. It is then clear that the Cauchy distribution μ is stationary and that the semigroup is ergodic. Given $\lambda > 0$ we let

$$f_\lambda(x) = \frac{\lambda(\lambda+1)}{\lambda^2 + x^2}.$$

Using (2.6) it is easily seen that $\int_{\mathbb{R}} f_\lambda \, d\mu = 1$ and that for every $t \geq 0$ we have $P_t(f_\lambda) = f_{\lambda_t}$, where $\lambda_t = \mathrm{e}^t(\lambda+1) - 1$. This shows that for every $\lambda > 0$ and every $t \geq 0$ the function f_λ belongs to the image of the operator P_t. On the other hand, given $r > 1$, and choosing $\lambda = 1/2(r-1)$ we get after some computation

$$\mu(\{f_\lambda \geq r\}) = \mu\left(\left\{|x| \leq \frac{1}{2\sqrt{r(r-1)}}\right\}\right) \sim_{r\to\infty} \frac{1}{\pi r}.$$

This shows that this semigroup does not satisfy any kind of L^1-regularization.

Chapter 3

Sampling from a log-concave distribution

This chapter is based on the article [14], which is a joint work with S. Bubeck, and R. Eldan.

3.1 Introduction and statement of the result

Let K be a convex body in $\mathbb{R}^n$, that is to say a convex compact set having non empty interior. In this chapter, we describe and analyze an algorithm that produces a random point uniformly distributed on K. This is a practical question that has many applications in theoretical computer science, in particular in machine learning, although we shall not describe these here. Some normalization is needed in order for the problem to be tractable, and we will assume that K contains the Euclidean ball B_2^n, and is contained in the ball of radius R, where R is a fixed parameter larger than 1. Of course the following naive rejection algorithm produces a uniform sample from K:

- Produce X uniform on RB_2^n;
- If $X \in K$, return X;
- Else start over.

However, the expected running time is $|RB_2^n|/|K|$ which is typically exponential in n. In most applications the dimension n is huge, so this does not work in practice. The real challenge is to produce a random point in K in polynomial time in n.

We actually study a more general situation where a convex potential is added. Let $V\colon K \to \mathbb{R}$ be a convex function and let μ be the probability measure given by

$$\mu(dx) = Z^{-1}\, \mathrm{e}^{-V(x)} 1_{\{x\in K\}}\, dx,$$

where Z is just the normalization constant. Again some hypothesis is needed and we assume that V is Lipschitz with constant L and that its gradient is Lipschitz with constant β.

We now describe the algorithm and state the main result. Let η be a positive parameter and let $\xi_1, \xi_2, \dots$ be an i.i.d. sequence of standard Gaussian random vectors in $\mathbb{R}^n$. We study the following Markov chain, which we call *Projected Langevin Monte Carlo*:

$$X_{k+1} = \mathcal{P}_K \left(X_k - \frac{\eta}{2} \nabla V(X_k) + \sqrt{\eta}\, \xi_{k+1} \right), \tag{3.1}$$

where $\mathcal{P}_K$ is the Euclidean projection on K, that is to say $\mathcal{P}_K(x)$ is the point in K that minimizes the Euclidean distance to x. One step of the algorithm thus consists in doing a Gaussian jump, a gradient descent step and projecting back on K. Note that the variance of the random jump is equal (up to a factor 2) to the size of the gradient descent step. Standard results in approximation theory, see for instance Robbins and Monro [56], show that if the variance of the noise were of smaller order, then the iterates (X_k) would converge to the minimum of the potential V on K. Our main result states as follows.

Theorem 3.1. *Let $\epsilon > 0$. If the parameter η and the number of steps N satisfy*

$$\eta \approx \frac{R^2}{N}, \quad N \approx \frac{R^6 \max(n, RL, R\beta)^{12}}{\epsilon^{12}},$$

then we have

$$\mathrm{TV}(X_N, \mu) \leq \epsilon.$$

Moreover, we have a slightly better result when V is constant, in other words when μ is the uniform measure on K: the same conclusion holds with $N \approx \frac{R^6 n^7}{\epsilon^8}$.

The symbol $\approx$ hides universal constants and possibly logarithmic dependences that we will not give precisely. In any case, what the theorem really says is that the projected Langevin Monte Carlo algorithm produces a random point arbitrarily close to the target distribution in a time that is polynomial in all the parameters of the problem, including the dimension. We do not address (nor here and neither in the article [14]) the question of the computation of $\mathcal{P}_K$, we treat it as a computational unit. Let us just say that minimizing the Euclidean distance over a convex set is usually a tractable problem.

There is a long line of works in theoretical computer science proving similar results. We shall only cite two of them and refer to the survey of Vempala [64] for a more accurate history of the problem. The first polynomial time algorithm for sampling from a convex set is due to Dyer, Frieze and Kannan [22]. Essentially their method was to approximate the convex body by its intersection with a lattice and then run a random walk on this graph. The best estimate as of today is based on a random walk called *hit-and-run*. A step of the hit-and-run consists in choosing uniformly a direction θ and then jump uniformly on the cord $(x + \mathbb{R}\theta) \cap K$, where x is the current position. In [49], Lovasz and Vempala showed that essentially n^4 steps of this walk are enough to approximate the uniform measure on K.

Of course our result is a lot worse than the state of the art bound of Lovasz and Vempala. On the other hand, we prove that (3.1) converges in polynomial time, which was not known before and which is a natural question to ask. Indeed, as we shall see below, the process given by (3.1) is just the Euler scheme associated to a Langevin dynamics with reflecting boundary conditions. At a more practical level, it is very likely that our bound is off the truth by several powers of n, so it could be that (3.1) actually performs as well as the hit-and-run. We did some numerical experiments which suggested that this is indeed the case.

3.2 Dalalyan's argument

The work presented in this chapter is inspired by Dalalyan's article [20] which treats the unconstrained case ($K = \mathbb{R}^n$) and assumes furthermore that the potential is uniformly convex. More precisely, let $V : \mathbb{R}^n \to \mathbb{R}$ be a smooth function satisfying

$$\alpha I_n \leq \nabla^2 V \leq \beta I_n$$

pointwise, where α, β are **positive** parameters and let μ be the measure given by

$$\mu(dx) = Z^{-1} \mathrm{e}^{-V(x)}\, dx$$

where Z is the normalization constant, note that the hypothesis implies that e^{-V} is integrable. Dalalyan shows that the algorithm

$$X_{k+1} = X_k - \frac{\eta}{2} \nabla V(X_k) + \sqrt{\eta}\, \xi_{k+1} \tag{3.2}$$

allows to sample from μ in polynomial time. Let us describe briefly his argument. Let (W_t) be a standard Brownian motion on $\mathbb{R}^n$ and consider the Langevin diffusion associated to the potential V:

$$dY_t = dW_t - \frac{1}{2} \nabla V(Y_t)\, dt.$$

The measure μ is stationary and ergodic: $Y_t \to \mu$ as t tends to $+\infty$. Now fix a positive parameter η and consider the Euler scheme associated to the Langevin diffusion:

$$d\widetilde{Y}_t = dW_t - \frac{1}{2} \nabla f(\widetilde{Y}_{\lfloor t/\eta \rfloor \eta})\, dt,$$

where $\lfloor \cdot \rfloor$ denotes the integer part. Note that for every integer k

$$\widetilde{Y}_{(k+1)\eta} = \widetilde{Y}_{k\eta} + W_{(k+1)\eta} - W_{k\eta} - \frac{\eta}{2} \nabla V(\widetilde{Y}_{k\eta}).$$

This shows the law of the sequence $(\widetilde{Y}_{k\eta})$ coincides with that of the Markov chain given by (3.2). Next write:

$$\mathrm{TV}(\widetilde{Y}_t, \mu) \leq \mathrm{TV}(\widetilde{Y}_t, Y_t) + \mathrm{TV}(Y_t, \mu). \tag{3.3}$$

Let us deal with the second term first. By the Bakry-Émery criterion, which we already mentioned in Chapter 1, the hypothesis $\nabla^2 V \geq \alpha\, I_n$ with α positive yields a log-Sobolev inequality, which, as we saw in the introduction, gives an exponential decay of the relative entropy $\mathrm{H}(Y_t \mid \mu)$. Together with Pinsker's inequality we get

$$\mathrm{TV}(Y_t, \mu) \leq \mathrm{e}^{-\alpha t/2} \sqrt{\mathrm{H}(Y_0 \mid \mu)}.$$

Up to a preprocessing making sure that $\mathrm{H}(Y_0 \mid \mu)$ is finite, this is called a *warm start* in the literature, we thus have a control of the second term of (3.3). The first term is dominated by the total variation between the Brownian motion W and the process $\widetilde{W}$ given by

$$d\widetilde{W}_t = dW_t + \frac{1}{2} \nabla V(\widetilde{Y}_t)\, dt - \frac{1}{2} \nabla V(\widetilde{Y}_{\lfloor t/\eta \rfloor \eta})\, dt.$$

Using the hypothesis $\nabla^2 V \leq \beta\, I_n$ and combining it with Proposition 1.2 from the first chapter of this notes, one can actually bound the relative entropy of $\widetilde{W}$ with respect to W. Putting everything together Dalalyan shows that essentially n^3 steps of the algorithm (3.2) are enough to approximate the measure μ.

3.3 Analysis of the algorithm

Now we want to adapt the previous argument to the case where the measure μ is supported on a convex body K. For simplicity, we shall only consider the constant potential case. Thus the measure μ is uniform on K and the algorithm reads

$$X_{k+1} = \mathcal{P}_K\left(X_k + \sqrt{\eta}\,\xi_{k+1}\right). \tag{3.4}$$

The first step is to understand the underlying continuous process. Let $w\colon [0,T) \to \mathbb{R}^n$ be a path with $w(0) \in K$. We say that a couple of paths (x,ϕ) solves the *Skorokhod* problem associated to w if

- $x(t) \in K$, for all $t < T$.
- $x = w + \phi$
- The path ϕ satisfies $\phi(t) = -\int_0^t \nu_s\, L(ds)$ where L is a measure on $[0,T]$ supported on the set $\{t \in [0,T)\colon x(t) \in \partial K\}$ and for every such t, the vector ν_t is an outer unit normal at $x(t)$.

Tanaka [63] showed that for every piecewise continuous w there is a unique solution (x,ϕ) to the Skorokhod problem. The path x is called the reflection of w at the boundary of K and L is called the local time of x at the boundary. Let (W_t) be a standard Brownian motion and let (Y_t) be its reflection at the boundary of K. One can show that (Y_t) is Markov and that μ (the normalized uniform measure on K) is stationary and ergodic. Now fix a parameter $\eta > 0$ and let $\widetilde{W}_t = W_{\lfloor t/\eta \rfloor \eta}$ be the discretized Brownian motion. Since $(\widetilde{W}_t)$ is piecewise constant one can solve the associated Skorokhod problem explicitly. Namely, letting $\widetilde{Y}$ be the reflection $\widetilde{W}$, it is not hard to see that $\widetilde{Y}$ is constant on intervals $[k\eta, (k+1)\eta)$ and that the sequence $(\widetilde{Y}_{k\eta})$ satisfies

$$\widetilde{Y}_{(k+1)\eta} = \mathcal{P}_K\left(\widetilde{Y}_{k\eta} + W_{(k+1)\eta} - W_{k\eta}\right).$$

As a result the sequence $(\widetilde{Y}_{k\eta})$ has the same law as the Markov chain given by (3.4). Now our goal is to bound $\mathrm{TV}(\widetilde{Y}_t, \mu)$. Following Dalalyan we write

$$\mathrm{TV}(\widetilde{Y}_t, \mu) \le \mathrm{TV}(\widetilde{Y}_t, Y_t) + \mathrm{TV}(Y_t, \mu). \tag{3.5}$$

For the second term, Bakry-Émery does not apply anymore, since the potential is not uniformly convex. Instead we use a coupling argument to estimate directly the mixing time of (Y_t). Let us describe it briefly. Let x and x' in K, let (W_t) and (W'_t) be two Brownian motions started from x and x' respectively and let (Y_t) and (Y'_t) be their respective reflections at the boundary of K. We couple (W_t) and (W'_t) in such a way that the increment dW'_t is the reflection of dW_t with respect to the hyperplane median to $[Y_t, Y'_t]$. Actually we only do this up to the coupling time $\tau = \inf\{t > 0\colon Y_t = Y'_t\}$. For $t \ge \tau$ we just set $Y_t = Y'_t$. This is called *mirror coupling*. Then using the convexity of K, it is easily seen that the reflection at the boundary of K is only helping us. We thus get

$$\mathsf{P}(\tau > t) \le \frac{|x - x'|}{\sqrt{2\pi t}},$$

which is what we would obtain if there were no reflection at all. This implies that the mixing time of the process (Y_t) is at most R^2. More precisely choosing $t \approx R^2/\epsilon^2$ yields $\mathrm{TV}(Y_t, \mu) \leq \epsilon$.
Now we need to deal with the first term of (3.5) and again Dalalyan's argument fails, just because no matter how small η is, the total variation between the Brownian motion W and its discretization $\widetilde{W}$ is always 1. This part of the article [14] is the most technical so we will not give much details here. Let us just give the overall strategy. Using a deterministic inequality of Tanaka, one can actually bound the expected distance between Y_t and $\widetilde{Y}_t$:

$$\mathsf{E}\left[|Y_t - \widetilde{Y}_t|\right] \lesssim n^{3/4} t^{1/2} \eta^{1/4}.$$

Then we need to pass from this transport cost estimate to a total variation estimate. This is done using the mirror coupling again, alongside with an estimate of the hitting time of the boundary of K for a Brownian motion started from a uniform point in K.

3.4 Further comments

We have presented a relatively simple argument showing that the projected Langevin Monte Carlo algorithm converges in polynomial time. On the other hand, our estimate is probably far from being optimal. For instance we prove that the mixing time of the Brownian motion reflected at the boundary of K is at most R^2. This estimate is sharp when K is close to being a line segment, but is very wasteful for typical bodies K. If K is a ball of radius $\sqrt{n}$ for instance, the mixing time is order 1 rather than n. If K is the cube $[-1, 1]^n$ the mixing time is order $\log n$ whereas $R = \sqrt{n}$. A natural attempt to use the geometry of K more would be to assume that K is in isotropic position. Recall that K is in isotropic position if a random vector uniform on K has expectation 0 and covariance equal to the identity matrix. The issue is that estimating the mixing time of a generic isotropic convex body in $\mathbb{R}^n$ is a hard problem. We shall come back to this question in more details in the next chapter, but let us say very quickly that the Kannan-Lovasz-Simonovits [39] conjecture, which is 20 years old, asserts that the Poincaré constant of such a body is bounded by a universal constant. The current best estimate, which is obtained combining two deep results of Guédon and Milman [35] and of Eldan [23], is essentially $n^{2/3}$. And this is only for Poincaré. Finding a warm start procedure allowing to pass from Poincaré to mixing in the total variation sense without loosing too many powers of n is not an easy task either.

Chapter 4

Stochastic localization and norms of log-concave vectors

This chapter is based on our joint work with R. Eldan [26].

4.1 Eldan's stochastic localization

In this section we describe a random process introduced by Eldan in [23] which is called *stochastic localization*. Let us start by describing a simplified version of the process. Let μ be a probability measure on $\mathbb{R}^n$ and let (B_t) be a standard Brownian motion. For $t \geq 0$, let ρ_t be the random function given by

$$\rho_t(x) = \exp\left(\langle x, B_t\rangle - \frac{t}{2}|x|^2\right)$$

and let μ_t be the random measure having density ρ_t with respect to μ. Observe that for any fixed x the process $(\rho_t(x))_{t\geq 0}$ is a martingale, so that in particular $\mathsf{E}[\rho_t(x)] = \rho_0(x) = 1$. This shows that in average the random measure μ_t equals μ. On the other hand, because of the factor $\mathrm{e}^{-t|x|^2/2}$ the measure μ_t gets more concentrated as t increases. Hence the name *localization*. The real process, which we present now, is a lot more complicated, but the idea stays the same. We consider the following infinite system of stochastic differential equations, where the unknown is a function valued process (ρ_t):

$$\begin{cases} \rho_0(x) = 1 & \forall x \in \mathbb{R}^n \\ d\rho_t(x) = \rho_t(x)\langle A_t^{-1/2}(x - a_t), dB_t\rangle & \forall x \in \mathbb{R}^n \\ a_t = \int_{\mathbb{R}^n} x\, \rho_t(x)\, \mu(dx) & \\ A_t = \int_{\mathbb{R}^n} (x - a_t) \otimes (x - a_t)\, \rho_t(x)\, \mu(dx) & \end{cases}$$

Eldan shows that under suitable conditions on μ this system has a unique strong solution defined for all time. The first two equations show that ρ_t remains positive almost surely and we let μ_t be the measure having density ρ_t with respect to μ. Note that (ρ_t) is a martingale, in particular $\mathsf{E}[\rho_t] = 1$ for all t, which shows that μ_t equals μ in average. One advantage of this upon the simplified process presented at the beginning of this section is that μ_t is a probability measure for

all t, almost surely. Indeed we have $\mu_0 = \mu$ and, at least formally, the Itô derivative of the total mass of μ_t reads

$$d\mu_t(\mathbb{R}^n) = \int_{\mathbb{R}^n} d\rho_t(x)\, \mu(dx) = \left\langle \int_{\mathbb{R}^n} A_t^{-1/2}(x - a_t)\, \mu_t(dx), dB_t \right\rangle$$

which is 0 by definition of a_t. Of course we have taken the Itô derivation in and out of the integral on x and this should be properly justified, but let us leave these technical details aside in these notes. Note that a_t and A_t are the barycenter and covariance matrix of the random probability measure μ_t. Applying Itô's formula to $\log \rho_t(x)$ we get

$$\log \rho_t(x) = \int_0^t \langle A_s^{-1/2}(x - a_s), dB_s \rangle - \frac{1}{2} \int_0^t \langle A_s^{-1}(x - a_s), x - a_s \rangle\, ds.$$

So $-\log \rho_t$ is a quadratic form in x, the quadratic term being $\langle H_t x, x \rangle$, where $H_t = \int_0^t A_s^{-1} ds$. In particular, if μ is log-concave then μ_t has density of the form $\exp\left(-\frac{1}{2}\langle H_t x, x\rangle - V_t(x)\right)$ where V_t is a convex function. So μ_t is a log-concave perturbation of a Gaussian measure. This will play a crucial role in the next section.

4.2 Thin-shell implies KLS

In order to illustrate the power of the previous construction let us sketch the proof of Eldan's theorem relating two famous conjectures from asymptotic convex geometry: *KLS* and *thin-shell*. Let us recall the two statements first. Given a random vector X on $\mathbb{R}^n$ we let C_X be its Poincaré constant, namely the smallest constant such that the inequality

$$\mathrm{Var}(f(X)) \leq C_X\, \mathsf{E}[|\nabla f(X)|^2]$$

holds for every Lipschitz (say) function f. Define

$$C_n = \sup\{C_X;\ X \text{ log-concave and isotropic on } \mathbb{R}^n\}$$

Recall that X is said to be isotropic if $\mathsf{E}[X] = 0$ and $\mathsf{E}[XX^T] = I_n$. Note that the Poincaré constant of an isotropic Gaussian vector is 1. In [39], Kannan, Lovasz and Simonovits (KLS in short) conjecture that the Gaussian behaviour is the worst case, up to a universal constant. In other words they conjecture that $C_n = O(1)$. Using a (deterministic) localization method they were able to prove that $C_n = O(n)$. The *thin-shell* conjecture, which is a priori much weaker, is obtained by restricting Poincaré to the function f being the Euclidean norm. More precisely letting

$$D_n = \sup\{\mathrm{Var}(|X|);\ X \text{ log-concave and isotropic on } \mathbb{R}^n\},$$

the thin-shell conjecture asserts that $D_n = O(1)$. Note that if X is isotropic then $\mathrm{Var}(|X|) \leq \mathsf{E}[|X|^2] = n$. So it is obvious that $D_n \leq n$. A deep result of Klartag [41] asserts that $D_n = o(n)$. Shortly after, Fleury, Guédon and Paouris [31] found another proof of this fact. This result was refined several times and the current best estimate, due to Guédon and E. Milman [35] reads

$D_n = O(n^{2/3})$. As we already said, the thin-shell conjecture is a priori much weaker than KLS. The remarkable result of Eldan asserts that actually

$$C_n = O(D_n^* \log n)$$

where $D_n^* = \sum_{k \leq n} D_k/k$. In particular, if $D_n = O(1)$ then $C_n = O((\log n)^2)$. In words, thin-shell implies KLS, up to a logarithmic factor. Note also that combining Eldan's theorem with the result of Guédon and Milman yields

$$C_n = O(n^{2/3} \log n)$$

which is the best estimate as of today. Lastly, let us mention an earlier result of the same nature, due to Bobkov [9]. He proved that if X is log-concave and isotropic then

$$C_X \leq K n^{1/2} \mathrm{Var}(|X|)^{1/2},$$

where K is a universal constant. This result has one advantage upon that of Eldan: the latter requires a bound on $\mathrm{Var}(Y)$ for every isotropic Y to provide information on C_X. The drawback with Bobkov's result is of course the $\sqrt{n}$ factor. For instance, his inequality only insures $C_n = O(\sqrt{n})$ under thin-shell, in contrast with Eldan's logarithmic dependence.

Now we sketch Eldan's argument. We adopt a slightly different strategy than his but the idea is the same. Let μ be an isotropic log-concave measure, let (μ_t) be the stochastic localization process started from μ and let f be a smooth function satisfying $\int f d\mu = 0$. We want to bound $\int f^2 \, d\mu$. Since μ_t equals μ is average we have

$$\begin{aligned} \int_{\mathbb{R}^n} f^2 \, d\mu &= \mathsf{E}\left[\int_{\mathbb{R}^n} f^2 d\mu_t\right] \\ &= \mathsf{E}\left[\mathrm{Var}_{\mu_t}(f^2)\right] + \mathsf{E}\left[\left(\int_{\mathbb{R}^n} f \, d\mu_t\right)^2\right] \end{aligned} \tag{4.1}$$

Note that $d \int_{\mathbb{R}^n} f \, d\mu_t = \langle v_t, dB_t \rangle$ where

$$v_t = \int_{\mathbb{R}^n} f(x) A_t^{-1/2}(x - a_t) \, \mu_t(dx).$$

Using the Cauchy-Schwarz's inequality and the fact that A_t is the covariance matrix of μ_t we easily get $|v_t|^2 \leq \int f^2 \, d\mu_t$ almost surely, hence

$$\mathsf{E}[|v_t|^2] \leq \int_{\mathbb{R}^n} f^2 \, d\mu, \quad \forall t,$$

Therefore $\mathsf{E}\left[\left(\int_{\mathbb{R}^n} f \, d\mu_t\right)^2\right] \leq t \int_{\mathbb{R}^n} f^2 \, d\mu$ and (4.1) becomes

$$\int_{\mathbb{R}^n} f^2 \, d\mu = \frac{1}{1-t} \mathsf{E}\left[\mathrm{Var}_{\mu_t}(f)\right].$$

So we are left with estimating the Poincaré constant of μ_t, let us call it C_t. Actually it is enough to estimate C_t in average. Indeed, on the one hand we can always write

$$\mathsf{E}\left[\mathrm{Var}_{\mu_t}(f)\right] \leq \mathsf{E}\left[C_t \int_{\mathbb{R}^n} |\nabla f|^2 \, d\mu_t\right] \leq \mathsf{E}[C_t] \, \|f\|_{\mathrm{Lip}}^2.$$

On the other hand, a theorem of E. Milman [53] asserts that for a log-concave measure μ, the Poincaré constant C_μ is within a universal multiple of the best constant C in the weak Poincaré inequality $\mathrm{Var}_\mu(f) \leq C\|f\|^2_{\mathrm{Lip}}$. Now, as we have seen at the end of the previous section, μ_t is a log-concave perturbation of a Gaussian measure. So by Bakry-Émery's criterion its Poincaré constant is dominated by the Poincaré constant of that particular Gaussian, namely the inverse of the smallest eigenvalue of $H_t = \int_0^t A_s^{-1}\, ds$. So, at the end of the day, all we need to control is the average operator norm of A_t. This is the most technical part of the proof. Without going into the details, let us say that the Itô derivative of the process A_t involves third moments of the measure μ_t. Using the thin shell constant D_n one can bound these in terms of second moments of μ_t, hence in terms of A_t itself. After a couple of pages of computations one can get the inequality

$$\mathsf{E}[\|A_t\|_{\mathrm{op}}] \leq K\, D_n^* \log n\, \mathrm{e}^{-t} \tag{4.2}$$

where K is a universal constant. From this inequality and what is above we easily get $\mathsf{E}[C_t] \leq K\, t^{-2} D_n^* \log n$. Eldan's theorem then follows by choosing $t = 1/2$ (say) and putting everything together.

4.3 Bounding the norm of log-concave vectors

The main result in [26] reads as follows.

Theorem 4.1. *Let X be an isotropic log-concave random vector in $\mathbb{R}^n$, let G be a standard Gaussian vector on $\mathbb{R}^n$ and let $\|\cdot\|$ be a norm. There is a universal constant C such that*

$$\mathsf{E}[\|X\|] \leq C\, (D_n^* \log n)^{1/2}\, \mathsf{E}[\|G\|]. \tag{4.3}$$

Using Talagrand's generic chaining, Bourgain proved in [12] that

$$\mathsf{E}[\|X\|] \leq C' n^{1/4} \mathsf{E}[\|G\|]. \tag{4.4}$$

Let us note that combining our main theorem with Guédon-Milman's estimate $D_n^* = O(n^{2/3})$ we obtain a worse bound than Bourgain's. On the other hand, if thin-shell is true, then $D_n^* = O(\log n)$ and the theorem allows to replace $n^{1/4}$ by $\log n$ in (4.4). These inequalities have a connection with yet another famous conjecture from asymptotic convex geometry, the *hyperplane conjecture*.

The hyperplane (or slicing) conjecture asserts that there exists a universal constant $c > 0$ such that every convex body K in $\mathbb{R}^n$ of Lebesgue measure 1 admits a hyperplane section whose $(n-1)$-dimensional Lebesgue measure is at least c. This conjecture can be rephrased in terms of isotropic constant. Recall that the isotropic constant of an isotropic log-concave vector X in $\mathbb{R}^n$ is $L_X = \rho(0)^{1/n}$, where ρ is the density of X. Letting

$$L_n = \sup\{L_X;\ X \text{ isotropic and log-concave on } \mathbb{R}^n\},$$

one can show that the slicing conjecture is equivalent to the statement $L_n = O(1)$. In [12], Bourgain derives from (4.4) the estimate $L_n = O(n^{1/4} \log n)$. Up to the factor $\log n$ this is still

the best general estimate as of today, the logarithmic improvement being due to Klartag [40]. Following Bourgain, we derive from Theorem 4.1 the following inequality

$$L_n = O\left((D_n^*)^{1/2}(\log n)^{3/2}\right).$$

In particular, if thin-shell holds true then so does slicing, up to a $(\log n)^2$ factor. This corollary of our main theorem is worth noticing, but it is actually nothing new. Indeed, in [24], Eldan and Klartag proved by a completely different method that actually $L_n = O((D_n)^{1/2})$.

Now let us say a few words about the proof of Theorem 4.1. As in the previous section, we use stochastic localization, but in a rather different way. Let μ be an isotropic log-concave measure and let (μ_t) be the stochastic localization process starting from μ. Recall that a_t and A_t denote the barycenter and the covariance matrix of μ_t, respectively. Using Itô's formula it is easily seen that

$$da_t = A_t^{1/2} dB_t,$$

and that $(\mathrm{e}^t A_t)$ is a martingale, in particular $\mathsf{E}[A_t] = \mathrm{e}^{-t} A_0 = \mathrm{e}^{-t} I_n$. This implies that the martingale (a_t) is bounded in L^2, hence convergent by Doob's theorem. Moreover, since μ_t equals μ in average and since the covariance of μ_t tends to 0, it is not hard to see that the limit a_∞ has law μ. So we have constructed a martingale (a_t) that converges to μ in law. Its quadratic covariation is $[a]_t = \int_0^t A_s \, ds$, and we have seen in the previous section that this is controlled by the thin-shell constant D_n^*. More specifically, it follows from (4.2) that

$$[a]_t \leq K' \, D_n^* \log n \, I_n, \quad \forall t > 0,$$

with probability at least 0.9 (say).

We combine this information with the following correlation inequality, which we learnt from Maurey. Let $(M_t)_{t \geq 0}$ be a continuous martingale taking values in $\mathbb{R}^n$. Assume that $M_0 = 0$ and that the quadratic variation of M satisfies

$$[M]_t \leq I_n, \quad \forall t > 0,$$

almost surely. Then (M_t) converges almost surely, and the law μ of M_∞ satisfies

$$\int_{\mathbb{R}^n} \phi \, d\mu \leq \int_{\mathbb{R}^n} \phi \, d\gamma_n,$$

for every convex function $\phi \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. We abbreviate this inequality as $\mu \prec \gamma_n$ in the sequel. This relation is called *convex ordering of measures* in the literature. A famous theorem of Strassen asserts that $\mu \prec \nu$ if and only if there exists a coupling (X, Y) of (μ, ν) such that $\mathsf{E}[Y \mid X] = X$. Let us sketch Maurey's argument briefly. Observe first that the convergence is given by Doob's theorem. The idea is to simply add the missing covariation. More precisely, let $\mathcal{F}$ be the σ-field generated by (M_t) and let G be a standard Gaussian vector on $\mathbb{R}^n$ that is independent of $\mathcal{F}$. We claim that

$$X = M_\infty + (I_n - [M]_\infty)^{1/2} G$$

is a standard Gaussian. Indeed, given $x \in \mathbb{R}^n$ we have

$$\mathsf{E}\left[\mathrm{e}^{\mathrm{i}\langle X,x\rangle} \mid \mathcal{F}\right] = \exp\left(\mathrm{i}\langle M_\infty, x\rangle + \frac{1}{2}\langle [M]_\infty x, x\rangle - \frac{1}{2}|x|^2\right).$$

Now Itô's formula shows that $\exp\left(\mathrm{i}\langle x, M_t\rangle + \frac{1}{2}\langle [M]_t x, x\rangle\right)$ is a martingale, so that taking expectation in the previous equation yields the claim. On the other hand, since G has mean 0 and his independent of $\mathcal{F}$, we have $\mathsf{E}[X \mid \mathcal{F}] = M_\infty$. Maurey's inequality then follows by Jensen's inequality (or the easy part in Strassen's theorem).

Let us come back to our martingale (a_t). Maurey's inequality does not apply directly because we only have a control on the quadratic variation of (a_t) with high probability. On the other hand, we only want the conclusion for norms and not for all convex functions ϕ. Playing around a bit with Maurey's proof, we obtain the desired inequality in a few lines that we omit here.

We close this section with another inequality obtained almost for free with the same method. If μ is a log-concave perturbation of γ_n, namely if μ has a log-concave density with respect to γ_n, then one can show that almost surely $A_t \leq \mathrm{e}^{-t} I_n$ for all t. We will omit the proof here, but let us mention that it is much easier to prove than (4.2). This implies that $[a]_t \leq I_n$ for all t, and combining this with Maurey's inequality we get the following correlation inequality due to Hargé [36].

Theorem 4.2. *Let μ be a probability measure which is a log-concave perturbation of γ_n having its barycenter at 0. Then $\mu \prec \gamma_n$.*

Let us conclude this chapter by saying that the stochastic localization is a really powerful method that has not been much exploited yet. We expect further applications to be discovered in a near future.

Chapter 5

Cover time and generic chaining

This chapter is based on the article [46].

5.1 Gaussian processes and generic chaining

Let $(X_t)_{t\in T}$ be a Gaussian process. It is known since Kolmogorov that estimating $\mathsf{E}\left[\sup_{t\in T}\{X_t\}\right]$ amounts to studying the metric space (T, d_X) where d_X is the distance given by

$$d_X(s,t)^2 = \mathsf{E}[(X_s - X_t)^2]$$

for every $s, t \in T$. For instance, Slepian's lemma [58] asserts that if (X_t) and (Y_t) are two centered Gaussian processes indexed by the same set T and are such that $d_X \leq d_Y$, then

$$\mathsf{E}[\sup_{t\in T}\{X_t\}] \leq \mathsf{E}[\sup_{t\in T}\{Y_t\}].$$

Here and in the sequel we say that a process (X_t) is centered if $\mathsf{E}[X_t] = 0$ for all t. Many authors, such as Sudakov, Dudley and Fernique, to name only a few of them, have contributed to this field of research but the most important result is surely Talagrand's *majorizing measure* theorem [59]. Let us present it in its later reformulation, still due to Talagrand, which is called *generic chaining*. We refer to the book [62] for a more accurate history of the problem.

Throughout we let $(N_n)_{n\geq 0}$ be the following sequence of integers:

$$N_0 = 1, \quad \text{and} \quad N_n = 2^{2^n},\ \forall n \geq 1. \tag{5.1}$$

Given a set T, a sequence $(T_n)_{n\geq 0}$ of subsets of T is called *admissible* if it is increasing (for the inclusion) and if $\mathrm{card}(T_n) \leq N_n$ for every $n \geq 0$. An admissible sequence (T_n) should be thought as sequence of approximations of T. Given a distance d we measure how good is this sequence by setting

$$J_2((T_n)) = \sup_{t\in T}\left\{\sum_{n=0}^{+\infty} 2^{n/2} d(t, T_n)\right\},$$

where as usual $d(t, T_n) = \min_{s\in T_n}\{d(t,s)\}$. Note that the factor $2^{n/2}$ is precisely $\sqrt{\log N_n}$ which is the cardinality of the set T_n. Next we optimize over admissible sequences and we set

$$\gamma_2(T,d) = \inf\left\{J_2((T_n));\ (T_n) \text{ admissible}\right\}.$$

This quantity should be thought as a measure of complexity of the metric space (T, d). The fundamental result of Talagrand asserts that if $(X_t)_{t\in T}$ is a a centered Gaussian process then $\mathsf{E}\left[\sup_{t\in T}\{X_t\}\right]$ and $\gamma_2(T, d_X)$ coincide, up to a universal factor:

$$\frac{1}{L}\gamma_2(T, d_X) \leq \mathsf{E}\left[\sup_{t\in T}\{X_t\}\right] \leq L\gamma_2(T, d_X) \tag{5.2}$$

where L is a universal constant. The upper bound is not specific to Gaussian processes. A process $(X_t)_{t\in T}$ is called *subgaussian* if it satisfies

$$\mathsf{P}(|X_s - X_t| \geq u) \leq \mathrm{e}^{-u^2/2d(s,t)^2}$$

for all $s, t \in T$, for all $u > 0$ and for some distance d. Observe that a Gaussian process satisfies this inequality with $d = d_X$. Using a union bound it is not hard to see that if (X_t) is subgaussian then

$$\mathsf{E}\left[\sup_{t\in A}\{X_t\}\right] \leq C\sqrt{\log \operatorname{card}(A)}\max_{s,t\in A}\{d(s, t)\}, \tag{5.3}$$

for every finite subset A of T, where C is a universal constant. Applying (5.3) repeatedly along the sets of an admissible sequence, one can then establish the right hand side of (5.2), see [62, section 2.2.]. This is what Talagrand calls *chaining*. The lower bound is another story, it is specific to Gaussian processes and much more difficult to prove. The main tool is the Sudakov inequality: if $(X_t)_{t\in T}$ is a centered Gaussian process then for every finite subset A of T we have

$$\mathsf{E}\left[\sup_{t\in T}\{X_t\}\right] \geq c\sqrt{\log \operatorname{card}(A)}\min_{s\neq t\in A}\{d_X(s, t)\} \tag{5.4}$$

where c is a universal constant. This inequality can be derived from Slepian's lemma.

5.2 Commute distance and cover times

Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain on a finite state space M. Given $A \subset M$ let

$$T(A) = \inf\{n \geq 0 \colon X_n \in A\}$$

be the first time the chain hits A and let $T_{\mathrm{cov}}(A) = \sup_{x\in A}\{T(x)\}$ be the first time the chain X has visited every point of A. The cover time of A is by definition

$$\mathrm{cov}(A) = \sup_{x\in A}\left\{\mathsf{E}_x[T_{\mathrm{cov}}(A)]\right\},$$

where E_x stands for conditional expectation given $X_0 = x$. Similarly P_x stands for conditional probability given $X_0 = x$ in the sequel. Using the strong Markov property it is easily seen that given x, y, z in M, the quantity $\mathsf{E}_x[T(y)] + \mathsf{E}_y[T(z)]$ is the expectation, starting from x, of the first time the chain has visited y and z, in this order. This implies that $\mathsf{E}_x[T(y)] + \mathsf{E}_y[T(z)] \geq \mathsf{E}_x[T(z)]$. As a result the commute time

$$d(x, y) = \mathsf{E}_x[T(y)] + \mathsf{E}_y[T(x)]$$

is a distance on M. The following inequalities are due to Matthews [51]. For every subset A of M we have

$$\begin{aligned} \mathrm{cov}(A) &\leq 2 \log \mathrm{card}(A) \max_{x,y \in A} \{\mathsf{E}_x[T(y)]\} \\ \mathrm{cov}(A) &\geq \log \mathrm{card}(A) \min_{x \neq y \in A} \{\mathsf{E}_x[T(y)]\} . \end{aligned} \tag{5.5}$$

These inequalities are very much reminiscent of (5.3) and (5.4). This analogy suggests that a result similar to Talagrand's theorem for suprema of Gaussian processes should hold for cover times of Markov chains. This question was investigated by several authors, see for instance [6] and the references therein. An arguably definitive answer was given by Ding, Lee and Peres in [21]. Namely, they proved that for any reversible and irreducible Markov chain on a finite state space M we have

$$\frac{1}{L} \left(\gamma_2(M, \sqrt{d}) \right)^2 \leq \mathrm{cov}(M) \leq L \left(\gamma_2(M, \sqrt{d}) \right)^2 , \tag{5.6}$$

where d is the commute distance and L is a universal constant. Their proof relies on the Ray-Knight isomorphism theorem, which gives an identity in law between local times of the chain and its Gaussian free field. This allows to translate the problem in terms of the Gaussian free field, and to work things out at the level of this Gaussian process. The whole argument is very clever but somewhat frustrating, there should be a simpler proof of (5.6) based only on elementary hitting times estimates such as Matthews' bound and Talagrand's generic chaining.

The article [46] is an attempt to provide such a proof. Unfortunately it fails to recover (5.6) completely. Here is what we prove, the statement involves γ_1, which is obtained by minimizing

$$J_1((T_n)) = \sup_{t \in T} \left\{ \sum_{n=0}^{+\infty} 2^n \, d(t, T_n) \right\}$$

over admissible sequences (T_n).

Theorem 5.1. *For every irreducible Markov chain on a finite state space M we have*

$$\mathrm{cov}(M) \leq L \left(\gamma_2(M, \sqrt{d}) \right)^2 , \tag{5.7}$$

where d is the commute distance and L is a universal constant. If in addition the chain is reversible then

$$\mathrm{cov}(M) \geq \frac{1}{L} \gamma_1(M, d).$$

Our upper bound is slightly stronger than that of Ding , Lee and Peres since it does not require reversibility. Using the inequality

$$\sum_{i \leq n} x_i^2 \leq \left(\sum_{i \leq n} x_i \right)^2 \leq n \sum_{i \leq n} x_i^2$$

it is easily seen that

$$\gamma_1(M, d) \leq \left(\gamma_2(M, \sqrt{d}) \right)^2 \leq C \, \log \log \mathrm{card}(M) \, \gamma_1(M, d),$$

where C is a universal constant. So while our lower bound is weaker than that of Ding, Lee and Peres it can only off the truth by a factor $\log\log \mathrm{card}(M)$ at most. We give an example in [46] showing that this is sharp, there exists a Markov chain for which the gap is indeed $\log\log \mathrm{card}(M)$. Let us also mention that the reversibility assumption is necessary. Indeed, consider the discrete torus $\mathbb{Z}_N$ and the Markov kernel given by $P(x, x+1) = 1$ for every x. Then clearly $d(x, y) = N$ for all $x \neq y$, which implies that $\gamma_1(\mathbb{Z}_N, d) \approx N \log N$. On the other hand $T_{cov}(\mathbb{Z}_N) = N$ almost surely (for every starting point).

5.3 Sketch of proof

Let us say a few words about the upper bound. Since the process of hitting times $(T(x))_{x\in M}$ is not subgaussian in general, Talagrand's generic chaining does not apply directly and one has to build a chaining procedure *ad hoc*. Actually this already exists in the literature, and we simply improve a result of Barlow, Ding, Nachmias and Peres [6], which is essentially a weaker version of (5.7), obtained by swapping the supremum in t and the sum in n in the definition of γ_2. We will not say more about this part here. Instead we focus on the lower bound, which is more interesting.

To bound γ_2 (or γ_1) from above requires to actually construct a good admissible sequence (T_n). Talagrand provides a general strategy to do that, which is the backbone of the generic chaining theory. Let (T, d) be a metric space, a functional $F\colon \mathcal{P}(T) \to \mathbb{R}_+$ is said to satisfy the growth condition if for every scale $a > 0$ and every integer m the followings holds: for every sequence $H_1, \dots, H_m$ of non-empty subsets of T satisfying $d(H_i, H_j) \geq a$ for all $i \neq j$ and $\mathrm{diam}(H_i) \leq a/r$ for all i we have

$$F(\cup_{i\leq m} H_i) \geq c\sqrt{\log m} + \min_{i\leq m} F(H_i), \tag{5.8}$$

where r and c are positive parameters. This has to be thought as a generalized abstract Sudakov inequality. The key result of Talagrand is that if F satisfies the growth property then essentially $\gamma_2(T, d) \lesssim F(T)$. The precise result depends on the parameters r and c but we treat these as absolute constants here. For instance if $(X_t)_{t\in T}$ is a Gaussian process, combining the Sudakov inequality with the Gaussian concentration property it is pretty straightforward to show that the functional $F(A) = \mathsf{E}[\sup_{t\in A}\{X_t\}]$ satisfies the growth condition on the metric space (T, d_X). Hence the majorizing measure theorem $\gamma_2(T, d_X) \lesssim \mathsf{E}[\sup_{t\in T}\{X_t\}]$.

The theorem also works for γ_1, the only difference is that $\sqrt{\log m}$ should be replaced by $\log m$ in the definition of the growth condition (5.8). Proving the lower bound in our main theorem thus amounts to proving that if (X_n) is a reversible Markov chain on a finite state space M then the functional $F(A) = \mathrm{cov}(A)$ satisfies the γ_1-growth condition. Let us sketch the argument. We start with a short proof of Matthews' (lower) bound. Let A be a subset of M, set $m = \mathrm{card}(A)$ and assume that

$$\mathsf{E}_x[T(y)] \geq a, \quad \forall x \neq y \in A. \tag{5.9}$$

Let $x \in A$ and let y be the element of A whose probability to be visited last starting from x is the largest. Then set $T = T_{cov}(A)$ and $S = T_{cov}(A\backslash\{y\})$. At time S, either we have already visited y, in which case $S = T$, or we have yet to visit y. By definition of y the second situation happens with probability $1/(m-1)$ at least, and the visit to y takes at least time a in average, by (5.9).

Therefore

$$\mathsf{E}_x[T] \geq \mathsf{E}_x[S] + \frac{a}{m-1}.$$

An obvious induction then shows that

$$\mathsf{E}_x[T_{\mathrm{cov}}(A)] \geq \sum_{k=1}^{m-1} \frac{a}{k} \geq a \log m,$$

which is the result. The first improvement we need to make is to establish the same conclusion assuming that $d(x,y) \geq a$ for every $x \neq y \in A$, rather than (5.9). This is where reversibility enters the picture. Recall that

$$d(x,y) = \mathsf{E}_x[T(y)] + \mathsf{E}_y[T(x)].$$

While there is no guarantee that the two terms have the same order of magnitude for every x, y in A, reversibility insures that this is the case for a constant proportion of elements of A. We will not detail this part of the argument here, let us just say that we borrowed it from the article [38]. This shows that the functional $\mathrm{cov}(A)$ satisfies Sudakov's inequality: if $d(x,y) \geq a$ for every $x \neq y \in A$ then $\mathrm{cov}(A) \geq c\, a \log \mathrm{card}(A)$, where c is a universal constant. Having observed this, it is only a matter of perturbing the proof of Matthews' bound given above to establish the full growth condition.

As we can see, elementary hitting estimates à la Matthews fit so perfectly the Talagrand machinery that it is really disappointing to get a bound that is slightly off the correct one at the end. We have tried to establish the correct bound with the same method but we could not prove the corresponding growth condition. Of course this does not mean that it is impossible, maybe we just missed something.

Chapter 6

Functional Santaló and a reversed log-Sobolev inequality

This chapter is based on our joint work with U. Caglar, M. Fradelizi, O. Guédon, C. Schütt, and E. Werner [15].

6.1 The Blaschke-Santaló inequality

If A is a subset of $\mathbb{R}^n$ we let $A^\circ = \{y \in \mathbb{R}^n : \langle x, y\rangle \leq 1,\ \forall x \in A\}$ be its polar. Restricted to convex bodies containing 0 in their interior, this operation becomes an involution. Its only fixed point is the Euclidean unit ball B_2^n. The Blaschke-Santaló inequality asserts that if K is a convex body having its center of mass at 0, i.e. $\int_K x\,dx = 0$, then

$$|K|\,|K^\circ| \leq |B_2^n|^2, \tag{6.1}$$

where $|\cdot|$ denotes the Lebesgue measure. Note that when 0 is on the boundary of K then K° contains a halfspace, so the inequality cannot be true for all bodies K and some centering hypothesis is needed. There is equality in (6.1) if and only if K is an ellipsoid centered at 0. Blaschke proved the inequality, as well as the equality case, in dimension 2 and 3. His argument was extended to arbitrary n by Santaló [57]. In [52], Meyer and Pajor give an elegant proof based on Steiner's symmetrization.

Let us present now a functional form of Santaló's inequality. Recall that the Legendre transform of a function ϕ is

$$\phi^*(y) = \sup_{x \in \mathbb{R}^n} \left\{\langle x, y\rangle - \phi(x)\right\}.$$

Restricted to proper lower semi-continuous convex functions this operation is again an involution. It also has a unique fixed point, namely $\phi(x) = |x|^2/2$. The functional Santaló inequality asserts that for every convex function ϕ satisfying $\int_{\mathbb{R}^n} x\,\mathrm{e}^{-\phi(x)}\,dx = 0$ we have

$$\int_{\mathbb{R}^n} \mathrm{e}^{-\phi(x)}\,dx \int_{\mathbb{R}^n} \mathrm{e}^{-\phi^*(y)}\,dy \leq \left(\int_{\mathbb{R}^n} \mathrm{e}^{-|x|^2/2}\,dx\right)^2 = (2\pi)^n. \tag{6.2}$$

Moreover there is equality if and only if $\phi(x) = \langle Ax, x\rangle$ for some symmetric positive definite matrix A. Applying this to $\phi(x) = \frac{1}{2}\|x\|_K^2$, where

$$\|x\|_K = \inf\{r > 0 \colon x \in rK\}$$

is the gauge associated to a convex body K, one can recover the Blaschke-Santaló inequality in a few lines which we omit here. The inequality (6.2) was first proved by Keith Ball [3] under the additional assumption that ϕ is even. The above statement is due to Artstein, Klartag and Milman in [1]. Both proofs rely on the usual Santaló inequality, by applying it to level sets of the function ϕ. Let us also mention that in the article [43], which was part of our Ph.D. thesis, we gave a direct proof of (6.2) by induction on the dimension, hence providing a new proof of the Blaschke-Santaló inequality.

6.2 The reversed log-Sobolev inequality

Recall first the logarithmic Sobolev inequality for the Gaussian measure

$$\mathrm{H}(\mu \mid \gamma_n) \leq \frac{1}{2}\,\mathrm{I}(\mu \mid \gamma_n) \tag{6.3}$$

Let us translate this inequality in terms of the Lebesgue measure. Let ρ be the density of μ with respect to the Lebesgue measure, let $\mathrm{S}(\mu) = -\int_{\mathbb{R}^n} \log \rho \, d\mu$ be its Shannon entropy and $\mathrm{J}(\mu) = \int_{\mathbb{R}^n} |\nabla \log \rho|^2 \, d\mu$ be its Fisher information. It is easily seen that (6.3) is equivalent to

$$\mathrm{S}(\gamma_n) - \mathrm{S}(\mu) \leq \frac{n}{2}\,(\mathrm{J}(\mu) - n)\,.$$

Note the different sign convention between S and H. The quantity $\mathrm{S}(\gamma_n) - \mathrm{S}(\mu)$ is called the *entropy gap*. Observe that if X is a random vector and λ is a positive real number then $\mathrm{S}(\lambda X) = \mathrm{S}(X) + n \log \lambda$ and $\mathrm{J}(\lambda X) = \lambda^{-2}\mathrm{J}(X)$. So the previous inequality is not invariant under scaling. As a result we can apply it to λX and optimize in λ. This yields

$$\mathrm{S}(\gamma_n) - \mathrm{S}(\mu) \leq \frac{n}{2} \log\left(\frac{\mathrm{J}(\mu)}{n}\right). \tag{6.4}$$

This improved version of the log-Sobolev inequality is sometimes referred to as the Euclidean log-Sobolev inequality in the literature. In [15], we derive from the functional Santaló inequality a reversed form of this inequality for log-concave measures. Let us describe this short argument.

Let ϕ be a convex function, assume that $\int_{\mathbb{R}^n} \mathrm{e}^{-\phi}\, dx = 1$ and let μ be the probability measure having density $\mathrm{e}^{-\phi}$ with respect to the Lebesgue measure. Assuming furthermore that μ has its barycenter at 0, we have by the functional Santaló inequality

$$\int_{\mathbb{R}^n} \mathrm{e}^{-\phi^*(y)}\, dy \leq (2\pi)^n. \tag{6.5}$$

Recall that $\phi(x)+\phi^*(y) \geq \langle x, y\rangle$ for every $x, y \in \mathbb{R}^n$, with equality if y belongs to the subdifferential of ϕ at x. Assuming that $\nabla\phi$ is a diffeomorphism from the domain of ϕ to that of ϕ^* we get using

the change of variable formula

$$\int_{\mathbb{R}^n} \mathrm{e}^{-\phi^*(y)}\, dy = \int_{\mathbb{R}^n} \mathrm{e}^{-\langle x, \nabla\phi(x)\rangle + \phi(x)} \det(\nabla^2\phi(x))\, dx$$
$$= \int_{\mathbb{R}^n} \mathrm{e}^{-\langle x, \nabla\phi(x)\rangle + 2\phi(x)} \det(\nabla^2\phi(x))\, \mu(dx).$$

Taking the logarithm, using Jensen's inequality and the inequality (6.5) we obtain

$$-\int_{\mathbb{R}^n} \langle x, \nabla\phi(x)\rangle\, \mu(dx) + 2\,\mathrm{S}(\mu) + \int_{\mathbb{R}^n} \log\det(\nabla^2\phi)\, \mu(dx) \le n\log(2\pi).$$

Lastly an integration by parts shows that $\int_{\mathbb{R}^n} \langle x, \nabla\phi\rangle\, d\mu = n$. Putting everything together we get the following result, which was obtained first by Artstein, Klartag, Schütt and Werner in [2].

Theorem 6.1. *Let μ be a log-concave probability measure and let ϕ be its potential, namely $\phi = -\log\rho$ where ρ is the density of μ. Then*

$$\frac{1}{2}\int_{\mathbb{R}^n} \log\det(\nabla^2\phi)\, d\mu \le \mathrm{S}(\gamma_n) - \mathrm{S}(\mu).$$

The inequality was proved under the additional assumption that μ has its barycenter at 0, but since both terms of the conclusion are invariant under translations of μ, this assumption is omitted in the statement of the theorem. Let us note that the original statement of Artstein, Klartag, Schütt and Werner contained additional regularity assumptions on ϕ. Also, in the proof given above $\nabla\phi$ is assumed to be smooth and one to one. Without going into the details, let us say that using a change of variable formula for gradients of convex functions due to McCann (see the appendix of [50]), it can be shown that the theorem holds without any regularity assumption on ϕ, provided the Hessian of ϕ is understood in the Aleksandrov sense. Let us remark also that an integration by parts gives $\mathrm{J}(\mu) = \int_{\mathbb{R}^n} \Delta\phi\, d\mu$. So our lower bound for the entropy gap is obtained from the upper bound in the Euclidean log-Sobolev inequality (6.4) by putting the logarithm inside the integral and replacing $\Delta\phi/n$ by $\det(\nabla^2\phi)^{1/n}$, which is smaller by the arithmetic mean/geometric mean inequality.

The argument of Artstein, Klartag, Schütt and Werner is much longer and much more technical than ours. Their main tool is an inequality called *affine isoperimetric inequality* which they apply to level sets of the function ϕ. The theorem can thus be interpreted as a functional form of the affine isoperimetric inequality.

The rest of the article [15] contains variations upon the above argument, about which we will only say a few words. As we have seen, the reversed log-Sobolev inequality is obtained by taking the logarithm of the functional Santaló inequality and using Jensen's inequality. The choice of the logarithm is rather arbitrary and one can try other concave functions such as x^α for $\alpha \in (0,1)$. We also apply the same scheme to the other functional forms of Santaló's inequality that exist in the literature. This gives rise to a family of inequalities which we call *functional affine isoperimetric inequalities*.

Bibliography

[1] S. Artstein-Avidan, B. Klartag, and V. Milman. The Santaló point of a function, and a functional form of Santaló inequality. *Mathematika* 51 (2005), no. 1–2, 33–48.

[2] S. Artstein-Avidan, B. Klartag, C. Schütt, and E. Werner. Functional affine-isoperimetry and an inverse logarithmic Sobolev inequality. *J. Funct. Anal.* 262 (2012), no. 9, 4181–4204.

[3] K. Ball. Isometric problems in ℓ_p and sections of convex sets. Doctoral thesis, University of Cambridge, 1986.

[4] D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators. Grundlehren der Mathematischen Wissenschaften, 348. *Springer, Cham*, 2014.

[5] K. Ball, F. Barthe, W. Bednorz, K. Oleszkiewicz, and P. Wolff. L_1-smoothing for the Ornstein-Uhlenbeck semigroup. *Mathematika* 59 (2013), no. 1, 160–168.

[6] M.T. Barlow, J. Ding, A. Nachmias, and Y. Peres, The evolution of the cover time, *Combin. Probab. Comput.* 20 (2011), no. 3, 331–345.

[7] F. Barthe. On a reverse form of the Brascamp-Lieb inequality, *Invent. Math.* 134 (1998), no 2, 335–361.

[8] F. Baudoin. Conditioned stochastic differential equations: theory, examples and application to finance. *Stochastic Process. Appl.* 100 (2002), 109–145.

[9] S.G. Bobkov. On isoperimetric constants for log-concave probability distributions. *Geometric aspects of functional analysis*, 81–88, Lecture Notes in Math., 1910, *Springer, Berlin*, 2007.

[10] C. Borell. Diffusion equations and geometric inequalities. *Potential Anal.* 12 (2000), no. 1, 49–71.

[11] M. Boué, and P. Dupuis. A variational representation for certain functionals of Brownian motion, *Ann. Probab.* 26 (1998), no. 4, 1641–1659.

[12] J. Bourgain. On the distribution of polynomials on high-dimensional convex sets. *Geometric aspects of functional analysis*, 127–137, Lecture Notes in Math., 1469, *Springer, Berlin*, 1991.

[13] H.J. Brascamp, and E.H. Lieb. Best constants in Young's inequality, its converse and its generalization to more than three functions. *Adv. Math.* 20 (1976), 151–173.

[14] S. Bubeck, R. Eldan, and J. Lehec. Sampling from a log-concave distribution with Projected Langevin Monte Carlo. Preprint, arXiv: 1507.02564, 2015.

[15] U. Caglar, M. Fradelizi, O. Guédon, J. Lehec, C. Schütt, and E. Werner. Functional Versions of L_p-Affine Surface Area and Entropy Inequalities. *Int. Math. Res. Not. IMRN* 2016 (2016), no. 4, 1223–1250.

[16] E.A. Carlen, E.H. Lieb, and M. Loss. A sharp analog of Young's inequality on $\mathbb{S}^N$ and related entropy inequalities. *J. Geom. Anal. 14* (2004), no. 3, 487–520.

[17] W-K. Chen, N. Dafnis, and G. Paouris. Improved Hölder and reverse Hölder inequalities for Gaussian random vectors. *Adv. Math.* 280 (2015), 643–689.

[18] D. Cordero-Erausquin, and B. Maurey. Some extensions of the Prékopa-Leindler inequality using Borell's stochastic approach. Preprint, arXiv: 1512.05131, 2015.

[19] D. Cordero-Erausquin, R.J. McCann, and M. Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.* 146 (2001), no. 2, 219-257.

[20] A. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. preprint, arXiv: 1412.7392, 2014.

[21] J. Ding, J.R. Lee, and Y. Peres. Cover times, blanket times, and majorizing measures. *Ann. of Math.* 175 (2012), no. 3, 1409–1471.

[22] M. Dyer, A. Frieze, and R. Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM)* 38 (1991), no. 1, 1–17.

[23] R. Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geom. Funct. Anal.* 23 (2013), no. 2, 532–569.

[24] R. Eldan, and B. Klartag. Approximately Gaussian marginals and the hyperplane conjecture. Concentration, functional inequalities and isoperimetry, 55–68, Contemp. Math., 545, *Amer. Math. Soc., Providence, RI*, 2011.

[25] R. Eldan, and J.R. Lee. Regularization under diffusion and anti-concentration of temperature. Preprint, arXiv: 1410.3887, 2015.

[26] R. Eldan, and J. Lehec. Bounding the norm of a log-concave vector via thin-shell estimate. *Geometric Aspects of Functional Analysis*, 107–122, Lecture Notes in Math. 2116, *Springer, Cham*, 2014.

[27] M. Émery. En cherchant une caractérisation variationnelle des martingales. *Séminaire de Probabilités, XXII*, 147–154, Lecture Notes in Math., 1321, *Springer, Berlin*, 1988.

[28] M. Fathi, E. Indrei, and M. Ledoux. Quantitative logarithmic Sobolev inequalities and stability estimates. Preprint, arXiv: 1410.6922, 2014.

[29] D. Feyel, and A.S. Üstünel. Measure transport on Wiener space and the Girsanov theorem. *C. R. Math. Acad. Sci. Paris* 334 (2002), no. 11, 1025–1028.

[30] W.H. Fleming, and H.M. Soner. Controlled Markov processes and viscosity solutions. Second edition. Stochastic Modelling and Applied Probability, 25. *Springer, New York*, 2006.

[31] B. Fleury, O. Guédon, G. Paouris. A stability result for mean width of L_p-centroid bodies. *Adv. Math.* 214 (2007), no. 2, 865–877.

[32] H. Föllmer. Time reversal on Wiener space. *Stochastic processes-mathematics and physics (Bielefeld, 1984)*, 119–129, Lecture Notes in Math. 1158, *Springer, Berlin*, 1986.

[33] I. Gentil. Inégalités de Sobolev logarithmiques et hypercontractivité en mécanique statistique et en EDP. Thèse de Doctorat, Université Paul Sabatier, 2001.

[34] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.* 97 (1975), no. 4, 1061–1083.

[35] O. Guédon, and E. Milman. Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures. *Geom. Funct. Anal.* 21 (2011), no. 5, 1043–1068.

[36] G. Hargé. A convex/log-concave correlation inequality for Gaussian measure and an application to abstract Wiener spaces. *Probab. Theory Related Fields* 130 (2004), no. 3, 415–440.

[37] N. Ikeda, and S. Watanabe. Stochastic differential equations and diffusion processes. Second edition. North-Holland Mathematical Library, 24. *North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo*, 1989.

[38] J. Kahn, J.H. Kim, L. Lovász, and V.H. Vu, The cover time, the blanket time, and the Matthews bound. 41st Annual Symposium on Foundations of Computer Science, 467–475, *IEEE Comput. Soc. Press, Los Alamitos*, 2000.

[39] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete Comput. Geom.* 13 (1995), no. 3–4, 541–559.

[40] B. Klartag. On convex perturbations with a bounded isotropic constant. *Geom. Funct. Anal.* 16 (2006), no. 6, 1274–1290.

[41] B. Klartag. A central limit theorem for convex sets. *Invent. Math.* 168 (2007), 91–131.

[42] M. Ledoux. The concentration of measure phenomenon. Mathematical Surveys and Monographs, 89. *American Mathematical Society, Providence, RI*, 2001.

[43] J. Lehec. A direct proof of the functional Santaló inequality. *C. R. Math. Acad. Sci. Paris* 347 (2009), no. 1–2, 55–58.

[44] J. Lehec. Representation formula for the entropy and functional inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.* 49 (2013), no. 3, 885–899.

[45] J. Lehec. A short probabilistic proof of the Brascamp-Lieb and Barthe theorems. *Canad. Math. Bull.* 57 (2014), no. 3, 585–597.

[46] J. Lehec. Cover times and generic chaining. *J. Appl. Probab.* 51 (2014), no. 1, 247–261.

[47] J. Lehec. Regularization in L_1 for the Ornstein-Uhlenbeck semigroup. *Ann. Fac. Sci. Toulouse Math. (6)* 25 (2016), no. 2, 191–204.

[48] J. Lehec. Borell's formula for a Riemannian manifold and applications. Preprint, arXiv: 1512.05992, 2015.

[49] L. Lovász, and S. Vempala. Hit-and-run from a corner. *SIAM J. Comput.* 35 (2006), no. 4, 985–1005.

[50] R.J. McCann. A Convexity principle for interacting gases. *Adv. Math.* 128 (1997), no. 1, 153–179.

[51] P. Matthews, Covering problems for Brownian motion on spheres, *Ann. Probab.* 16 (1988), no. 1, 189–199.

[52] M. Meyer, and A. Pajor. On Santaló's inequality. *Geometric aspects of functional analysis (1987–88)*, 261–263, Lecture Notes in Math., 1376, *Springer, Berlin*, 1989.

[53] E. Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* 177 (2009), no. 1, 1–43.

[54] E. Nelson. The free Markoff field. *J. Functional Analysis* 12 (1973), 211–227.

[55] J. Neveu. Sur l'espérance conditionnelle par rapport à un mouvement brownien. *Ann. Inst. H. Poincaré Sect. B (N.S.)* 12 (1976), no. 2, 105–109.

[56] H. Robbins, and S. Monro. A stochastic approximation method, *Ann. Math. Statistics* 22, (1951), 400–407.

[57] L.A. Santaló. Un invariante afin para los cuerpos convexos del espacio de n dimensiones. *Portugal Math.* 8 (1949), no. 4, 155–161.

[58] D. Slepian. The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.* 41 (1962), 463–501.

[59] M. Talagrand. Regularity of Gaussian processes. *Acta Math.* 159 (1987), no. 1–2, 99–149.

[60] M. Talagrand. A conjecture on convolution operators, and a non-Dunford-Pettis operator on L_1. *Israel J. Math.* 68 (1989), no. 1, 82–88.

[61] M. Talagrand. Transportation cost for Gaussian and other product measures, *Geom. Funct. Anal.* 6 (1996), no. 3, 587–600.

[62] M. Talagrand. Upper and lower bounds for stochastic processes. Modern methods and classical problems. A Series of Modern Surveys in Mathematics, 60. *Springer, Heidelberg*, 2014.

[63] H. Tanaka. Stochastic differential equations with reflecting boundary condition in convex regions. *Hiroshima Math. J.* 9 (1979), no. 1, 163–177.

[64] S. Vempala. Geometric random walks: a survey. *Combinatorial and computational geometry*, 577–616, Math. Sci. Res. Inst. Publ. 52, *Cambridge Univ. Press, Cambridge*, 2005.