Phenomènes de concentration en grande dimension, transport de mesure et inegalites fonctionnelles.

Paul-Marie Samson

Université Paris-Est Marne-la-Vallée
Laboratoire d'Analyse et de Mathématiques Appliquées

Phénomènes de concentration en grande dimension, transport de mesure et inégalités fonctionnelles.

Paul-Marie Samson

Synthèse des travaux en vue de l'obtention de
l'**Habilitation à Diriger des Recherches**

Habilitation à Diriger des Recherches soutenue le **27 mai 2016** devant le jury composé de

Franck BARTHE,	Université de Toulouse 3
Djalil CHAFAÏ,	Université Paris- Dauphine
Michel LEDOUX,	Université de Toulouse 3
Christian LÉONARD,	Université Paris-Ouest Nanterre
Florence MERLEVÈDE,	Université Paris-Est Marne-la-Vallée
Laurent MICLO,	Université de Toulouse 3
Krzysztof OLESZKIEWICZ,	Université de Varsovie
Emmanuel RIO,	Université Paris-Saclay, Versailles

au vu des rapports de

Wilfrid GANGBO,	Université Georgia-Tech, Atlanta
Laurent MICLO,	Université de Toulouse 3
Krzysztof OLESZKIEWICZ,	Université de Varsovie

Table des matières

Remerciements **5**

Travaux présentés **7**

Préambule **9**

Présentation **11**

I Principe de concentration de la mesure **15**
I.1 Formulation ensembliste 15
Les fonctions de coût usuelles 16
Les fonctions de coût barycentriques 17
Les fonctions de coût universelles 17
I.2 Formulation fonctionnelle 18

II Inégalités de transport-entropie **21**
II.1 Introduction : concentration sur les espaces produits et propriétés isopérimétriques 21
II.2 Inégalités de transport et propriété de concentration 23
II.3 Formulation fonctionnelle des inégalités de transport, Théorème de Kantorovich 27
II.4 Tensorisation - Caractérisation par concentration adimensionnelle 29
II.5 Liens avec les inégalités de Sobolev logarithmiques 31

III Résultats autour des coûts barycentriques **37**
III.1 Coûts barycentriques et ordre convexe 37
III.2 Inégalités de transport barycentriques et inégalités de Sobolev logarithmiques 39
III.3 Inégalités de transport barycentriques pour la loi binômiale et la loi de Poisson 41
III.4 Plan de transports optimaux pour les coûts barycentriques sur $\mathbb{R}$ 42

III.5 Caractérisation des probabilités sur ℝ satisfaisant une inégalité de transport barycentrique 44

IV Inégalités de transport universelles 47

IV.1 En dimension 1 47

IV.2 Inégalités de transport pour des variables aléatoires faiblement dépendantes 50

IV.3 Inégalités de type Bernstein pour des suprema de processus empiriques indépendants 52

IV.4 Inégalités de transport pour la loi uniforme sur le groupe symétrique. 55

V Inégalité de Poincaré et concentration adimensionnelle 61

VI Notion de courbure 65

VI.1 Concentration et inégalité de Sobolev logarithmi-que sous hypothèse de courbure 66

VI.2 Notion de courbure dans les espaces discrets 70

VII Améliorations de l'inégalité de Sobolev logarithmique gaussienne 73

Remerciements

Je tiens tout d'abord à remercier Wilfrid Gangbo, Laurent Miclo et Krzysztof Oleskiewicz qui ont accepté de faire le rapport de ce mémoire, ainsi que Franck Barthe, Djalil Chafaï, Michel Ledoux, Christian Léonard, Florence Merlevède et Emmanuel Rio qui me font l'honneur de participer à mon jury. Je remercie Mathieu Meyer qui m'a proposé de relire la première version de ce mémoire.

Je remercie Michel Ledoux d'avoir orienté mes recherches en thèse sur une thématique à la croisée de nombreuses autres, qui alimentent aujourd'hui encore mes réflexions mathématiques.

Je remercie mes collaborateurs, Sergey Bobkov, Nathaël Gozlan, Cyril Roberto, Prasad Tetali pour les bons moments de recherche passés ensemble, source de confiance, de convialité, qui donnent sens au travail de recherche solitaire.

Je remercie les collègues du laboratoire pour les projets de recherche ou d'enseignement auxquels ils participent. Ces projets apportent la convivialité et le dynamisme à la vie notre laboratoire, ils m'ont porté au travail toutes ces années. Je tiens aussi à remercier mes collègues secrétaires et informaticiens pour leur attention au bon fonctionnement du laboratoire et de l'UFR, humainement et professionnellement. Leur présence quotidienne est le ciment de notre équipe. Je remercie aussi tous les animateurs occasionnels des pauses café.

Enfin mes derniers remerciements vont à mes proches, mes amis, mes amis collègues, ceux qui m'ont soutenu dans les aléas de la vie, qui me font confiance et me portent, voir parfois me supportent au quotidient. Ils me permettent aujourd'hui de présenter ce travail.

Paul-Marie Samson, Champs-sur-Marne, mai 2016.

Travaux présentés

[1] *Concentration of measure inequalities for Markov chains and Φ-mixing processes*. Ann. Probab. 28 (2000), no. 1, 416-461. (Dans la Thèse de Doctorat)

[2] *Concentration inequalities for convex functions on product spaces.* Stochastic inequalities and applications, 33-52, Progr. Probab., 56, Birkhäuser, Basel, 2003.

[3] *Infimum-convolution description of concentration properties of product probability measures, with applications.* Ann. Inst. H. Poincaré Probab. Statist. 43 (2007), no. 3, 321-338.

[4] *Isoperimetry for product of heavy tails distributions.* Progress in analysis and its applications, 470-478, World Sci. Publ., Hackensack, NJ, 2010. Avec N. Gozlan et C. Roberto.

[5] *From concentration to logarithmic Sobolev and Poincaré inequalities.* J. Funct. Anal. 260 (2011), no. 5, 1491-1522. Avec N. Gozlan et C. Roberto.

[6] *A new characterization of Talagrand's transport-entropy inequalities and applications.* Ann. Probab. 39 (2011), no. 3, 857-880. Avec N. Gozlan et C. Roberto.

[7] *Characterization of Talagrand's transport-entropy inequalities in metric spaces.* Ann. Probab. 41 (2013), no. 5, 3112-3139. Avec N. Gozlan et C. Roberto.

[8] *Hamilton Jacobi equations on metric spaces and transport entropy inequalities.* Rev. Mat. Iberoam. 30 (2014), no. 1, 133-163. Avec N. Gozlan et C. Roberto.

[9] *Displacement convexity of entropy and related inequalities on graphs.* Probab. Theory. Relat. Fields. (2013), 1-48. Avec N. Gozlan, C. Roberto et P. Tetali.

[10] *From dimension free concentration to Poincaré inequality.* Calculus of Variations and Partial Differential Equations. (2014), 1-22. Avec N. Gozlan et C. Roberto.

[11] *Bounds on the deficit in the logarithmic Sobolev inequality.* J. Funct. Anal. 267 (2014), no. 11, 4110–4138. Avec S. Bobkov, N. Gozlan et C. Roberto.

[12] *Kantorovich duality for general transport costs and applications.* Preprint hal-01098114v2, 2015 (soumis). Avec N. Gozlan, C. Roberto et P. Tetali.

[13] *Characterization of a class of weak transport-entropy inequalities on the line.* Preprint arXiv :1509.04202, 2015 (soumis). Avec N. Gozlan, C. Roberto, Y. Shu et P. Tetali.

[14] *Transport-entropy inequalities on locally acting groups of permutations.* Preprint.

Préambule

Ce document est une synthèse de mes recherches référencées précédemment. Voici un rapide descriptif du contenu de ces articles.

- Les articles [1], [2] et [3] établissent des inégalités de déviation pour des suprema de processus empiriques ou des normes de sommes de vecteurs aléatoires.
  - . L'article [1] établit ces résutats dans un cadre de dépendance faible. Les inégalités de déviation découlent d'inégalités de transport-entropie.
  - . Les articles [2] et [3] proposent des inégalités de déviations optimales dans le cadre indépendant, issues d'inégalités de moment exponentiel pour des opérateurs de type infimum-convolution.
- L'article [12], obtenu plus récemment, présente une extension du théorème de dualité de Kantorovich. Cette extension permet de donner la formulation fonctionnelle équivalente pour de nombreuses inégalités de transport-entropie, appelées inégalités de transport faibles. On y trouve de nouvelles inégalités de transport en particulier pour la loi binômiale et la loi de Poisson.
- L'article [13] présente une description des plans de transports optimaux entre deux mesures sur la droite réelle pour des coûts de transport faibles appelés coûts "barycentriques". Cette étude permet de caractériser les mesures sur $\mathbb{R}$ qui satisfont des inégalités de transport associées à ces coûts barycentriques.
- En préparation, l'article [14] présente une inégalité de transport-entropie faible pour la loi uniforme sur le groupe symétrique, associée à un principe de concentration introduit par Talagrand.
- L'article [5] étudie, dans les espaces métriques, les correspondances entre inégalités de Sobolev logarithmiques ou de Poincaré, et la propriété de concentration de la mesure sous des hypothèses de courbure.

– Dans [5] et [12], nous montrons par ailleurs que les inégalités de transport-entropie sont équivalentes à une propriété de concentration adimensionnelle. Dans le même ordre d'idées, dans [10], nous avons obtenu qu'une inégalité de Poincaré est satisfaite dès qu'un principe de concentration adimensionnelle faible est vérifié.
– Les articles [6], [7] et [8] concernent les relations entre les inégalités de type Sobolev logarithmiques et les inégalités de transport-entropie. En particulier, dans ces articles, une nouvelle caractérisation des inégalités de transport en termes d'inégalité de Sobolev logarithmique permet d'établir un résultat de stabilité des inégalités de transport, par perturbation bornée de la mesure considérée.
– Dans l'article [9], on s'intéresse à la notion de courbure sur les espaces discrets, les graphes, et les inégalités fonctionnelles qui en découlent.
– L'article [11] propose des versions améliorées de l'inégalité de Sobolev logarithmique gaussienne, qui quantifient l'erreur dans cette inégalité.
– Dans [4], on étudie la tensorisation des inégalités isopérimétriques pour les mesures à queues lourdes. Ce thème ne sera pas développé dans ce document car il est en marge de mes autres sujets de recherche.

Présentation

Mes recherches s'articulent autour des inégalités fonctionnelles associées au phénomène de concentration de la mesure, en particulier les inégalités de transport-entropie, dites aussi inégalités de transport. Les chapitres de ce document s'organisent autour de cette inégalité, des quantités qui la définissent, et de ses rapports avec d'autres inégalités fonctionnelles. La formulation générale et abstraite du principe de concentration de la mesure et des inégalités de transport qui lui sont associées, est ici choisie pour donner une unité aux différents résultats obtenus. Par ailleurs, certains éléments de preuve sont présentés, lorsqu'ils sont simples, pour une meilleure compréhension des outils mathématiques utilisés.

Le premier chapitre introduit le principe de concentration de la mesure. Sa formulation générale permet de comprendre et de mettre en parallèle de nombreux résultats de concentration suivant les types de fonctions de coût considérés. Je présente deux écritures strictement équivalentes de ce principe de concentration, l'une ensembliste (section I.1), l'autre fonctionnelle, associée à un nouveau type d'opérateur d'infimum-convolution (section I.2). Dans ce document, on distingue trois catégories de fonctions de coût, les fonctions de coût "usuelles", "barycentriques" et "universelles".

Le deuxième chapitre est consacré aux inégalités de transport-entropie dites "faibles" associées à la formulation générale du principe de concentration de la mesure.

Historiquement, dans les années 70, les propriétés de concentration en grande dimension ont été mises en avant par V. Milman, comme conséquences de propriétés isopérimétriques souvent difficiles à établir (voir section II.1). Les inégalités de transport fournissent une voie alternative pour établir ces propriétés de concentration.

Dans la section II.2, nous expliquons comment les propriétés de tensorisation des inégalités de transport en font un outil performant pour établir des propriétés de concentration en grande dimension, en particulier pour des probabilités produits. Ces arguments de tensorisation peuvent être adaptés pour démontrer des inégalités de transport sous des conditions de dépendances faibles des marginales des mesures étudiées.

La section II.3 du deuxième chapitre présente la forme duale fonctionnelle des inégalités de transport faibles (Proposition II.3.1). Cette formulation fonctionnelle est une conséquence de l'extension du théorème de dualité de Kantorovich obtenu dans l'article [12]. Ces résultats s'appuient sur une nouvelle définition de l'opérateur d'infimum-convolution, associée aux fonctions de coût faibles.

Dans la section II.4, nous donnons une démonstration simple de l'équivalence entre inégalité de transport faible et le principe de concentration adimensionnelle (Proposition II.4.2), en s'appuyant sur la forme fonctionnelle des inégalités de transport faibles.

Une partie de mes recherches concerne les liens entre inégalités de Sobolev logarithmiques et inégalités de transport. Les résultats obtenus en collaboration à ce sujet, présentés en section II.5, ont donné lieu à trois publications [6], [7] et [8] dont voici les grandes lignes.

D'après le Théorème d'Otto-Villani (Théorème II.5.1), l'inégalité de transport $\mathbf{T}_2$ de Talagrand est une conséquence stricte de l'inégalité de Sobolev logarithmique. L'article [6] propose une caractérisation de l'inégalité de transport $\mathbf{T}_2$ en termes d'inégalité de Sobolev logarithmique restreinte à une classe de fonctions semi-convexes. Un ingrédient essentiel de la preuve est l'expression des solutions de l'équation d'Hamilton-Jacobi sous la forme d'opérateur d'infimum-convolution, dite formule de Hopf-Lax.

On sait que si une probabilité satisfait l'inégalité de Sobolev logarithmique, alors une perturbation bornée de cette mesure la satisfait encore. La caractérisation des inégalités de transport-entropie en termes d'inégalités de Sobolev logarithmiques nous a permis d'étendre ce résultat de perturbation aux inégalités de transport (voir Corollaire II.5.1).

L'article [7] élargit aux espaces métriques le Théorème d'Otto-Villani et les résultats de [6] avec des fonctions de coût plus générales (voir Théorème II.5.2). Nous proposons une technique de démonstration robuste, très différente de l'article [6], qui s'appuie sur les propriétés de tensorisation des inégalités de Sobolev logarithmiques.

L'article [8] établit une formule de Hopf-Lax pour les solutions de l'equation d'Hamilton-Jacobi dans le cadre des espaces métriques, qui permet de montrer l'équivalence d'une inégalité de type Sobolev logarithmique avec une propriété d'hypercontractivité du semi-groupe d'Hamilton-Jacobi. Grâce à cette formule, nous retrouvons les résultats de l'article [7] sur les espaces métriques, en suivant l'approche de [5].

Dans le troisième chapitre, nous présentons divers résultats liés aux fonctions de coût dites barycentriques.

Nous exprimons le coût de transport optimal barycentrique en fonction du coût de transport optimal usuel, en considérant la notion d'ordre convexe sur les mesures (voir Proposition III.1.1).

Plusieurs des résultats du paragraphe II.5 sur les inégalités de transport avec des fonctions de coût usuels restent valables avec des fonctions de coûts barycentriques. En particulier, nous montrons dans [12] que les inégalités de transport $\overline{\mathbf{T}}_2^+$ ou $\overline{\mathbf{T}}_2^-$, associées à un coût barycentrique quadratique, sont équivalentes soit à une inégalité de Sobolev logarithmique restreinte à la classe des fonctions convexes, soit à une inégalité de Sobolev logarithmique restreinte à une classe de fonctions concaves (voir section III.2).

Rappelons qu'une mesure discrète sur $\mathbb{R}$ ne peut pas satisfaire l'inégalité de transport $\mathbf{T}_2$ de Talagrand. Ces dernières années, suite à l'étude de la notion de courbure sur les espaces discrets, d'autres inégalités de transport pour les mesures discrètes ont été proposées, en particulier dans les travaux de Erbar-Maas [Maa11, EM12]. Cependant, en raison du caractère différent de la définition du coût de transport, le phénomène de concentration associé à ces coûts est très abstrait. Les coûts de transport barycentriques, plus faibles que les coûts usuels, sont adaptés pour établir des inégalités de transport pour les mesures discrètes. Dans la section III.3, à titre d'exemples, nous donnons les inégalités de transport "barycentriques" obtenues pour la loi binômiale et la loi de Poisson.

La partie III.4 est consacrée à l'étude des coûts de transport barycentriques optimaux sur la droite réelle. Dans l'article [13], nous montrons l'existence d'un couplage optimal indépendant de la fonction convexe qui définit le coût. Ce résultat est l'analogue du résultat bien connu sur $\mathbb{R}$ pour les coûts de transport usuels : le transport monotone construit à partir des fonctions de répartition des lois considérées est optimal quelle que soit la fonction de coût convexe considérée.

L'indépendance du plan de transport optimal et de la fonction de coût permet de caractériser les mesures qui satisfont des inégalités de transport barycentriques sur la droite réelle, par le comportement des queues de leur distribution (voir Théorème III.5.1). En outre, dans l'article [13], nous avons montré l'équivalence entre une inégalité de Poincaré restreinte aux fonctions convexes sur $\mathbb{R}$, et une inégalité de transport barycentrique pour une fonction de coût bien choisie (voir Théorème III.5.2).

Les développements du chapitre IV s'appuient sur des exemples d'inégalités de transport "universelles", comme celle de Pinsker-Csizár-Kullback. Leur universalité réside dans le fait qu'elles sont vérifiées par toutes les probabilités. La plus emblématique, dans ce document, est l'inégalité de transport de Marton avec son coût faible $\widetilde{\mathcal{T}}_2$ [Mar96b]. Par tensorisation, cette inégalité donne un principe de concentration adimensionnelle qui fournit des inégalités de déviation de type Bernstein pour les supréma de processus empiriques.

Dans [1], l'inégalité de transport de Marton est démontrée pour des probabilités dont les marginales sont faiblement dépendantes. Dans [2] et [3], le coût faible de Marton $\widetilde{\mathcal{T}}_2$ est modifié et amélioré pour démontrer des inégalités de déviation de type Bernstein optimales, pour les suprema de processus empiriques bornés indépendants. Cette méthode de preuve est une alternative à la méthode dite de Herbst, utilisée pour la première fois par Ledoux pour obtenir les déviations des suprema de processus empiriques [Led97], puis largement développée par de nombreux auteurs.

La dernière section IV.4 présente les inégalités de transport obtenues récemment pour la loi uniforme sur le groupe symétrique. Elles découlent des inégalités de transport universelles comme l'inégalité de Pinsker-Csizár-Kullback ou encore l'inégalité $\widetilde{\mathbf{T}}_2$, par différentes techniques de tensorisation, proche des techniques de martingales, inspirées des preuves des résultats de concentration de Talagrand pour la loi uniforme sur le groupe symétrique [Tal95].

Dans le chapitre V sont étudiées l'inégalité de Poincaré et les propriétés de concentration qui lui sont associées. Au même titre que certaines inégalités de transport sont équivalentes à une propriété de concentration adimensionnelle, nous avons obtenu que l'inégalité de Poincaré est une conséquence d'un principe de concentration adimensionnelle très faible [10]. Ce résultat est encore valable si on restreint l'inégalité de Poincaré aux fonctions convexes, et dans ce cas, le principe de concentration adimensionnelle est issu d'un coût de transport barycentrique.

Le chapitre VI concerne la notion de courbure des espaces métriques. L'hypothèse de minoration de la courbure de Ricci sur les variétés Riemaniennes ou dans les espaces métriques géodésiques s'écrit en termes de propriété de convexité de l'entropie sur l'espace de Wasserstein. Lorsque la courbure est strictement positive, cette propriété implique immédiatement une inégalité de Sobolev logarithmique et une inégalité de transport, et par conséquent des propriétés de concentration. Autour de cette notion de courbure, nous avons obtenu des résultats dans deux directions.

Dans [5], nous montrons comment, inversement, une propriété de concentration implique une inégalité de Sobolev logarithmique ou une inégalité de Poincaré sous une hypothèse de courbure de Ricci minorée, éventuellement négative, dans les espaces géodésiques.

Dans une autre direction, dans [9], nous définissons une propriété de convexité de l'entropie sur les espaces discrets, l'objectif étant d'élargir la notion de courbure de Ricci minorée aux espaces discrets. Dans la seconde section du chapitre VI, nous présentons la construction naturelle de chemin de probabilité sur laquelle s'appuie cette définition. A titre d'exemple, nous donnons la propriété de convexité satisfaite par les probabilités produits sur le cube discret $\{0,1\}^n$.

Le chapitre VII est consacré à des versions améliorées de l'inégalité de Sobolev logarithmique et de l'inégalité HWI pour la mesure gaussienne standard sur $\mathbb{R}^n$. Ces améliorations visent à quantifier l'erreur dans l'inégalité de Sobolev logarithmique gaussienne. En particulier, en utilisant des résultats classiques de la théorie de l'information, nous avons amélioré l'inégalité HWI gaussienne. Ceci fournit une nouvelle preuve de l'inégalité HWI gaussienne. D'autres résultats permettent de retrouver les cas d'égalité dans l'inégalité de Sobolev logarithmique gaussienne (voir [11]).

CHAPITRE I

Principe de concentration de la mesure

Le phénomène de concentration de la mesure a été mis en avant par V. Milman au cours des années 70, dans l'étude de la géométrie asymptotique des espaces de Banach. Ce principe a des applications dans de nombreux domaines des mathématiques. Le livre de M. Ledoux [Led01], consacré à ce sujet, présente de nombreux exemples et des techniques probabilistes, analytiques et géométriques liées à cette notion.

Je souhaite ici décrire le principe de concentration de la mesure en le formulant de façon assez générale pour englober les divers phénomènes de concentration étudiés dans mes recherches. La formulation ensembliste de ce principe a ses origines dans les papiers de M. Talagrand [Tal95, Tal96b, Tal96a]. La formulation fonctionnelle présentée ensuite a été rigoureusement introduite dans [12].

I.1 Formulation ensembliste

Soit $(\mathcal{X}, d)$ un espace métrique polonais (séparable, complet), muni de sa tribu borélienne $\mathcal{B}(\mathcal{X})$. On notera $\mathcal{P}(A)$ l'ensemble des probabilités sur un borélien $A \in \mathcal{B}(\mathcal{X})$. Soit $q \geq 1$, on note $\mathcal{P}_q(A)$ l'ensemble des probabilités p sur A telles que pour un point x_0 on ait $\int d(x_0, y)^q dp(y) < +\infty$.

On suppose donnée une notion de pseudo-distance d'un point $x \in \mathcal{X}$ à un sous-ensemble $A \subset \mathcal{X}$, notée $c(x, A)$, et telle que $c(x, A) \in [0, +\infty]$ et $c(x, A) = 0$ si $x \in A$. L'exemple usuel est

$$c(x, A) = d(x, A) = \inf_{y \in A} d(x, y).$$

On peut aussi choisir

$$c(x, A) = \alpha(d(x, A)),$$

où α est une fonction positive telle que $\alpha(0) = 0$.

Pour $r \geq 0$, *l'élargissement* de l'ensemble A associé à cette pseudo-distance est noté

$$A_{r,c} = \{x \in \mathcal{X}, c(x, A) \leq r\}.$$

Définition I.1.1. *Soit $\beta : \mathbb{R}^+ \to [0, +\infty]$ une fonction croissante telle que $\beta(0) = 0$. On dit qu'une mesure de probabilité μ sur $\mathcal{X}$ satisfait un principe de concentration de profil β et de fonction de coût c s'il existe $a_1, a_2 > 0$ tels que pour tout $A \in \mathcal{B}(\mathcal{X})$ et tout $r \geq 0$,*

$$\mu(A)^{a_1}\mu(\mathcal{X} \setminus A_{r,c})^{a_2} \leq e^{-\beta(r)}.$$

Si μ satisfait le principe de concentration de profil β, alors pour tout $A \in \mathcal{B}(\mathcal{X})$ tel que $\mu(A) \geq 1/2$, on a

$$\mu(\mathcal{X} \setminus A_{r,c}) \leq 2^{a_1/a_2}e^{-\beta(r)/a_2} = e^{-\tilde{\beta}(r)}, \qquad \forall r \geq 0.$$

Cette dernière propriété est la formulation usuelle du principe de concentration de la mesure. En fait, ces deux formulations sont équivalentes, à des constantes près, dès que

$$A \subset \mathcal{X} \setminus (\mathcal{X} \setminus A_{r,c})_r, \qquad \forall r \geq 0, \forall A \in \mathcal{B}(\mathcal{X}),$$

(voir Lemme 5.6, [12]). Cette inclusion dépend cependant du type d'élargissement considéré, et n'est vérifiée que pour certains élargissements.

Dans toute la suite de ce document, la pseudo-distance $c(x, A)$ est définie à partir d'une *fonction de coût*.

Définition I.1.2. *On appelle fonction de coût toute fonction*

$$c : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \to [0, +\infty],$$

telle que $p \mapsto c(x, p)$ est convexe, et $c(x, \delta_x) = 0$ pour tout $x \in \mathcal{X}$ (δ_x désigne la mesure de Dirac en x). On définit alors la pseudo-distance c par

$$c(x, A) = \inf_{p \in \mathcal{P}(A)} c(x, p), \qquad x \in \mathcal{X}, \quad A \in \mathcal{B}(\mathcal{X}).$$

Ici, pour simplifier, on adopte la même notation c pour la pseudo-distance et la fonction de coût.

Dans ce document, on distingue trois sortes de fonctions de coût dont voici les définitions.

Les fonctions de coût usuelles

Généralement, la fonction de coût est définie sur $\mathcal{X} \times \mathcal{X}$ plutôt que sur $\mathcal{X} \times \mathcal{P}(\mathcal{X})$.

Définition I.1.3. *Une fonction de coût $c : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \to [0, +\infty]$ est dite usuelle s'il existe une fonction mesurable $\omega : \mathcal{X} \times \mathcal{X} \to [0, +\infty]$ telle que pour tout $x \in \mathcal{X}$ et tout $p \in \mathcal{P}(\mathcal{X})$, $\omega(x, x) = 0$ et*

$$c(x, p) = \int \omega(x, y)dp(y).$$

Dans ce cas, $p \mapsto c(x,p)$ est affine et la pseudo-distance $\inf_{p\in\mathcal{P}(A)} c(x,p)$ est atteinte aux points extrémaux du convexe $\mathcal{P}(A)$, qui sont les mesures de Dirac. Par conséquent, on a

$$c(x,A) = \inf_{y\in A} \omega(x,y).$$

Cette pseudo-distance redonne ainsi la pseudo-distance usuelle associée à la fonction de coût ω définie sur $\mathcal{X}\times\mathcal{X}$. Les fonctions de coût ω les plus étudiées sont de la forme $\omega(x,y) = d(x,y)^q$, $q > 0$ ou encore $\omega(x,y) = \alpha(d(x,y))$, avec $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ et $\alpha(0) = 0$.

Dans le cas où $\omega(x,y) = d(x,y)$ et par conséquent $c(x,A) = d(x,A)$, on notera simplement A_r le grossissement $A_{r,c}$, $r \geq 0$.

Les fonctions de coût barycentriques

Ces fonctions de coût sont définies sur l'espace euclidien, $(\mathcal{X}, d) = (\mathbb{R}^n, |\cdot|)$.

Définition I.1.4. *Une fonction de coût $c : \mathcal{X}\times\mathcal{P}_1(\mathcal{X}) \to [0,+\infty]$ est dite barycentrique s'il existe une fonction convexe $\theta : \mathbb{R}^n \to [0,+\infty]$ semi-continue inférieurement, telle que $\theta(0) = 0$ et pour tout $x \in \mathbb{R}^n$ et tout $p \in \mathcal{P}_1(\mathbb{R}^n)$,*

$$c(x,p) = \theta\left(x - \int y\, dp(y)\right).$$

Remarquons que les propriétés de concentration associées aux fonctions de coût barycentriques sont plus faibles que celles associées aux fonctions de coût usuelles avec $\omega(x,y) = \theta(x-y), x, y \in \mathbb{R}^n$, puisque d'après l'inégalité de Jensen,

$$c(x,p) \leq \int \theta(x-y) dp(y).$$

Les fonctions de coût universelles

Ces fonctions de coût ont été introduites par Marton [Mar96b] et par Talagrand [Tal96a] pour étudier divers problèmes de concentration, par exemple les déviations des processus empiriques, de la plus grande sous-suite croissante, le “bin” packing problem, etc... Ils sont l'outil de base de [1], [2] et [3].

On note $\mathbb{1}_{x\neq y}$ *la distance de Hamming* entre deux points x et y de $\mathcal{X}$, par définition

$$\mathbb{1}_{x\neq y} = \gamma(d(x,y)),$$

où $\gamma : \mathbb{R}^+ \to \mathbb{R}^+$ est la fonction définie par $\gamma(h) = 1$ si $h \neq 0$ et $\gamma(0) = 0$.

Définition I.1.5. *Soit $\alpha : \mathbb{R}^+ \to [0,+\infty]$ une fonction convexe semi-continue inférieurement telle que $\alpha(0) = 0$ et μ_0 une probabilité sur $\mathcal{X}$. Nous définissons deux classes de fonctions de coût universelles.*

— *Une fonction de coût $c : \mathcal{X}\times\mathcal{P}(\mathcal{X}) \to [0,+\infty]$ est dite universelle, associée à la fonction α, si elle s'écrit sous la forme*

$$c(x,p) = \alpha\left(\int \mathbb{1}_{x\neq y}\, dp(y)\right),$$

pour tout $x \in X$ et $p \in \mathcal{P}(\mathcal{X})$.

— *Une fonction de coût* $c : \mathcal{X}\times\mathcal{P}(\mathcal{X}) \to [0,+\infty]$ *est dite universelle, associée à la fonction* α *et à la probabilité* μ_0*, si on a*

$$c(x,p) = \int \alpha\left(\mathbb{1}_{x\neq y}\frac{dp}{d\mu_0}(y)\right) d\mu_0(y),$$

pour tout $(x,p) \in \mathcal{X}\times\mathcal{P}(\mathcal{X})$ *tel que* p *est absolument continue par rapport à* μ_0 *sur l'ensemble* $X \setminus \{x\}$*, et* $c(x,p) = +\infty$ *sinon.*

Ces fonctions de coût ne dépendent pas de la distance d sur $\mathcal{X}$, donc de la géométrie de l'espace $\mathcal{X}$. Dans [12], pour établir rigoureusement le théorème général de dualité de Kantorovich associé à ces fonctions coût universelles sur un espace métrique compact $(\mathcal{X}, d)$, un point essentiel est la semi-continuité inférieure de la fonction γ qui définit $\mathbb{1}_{x\neq y} = \gamma(d(x,y))$, $x, y \in \mathcal{X}$.

I.2 Formulation fonctionnelle

La formulation fonctionnelle du principe de concentration fait intervenir *l'opérateur de type infimum-convolution* suivant, introduit dans [3] : pour toute fonction mesurable $\varphi : \mathcal{X} \to \mathbb{R}\cup\{\infty\}$ bornée inférieurement

$$R_c\varphi(x) = \inf_{p\in\mathcal{P}(\mathcal{X})}\left\{\int \varphi dp + c(x,p)\right\}, \qquad x \in \mathcal{X}. \tag{I.1}$$

Comme $c(x,\delta_x) = 0$, on a $R_c\varphi(x) \leq \varphi(x)$.

Observons que pour une fonction de coût usuelle, $c(x,p) = \int \omega(x,y)dp(y)$, du fait que $p \mapsto \int \varphi dp + c(x,p)$ est affine , l'opérateur $R_c\varphi$ est l'opérateur d'infimum-convolution habituel associé à la fonction de coût ω,

$$R_c\varphi(x) = \inf_{y\in\mathcal{X}}\{\varphi(y) + \omega(x,y)\} = Q_\omega\varphi(x).$$

La formulation fonctionnelle du principe de concentration est donnée par la proposition suivante.

Proposition I.2.1. *Soit* $a_1, a_2 > 0$ *et* $\beta : \mathbb{R}^+ \to [0,+\infty]$ *une fonction. Les assertions suivantes sont équivalentes.*

(1) Pour tout sous-ensemble $A \in \mathcal{B}(\mathcal{X})$*, pour tout* $r \geq 0$,

$$\mu(A)^{a_1}\mu(\mathcal{X}\setminus A_{r,c})^{a_2} \leq e^{-\beta(r)}.$$

(2) Pour toute fonction mesurable $\varphi : \mathcal{X} \to \mathbb{R}\cup\{\infty\}$ *bornée inférieurement,*

$$\mu(\varphi \leq m)^{a_1}\mu(R_c\varphi > m+r)^{a_2} \leq e^{-\beta(r)}, \qquad \forall m \in \mathbb{R}, \forall r \geq 0.$$

Démonstration. Etant donné $A \in \mathcal{B}(\mathcal{X})$, soit i_A la fonction nulle sur A et égale à $+\infty$ sur $\mathcal{X}\setminus A$. En appliquant (2) à $\varphi = i_A$ et à $m = 0$, on obtient le point (1) car $\{\varphi \leq 0\} = A$ et $R_ci_A(x) = c(x,A)$, $x \in \mathcal{X}$.

Inversement, étant donnée une fonction φ, on applique (1) à $A = \{\varphi \leq m\}$. L'assertion (2) résulte du fait que $\{R_c\varphi > m+r\} \subset \mathcal{X}\setminus A_{r,c}$. En effet, si $x \in A_{r,c}$, pour tout $\varepsilon > 0$, il existe $p_\varepsilon \in \mathcal{P}(A)$ tel que $c(x,p_\varepsilon) \leq r+\varepsilon$. Puisque

$$R_c\varphi(x) \leq \int \varphi dp_\varepsilon + c(x,p_\varepsilon) \leq m+r+\varepsilon,$$

en faisant tendre ε vers 0, on obtient que $x \in \{R_c\varphi \leq m + r\}$. Par conséquent $A_{r,c} \subset \{R_c\varphi \leq m + r\}$ et $\mu(R_c\varphi > m + r) \leq \mu(\mathcal{X} \setminus A_{r,c})$. $\square$

Pour une fonction de coût usuelle du type

$$c(x,p) = \int \alpha(d(x,y))dp(y), \qquad x \in \mathcal{X},\, p \in \mathcal{P}(\mathcal{X}),$$

où $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ est une bijection, il est courant d'écrire le phénomène de concentration en considérant les grossissements naturels $A_t = \{x \in \mathcal{X}, d(x,A) \leq t\}$ associés à la distance d. Puisque $A_{r,c} = A_t$ lorsque $r = \alpha(t)$, (1) peut être reformulé de la manière suivante : pour tout $A \in \mathcal{B}(\mathcal{X})$ et tout $t \geq 0$,

$$\mu(A)^{a_1}\mu(\mathcal{X} \setminus A_t)^{a_2} \leq e^{-\beta(\alpha(t))}.$$

Dans ce cas, comme le suggéraient à l'origine les idées de P. Lévy, cette propriété de concentration s'énonce en utilisant la classe des fonctions f 1-lipschitziennes (cf. [Led01]). Pour cela, comme dans la preuve précédente, en choisissant $A = \{f \leq m\}$, $m \in \mathbb{R}$, on montre que $A_t \subset \{f \leq m+t\}$, $t \geq 0$. Ceci donne la formulation fonctionnelle équivalente suivante, pour toute fonction f 1-lipschitzienne,

$$\mu(f \leq m)^{a_1}\mu(f > m + t)^{a_2} \leq e^{-\beta(\alpha(t))}, \qquad \forall m \in \mathbb{R},\, \forall t \geq 0. \tag{I.2}$$

Supposons de plus que α est convexe. La formulation (I.2) découle alors également de (2) appliqué à la fonction $\varphi = \lambda f$, $\lambda > 0$, en supposant dans un premier temps f bornée inférieurement. Puisque f est 1-lipschitzienne, $R_c\varphi$ est proche de φ et cette proximité est controlée par λ. Plus précisément, pour tout $x \in \mathcal{X}$

$$\begin{aligned} R_c\varphi(x) &= \inf_{y\in\mathcal{X}} \{\varphi(y) + \alpha(d(x,y))\} \\ &\qquad\qquad \geq \varphi(x) - \sup_{y\in\mathcal{X}} \{\lambda d(x,y) - \alpha(d(x,y))\} \geq \varphi(x) - \alpha^*(\lambda), \end{aligned}$$

avec $\alpha^*(\lambda) = \sup_{v\geq 0}\{\lambda v - \alpha(v)\}$. Ainsi, en remplaçant m par λm, le point (2) donne, pour toute fonction f 1-lipschitzienne bornée inférieurement et pour tout $\lambda > 0$,

$$\mu(f \leq m)^{a_1}\mu(\lambda f > \lambda m + r + \alpha^*(\lambda))^{a_2} \leq e^{-\beta(r)}, \qquad \forall m \in \mathbb{R},\, \forall r \geq 0.$$

Puisque $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ est convexe bijective et $\alpha(0) = 0$, α est strictement croissante et pour tout $t > 0$, on a $\partial\alpha(t) \subset]0, +\infty[$, où $\partial\alpha(t)$ désigne le sous-différentiel de α au point t. Par suite l'inégalité précédente redonne la propriété (I.2) pour les fonctions 1-lipschitzienne bornées inférieurement en choisissant $r = \alpha(t)$ et $\lambda \in \partial\alpha(t)$ de sorte que $\lambda t = \alpha(t)+\alpha^*(\lambda)$. Ensuite, par un argument de convergence monotone la propriété (I.2) s'étend ensuite à toutes les fonctions 1-lipschitziennes.

Enfin, la propriété (I.2) implique la propriété usuelle de concentration des fonctions 1-lipschitziennes f autour de leur médiane m_f. En appliquant (I.2) à f ou à $-f$, et en choisissant $m = m_f$ ou $m = -m_f$, on obtient

$$\mu(|f - m_f| > t) \leq 2.2^{a_1/a_2}e^{-\beta(\alpha(t))/a_2},$$

pour tout $t \geq 0$.

CHAPITRE II

Inégalités de transport-entropie

Ce chapitre souligne l'intérêt des inégalités de transport-entropie dans l'étude des phénomènes de concentration de la mesure sur les espaces produits. Comme pour les inégalités de Sobolev logarithmiques, les propriétés de tensorisation des inégalités de transport en font un outil performant pour établir des propriétés de concentration dans les espaces produits. La dernière partie de ce chapitre aborde les liens établis entre ces inégalités de transport et les inégalités de Sobolev logarithmiques.

II.1 Introduction : concentration sur les espaces produits et propriétés isopérimétriques

Soit μ une mesure sur un espace métrique $(\mathcal{X}, d)$ muni de sa tribu borélienne $\mathcal{B}(\mathcal{X})$. Pour tout borélien A, on définit la *mesure du bord* de A par

$$\mu^+(\partial A) = \liminf_{t\to 0^+} \frac{\mu(A_t) - \mu(A)}{t}.$$

Résoudre le problème isopérimétrique pour la mesure μ consiste à déterminer quelle est la plus petite mesure de bord $\mu^+(\partial A)$ lorsque la mesure des ensembles A, $\mu(A)$, est fixée. Autrement dit, on cherche la plus grande fonction, notée $I_\mu : \mathbb{R}^+ \to \mathbb{R}^+$ telle que pour tout borélien A,

$$\mu^+(\partial A) \geq I_\mu(\mu(A)). \tag{II.1}$$

La fonction I_μ est appelée *profil isopérimétrique* de μ.

Lorsque $I_\mu \geq v' \circ v^{-1}$, où $v : \mathbb{R} \to [0, \mu(\mathcal{X})]$ est une fonction strictement croissante dérivable, *l'inégalité isopérimétrique* (II.1) fournit une estimation inférieure de la mesure de A_t (voir Proposition 2.1 [Led01]) : pour tout $t \geq 0$,

$$\mu(A_t) \geq v(v^{-1}(\mu(A)) + t). \tag{II.2}$$

Par conséquent si μ est une probabilité, on obtient la propriété de concentration suivante, pour tout $A \in \mathcal{B}(\mathcal{X})$ tel que $\mu(A) \geq 1/2$,

$$\mu(\mathcal{X} \setminus A_t) \leq 1 - v(v^{-1}(1/2) + t), \qquad \forall t \geq 0.$$

Dans les cas particuliers où $(\mathcal{X}, d) = (\mathbb{R}^n, |\cdot|)$ est l'espace euclidien et μ la mesure de Lebesgue, ou encore $\mathcal{X} = \mathbb{S}^n$ est la sphère euclidienne unité de $\mathbb{R}^{n+1}$ munie de sa distance géodésique d et $\mu = \sigma^n$ la loi uniforme sur $\mathbb{S}^n$, le profil isopérimétrique est donné par $I_\mu = v' \circ v^{-1}$ où pour tout r, $v(r)$ est la mesure d'une boule de de rayon r (voir [Lév51]). Par suite, si $\mu(A) = \mu(B) = v(r)$ où B est une boule de $\mathcal{X}$, alors

$$v(v^{-1}(\mu(A)) + t) = v(r + t) = \mu(B_t).$$

La propriété (II.2) établit donc que les boules sont des ensembles extrémaux au sens où : pour tout borélien A et pour toute boule B de mesure $\mu(B) = \mu(A)$,

$$\mu(A_t) \geq \mu(B_t),$$

pour tout $t \geq 0$. Cette dernière propriété est en fait équivalente à l'inégalité isopérimétrique car elle implique,

$$\mu^+(\partial A) = \liminf_{t \to 0^+} \frac{\mu(A_t) - \mu(A)}{t} \geq \liminf_{t \to 0^+} \frac{\mu(B_t) - \mu(B)}{t} = I_\mu(\mu(A)),$$

puisque $\mu(A) = \mu(B)$.

Ainsi, le profil de concentration de σ^n est donné par l'estimation des mesures des calottes sphériques : pour tout $A \in \mathcal{B}(\mathbb{S}^n)$ tel que $\sigma^n(A) \geq 1/2$,

$$\sigma^n(\mathcal{X} \setminus A_t) \leq 1 - \sigma^n(B_t) = v(v^{-1}(1/2) + t) \leq e^{-(n-1)t^2/2}, \quad \forall t \geq 0,$$

où B est une demi-sphère, $\sigma^n(B) = 1/2$. L'estimation donnée par la dernière inégalité est démontrée suite au Corollaire 2.2 dans [Led01].

Par dilatation, sachant que la loi uniforme sur la sphère de $\mathbb{R}^{n+1}$ de rayon $\sqrt{n}$ tend vers la mesure gaussienne canonique sur $\mathbb{R}^{\mathbb{N}}$, on obtient le profil isopérimétrique pour les lois gaussiennes (cf. [Led93, Led01]). Les demi-hyperplans sont les ensembles extrémaux pour la loi gaussienne standard γ^n sur $\mathbb{R}^n$ et on a

$$I_{\gamma^n}(s) = \varphi \circ \Phi^{-1}(s) \underset{0^+}{\sim} s\sqrt{2 \log \frac{1}{s}}, \quad \text{avec} \quad \Phi(r) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r} e^{-u^2/2} du, \ r \in \mathbb{R},$$

et $\varphi = \Phi'$, Φ est la fonction de répartition d'une loi gaussienne standard sur $\mathbb{R}$. La propriété de concentration qui découle de l'inégalité isopérimétrique gaussienne est la suivante : pour tout $A \in \mathcal{B}(\mathbb{R}^n)$ tel que $\gamma^n(A) \geq 1/2$, on a

$$\gamma^n(\mathbb{R}^n \setminus A_t) \leq 1 - \Phi(t) \leq e^{-t^2/2}, \quad \forall t \geq 0. \tag{II.3}$$

Sur le cube discret, $\mathcal{X} = \{0, 1\}^n$, muni de la loi uniforme μ^n et de la distance de Hamming définie par

$$d(x, y) = \sum_{i=1}^{n} \mathbb{1}_{x_i \neq y_i}, \qquad x, y \in \{0, 1\}^n,$$

les ensembles extrémaux A qui minimisent $\mu^n(A_t)$ lorsque $\mu^n(A) \geq 1/2$, ont étés identifiés , ce qui permet de donner une propriété de concentration pour la loi μ^n uniforme sur $\{0,1\}^n$ (voir [Har66, WW77]) : pour tout $A \subset \{0,1\}^n$ tel que $\mu^n(A) \geq 1/2$,

$$\mu^n(\mathcal{X} \setminus A_t) \leq e^{-2t^2/n}, \quad \forall t \geq 0. \tag{II.4}$$

Dans les trois exemples précédents, selon les cas, la dimension n accentue ou non les propriétés de concentration de la mesure. En grande dimension, les exemples pour lesquels le problème isopérimétrique est résolu sont peu nombreux.

Remarquons par ailleurs que la propriété de concentration ne se tensorise pas correctement. Par exemple, pour la mesure gaussienne standard produit $\gamma^n = \gamma^1 \otimes \cdots \otimes \gamma^1$, la propriété de concentration adimensionnelle (II.3) ne peut pas, à notre connaissance, se déduire directement de la même propriété de concentration en dimension 1 pour γ^1.

Les méthodes “entropiques” que nous présentons ici sont une alternative efficace à la méthode isopérimétrique. Elles permettent d'élargir considérablement la classe des exemples de propriétés de concentration sur les espaces de grande dimension, grace aux propriétés de tensorisation de l'entropie.

Les deux méthodes entropiques les plus utilisées dans ce domaine sont la méthode liée aux inégalités de type Sobolev logarithmiques avec “l'argument de Herbst” (cf. chapitre 5. [Led01]) et la méthode de transport liée aux inégalités de type transport-entropie avec “l'argument de Marton” (cf. chapitre 6. [Led01]). On s'intéresse ici plus particulièrement à cette deuxième méthode. Par exemple, les résultats de concentration (II.3) pour la mesure gaussienne canonique sur $\mathbb{R}^n$, ou (II.4) pour la loi uniforme sur le cube discret, se déduisent simplement de la tensorisation d'inégalités de transport.

II.2 Inégalités de transport et propriété de concentration

Dans ce document, $\Pi(\mu,\nu)$ désigne l'ensemble des probabilités sur l'espace polonais produit $\mathcal{X} \times \mathcal{X}$ muni de la tribu borélienne produit, ayant pour première marginale μ et pour seconde marginale ν. L'espace des probabilité $\mathcal{P}(\mathcal{X})$ est muni de la tribu engendrée par les applications

$$\begin{array}{ccc} \mathcal{P}(\mathcal{X}) & \to & ([0,1],\mathcal{B}) \\ \nu & \mapsto & \nu(A), \end{array} .$$

où A est un borélien de $\mathcal{X}$ et $\mathcal{B}$ la tribu borélienne sur $[0,1]$. Toute probabilité $\pi \in \Pi(\mu,\nu)$ admet une décomposition

$$d\pi(x,y) = d\mu(x)dp_x(y),$$

où $p : x \in \mathcal{X} \mapsto p_x \in \mathcal{P}(\mathcal{X})$ est une application mesurable μ-presque-sûrement uniquement déterminée, c'est à dire un noyau de probabilité satisfaisant

$$\mu p(A) = \int p_x(A)d\mu(x) = \nu(A), \qquad \forall A \in \mathcal{B}(\mathcal{X}).$$

Nous définissons dans [12] le coût de transport optimal entre deux probabilités, associé à une fonction de coût c de la manière suivante.

Définition II.2.1. *Le coût de transport optimal associé à une fonction de coût $c : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \to [0, +\infty]$, entre deux probabilités μ et ν sur $\mathcal{X}$ est la quantité*

$$T_c(\nu|\mu) = \inf_{\pi \in \Pi(\mu,\nu)} \int c(x, p_x) d\mu(x),$$

où, étant donné $\pi \in \Pi(\mu, \nu)$, le noyau $p = (p_x)_{x \in X}$ est tel que $d\pi(x, y) = d\mu(x) dp_x(y)$.

A priori, le coût optimal $T_c(\nu|\mu)$ n'est pas une fonction symétrique de μ et ν. A titre d'exemple, considérons une fonction de coût usuelle,

$$c(x, p_x) = \int \omega(x, y) dp_x(y), \qquad x \in \mathcal{X}.$$

$T_c(\nu|\mu)$ correspond alors au *coût de transport optimal usuel* associé à la fonction de coût ω,

$$\mathcal{T}_\omega(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \iint \omega(x, y) d\pi(x, y) = \inf_{\pi \in \Pi(\mu,\nu)} \int c(x, p_x) d\mu(x) = T_c(\nu|\mu).$$

Si la fonction ω est symétrique, $\omega(x, y) = \omega(y, x)$ pour tout $x, y \in \mathcal{X}$, alors $\mathcal{T}_\omega$ et donc T_c est symétrique, $\mathcal{T}_\omega(\mu, \nu) = \mathcal{T}_\omega(\nu, \mu)$.

Nous allons maintenant présenter une version générale des inégalités de transport-entropie associées aux coût de transport optimaux $T_c(\nu|\mu)$. L'argument de Marton, rappelé plus loin, permet d'en déduire les propriétés de concentration de la Proposition I.2.1 du paragraphe précédent.

Définition II.2.2. *Soit $a_1, a_2 > 0$ et soit $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ une fonction croissante. On dit que la probabilité $\mu \in \mathcal{P}(\mathcal{X})$ satisfait l'inégalité de transport $\mathbf{T}_{c,\beta}(a_1, a_2)$ si*

$$\mathbf{T}_{c,\beta}(a_1, a_2) : \qquad \beta\left(T_c(\nu_1|\nu_2)\right) \le a_1 H(\nu_1|\mu) + a_2 H(\nu_2|\mu), \qquad \forall \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X})$$

où $H(\nu_1|\mu)$ est l'entropie relative de ν_1 par rapport à μ définie par

$$H(\nu_1|\mu) = \int \log \frac{d\nu_1}{d\mu} \, d\nu_1,$$

si ν_1 est absolument continue par rapport à μ ($\nu_1 << \mu$), et $H(\nu_1|\mu) = +\infty$ sinon.

En toute généralité, nous appelons inégalité de transport faible une inégalité du type $\mathbf{T}_{c,\beta}(a_1, a_2)$, et dans des cas particuliers, suivant le profil de la fonction de coût $c : \mathcal{X} \times \times \mathcal{P}(\mathcal{X}) \to [0, +\infty]$, cette inégalité de transport sera parfois appelée inégalité de transport barycentrique (dans le Chapitre III) ou encore inégalité de transport universelle (dans le Chapitre IV).

Remarquons que les inégalités de transport $\mathbf{T}_{c,\beta}(0, a_2)$ ou $\mathbf{T}_{c,\beta}(a_1, 0)$ impliqueraient $\beta(T_c(\nu_1|\mu)) = 0, \forall \nu_1 \in \mathcal{P}(\mathcal{X})$, ou encore $\beta(T_c(\mu|\nu_2)) = 0, \forall \nu_2 \in$

$\mathcal{P}(\mathcal{X})$. Elles ne sont donc jamais vérifiées sauf dans des cas dégénérés, comme lorsque la fonction de coût c ou encore la fonction β est identiquement nulle. Cependant, avec la convention $0.\infty = 0$, l'inégalité de transport $\mathbf{T}_{c,\beta}(b,\infty)$ correspond à l'inégalité de transport

$$\mathbf{T}^+_{c,\beta}(b) : \qquad \beta\left(T_c(\nu|\mu)\right) \leq bH(\nu|\mu), \qquad \forall \nu \in \mathcal{P}(\mathcal{X}),$$

et $\mathbf{T}_{c,\beta}(\infty,b)$ correspond à l'inégalité de transport

$$\mathbf{T}^-_{c,\beta}(b) : \qquad \beta\left(T_c(\mu|\nu)\right) \leq bH(\nu|\mu), \qquad \forall \nu \in \mathcal{P}(\mathcal{X}).$$

Ces deux dernières inégalités de transport sont identiques pour un coût de transport optimal symétrique. Lorsque β est l'identité, $\beta(r) = r, r \geq 0$, les inégalités de transport sont notées plus simplement $\mathbf{T}_c(a_1,a_2)$, $\mathbf{T}^+_c(b)$ et $\mathbf{T}^-_c(b)$.

L'argument de Marton est le suivant. Etant donné $A \in \mathcal{B}(\mathcal{X})$, si ν_1 est la restriction de μ à A et ν_2 la restriction de μ à $B = \mathcal{X} \setminus A_{r,c}$, $r \geq 0$,

$$\frac{d\nu_1}{d\mu} = \frac{\mathbb{1}_A}{\mu(A)} \quad \text{et} \quad \frac{d\nu_2}{d\mu} = \frac{\mathbb{1}_B}{\mu(B)},$$

alors

$$H(\nu_1|\mu) = \log\frac{1}{\mu(A)}, \quad H(\nu_2|\mu) = \log\frac{1}{\mu(B)}, \quad \text{et} \quad T_c(\nu_1|\nu_2) \geq r.$$

Par suite, puisque β est une fonction croissante, l'inégalité de transport $\mathbf{T}_{c,\beta}(a_1,a_2)$ implique que pour tout $r \geq 0$,

$$\beta(r) \leq \log\left(\mu(A)^{-a_1}\right) + \log\left(\mu(B)^{-a_2}\right),$$

ce qui correspond à la propriété de concentration du point (1) de la Proposition I.2.1.

Quelques exemples permettent de mieux comprendre la formulation générale des inégalités de transport introduite dans la définition précédente.

Si c est la fonction de coût usuelle $c(x,p) = \int d(x,y)^q dp(y)$, $q \geq 1$, alors $T_c(\nu|\mu) = T_c(\mu|\nu)$ s'écrit en fonction de la distance de Wasserstein W_q d'ordre q,

$$T_c(\nu|\mu) = W_q^q(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \iint d(x,y)^q d\pi(x,y).$$

L'inégalité de transport $\mathbf{T}_2(b)$ de Talagrand, satisfaite par la mesure gaussienne canonique $\mu = \gamma^n$ sur $\mathbb{R}^n$ pour $b = 2$, correspond aux inégalités de transport $\mathbf{T}^+_c(b)$ ou $\mathbf{T}^-_c(b)$ avec le coût usuel précédent pour $q = 2$.

$$\mathbf{T}_2(b) : \qquad W_2^2(\mu,\nu) \leq bH(\nu|\mu), \qquad \forall \nu \in \mathcal{P}(\mathcal{X}).$$

Une particularité de l'inégalité $\mathbf{T}_2(b)$ est qu'elle est équivalente à la famille d'inégalités $\mathbf{T}_c(b/t, b/(1-t))$, pour $t \in]0,1[$. En effet, si μ satisfait $\mathbf{T}_2(b)$, alors d'après l'inégalité triangulaire pour la distance de Wasserstein,

$$W_2^2(\nu_1,\nu_2) \leq \left(W_2(\nu_1,\mu) + W_2(\mu,\nu_2)\right)^2 \leq b\left(\sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)}\right)^2.$$

Comme $\left(\sqrt{u}+\sqrt{v}\right)^2 = \inf_{t\in]0,1[}\left\{\frac{u}{t}+\frac{v}{1-t}\right\}$, on obtient que pour tout $t\in]0,1[$, μ satisfait $T_c(b/t, b/(1-t))$,

$$W_2^2(\nu_1,\nu_2) \leq \frac{b}{t}H(\nu_1|\mu) + \frac{b}{1-t}H(\nu_2|\mu), \quad \forall \nu_1,\nu_2 \in \mathcal{P}(\mathcal{X}).$$

Inversement si μ satisfait $T_c(b/t, b/(1-t))$ pour tout $t\in]0,1[$, alors en choisissant $\nu_2=\mu$ puis en faisant tendre t vers 1 on retrouve l'inégalité de transport $\mathbf{T}_2(b)$.

Plus généralement, supposons que c est la fonction de coût usuelle $c(x,p) = \int \alpha(d(x,y))p(dy)$, où $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ est une fonction convexe non nulle. On note alors $\mathcal{T}_\alpha(\mu,\nu) = T_c(\nu|\mu) = T_c(\mu|\nu)$. Si de plus α est strictement croissante, $\alpha(0)=\alpha'_+(0)=0$ et α vérifie la *condition* Δ_2 suivante, [RR91] : il existe une constante positive C telle que

$$\alpha(2h) \leq C\alpha(h), \qquad \forall h \geq 0, \tag{II.5}$$

(nécessairement $C \geq 2$), alors on peut utiliser le lemme de changement de métrique suivant donné dans [7].

Lemme II.2.1. *Sous les conditions précédentes, en posant* $p_\alpha = \sup_{h>0}\frac{h\alpha'_+(h)}{\alpha(h)}$, *la fonction* $h\mapsto \alpha^{1/p_\alpha}(h)$ *est sous-additive, c'est à dire*

$$\alpha^{1/p_\alpha}(h+k) \leq \alpha^{1/p_\alpha}(h) + \alpha^{1/p_\alpha}(k), \qquad \forall h,k\in\mathbb{R}^+.$$

Par conséquent, $d_\alpha(x,y) = \alpha^{1/p_\alpha}(d(x,y)), x,y\in\mathcal{X}$, *définit une distance sur* $\mathcal{X}$.

Remarquons que sous la condition (II.5), $p_\alpha \leq C-1$. Grâce à ce lemme et à l'inégalité triangulaire, on obtient, pour toutes mesures $\nu_1,\nu_2\in\mathcal{P}(\mathcal{X})$,

$$\begin{aligned}\mathcal{T}_\alpha(\nu_1,\nu_2) &= W_{p_\alpha}^{p_\alpha}(\nu_1,\nu_2)\\ &\leq \left(W_{p_\alpha}(\nu_1,\mu)+W_{p_\alpha}(\mu,\nu_2)\right)^{p_\alpha} = \left(\mathcal{T}_\alpha^{1/p_\alpha}(\nu_1,\mu)+\mathcal{T}_\alpha^{1/p_\alpha}(\mu,\nu_2)\right)^{p_\alpha},\end{aligned}$$

où la distance de Wasserstein W_{p_α} est définie avec la distance d_α du Lemme II.2.1. Ensuite, en remarquant que $p_\alpha > 1$, lorsque la fonction α n'est pas linéaire, et en utilisant l'identité

$$\left(u^{1/p}+v^{1/p}\right)^p = \inf_{t\in]0,1[}\left\{\frac{u}{t^{p-1}}+\frac{v}{(1-t)^{p-1}}\right\}, \qquad p>1,$$

on montre que μ satisfait l'inégalité de transport usuelle

$$\mathcal{T}_\alpha(\mu,\nu) \leq H(\nu|\mu), \qquad \forall \nu\in\mathcal{P}(\mathcal{X}),$$

si et seulement si μ satisfait les inégalités de transport suivantes : pour tout $t\in]0,1[$,

$$\mathcal{T}_\alpha(\nu_1,\nu_2) \leq \frac{H(\nu_1|\mu)}{t^{p_\alpha-1}} + \frac{H(\nu_2|\mu)}{(1-t)^{p_\alpha-1}}, \qquad \forall\nu_1,\nu_2\in\mathcal{P}(\mathcal{X}). \tag{II.6}$$

Voici un autre exemple d'inégalité de transport. Lorsque

$$c(x,p) = 2\int \mathbb{1}_{x\neq y}\,dp(y),$$

le coût de transport optimal universel $T_c(\nu|\mu)$ est en fait la distance en variation totale entre μ et ν

$$T_c(\nu|\mu) = \|\mu-\nu\|_{TV} = 2\sup_{A\subset\mathcal{X}}|\mu(A)-\nu(A)|.$$

L'inégalité de Csizár-Kullback-Pinsker [Pin64, Csi67, Kul67] satisfaite quel que soit la probabilité μ

$$\|\mu-\nu\|_{TV}^2 \leq 2\,H(\nu|\mu), \qquad \forall \nu\in\mathcal{P}(\mathcal{X}).$$

correspond alors à l'inégalité de transport $\mathbf{T}^+_{c,\beta}(b)$ ou $\mathbf{T}^-_{c,\beta}(b)$ avec $\beta(r)=r^2/2$, $r\geq 0$. Ici encore, cette inégalité est équivalente à $\mathbf{T}_{c,\beta}(b/t, b/(1-t))$ pour tout $t\in]0,1[$. Cette inégalité de transport et ses généralisations sont connues pour leurs nombreuses applications en probabilité, en analyse et en théorie de l'information (cf. [Vil09], page 636).

Comme dernier exemple, considérons la fonction de coût universelle

$$c(x,p) = \left(\int \mathbb{1}_{x\neq y}\,dp(y)\right)^2, \qquad x\in\mathcal{X}, \quad p\in\mathcal{P}(\mathcal{X}).$$

Les inégalités de transport $\mathbf{T}^+_c(2)$ et $\mathbf{T}^-_c(2)$ correspondent alors aux inégalités de transport faibles introduites par Marton [Mar96b]. Au même titre que l'inégalité de Csizár-Kullback-Pinsker, ces inégalités sont satisfaites pour toute probabilité μ. Dans ce cas, les inégalités de transport $\mathbf{T}^+_c(b)$ et $\mathbf{T}^-_c(b)$ sont équivalentes à la famille d'inégalités $\mathbf{T}_c(b/t, b/(1-t))$ pour tout $t\in]0,1[$, car d'après [Mar97], le coût de transport faible T_c, noté $\widetilde{\mathcal{T}_2}$ satisfait l'inégalité triangulaire

$$\sqrt{\widetilde{\mathcal{T}_2}(\nu_1|\nu_2)} \leq \sqrt{\widetilde{\mathcal{T}_2}(\nu_1|\mu)} + \sqrt{\widetilde{\mathcal{T}_2}(\mu|\nu_2)}, \qquad \forall \mu,\nu_1,\nu_2\in\mathcal{P}(\mathcal{X}). \tag{II.7}$$

II.3 Formulation fonctionnelle des inégalités de transport, Théorème de Kantorovich

La forme fonctionnelle duale des inégalités de transport usuelles a été obtenue par Bobkov et Götze [BG99] puis étendue dans l'article [GL10].

Le premier argument qui permet d'obtenir cette forme duale est la dualité de l'entropie relative avec la transformée log-Laplace : pour toute fonction $g:\mathcal{X}\to\mathbb{R}$ continue bornée,

$$\log\int e^g\,d\mu = \sup_{\nu\in\mathcal{P}(\mathcal{X})}\left\{\int g\,d\mu - H(\nu|\mu)\right\}. \tag{II.8}$$

(une preuve simple de cette égalité est donnée dans [GL10] et une plus générale dans [5])

Le second argument est le théorème de dualité de Kantorovich. Ce théorème est connu dans le cas usuel pour des fonctions de coût $\omega : \mathcal{X}\times\mathcal{X} \to]-\infty,+\infty]$ semi-continues inférieurement (cf.[Vil09])

$$\mathcal{T}_\omega(\nu_1,\nu_2) = \sup_{\varphi\in\mathcal{C}_b(\mathcal{X})} \left\{\int Q_\omega\varphi\, d\nu_2 - \int \varphi d\nu_1\right\}, \qquad \nu_1,\nu_2\in\mathcal{P}(\mathcal{X}) \qquad \text{(II.9)}$$

où $\mathcal{C}_b(\mathcal{X})$ est l'espace des fonctions continues bornées sur $\mathcal{X}$ et

$$Q_\omega\varphi(y) = \inf_{x\in\mathcal{X}} \{\varphi(x)+\omega(x,y)\}, \qquad y\in\mathcal{X}.$$

Dans [12] (Théorème 3.5), nous généralisons ce théorème pour des fonctions de coûts générales $c : \mathcal{X}\times\mathcal{P}(\mathcal{X}) \to [0,+\infty]$, sous des hypothèses de régularité peu restrictives sur c. Le résultat est globalement le suivant

$$T_c(\nu_1|\nu_2) = \sup_{\varphi\in\mathcal{C}_b(\mathcal{X})} \left\{\int R_c\varphi\, d\nu_2 - \int \varphi d\nu_1\right\}, \qquad \nu_1,\nu_2\in\mathcal{P}(\mathcal{X}), \qquad \text{(II.10)}$$

où $R_c\varphi$ est l'opérateur d'infimum-convolution (I.1) défini au premier chapitre ,

$$R_c\varphi(x) = \inf_{p\in\mathcal{P}(\mathcal{X})} \left\{\int \varphi dp + c(x,p)\right\}.$$

Ici, pour être exact, il faudrait modifier l'ensemble des probabilités et la classe des fonctions continue suivant le coût choisi (voir [12]).

Les deux arguments de dualité (II.8) et (II.10) permettent de démontrer la formulation fonctionnelle duale de l'inégalité de transport $\mathbf{T}_{c,\beta}(a_1,a_2)$ suivante.

Proposition II.3.1. *Soit $\mu\in\mathcal{P}(\mathcal{X})$ et $\beta : \mathbb{R}^+\to[0,+\infty]$ une fonction semi-continue inférieurement convexe telle que $\beta(0)=0$. Les assertions suivantes sont équivalentes.*

(1) La probabilité μ satisfait $\mathbf{T}_{c,\beta}(a_1,a_2)$.

(2) Pour toute fonction $\varphi\in\mathcal{C}_b(\mathcal{X})$ et pour tout $\lambda\geq 0$,

$$\left(\int e^{\frac{\lambda R_c\varphi}{a_2}} d\mu\right)^{a_2} \left(\int e^{-\frac{\lambda\varphi}{a_1}} d\mu\right)^{a_1} \leq e^{\beta^*(\lambda)},$$

avec $\beta^*(\lambda) = \sup_{t\geq 0}\{\lambda t - \beta(t)\}$.

Cette formulation fonctionnelle est en fait une extension de la propriété-(τ) de Maurey [Mau91], qui fut le premier à introduire l'opérateur d'infimim-convolution pour retrouver les résultats de Talagrand pour la mesure gaussienne, exponentielle, et les mesures à support borné.

Idée de la preuve (1) $\Rightarrow$ (2). On a $\beta(t) = \sup_{\lambda\geq 0}\{\lambda t - \beta^*(\lambda)\}$, $\forall t\geq 0$. Si μ satisfait $\mathbf{T}_{c,\beta}(a_1,a_2)$, alors par le théorème de dualité de Kantorovich généralisé (II.10), pour tout $\varphi\in\mathcal{C}_b(\mathcal{X})$ et tout $\lambda\geq 0$,

$$\lambda\left(\int R_c\varphi\, d\nu_2 - \int\varphi d\nu_1\right) - \beta^*(\lambda) \leq a_2 H(\nu_2|\mu) + a_1 H(\nu_1|\mu), \qquad \forall \nu_1,\nu_2\in\mathcal{P}(\mathcal{X}).$$

Le point (2) de la proposition découle en réarrangeant les termes de cette inégalité, en optimisant sur les mesures ν_1 et ν_2, puis en appliquant la formule de dualité de l'entropie (II.8), avec la fonction $g = \lambda R_c\varphi/a_2$ et avec la fonction $g = -\lambda\varphi/a_1$. □

Par densité de l'espace des fonctions continues bornées dans $L^1(\mu)$, puis par convergence monotone, (2) s'étend à toutes les fonctions mesurables $\varphi : \mathcal{X} \to]-\infty, +\infty]$ bornées inférieurement.

Etant donné $A \in \mathcal{B}(\mathcal{X})$, considérons la fonction i_A, nulle sur A, égale à $+\infty$ en dehors de A. En appliquant (2) à la fonction i_A, comme $R_c i_A(x) = c(x, A)$, $x \in \mathcal{X}$, l'inégalité de transport $\mathbf{T}_{c,\beta}(a_1, a_2)$ redonne la formulation de Talagrand des propriétés de concentration pour la mesure μ sous la forme générale suivante (cf. [Tal95, Tal96a, Tal96b]) :

$$\int e^{\frac{\lambda c(x,A)}{a_2}} d\mu \leq \frac{e^{\beta^*(\lambda)/a_2}}{\mu(A)^{a_1/a_2}}, \qquad \forall \lambda \geq 0, \quad \forall A \in \mathcal{P}(\mathcal{X}).$$

Dans le cas particulier où β est la fonction identique, on a $\beta^* = i_{(-\infty,1]}$, et la Proposition II.3.1 s'écrit :

Proposition II.3.2. *[12] Les assertions suivantes sont équivalentes.*

(1) La probabilité μ satisfait $\mathbf{T}_c(a_1, a_2)$.

(2) Pour toute fonction $\varphi \in \mathcal{C}_b(\mathcal{X})$,

$$\left(\int e^{\frac{R_c \varphi}{a_2}} d\mu\right)^{a_2} \left(\int e^{-\frac{\varphi}{a_1}} d\mu\right)^{a_1} \leq 1.$$

II.4 Tensorisation - Caractérisation par concentration adimensionnelle

Les inégalités de transport se tensorisent, ce qui fournit en retour des résultats de concentration en grande dimension.

Proposition II.4.1. *Soit $\mathcal{X}_1$ et $\mathcal{X}_2$ des espaces métriques polonais. Soit $\beta_1 : \mathbb{R}^+ \to \mathbb{R}^+$, $\beta_2 : \mathbb{R}^+ \to \mathbb{R}^+$ des fonctions convexes et soit $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ définie par*

$$\beta(t) = \beta_1 \square \beta_2(t) = \inf\{\beta_1(t_1) + \beta_2(t_2), t = t_1 + t_2\}, \qquad t \geq 0.$$

Si $\mu_1 \in \mathcal{P}(\mathcal{X}_1)$ et $\mu_2 \in \mathcal{P}(\mathcal{X}_2)$ satisfont respectivement les inégalités de transport $\mathbf{T}_{c_1,\beta_1}(a_1, a_2)$ et $\mathbf{T}_{c_2,\beta_2}(a_1, a_2)$, $a_1, a_2 > 0$, alors $\mu_1 \otimes \mu_2 \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ satisfait l'inégalité de transport $\mathbf{T}_{c,\beta}(a_1, a_2)$ avec pour tout $x = (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, pour tout $p \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ de marginales $p_1 \in \mathcal{P}(\mathcal{X}_1)$ et $p_2 \in \mathcal{P}(\mathcal{X}_2)$

$$c(x, p) = c_1 \oplus c_2(x, p) = c_1(x_1, p_1) + c_2(x_2, p_2).$$

La preuve de cette proposition suit exactement celle du Théorème 4.11, [12], pour lequel $\beta_1(t) = \beta_2(t) = \beta(t) = t$, $t \geq 0$. Nous pourrions aussi adapter la preuve donnée dans [3] sur la forme fonctionnelle de ces inégalités de transport.

Cette propriété de tensorisation repose sur les propriétés de tensorisation de l'entropie relative et du coût de transport optimal : pour toute mesure $\nu \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ dont la décomposition s'écrit $d\nu(x_1, x_2) = d\nu_1(x_1) d\nu_2^{x_1}(x_2)$,

$$H(\nu|\mu) = H(\nu_1|\mu_1) + \int H(\nu_2^{x_1}|\mu_2) d\nu_1(x_1),$$

et pour toute autre mesure $\nu' \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ de décomposition $d\nu'(x_1', x_2') = d\nu_1(x_1') d\nu_2^{x_1'}(x_2')$, pour tout $\varepsilon \geq 0$, il existe $\pi_1^\varepsilon \in \Pi(\nu_1, \nu_1')$ tel que

$$T_c(\nu|\nu') \leq T_{c_1}(\nu_1|\nu_1') + \iint T_{c_2}(\nu_2^{x_1}|\nu_2'^{x_1'})\, d\pi_1^\varepsilon(x_1, x_1') + \varepsilon,$$

où $c = c_1 \oplus c_2$. Le terme d'erreur ε peut être choisi égal à zéro quand $\mathcal{X}_1$ et $\mathcal{X}_2$ sont compacts.

Ainsi si $\mu \in \mathcal{P}(\mathcal{X})$ satisfait l'inégalité transport $\mathbf{T}_{c,\beta}(a_1, a_2)$, alors $\mu^n = \mu \otimes \cdots \otimes \mu \in \mathcal{P}(\mathcal{X}^n)$ satisfait $\mathbf{T}_{c^n, \beta^{\Box n}}(a_1, a_2)$, avec pour tout $p \in \mathcal{P}(\mathcal{X}^n)$ de marginales $p_i, i \in \{1, \ldots, n\}$,

$$c^n(x, p) = c^{\oplus n}(x, p) := c(x_1, p_1) + \cdots + c(x_n, p_n), \qquad x = (x_1, \ldots, x_n) \in \mathcal{X}^n,$$

et puisque β est une fonction convexe,

$$\beta^{\Box n}(t) := \beta \Box \cdots \Box \beta(t) = n\beta(t/n), \qquad t \geq 0.$$

De l'inégalité de transport $\mathbf{T}_{c^n, \beta^{\Box n}}(a_1, a_2)$, on déduit que μ^n satisfait la propriété de concentration suivante, pour tout $A \in \mathcal{B}(\mathcal{X}^n)$,

$$\mu^n(A)^{a_1} \mu^n(\mathcal{X} \setminus A_{r,c^n})^{a_2} \leq e^{-n\beta(r/n)}, \qquad \forall r \geq 0.$$

Le profil de concentration dans le membre de droite n'est indépendant de n que si β est linéaire. On dit alors que μ satisfait *une propriété de concentration adimensionnelle.*

Définition II.4.1. *On dit qu'une mesure $\mu \in \mathcal{P}(\mathcal{X})$ satisfait la propriété de concentration adimensionnelle associée à la fonction de côut $c : \mathcal{X} \times \mathcal{P}(\mathcal{X})$, si pour tout $n \geq 1$ et pour tout $A \in \mathcal{B}(\mathcal{X}^n)$,*

$$\mu^n(A)^{a_1} \mu^n(\mathcal{X} \setminus A_{r,c^n})^{a_2} \leq e^{-r}, \qquad \forall r \geq 0.$$

Pour les grossissements usuels correspondant à une fonction de coût $\omega(x, y) = \alpha(d(x, y))$, Gozlan a démontré que cette propriété de concentration adimensionnelle était en fait équivalente à une inégalité de transport. La démonstration repose sur des théorèmes de grandes déviations (voir [Goz09]). Dans [5], nous proposons une démonstration moins technique. Suite à une remarque de Ledoux, cette démonstration s'adapte en travaillant directement sur les formulations fonctionnelles de la propriété de concentration adimensionnelle et de l'inégalité de transport. Dans [12], la souplesse de cette approche permet d'étendre aux inégalités de transport $\mathbf{T}_c(a_1, a_2)$ la caractérisation de ces inégalités en terme de concentration adimensionnelle. Le résultat, écrit dans sa version fonctionnelle, est le suivant.

Proposition II.4.2. *[12] Les assertions suivantes sont équivalentes.*

(i) μ satisfait $\mathbf{T}_c(a_1, a_2)$: pour toute fonction $\psi \in \mathcal{C}_b(\mathcal{X})$,

$$\left(\int e^{\frac{R_c \psi}{a_2}} d\mu\right)^{a_2} \left(\int e^{-\frac{\psi}{a_1}} d\mu\right)^{a_1} \leq 1.$$

(ii) Pour tout entier $n \geq 1$ et pour toute fonction $\varphi \in \mathcal{C}_b(\mathcal{X}^n)$,

$$\mu^n(\varphi \leq m)^{a_1} \mu^n(R_{c^n}\varphi > m + r)^{a_2} \leq e^{-r}, \qquad \forall m \in \mathbb{R}, \forall r \geq 0. \quad \text{(II.11)}$$

Idée de la démonstration. Comme nous l'avons vu, $(i) \Rightarrow (ii)$ résulte des propriétés de tensorisation de $\mathbf{T}_c(a_1, a_2)$.

Pour montrer (i) à partir de (ii), on estime le produit des moments exponentiels de $R_{c^{\oplus n}}\varphi$ et $-\varphi$ grâce aux estimations de queues de distribution données par (ii). Plus précisément (ii) donne : pour tout $\varepsilon > 0$

$$\left(\int e^{\frac{R_{c^n}\varphi}{(1+\varepsilon)a_2}}\, d\mu^n\right)^{a_2} \left(\int e^{-\frac{\varphi}{(1-\varepsilon)a_1}}\, d\mu^n\right)^{a_1} \leq K(\varepsilon, a_1, a_2),$$

où $K(\varepsilon, a_1, a_2)$ est une constante indépendante de n. On souhaite "tendre" cette inégalité en remplaçant cette constante par 1. Pour cela, on choisit l'application $\varphi(x) = \psi(x_1) + \cdots + \psi(x_n)$, $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$, pour laquelle

$$R_{c^n}\varphi(x) = R_c\psi(x_1) + \cdots + R_c\psi(x_n).$$

Par indépendance, l'inégalité précédente s'écrit alors

$$\left(\int e^{\frac{R_c\psi}{(1+\varepsilon)a_2}}\, d\mu\right)^{a_2} \left(\int e^{-\frac{\psi}{(1-\varepsilon)a_1}}\, d\mu\right)^{a_1} \leq K(\varepsilon, a_1, a_2)^{1/n}.$$

Le résultat découle de cette inégalité lorsque n tend vers $+\infty$ puis ε vers 0. $\square$

II.5 Lien avec les inégalités de Sobolev logarithmiques - semi groupe d'Hamilton-Jacobi

Dans ce paragraphe, sont présentés brièvement les résultats des articles [6], [7], [8] qui concernent les liens entre les inégalités de transport usuelles et les inégalités de type Sobolev logarithmiques modifiées.

Les coûts de transport considérés sont des coût usuels sur un espace métrique $(\mathcal{X}, d)$,

$$c(x, p) = \int \alpha(d(x, y))dp(y), \qquad x \in \mathcal{X}, p \in \mathcal{P}(\mathcal{X}),$$

ou encore en dimension n

$$c^n(x, p) := \int \sum_{i=1}^n \alpha(d(x_i, y_i))dp(y), \qquad x \in \mathcal{X}^n, p \in \mathcal{P}(\mathcal{X}^n),$$

où $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ est une fonction convexe telle que $\alpha(0) = \alpha'(0) = 0$. Dans ce cadre, on note $\mathcal{T}_\alpha(\mu, \nu) = T_c(\mu|\nu)$ ou encore $\mathcal{T}_\alpha(\mu, \nu) = T_{c^n}(\mu|\nu)$ les coûts de transport optimaux. On note aussi $\mathbf{T}_\alpha(b)$ l'inégalité de transport $\mathbf{T}_c^+(b)$ qui coincide avec $\mathbf{T}_c^-(b)$, et $\mathbf{T}_{\alpha^n}(b)$ l'inégalité de transport $\mathbf{T}_{c^n}^+(b)$ ou $\mathbf{T}_{c^n}^-(b)$ en dimension n.

Ecrire l'inégalité de Sobolev logarithmique nécessite de définir le gradient : pour $f : \mathcal{X} \to \mathbb{R}$ localement lipschitzienne, on pose pour tout $x \in \mathcal{X}$,

$$|\nabla^+ f|(x) = \limsup_{y \to x} \frac{[f(y) - f(x)]_+}{d(x, y)}, \quad |\nabla^- f|(x) = \limsup_{y \to x} \frac{[f(y) - f(x)]_-}{d(x, y)},$$

et $|\nabla^+ f|(x) = |\nabla^- f|(x) = 0$ si x est un point isolé. Si $\mathcal{X}$ est une variété riemannienne, lorsque f est différentiable, $|\nabla^+ f|(x)$ et $|\nabla^- f|(x)$ correspondent à la norme du vecteur gradient $\nabla f(x)$ dans l'espace tangent $T_x\mathcal{X}$ au point x.

Plus généralement, lorsque $(\mathcal{X}, d)$ est un espace complet, Ambrosio-Gigli-Savaré on démontré dans [AGS14] (Remarque 6.5), que pour toute fonction localement lipschitzienne, pour μ-presque tout x

$$|\nabla^+ f|(x) = |\nabla^- f|(x),$$

sous l'hypothèse très faible suivante sur la mesure de Borel μ : il existe $x_0 \in \mathcal{X}$ et $\kappa > 0$ tel que pour tout $r > 0$,

$$\mu(B_r(x_0)) \leq e^{\kappa r^2},$$

où $B_r(x_0) = \{x \in \mathcal{X}, d(x, x_0) < r\}$.

En dimension n, pour toute fonction $f : \mathcal{X}^n \to \mathbb{R}$ telle que l'application partielle $f_i^x, x \in \mathcal{X}^n$ soit localement lipschitzienne, en posant

$$f_i^x(u) = f(x_1, \dots, x_{i-1}, u, x_{i+1}, \dots, x_n), \qquad \forall u \in \mathcal{X}, \quad 1 \leq i \leq n,$$

on définit

$$|\nabla_i^+ f|(x) = \limsup_{y_i \to x_i} \frac{[f_i^x(y_i) - f(x)]_+}{d(x_i, y_i)}, \quad |\nabla_i^- f|(x) = \limsup_{y_i \to x_i} \frac{[f_i^x(y_i) - f(x)]_-}{d(x_i, y_i)}.$$

Définition II.5.1. *Une probabilité $\mu \in \mathcal{P}(\mathcal{X})$ satisfait l'inégalité de Sobolev logarithmique modifiée $\mathbf{LogSob}_\alpha^+(b)$, $b \geq 0$, associée à la fonction de coût α, si pour toute fonction $f : \mathcal{X} \to \mathbb{R}$ localement lipschitzienne, on a*

$$\mathbf{LogSob}_\alpha^+(b) : \qquad \mathrm{Ent}_\mu(e^f) \leq b \int \alpha^*(|\nabla^+ f|) e^f d\mu,$$

où $\alpha^*(h) = \sup_{t \geq 0}\{ht - \alpha(t)\}$ et pour toute fonction $g : \mathcal{X} \to \mathbb{R}^+$,

$$\mathrm{Ent}_\mu(g) = \int g \log g \, d\mu - \int g \, d\mu \log \int g d\mu.$$

En dimension n, une probabilité $\mu \in \mathcal{P}(\mathcal{X}^n)$ satisfait l'inégalité de Sobolev logarithmique modifiée $\mathbf{LogSob}_{\alpha^n}^+(b)$, si pour toute fonction $f : \mathcal{X}^n \to \mathbb{R}$ localement lipschitzienne dans chacune des variables

$$\mathbf{LogSob}_{\alpha^n}^+(b) : \qquad \mathrm{Ent}_\mu(e^f) \leq b \int \sum_{i=1}^n \alpha^*(|\nabla_i^+ f|) e^f d\mu.$$

De la même façon, on définit l'inégalité de Sobolev logarithmique $\mathbf{LogSob}_\alpha^-(b)$ en remplaçant $|\nabla^+ f|$ par $|\nabla^- f|$. Lorsque pour toute fonction localement lipschitzienne $f : \mathcal{X} \to \mathbb{R}$, $|\nabla^+ f| = |\nabla^- f|$ μ-presque surement, on note $\mathbf{LogSob}_\alpha(b)$. Lorsque α est le coût quadratique, $\alpha(t) = t^2$, $t \geq 0$, $\alpha^*(h) = h^2/4$, $h \geq 0$, l'inégalité de Sobolev logarithmique est notée $\mathbf{LogSob}_2(b)$.

Dans les articles [6], [7], [8], sont démontrés des extensions ou des compléments du théorème d'Otto-Villani suivant.

Théorème II.5.1. *[OV00] Soit $\mathcal{X}$ une variété riemannienne. Si $\mu \in \mathcal{P}_2(\mathcal{X})$ satisfait l'inégalité de Sobolev logarithmique* $\mathbf{LogSob}_2(b)$ *alors μ satisfait l'inégalité de transport de Talagrand* $\mathbf{T}_2(b)$.

Dans la partie heuristique de [OV00], Otto et Villani expliquent leur preuve de ce résultat en interprétant l'espace de Wasserstein $(\mathcal{P}_2(X), W_2)$ comme une variété riemanienne et en considérant le flot gradient de l'entropie relative $\nu \mapsto H(\nu|\mu)$.

Bobkov, Gentil et Ledoux [BGL01] ont proposé une autre démontration s'appuyant sur la formule de Hopf-Lax pour les solutions de l'équation d'Hamilton-Jacobi. Plus précisément sur une variété Riemannienne $\mathcal{X}$, l'opérateur d'infimum-convolution

$$v(x,t) = Q_t f(x) = \inf_{y \in \mathcal{X}} \left\{ f(y) + \frac{1}{2t} d(x,y)^2 \right\}, \qquad x \in \mathcal{X}, t > 0,$$

est un semi-groupe, solution de l'équation d'Hamilton-Jacobi

$$\frac{\partial v}{\partial t} = -\frac{1}{2}|\nabla v|^2, \qquad \text{avec} \quad v(x,0) = f(x), \ \forall x \in \mathcal{X}.$$

Cattiaux et Guillin ont montré que la réciproque de ce théorème peut être fausse, en proposant un contre exemple [CG06] (voir aussi [Goz07]). L'article [6] compléte ce théorème ; nous avons obtenu que l'inégalité de transport $\mathbf{T}_2$ est en fait équivalente (à des constantes près) à l'inégalité de Sobolev logarithmique, lorsqu'elle est restreinte à une classe de fonctions K-semi-convexes ($K > 0$), dans le cadre euclidien, $\mathcal{X}^n = \mathbb{R}^n$. Rappelons qu'une fonction $f : \mathbb{R}^n \to \mathbb{R}$ est K-semi-convexe si pour tout $\lambda \in [0,1]$ et pour tout $x, y \in \mathbb{R}^n$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) + K\lambda(1-\lambda)|x-y|^2.$$

Cette caractérisation est en fait valable pour toute fonction de coût α de dérivée concave. L'exemple type est $\alpha = \alpha_{p_1,p_2}$ avec $p_1 = 2$ et $p_2 = p \in [1,2]$ où

$$\alpha_{p_1,p_2}(h) := \begin{cases} h^{p_1} & \text{si } 0 \leq h \leq 1 \\ \frac{p_1}{p_2} h^{p_2} + 1 - \frac{p_1}{p_2} & \text{si } h \geq 1 \end{cases}.$$

Les exemples génériques de mesures satisfaisant une inégalité de transport $\mathbf{T}_{\alpha_{2,p}}(b)$ sur $\mathbb{R}$ sont : la mesure gaussienne pour $p = 2$ [Tal96c], la mesure exponentielle pour $p = 1$ [Tal91, BK08], et plus largement les probabilités $d\mu_p = e^{-|t|^p}/Z_p dt$, $p \in [1,2]$ [GGM05]. Remarquons que ces probabilités μ_p satisfont aussi une inégalité de Sobolev logarithmique $\mathbf{LogSob}_{\alpha_{2,p}}(b)$ (voir [GGM07, BR08, Goz07]).

Une application majeure de cette nouvelle caractérisation des inégalités de transport est le résultat de perturbation suivant.

Corollaire II.5.1. *[6] Soit $\mu \in \mathcal{P}(\mathcal{X}^n)$ et $\tilde{\mu} \in \mathcal{P}(\mathcal{X}^n)$ de densité e^{ϕ} par rapport à μ, $\phi : \mathcal{X}^n \to \mathbb{R}$. Supposons que α est une fonction convexe sur $\mathbb{R}^+$ telle que $\alpha(0) = \alpha'(0) = 0$ et α' est concave. Si μ satisfait $\mathbf{T}_{\alpha^n}(C)$ alors $\tilde{\mu}$ satisfait $\mathbf{T}_{\alpha^n}\left(8Ce^{\operatorname{Osc}\phi}\right)$ avec* $\operatorname{Osc}\phi = \sup\phi - \inf\phi$.

Ce type de résultat de perturbation avait éte établi par Holley et Stroock [HS87] pour les inégalités de Sobolev logarithmiques. Appliqués aux inégalités de Sobolev logarithmiques restreintes à une classe de fonctions, leurs arguments donnent le corollaire précédent. Nous avons appris de Villani, que ce type de résultat était l'une des questions restée ouverte, qui avait motivé l'article [OV00].

Dans [7] et [8], nous élargissons cette caractérisation des inégalités de transport et les résultats de perturbation qui en découlent, dans deux directions. La fonction de coût α telle que $\alpha(0) = \alpha'(0) = 0$ satisfait uniquement la condition Δ_2 (II.5). Cette fois-ci, toutes les fonctions génériques $\alpha_{p_1,p_2}, p_1 \geq 2, p_2 \geq 1$ satisfont cette condition. Par ailleurs, nous sortons du cadre Riemannien : les résultats sont valables dans des espaces métriques séparables complets.

L'extension aux fonctions de coût satisfaisant la condition Δ_2 a été possible dès lors que nous avons compris que la classe des fonctions $K-\alpha$-convexe, $K > 0$, est celle sur laquelle il faut restreindre l'inégalité de Sobolev logarithmique. Plus précisément, une fonction $f : \mathcal{X} \to \mathbb{R}$ est $K - \alpha$-convexe s'il existe une fonction $h : \mathcal{X} \to \mathbb{R}$ telle que

$$f(x) = \sup_{y\in\mathcal{X}} \{h(y) - K\alpha(d(x, y))\} = P_\alpha^K h(x), \qquad \forall x \in \mathcal{X}.$$

Ces fonctions apparaissent naturellement dans l'étude du transport optimal de mesure associé à la fonction de coût α (voir [Vil09]).

En fait, sur l'espace euclidien, lorsque α est la fonction quadratique $\alpha(h) = h^2$, $h \geq 0$, les fonctions $K-\alpha$-convexes coincides exactement avec les fonctions semi-convexes, ce qui n'est pas vrai pour d'autres fonctions α.

Sous sa forme générale la plus simple, la caractérisation des inégalités de transport s'énonce ainsi (voir Théorème 1.12 [7] et Théorème 5.1 [8]) dans les espaces $(\mathcal{X}, d)$ géodésiques. Rappelons que $(\mathcal{X}, d)$ est un *espace géodésique* si pour tous $x, y \in \mathcal{X}$, il existe un chemin $(\gamma_t)_{t\in[0,1]}$ tracé dans $\mathcal{X}$, telle que $\gamma_0 = x$, $\gamma_1 = y$ et $d(\gamma_s, \gamma_t) = |t - s|d(x, y)$, pour tout $s, t \in [0, 1]$.

Théorème II.5.2. *Soit $(\mathcal{X}, d)$ un espace géodésique. Soit $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ une fonction convexe satisfaisant la condition Δ_2* (II.5) *et telle que $\alpha(0) = \alpha'(0) = 0$. Les propriétés suivantes sont équivalentes.*

(1) Il existe une constante $C_1 > 0$, telle que μ satisfait l'inégalité de transport $\mathbf{T}_\alpha(C_1)$.

(2) Il existe des constantes $C_2 > 0$ et $\lambda > 0$, telles que μ satisfait l'inégalité (τ)-log-Sobolev suivante : pour toute fonction $f : \mathcal{X} \to \mathbb{R}$, localement lipschitzienne,

$$(\tau) - \mathbf{LogSob}_\alpha(C_2, \lambda) \qquad \mathrm{Ent}_\mu(e^f) \leq C_2 \int (f - Q_\alpha^\lambda f)e^f d\mu,$$

où $Q_\alpha^\lambda f(x) := \inf_{y\in\mathcal{X}} \{f(y) + \lambda\alpha(d(x, y))\}$, $x \in \mathcal{X}$.

(3) Il existe des constantes $C_3 > 0$ et $\lambda > 0$, telles que μ satisfait l'inégalité de Sobolev logarithmique modifiée restreinte suivante : pour toute fonction $f : \mathcal{X} \to \mathbb{R}$, $K - \alpha$-convexe avec $0 \leq K < \lambda$

$$\mathbf{r} - \mathbf{LogSob}_\alpha(C_3, \lambda) \qquad \mathrm{Ent}_\mu(e^f) \leq C_3 \int \alpha^*(|\nabla^+ f|)e^f d\mu.$$

Nous introduisons au point (2) une nouvelle classe d'inégalités de Sobolev logarithmiques $(\tau) - \mathbf{LogSob}_\alpha(C_2, \lambda)$. La notation "$(\tau)$" fait référence à la propriété-(τ) de Maurey [Mau91] où intervient aussi l'opérateur d'infimum-convolution. Dans le membre de droite de cette nouvelle inégalité de Sobolev (la "forme de Dirichlet"), la quantité $(f - Q_\alpha^\lambda f)$ peut être interprétée comme un gradient de la fonction f associé au coût α. L'idée pour passer de (2) à (3) est de comparer ce type de gradient à $\alpha^*(|\nabla^+ f|)$ lorsque la fonction f est $K - \alpha$-convexe pour $K < \lambda$ (voir Lemme 5.3, [8]).

Idée de la preuve. Le passage de (1) à (2) repose simplement sur l'inégalité de Jensen (voir Théorème 2.1, [7]). Cet argument avait été utilisé une première fois dans [1] pour obtenir des inégalités de Sobolev logarithmiques restreintes aux fonctions convexes, pour des mesures à support borné dépendantes.

Le passage de (2) à (3) a été expliqué ci-dessus.

Le passage de (3) à (1) s'effectue de manière totalement différente dans l'article [7] et dans l'article [8].

Dans [7], un des éléments clefs de la preuve est la propriété de tensorisation de l'inégalité de Sobolev logarithmique, qui repose sur l'inégalité suivante : pour toute fonction $h : \mathcal{X}^n \to \mathbb{R}$

$$\mathrm{Ent}_{\mu^n}(h) \leq \int \sum_{i=1}^n \mathrm{Ent}_\mu(h_i^x) d\mu^n(x),$$

où h_i^x est la i-ème application partielle associée à h. Puis, par l'argument de Herbst, l'inégalité de Sobolev logarithmique restreinte tensorisée fournit une borne supérieure pour le moment exponentiel de la fonction α^n-convexe $P_{\alpha^n} f$, $f : \mathcal{X}^n \to \mathbb{R}$, $f \geq 0$ (voir la preuve du Théorème 4.1, [7]) : il existe une constante $c \in]0, 1]$, indépendante de n, telle que

$$\int e^{P_{\alpha^n} f} d\mu^n \leq e^{\|f\|_\infty (1-c)} e^{c \int f d\mu^n},$$

où

$$P_{\alpha^n} f(x) := \sup_{y \in \mathcal{X}^n} \left\{ f(y) - \sum_{i=1}^n \alpha(d(x_i, y_i)) \right\}, \qquad x \in \mathcal{X}^n.$$

Ensuite, comme dans la preuve de la proposition II.4.2, on sait tendre cette inégalité lorsque n tend vers $+\infty$ pour retrouver la forme fonctionnelle de l'inégalité de transport $\mathbf{T}_\alpha(1/c)$ (cf. Corollaire 2.7, [7]).

Dans [8], la preuve est une adaptation de celle du théorème d'Otto-Villani présentée dans [BGL01], qui repose sur une formule de Hopf-Lax pour les solutions de l'équation de Hamilton-Jacobi (voir Théorème 1.1, [8]). Pour adapter cette preuve, nous avons au préalable démontré des formules de Hopf-Lax pour des solutions d'équations d'Hamilton-Jacobi dans les espaces métriques généraux.

Théorème II.5.3. *[8] Soit $(\mathcal{X}, d)$ un espace métrique séparable complet et $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ une fonction convexe. Pour toute fonction $f : \mathcal{X} \to \mathbb{R}$ semi-continue inférieurement et bornée,*

$$\frac{d}{dt_+} Q_t f(x) \leq -\alpha^* \left(|\nabla^- Q_t f| \right), \qquad \forall t > 0, \forall x \in \mathcal{X},$$

où $Q_t f(x) := \inf_{y\in X}\left\{f(y)+t\alpha\left(\frac{d(x,y)}{t}\right)\right\}$*. De plus, on a égalité si* $(\mathcal{X},d)$ *est un espace géodésique.*

Ce résultat avait été obtenu par Lott et Villani [LV07, Vil09] dans les espace géodésiques mesurés $(\mathcal{X},d,\mu)$, pour tout t, pour tout x en dehors d'un ensemble N_t de μ-mesure nulle, et avec des hypothèses supplémentaires sur μ (μ satisfait une condition de doublement et une inégalité de Poincaré locale). L'intérêt du Théorème II.5.3 est que ces hypothèses disparaissent ; il ne fait appel à aucune théorie de la mesure.

Le Théorème II.5.3 précédent a été obtenu simultanément par Ambrosio, Gigli et Savaré dans leur papier [AGS14] sur l'étude du flot gradient de l'entropie dans les espaces métriques. $\square$

Dans [7], nous présentons l'argument de changement de métrique qui permet de restreindre l'étude des inégalités de transport de fonction de coût α (satisfaisant la condition Δ_2 (II.5)), à celle des inégalités de transport de fonction de coût $\alpha_{p,p}$, $p>1$. Cet argument repose sur le Lemme II.2.1. D'après ce lemme, en posant $p_\alpha=\sup_{h>0}\frac{h\alpha'_+(h)}{\alpha(h)}$, on a $\alpha(d(x,y))=(d_\alpha(x,y))^{p_\alpha}$ où d_α est une distance sur $\mathcal{X}$. Ainsi, l'étude des l'inégalités de transport $\mathbf{T}_\alpha(b)$ dans l'espace métrique $(\mathcal{X},d)$ se réduit à l'étude des inégalités de transport $\mathbf{T}_{p_\alpha}(b)$ dans l'espace métrique $(\mathcal{X},d_\alpha)$.

Par ailleurs, les techniques développées sur les espaces métriques dans [7] et [8], ont permis d'étendre le Théorème d'Otto-Villani [OV00] (voir aussi [GGM05]) aux espaces métriques, et de caractériser les inégalités de Sobolev logarithmiques en termes d'hypercontractivité du semi-groupe d'Hamilton-Jacobi (voir [BGL01]).

CHAPITRE III

Résultats autour des coûts barycentriques

III.1 Coûts barycentriques et ordre convexe

Soit γ et ν deux probabilités sur $\mathbb{R}^n$. On dit que γ est *dominée par ν au sens de l'ordre convexe*, et on note $\gamma \preceq \nu$, si pour toute fonction convexe $f : \mathbb{R}^n \to \mathbb{R}$,

$$\int f d\gamma \leq \int f d\nu.$$

Le théorème de Strassen suivant donne une définition alternative de l'ordre convexe.

Théorème III.1.1. *[Str65] Soit γ et ν deux probabilités sur $\mathbb{R}^n$, alors $\gamma \preceq \nu$ si et seulement s'il existe une martingale (X, Y) pour laquelle X suit la loi γ et Y la loi ν. Autrement dit, si π est la loi du couple (X, Y), décomposée sous la forme*

$$d\pi(x, y) = d\gamma(x)dp_x(y),$$

où p est un noyau de Markov tel que $\gamma p = \nu$, alors on a pour γ-presque tout x,

$$\int y \, dp_x(y) = x.$$

Idée de la preuve. Ce résultat est une simple conséquence de la formule duale du coût de transport barycentrique optimal :

$$\begin{aligned}\overline{\mathcal{T}}_1(\nu|\gamma) &:= \inf_{\pi \in \Pi(\mu,\nu)} \int \left| x - \int y \, dp_x(y) \right| d\gamma(x) \\ &= \sup \left\{ \int f \, d\gamma - \int f \, d\nu \, ; \, f \text{ convexe, 1-lipschitzienne,} \right. \\ &\qquad\qquad \left. \text{bornée inférieurement} \right\},\end{aligned}$$

(voir Proposition 3.2. [12]). Ainsi, si $d\pi^*(x, y) = d\gamma(x)dp^*_x(y)$ est la loi d'une martingale (X, Y), de marginales γ et ν alors

$$0 \leq \overline{\mathcal{T}}_1(\nu|\gamma) \leq \int \left| x - \int y \, dp^*_x(y) \right| d\gamma(x) = 0.$$

Par suite $\overline{\mathcal{T}}_1(\nu|\gamma) = 0$ et la formule de dualité précédente implique que $\nu \preceq \gamma$. □

Soit $\theta : \mathbb{R}^n \to \mathbb{R}$ une fonction convexe. Soit μ une mesure de probabilité sur $\mathbb{R}^n$ absolument continue par rapport à la mesure de Lebesgue, de telle sorte que pour toute probabilité γ, il existe une application de transport optimale $S^* : \mathbb{R}^n \to \mathbb{R}^n$ telle que $S^*\#\mu = \gamma$ et

$$\mathcal{T}_\theta(\gamma, \mu) = \inf_{\pi\in\Pi(\mu,\gamma)} \int \theta(x-y)d\pi(x,y) = \int \theta(x - S^*(x))d\mu(x).$$

Le coût de transport barycentrique optimal $\overline{\mathcal{T}}_\theta(\nu|\mu)$ entre deux probabilités μ et ν sur $\mathbb{R}^n$ est défini par

$$\overline{\mathcal{T}}_\theta(\nu|\mu) := \inf_{\pi\in\Pi(\mu,\nu)} \int \theta\left(x - \int y\, dp_x(y)\right) d\mu(x), \qquad d\pi(x,y) = d\gamma(x)dp_x(y),$$

Ce coût s'exprime en fonction du coût de transport usuel $\mathcal{T}_\theta$ de la manière suivante.

Proposition III.1.1. *Sous les conditions précédentes sur μ, pour toute probabilité ν telle que $\overline{\mathcal{T}}_\theta(\nu|\mu) < \infty$, on a*

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf_{\gamma\in\mathcal{P}(\mathbb{R}^n),\gamma\preceq\nu} \mathcal{T}_\theta(\gamma, \mu).$$

Démonstration. Soit $\gamma \preceq \nu$. D'après le théorème de Strassen précédent, il existe un noyau p^* tel que $\gamma p^* = \nu$ et $x = \int y\, dp^*_x(y)$ γ-presque sûrement. Par ailleurs, par hypothèse, il existe une application de transport $S^* : \mathbb{R}^n \to \mathbb{R}^n$ telle que $S^*\#\mu = \gamma$ et

$$\mathcal{T}_\theta(\gamma, \mu) = \int \theta(x - S^*(x))d\mu(x).$$

Considérons le noyau $p_x(dy) = p^*_{S^*(x)}(dy)$. On vérifie simplement que $\mu p = \nu$ et pour μ-presque tout x,

$$\int y\, dp_x(y) = \int y\, dp^*_{S^*(x)}(y) = S^*(x).$$

Par conséquent

$$\mathcal{T}_\theta(\gamma, \mu) = \int \theta\left(x - \int y\, dp_x(y)\right) d\mu(x) \geq \overline{\mathcal{T}}_\theta(\nu|\mu),$$

et en optimisant sur toutes les probabilités γ, avec $\gamma \preceq \nu$, on obtient

$$\inf_{\gamma\in\mathcal{P}(\mathbb{R}^n),\gamma\preceq\nu} \mathcal{T}_\theta(\gamma, \mu) \geq \overline{\mathcal{T}}_\theta(\nu|\mu).$$

Pour établir l'inégalité inverse, considérons un noyau p tel que $\mu p = \nu$ et

$$\int \theta\left(x - \int y\, dp_x(y)\right) d\mu(x) < \infty.$$

Soit $S : \mathbb{R}^n \to \mathbb{R}^n$ l'application mesurable définie par

$$S(x) = \int y\, dp_x(y),$$

pour μ-presque tout x. Soit γ la mesure image de μ par l'application S, $\gamma = S\#\mu$. On a alors $\gamma \preceq \nu$ car d'après l'inégalité de Jensen, pour toute fonction convexe $f : \mathbb{R}^n \to \mathbb{R}$,

$$\int f\,d\gamma = \int f\Big(\int y\,dp_x(y)\Big)\,d\mu(x) \leq \iint f(y)\,dp_x(y)d\mu(x) = \int f\,d\nu;$$

et par ailleurs

$$\begin{aligned}\int \theta\Big(x - \int y\,dp_x(y)\Big)\,d\mu(x) &= \int (x - S(x))\,d\mu(x)\\ &\geq \mathcal{T}_\theta(\gamma,\mu) \geq \inf_{\gamma\in\mathcal{P}(\mathbb{R}^n),\gamma\preceq\nu} \mathcal{T}_\theta(\gamma,\mu).\end{aligned}$$

On obtient alors l'inégalité voulue en optimisant sur tous les noyaux p telle que $\mu p = \nu$. □

III.2 Inégalités de transport barycentriques et inégalités de Sobolev logarithmiques

Dans la deuxième partie (section II.5), nous avons expliqué que certaines inégalités de transport usuelles sont équivalentes à des inégalités de Sobolev restreintes à des classes de fonctions. Fixons dans le reste de ce paragraphe une norme $\|\cdot\|$ sur $\mathbb{R}^n$. Dans l'article [12], nous établissons le même type de résultats avec les inégalités de transport barycentriques $\mathbf{T}_c^+(b)$ et $\mathbf{T}_c^-(b)$ avec la fonction de coût

$$c(x,p) = \left\|x - \int y\,dp(y)\right\|^2, \qquad x\in\mathbb{R}^n, p\in\mathcal{P}(\mathbb{R}^n).$$

Pour ces coûts barycentriques, le théorème général de dualité de Kantorovich admet une version particulière, établie dans le Théorème 2.11 de l'article [12],

$$\begin{aligned}&T_c(\nu|\mu) = \overline{\mathcal{T}}_2(\nu|\mu)\\ &= \sup\left\{\int Q\varphi\,d\mu - \int \varphi\,d\nu\,;\,\varphi \text{ convexe, lipschitzienne, bornée inférieurement}\right\}.\end{aligned}$$

Dans le supremum, l'opérateur $Q\varphi$ est l'opérateur d'infimum-convolution usuel,

$$Q\varphi(x) = \inf_{y\in\mathbb{R}^n}\left\{\varphi(y) + \|x-y\|^2\right\}, \qquad x\in\mathbb{R}^n.$$

Ainsi, moyennant une restriction aux fonctions convexes, l'opérateur $Q\varphi$ remplace l'opérateur $R_c\varphi$ dans la formule de dualité (II.10).

Ce nouveau résultat de dualité donne une formulation fonctionnelle des inégalités de transport $\mathbf{T}_c^+(b)$ et $\mathbf{T}_c^-(b)$, dans laquelle n'intervient que l'opérateur d'infimum-convolution usuel $Q\varphi$, restreint aux fonctions convexes φ.

Par ailleurs, en appliquant par exemple la technique liée au semi-groupe d'Hamilton-Jacobi de l'article [8], nous avons établi une caractérisation des inégalités de transport $\mathbf{T}_c^+(b)$ et $\mathbf{T}_c^-(b)$ en termes d'inégalités de Sobolev logarithmiques restreintes à une classe de fonctions différente, pour chacune des inégalités $\mathbf{T}_c^+(b)$ et $\mathbf{T}_c^-(b)$. Ces résultats sont résumés dans les deux théorèmes suivants (voir [12] Théorème 8.8 et 8.15).

Théorème III.2.1. *[12] Soit $\mu \in \mathcal{P}_1(\mathbb{R}^n)$. Les propriétés suivantes sont équivalentes*

(1) Il existe une constante $C_1 > 0$ telle que μ satisfait $\mathbf{T}_2^+(C_1)$,

$$\overline{\mathcal{T}}_2(\nu|\mu) \leq C_1 H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}^n).$$

(2) Il existe une constante $C_2 > 0$ telle que pour toute fonction $\varphi : \mathbb{R}^n \to \mathbb{R}$ lipschitzienne convexe et bornée inférieurement,

$$\int e^{-\varphi/C_2} d\mu \leq e^{-\int Q\varphi/C_2 \, d\mu}.$$

(3) Il existe des constantes $C_3 > 0$ et $\lambda > 0$, telle que pour toute fonction $\psi : \mathbb{R}^n \to \mathbb{R}$, lipschitzienne, concave, bornée supérieurement et $\frac{\lambda}{2}\|\cdot\|^2$-convexe,

$$\mathrm{Ent}_\mu(e^\psi) \leq C_3 \int \|\nabla\psi\|_*^2 e^\psi d\mu,$$

où $\|\cdot\|_*$ est la norme duale de $\|\cdot\|$ sur $\mathbb{R}^n$.

Rappelons qu'une fonction $\psi : \mathbb{R}^n \to \mathbb{R}$ est $\frac{\lambda}{2}\|\cdot\|^2$-convexe s'il existe une fonction $g : \mathbb{R}^2 \to \mathbb{R}$ telle que pour tout $x \in \mathbb{R}^n$,

$$\psi(x) = \sup_{y \in \mathbb{R}^n} \left\{ g(y) - \frac{\lambda}{2}\|x-y\|^2 \right\}.$$

Si $\|\cdot\| = |\cdot|$ est la norme euclidienne, alors les fonctions ψ qui sont $\frac{\lambda}{2}\|\cdot\|^2$-convexes de classe $\mathcal{C}^2$ sont celles dont la matrice hessienne est minorée par $-\lambda I$ au sens des formes quadratiques.

Précisons que dans le théorème précédent, les points (1) et (2) sont équivalents avec la même constante $C_1 = C_2$.

Théorème III.2.2. *[12] Soit $\mu \in \mathcal{P}_1(\mathbb{R}^n)$. Les propriétés suivantes sont équivalentes*

(1) Il existe une constante $C_1 > 0$ telle que μ satisfait $\mathbf{T}_2^-(C_1)$,

$$\overline{\mathcal{T}}_2(\mu|\nu) \leq C_1 H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}^n).$$

(2) Il existe une constante $C_2 > 0$ telle que pour toute fonction $\varphi : \mathbb{R}^n \to \mathbb{R}$, lipschitzienne, convexe et bornée inférieurement,

$$\int e^{Q\varphi/C_2} d\mu \leq e^{\int \varphi/C_2 \, d\mu}.$$

(3) Il existe une constante $C_3 > 0$ telle que pour toute fonction $\varphi : \mathbb{R}^n \to \mathbb{R}$, lipschitzienne, convexe et bornée inférieurement,

$$\mathrm{Ent}_\mu(e^\varphi) \leq C_3 \int \|\nabla\varphi\|_*^2 e^\varphi d\mu,$$

où $\|\cdot\|_*$ est la norme duale de $\|\cdot\|$ sur $\mathbb{R}^n$.

(4) Il existe une constante $C_4 > 0$ telle que pour toute fonction $\varphi : \mathbb{R}^n \to \mathbb{R}$, lipschitzienne, convexe et bornée inférieurement,

$$\left\| e^{Q_t\varphi} \right\|_{a+t/(2C_4),(\mu)} \leq \| e^{\varphi} \|_{a,(\mu)} ,$$

où $Q_t\varphi(x) = \inf_{y\in\mathbb{R}^n} \left\{ \varphi(y) + \frac{1}{t}\|x-y\|^2 \right\}, x \in \mathbb{R}^n$, et pour tout $h : \mathbb{R}^n \to \mathbb{R}$, $p \geq 0$, $\|h\|_{p,(\mu)} = \left(\int |h|^p d\mu\right)^{1/p}$.

Les points (1) et (2) sont équivalents avec la même constante $C_1 = C_2$. Les points (3) et (4) sont équivalents avec la même constante $C_3 = C_4$. Les autres relations entre les constantes dans les deux derniers théorèmes sont données dans [12].

Dans le Théorème III.2.2, la propriété d'hypercontractivité (4) du semi-groupe d'Hamilton-Jacobi découle du point (3) par l'argument de Herbst usuel [BGL01] en appliquant l'inégalité de Sobolev logarithmique aux fonctions convexes $\lambda Q_t\varphi$, $\lambda > 0$. En effet, lorsque φ est convexe, la fonction $Q_t\varphi$ l'est aussi.

Les inégalités de Sobolev logarithmiques restreintes à la classe des fonctions convexes ont été étudiées par Adamaczak [Ada05] qui propose des conditions suffisantes sur la mesure μ pour qu'une telle inégalité soit vérifiée.

III.3 Inégalités de transport barycentriques pour la loi binômiale et la loi de Poisson

Les probabilités discrètes ne satisfont généralement pas l'inégalité de transport de Talagrand $\mathbf{T}_2$. Pour s'en convaincre, il suffit par exemple de considérer μ_ρ une combinaison de deux masses de Dirac en deux réels distincts a et b, $\mu_\rho = \rho\delta_a + (1-\rho)\delta_b$, $\rho \in]0,1[$. Dire que μ_ρ satisfait $\mathbf{T}_2(C)$, $C > 0$ s'écrit alors pour tout $q \in [0,1]$

$$W_2^2(\mu_\rho, \mu_q) = \mathcal{T}_2(\mu_\rho, \mu_q) = |a-b|^2|q-\rho| \leq CH(\mu_q|\mu_\rho) = q\log\frac{q}{\rho} + (1-q)\log\frac{1-q}{1-\rho}.$$

On obtient une contradiction lorsque q tend vers ρ en remarquant que $H(\mu_q|\mu_\rho) = o(|q-\rho|)$.

Cependant la mesure μ_ρ satisfait l'inégalité de transport

$$W_1^2(\mu_\rho, \nu) \leq \frac{d(a,b)^2}{2} H(\nu, \mu_\rho), \qquad \forall \nu << \mu_\rho.$$

Pour obtenir d'autres inégalités de transport, une stratégie est de remplacer le coût usuel par un coût barycentrique. Pour simplifier choisissons pour μ_ρ la mesure de Bernoulli de paramètre ρ ($a = 0$ et $b = 1$). Dans [2] et [12] (voir Proposition 7.1), nous présentons une fonction de coût convexe optimale $\theta_{\rho,t}$: $\mathbb{R} \to \mathbb{R}^+$, $t \in]0,1[$ telle que μ_ρ satisfasse l'inégalité de transport barycentrique $\mathcal{T}_c(1/(1-t), 1/t)$ avec

$$c(x,p) = \theta_{\rho,t}\left(x - \int y\, dp(y)\right), \qquad x \in \mathbb{R}, \quad p \in \mathcal{P}(\mathbb{R}),$$

c'est à dire

$$\overline{\mathcal{T}}_{\theta_{\rho,t}}(\nu_1|\nu_2) \leq \frac{1}{1-t}H(\nu_1|\mu_\rho) + \frac{1}{t}H(\nu_2|\mu_\rho), \qquad \forall \nu_1, \nu_2 << \mu_\rho.$$

Par tensorisation, cette inégalité procure une inégalité de transport pour la mesure produit $\mu_\rho^n = \mu_\rho \otimes \cdots \otimes \mu_\rho$ associée au coût

$$c^n(x,p) = c(x_1,p_1) + \cdots + c(x_n,p_n), \qquad x = (x_1,\ldots,x_n) \in \mathbb{R}^n, p \in \mathcal{P}_1(\mathbb{R}^n).$$

Par projection, en remarquant que par convexité

$$c^n(x,p) \geq n\, c\left(\frac{\sum_i^n x_i}{n}, \frac{\sum_i^n p_i}{n}\right),$$

on obtient une inégalité de transport barycentrique pour la loi binômiale $\mu_{n,\rho}$ de paramètre n et ρ (voir [12], Corollaire 7.7)

$$\overline{\mathcal{T}}_{\theta_{\rho,t,n}}(\nu_1|\nu_2) \leq \frac{1}{1-t}H(\nu_1|\mu_{n,\rho}) + \frac{1}{t}H(\nu_2|\mu_{n,\rho}), \qquad \forall \nu_1, \nu_2 << \mu_{n,\rho},$$

où

$$\theta_{\rho,t,n}(h) = n\theta_{\rho,t}(h/n), \qquad h \in \mathbb{R}.$$

Finalement, en utilisant le théorème de convergence faible de la loi binômiale μ_{n,ρ_n} vers la loi de Poisson p_λ de paramètre $\lambda > 0$ lorsque $\rho_n = \lambda/n$, on démontre l'inégalité de transport barycentrique pour la loi de Poisson

$$\overline{\mathcal{T}}_{c_{\lambda,t}}(\nu_1|\nu_2) \leq \frac{1}{1-t}H(\nu_1|p_\lambda) + \frac{1}{t}H(\nu_2|p_\lambda), \qquad \forall \nu_1, \nu_2 << p_\lambda,$$

avec

$$c_{\lambda,t}(h) = \lim_{n\to\infty} n\theta_{\rho_n,t}(h/n), \qquad h \in \mathbb{R}.$$

Une particularité de la fonction de coût $c_{\lambda,t}$ est d'être nulle pour $h \geq 0$ (voir [12], Proposition 7.11).

Une stratégie différente, issue d'une notion de courbure sur les espaces discrets introduite par Maas [Maa11], permet d'obtenir d'autres inégalités de transport pour les probabilités discrètes. Le coût de transport optimal est alors défini par une formule de type Benamou-Brenier. Ce coût n'est pas comparable à un coût barycentrique.

Par ailleurs, notons que pour les processus de Poisson, une inégalité de transport d'une autre nature de type W_1 a été proposée par Ma, Shen,Wang et Wu [MSWW11].

III.4 Plan de transports optimaux pour les coûts barycentriques sur $\mathbb{R}$

Cette partie concerne la construction d'un couplage optimal π^* qui réalise le transport optimal barycentrique sur la droite réelle (en dimension 1).

dans ce paragraphe, la fonction de coût $\theta : \mathbb{R} \to \mathbb{R}$ est supposée paire et convexe. Pour toute probabilité μ, on note F_μ sa fonction de répartition, $F_\mu(x) = \mu(]-\infty, x]))$, et F_μ^{-1} son inverse généralisé,

$$F_\mu^{-1}(u) = \inf\{x \in \mathbb{R}, F_\mu(x) \geq u\}, \quad u \in [0,1].$$

Soit μ et γ deux probabilités. Supposons que μ n'a pas d'atome ; il est alors bien connu que pour toute fonction de coût convexe θ, le transport optimal usuel $T_\theta(\mu, \gamma)$ est atteint par le plan de transport déterministe

$$d\pi^*(x,y) = d\mu(x) d\delta_{S^*(x)}(y),$$

où S^* est l'application de transport monotone définie par

$$S^*(x) = F_\gamma^{-1} \circ F_\mu(x), x \in \mathbb{R}.$$

Autrement dit, il existe une application de transport S^*, indépendante de θ et telle que

$$\mathcal{T}_\theta(\mu, \gamma) = \int \theta(x - S^*(x))\, d\mu(x).$$

On souhaite obtenir le même type de résultat d'indépendance de la fonction de coût θ, pour un couplage optimal du coût de transport barycentrique $\overline{\mathcal{T}}_\theta(\nu|\mu)$. Pour celà, on peut s'appuyer sur le résultat préliminaire suivant obtenu dans [13] (Théorème 1.3).

Théorème III.4.1. *Soit $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. Il existe $\hat{\gamma} \in \mathcal{P}_1(\mathbb{R})$ tel que $\hat{\gamma} \preceq \nu$ et pour toute fonction convexe θ paire, on a*

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\hat{\gamma}, \mu).$$

Ce résultat reste valable lorsque μ possède des atomes. D'après ce qui précède, lorsque μ ne possède pas d'atome,

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\hat{\gamma}, \mu) = \int \theta(x - S^*(x))\, d\mu(x), \quad \text{avec } S^* = F_{\hat{\gamma}}^{-1} \circ F_\mu,$$

et puisque $\hat{\gamma} \preceq \nu$, d'après le Théorème III.1.1 de Strassen, il existe un noyaux p^* tel que $\hat{\gamma} p^* = \nu$ et $\hat{\gamma}$-presque surement

$$\int y\, dp_x^*(y) = x.$$

Puisque $\hat{\gamma}$ est indépendant de θ, le noyau p^* est lui aussi indépendant de θ. Par ailleurs, puisque $S^* \# \mu = \hat{\gamma}$, on obtient pour μ-presque tout x,

$$\int y\, dp_{S^*(x)}^*(y) = S^*(x),$$

ce qui donne finalement

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \int \theta\left(x - \int y\, dp_{S^*(x)}^*(y)\right) d\mu(x).$$

Ceci montre que si μ n'a pas d'atomes, alors le coût de transport barycentrique $\overline{\mathcal{T}}_\theta(\nu|\mu)$ est atteint pour le couplage optimal

$$d\pi^*(x,y) = d\mu(x) dp_{S^*(x)}^*(y),$$

qui est indépendant de la fonction de coût θ.

III.5 Caractérisation des probabilités sur $\mathbb{R}$ satisfaisant une inégalité de transport barycentrique

On sait caractériser les probabilités sur $\mathbb{R}$ satisfaisant de nombreuses inégalités fonctionnelles comme l'inégalité de Poincaré [Muc72], l'inégalité de Sobolev logarithmique [BG99, BR03] (voir aussi chapitre 6, [ABC+00]), les inégalités de transport usuelles [Goz12]. La caractérisation est donnée par des critères de type Hardy sur les queues de distributions et les densités des lois considérées. Cette section concerne la caractérisation des inégalités de transport barycentriques. La démarche adoptée reprend celle de Gozlan [Goz12].

Soit τ la loi exponentielle symétrique sur $\mathbb{R}$, de densité $e^{-|x|}/2$. Pour toute probabilité μ sur $\mathbb{R}$, notons U_μ la fonction de transport optimale qui transporte la mesure exponentielle sur la mesure μ,

$$U_\mu = F_\mu^{-1} \circ F_\tau,$$

c'est à dire,

$$U_\mu(x) = \begin{cases} F_\mu^{-1}\left(1 - \frac{1}{2}e^{-|x|}\right) & \text{si } x \geq 0 \\ F_\mu^{-1}\left(\frac{1}{2}e^{-|x|}\right) & \text{si } x \leq 0 \end{cases}$$

Voici un résultat central de [Goz12] : une probabilité μ satisfait l'inégalité de transport $\mathbf{T}_2(C)$ avec $C > 0$ si et seulement si elle satisfait l'inégalité de Poincaré ainsi que la condition suivante : il existe $b \geq 0$ tel que

$$\sup_{x\in\mathbb{R}} \left(U_\mu(x+u) - U_\mu(x)\right) \leq b\sqrt{1+u}, \qquad \forall u \geq 0,$$

ce qui impose un comportement particulier des queues de distribution de μ.

Nous avons établi l'analogue de ce résultat pour les inégalités de transport barycentriques avec le coût optimal $\overline{\mathcal{T}}_\theta$ lorsque la fonction θ est convexe symétrique et quadratique au voisinage de 0 (Théorème 1.2 [13]). Pour tout $a > 0$, notons $\theta^a(t) = \theta(at)$, $t \in \mathbb{R}$.

Théorème III.5.1. *[13] Soit θ une fonction de coût convexe symétrique telle que $\theta(t) = t^2$ pour tout $|t| \leq t_0$, $t_0 > 0$. Une probabilité μ satisfait les inégalités de transport barycentriques*

$$\overline{\mathbf{T}}^-_{\theta^a}: \qquad \overline{\mathcal{T}}_{\theta^a}(\mu|\nu) \leq H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

et

$$\overline{\mathbf{T}}^+_{\theta^a}: \qquad \overline{\mathcal{T}}_{\theta^a}(\nu|\mu) \leq H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

pour une certaine constante $a > 0$ si et seulement s'il existe $b \geq 0$ tel que

$$\sup_x \left(U_\mu(x+u) - U_\mu(x)\right) \leq b\,\theta^{-1}\left(u + t_0^2\right), \qquad \forall u \geq 0. \tag{III.1}$$

Comme nous l'avons vu dans le paragraphe III.3, certaines mesures discrètes satisfont des inégalités de transport barycentriques du type précédent. Par ailleurs, on sait qu'une probabilité qui satisfait l'inégalité de Poincaré usuelle a nécessairement un support connexe. Ces mesures discrètes sont donc des mesures qui satisfont une inégalité de transport barycentrique sans satisfaire une inégalité de Poincaré classique.

Cependant, la condition (III.1) du théorème précédent implique la propriété suivante : il existe $h > 0$ tel que

$$\sup_{x\in\mathbb{R}} (U_\mu(x+1) - U_\mu(x)) \leq h. \tag{III.2}$$

Bobkov et Götze [BG99] ont montré que cette condition est équivalente à l'existence d'une inégalité de Poincaré restreinte aux fonctions convexes pour la mesure μ. Les mesures qui satisfont (III.1) satisfont donc nécessairement une inégalité de Poincaré "convexe".

Idée de la preuve du Théorème III.5.1. La demonstration s'appuie sur les deux arguments suivants. On montre d'abord que (III.2) est équivalent à l'existence d'une inégalité de transport barycentrique pour une fonction de coût θ_1 quadratique puis linéaire. On obtient ainsi, de surcroît, une caractérisation de l'inégalité de Poincaré convexe en termes d'inégalité de transport. Le résultat est le suivant (Théorème 1.5, [13]).

Théorème III.5.2. *[13] Soit μ une probabilité sur $\mathbb{R}$. La condition* (III.2) *est équivalente a chacune des propriétés suivantes.*

(a) *Il existe $C > 0$ tel que pour toute fonction convexe f sur $\mathbb{R}$,*

$$\mathrm{Var}_\mu(f) \leq C \int_{\mathbb{R}} f'^2 \, d\mu.$$

(b) *Il existe des constantes $a, t_0 > 0$ telles que*

$$\overline{\mathbf{T}}^-_{\theta_1^a} : \qquad \overline{\mathcal{T}}_{\theta_1^a}(\mu|\nu) \leq H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

et

$$\overline{\mathbf{T}}^+_{\theta_1^a} : \qquad \overline{\mathcal{T}}_{\theta_1^a}(\nu|\mu) \leq H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

où la fonction θ_1 est définie par

$$\theta_1(t) = \begin{cases} t^2 & si\ |t| \leq t_0, \\ 2|t|t_0 - t_0^2 & si\ |t| > t_0. \end{cases}$$

A partir de ce résultat intermédiaire, la preuve du Théorème III.5.1 est la même que dans [Goz12]. Soit θ la fonction convexe vérifiant les hypothèses du théorème et la condition (III.1) et $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. D'après le Théorème III.4.1, nous savons qu'il existe $\hat{\gamma} \in \mathcal{P}_1(\mathbb{R})$, $\hat{\gamma} \preceq \nu$, tel que indépendamment de la fonction de coût θ

$$\overline{\mathcal{T}}_{\theta^a}(\nu|\mu) = \mathcal{T}_{\theta^a}(\hat{\gamma}, \mu), \qquad \forall a > 0.$$

Soit U une variable aléatoire de loi uniforme sur $[0,1]$, et π^* la loi du couple

$$(F_{\hat{\gamma}}^{-1}(U), F_\mu^{-1}(U)).$$

Il est bien connu que ce couplage π^*, indépendant de la fonction θ, réalise le transport optimal $\mathcal{T}_\theta(\hat{\gamma}, \mu)$ quel que soit la fonction convexe θ. Par conséquent, si θ_1 et θ_2 sont des fonctions convexes telles que

$$\theta \leq \theta_1 + \theta_2,$$

alors

$$\mathcal{T}_{\theta^a}(\hat{\gamma}, \mu) \leq \mathcal{T}_{\theta_1^a}(\hat{\gamma}, \mu) + \mathcal{T}_{\theta_2^a}(\hat{\gamma}, \mu), \qquad \forall a > 0.$$

Ainsi, on obtient

$$\overline{\mathcal{T}}_{\theta^a}(\nu|\mu) \leq \overline{\mathcal{T}}_{\theta_1^a}(\nu|\mu) + \overline{\mathcal{T}}_{\theta_2^a}(\nu|\mu), \qquad \forall a > 0.$$

On choisit pour fonction θ_1 la partie quadratique de θ au voisinage de 0,

$$\theta_1(t) = t^2, \quad \forall |t| \leq t_0,$$

puis linéaire

$$\theta_1(t) = 2|t|t_0 - t_0^2, \quad \forall |t| \geq t_0.$$

Puisque la condition (III.2) est, elle aussi, satisfaite, d'après le Théorème III.5.2, il existe des constantes t_0 et a telles que les inégalités barycentriques $\overline{\mathbf{T}}^+_{\theta_1^a}$ et $\overline{\mathbf{T}}^-_{\theta_1^a}$ sont satisfaites.

On pose $\theta_2(t) = [\theta(t) - \theta_1(t)]_+, t \in \mathbb{R}$ qui est nulle au voisinage de 0, sur l'intervalle $[-t_0, t_0]$. La condition (III.1) s'écrit encore

$$\sup_{x \in \mathbb{R}} (U_\mu(x+u) - U_\mu(x)) \leq b\, \theta_2^{-1}(u), \qquad \forall u \geq 0.$$

Dans le Théorème 6.1 [13], nous établissons que cette condition est équivalente à l'existence d'une constante a telle que μ satisfait les inégalités de transport barycentriques $\overline{\mathbf{T}}^+_{\theta_2^a}$ et $\overline{\mathbf{T}}^-_{\theta_2^a}$.

Finalement, en ajustant correctement la constante a, puisque $\theta \leq \theta_1 + \theta_2$, on obtient

$$2\overline{\mathcal{T}}_{\theta^{a/2}}(\nu|\mu) \leq \overline{\mathcal{T}}_{\theta^a}(\nu|\mu) \leq \overline{\mathcal{T}}_{\theta_1^a}(\nu|\mu) + \overline{\mathcal{T}}_{\theta_2^a}(\nu|\mu) \leq 2H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}).$$

Par conséquent, μ satisfait l'inégalité de transport barycentrique $\overline{\mathbf{T}}^+_{\theta^{a/2}}$:

$$\overline{\mathcal{T}}_{\theta^{a/2}}(\nu|\mu) \leq H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}).$$

La preuve de l'inégalité barycentrique $\overline{\mathbf{T}}^-_{\theta^{a/2}}$ est analogue. $\square$

CHAPITRE IV

Inégalités de transport universelles

IV.1 En dimension 1

Soit $(\mathcal{X}, d)$ un espace métrique séparable complet. Dans ce document, les *inégalités de transport universelles* sont des inégalités de transport satisfaites par toute mesure $\mu \in \mathcal{P}(X)$.

La première d'entre elles, la plus connue, évoquée en II.2, est l'inégalité de Csizár-Kullback-Pinsker [Csi67, Kul67, Pin64],

$$\frac{1}{2}\|\mu - \nu\|_{TV}^2 \leq H(\nu|\mu), \qquad \forall \mu, \nu \in \mathcal{P}(\mathcal{X}), \tag{IV.1}$$

où $\|\mu - \nu\|_{TV}$ est la distance en variation totale entre les mesures μ et ν,

$$\|\mu - \nu\|_{TV} = 2 \inf_{\pi \in \Pi(\mu,\nu)} \iint \mathbb{1}_{x \neq y} \, d\pi(x, y).$$

La forme duale de l'inégalité de Csizár-Kullback-Pinsker est l'inégalité exponentielle suivante, pour toute fonction $f : \mathcal{X} \to \mathbb{R}$ telle que $\sup_{x,y \in \mathcal{X}} |f(x) - f(y)| \leq c$,

$$\int e^{tf} d\mu \leq e^{t \int f \, d\mu + t^2 c^2/8}, \qquad t \geq 0.$$

Cette inégalité exponentielle, très utilisée, donne l'inégalité de Hoeffding en appliquant l'inégalité de Markov,

$$\mu\Big(f \geq \int f \, d\mu + t\Big) \leq e^{-2t^2/c^2} \qquad t \geq 0.$$

L'inégalité de Csizár-Kullback-Pinsker (IV.1) peut être améliorée de différentes façons, par exemple en changeant la fonction de la distance en variation totale dans son membre de gauche (voir [FHT03, Gil10]).

Cette partie est plus particulièrement consacrée aux versions dérivées de l'inégalité de Csizár-Kullback-Pinsker qui mettent en jeu les fonctions coûts universelles présentées dans la Définition I.1.5. Plus précisément, étant donnés $\alpha : \mathbb{R}^+ \to [0, +\infty]$ une fonction convexe, et μ, ν_1, ν_2 des probabilités sur $\mathcal{X}$, notons

$$\widetilde{\mathcal{T}}_\alpha(\nu_1|\nu_2) = \inf_{\pi \in \Pi(\nu_2,\nu_1)} \left\{ \int c(x, p_x) d\nu_2(x), d\pi(x, y) = d\nu_2(x) dp_x(y) \right\},$$

lorsque

$$c(x,p) = \alpha\left(\int \mathbb{1}_{x\neq y}dp(y)\right), \qquad x \in \mathcal{X}, \quad p \in \mathcal{P}(\mathcal{X}),$$

et

$$\widehat{\mathcal{T}}_\alpha(\nu_1|\nu_2) = \inf_{\pi\in\Pi(\nu_2,\nu_1)} \left\{\int c(x,p_x)d\nu_2(x), d\pi(x,y) = d\nu_2(x)dp_x(y)\right\},$$

lorsque

$$c(x,p) = \int \alpha\left(\mathbb{1}_{x\neq y}\frac{dp}{d\mu}(y)\right)d\mu(y),$$

si (x,p) est tel que p est absolument continue par rapport à μ sur $\mathcal{X}\setminus\{x\}$, et $c(x,p) = +\infty$ sinon. Dans [3] (Théorème 1.1 et Théorème 1.2), sont établies les inégalités de transport universelles suivantes (voir aussi [13]).

Théorème IV.1.1. *[3] Soit $\mathcal{X}$ un espace métrique compact, $\mu \in \mathcal{P}(\mathcal{X})$ et $t \in]0,1[$.*

(a) Pour toutes probabilités ν_1 et ν_2 sur $\mathcal{X}$, on a

$$\widetilde{\mathcal{T}}_{\alpha_t}(\nu_1|\nu_2) \leq \frac{1}{1-t}H(\nu_1|\mu) + \frac{1}{t}H(\nu_2|\mu),$$

où la fonction de coût convexe $\alpha_t : \mathbb{R}^+ \to [0,+\infty]$ est définie par

$$\alpha_t(u) = \frac{t(1-u)\log(1-u) - (1-tu)\log(1-tu)}{t(1-t)}, \qquad 0 \leq u \leq 1.$$

et $\alpha_t(u) = +\infty$ si $u > 1$.
Par continuité, pour $t = 0$, on obtient

$$\widetilde{\mathcal{T}}_{\alpha_0}(\nu_1|\mu) \leq H(\nu_1|\mu),$$

avec $\alpha_0(u) = (1-u)\log(1-u) + u$ si $0 \leq u \leq 1$, et $\alpha_0(u) = +\infty$ si $u > 1$.
Par continuité, pour $t = 1$, on obtient

$$\widetilde{\mathcal{T}}_{\alpha_1}(\mu|\nu_2) \leq H(\nu_2|\mu),$$

avec $\alpha_1(u) = -\log(1-u) - u$, si $0 \leq u < 1$ et $\alpha_0(u) = +\infty$ si $u \geq 1$.

(b) Pour toutes probabilités ν_1 et ν_2 sur $\mathcal{X}$, on a

$$\widehat{\mathcal{T}}_{\beta_t}(\nu_1|\nu_2) \leq \frac{1}{1-t}H(\nu_1|\mu) + \frac{1}{t}H(\nu_2|\mu),$$

où la fonction de coût convexe $\beta_t : \mathbb{R}^+ \to [0,+\infty]$ est définie par

$$\beta_t(u) := \sup_{s\in\mathbb{R}^+}\{su - \beta_t^*(s)\}, \qquad u \in \mathbb{R}^+.$$

avec

$$\beta_t^*(s) = \frac{te^{(1-t)s} + (1-t)e^{-ts} - 1}{t(1-t)}, \qquad s \in \mathbb{R}^+.$$

Par continuité, pour $t = 0$, on obtient

$$\widehat{\mathcal{T}}_{\beta_0}(\nu_1|\mu) \leq H(\nu_1|\mu), \tag{IV.2}$$

avec $\beta_0(u) = (1+u)\log(1+u) - u$, $u \geq 0$, et pour $t = 1$,

$$\widehat{\mathcal{T}}_{\beta_1}(\mu|\nu_2) \leq H(\nu_2|\mu),$$

avec $\beta_1(u) = (1-u)\log(1-u) + u$, si $u \leq 1$ et $\beta_1(u) = +\infty$ si $u > 1$.

Puisque pour tout $t \in [0,1]$, $\alpha_t(u) \geq u^2/2 = \alpha(u)$, $u \geq 0$, les inégalités de transport du point (a) redonnent les inégalités de Marton [Mar96b] associées à la fonction de coût α :

$$\frac{1}{2}\widetilde{\mathcal{T}}_2(\nu_1|\mu) \leq H(\nu_1|\mu), \qquad \frac{1}{2}\widetilde{\mathcal{T}}_2(\mu|\nu_2) \leq H(\nu_2|\mu), \tag{IV.3}$$

ou encore pour tout $t \in (0,1)$,

$$\frac{1}{2}\widetilde{\mathcal{T}}_2(\nu_1|\nu_2) \leq \frac{1}{1-t}H(\nu_1|\mu) + \frac{1}{t}H(\nu_2|\mu).$$

En optimisant en t, cette inégalité s'écrit encore

$$\frac{1}{2}\widetilde{\mathcal{T}}_2(\nu_1|\nu_2) \leq \left(\sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)}\right)^2,$$

qui peut être améliorée en

$$\frac{1}{2}\left(\widetilde{\mathcal{T}}_2(\nu_1|\nu_2) + \widetilde{\mathcal{T}}_2(\nu_2|\nu_1)\right) \leq \left(\sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)}\right)^2, \qquad \forall \mu, \nu_1\nu_2 \in \mathcal{P}(\mathcal{X}), \tag{IV.4}$$

comme conséquence de la convexité de l'entropie (VI.4) présentée en section VI.2.

Comme il a été expliqué en section II.4, ces inégalités de transport faibles se tensorisent sur les espaces produits, fournissant ainsi des propriétés de concentration pour les mesures produits, voire des mesures faiblement dépendantes. Les principes de concentration associés à ces coûts de transport en dimension supérieure ont été introduits par Talagrand [Tal96a, Tal96c], en particulier pour établir des inégalités de déviation pour les suprema de moyennes empiriques de type Bernstein.

Dans la section IV.2 suivante, sont brièvement présentés les résultats de l'article [1] obtenus dans le cadre de dépendance faible. Ensuite, la section IV.3 concerne les résultats de l'article [3] à propos des inégalités de déviations pour des suprema de processus empiriques pour des variables aléatoires indépendantes. Le dernier paragraphe de ce chapitre, section IV.4, propose une autre forme de tensorisation des inégalités de Csizár-Kullback-Pinsker ou des inégalités de transport de Marton qui fournit de nouvelles inégalités de transport faibles pour la loi uniforme sur le groupe symétrique. Ces résultats récents du preprint [14] s'inspirent des travaux de Talagrand [Tal95].

IV.2 Inégalités de transport pour des variables aléatoires faiblement dépendantes

Dans l'article [1], est proposé un schéma de tensorisation des inégalités de Marton (IV.3) lorsque μ est une probabilité sur $\mathcal{X}^n$, dont les marginales dépendent faiblement les unes des autres ; en particulier lorsque μ est la loi des n premières variables aléatoires $X_1, \ldots, X_n$ d'un processus Φ-mélangeant. Les chaînes de Markov Doeblin récurrentes sont des cas particuliers de processus Φ-mélangeants.

Cette tensorisation de l'inégalité de transport repose sur la construction de couplages. Une construction similaire a été proposée ensuite par Marton [Mar03] afin de permettre d'autres applications de ces inégalités de transport.

Le résultat principal de [1] est le suivant, en supposant pour simplifier que $\mathcal{X}$ est fini. Etant donnée une suite de variables aléatoires $X_1, \ldots, X_n$ à valeurs dans $\mathcal{X}$, pour $1 \leq i < j \leq n$, notons

$$\mathcal{L}(X_j^n | X_1^{i-1} = x_1^{i-1}, X_i = x_i)$$

la loi de $(X_j, \ldots, X_n)$ sachant que $X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_i = x_i$, et $\Gamma = (\gamma_{i,j})_{1 \leq i,j \leq n}$ la matrice triangulaire supérieure définie par

$$\gamma_{ij}^2 = \sup_{x_1^{i-1}, x_i, y_i} \left\| \mathcal{L}(X_j^n | X_1^{i-1} = x_1^{i-1}, X_i = x_i) - \mathcal{L}(X_j^n | X_1^{i-1} = x_1^{i-1}, X_i = y_i) \right\|_{TV},$$

pour $i < j$ et $\gamma_{ii} = 1$.

Théorème IV.2.1. *[1] Avec les notations précédentes, pour toutes probabilités μ et ν sur $\mathcal{X}^n$, on a*

$$\widetilde{\mathcal{T}}_2(\nu|\mu) \leq 2\|\Gamma\|^2 H(\nu|\mu) \quad et \quad \widetilde{\mathcal{T}}_2(\mu|\nu) \leq 2\|\Gamma\|^2 H(\nu|\mu),$$

où $\|\Gamma\|$ désigne la norme d'opérateur de Γ de $(\mathbb{R}^n, |\cdot|)$ dans $(\mathbb{R}^n, |\cdot|)$.

Puisque le coût de Marton $\widetilde{\mathcal{T}}_2$ en dimension n satisfait l'inégalité triangulaire (II.7), les deux inégalités de transport de ce théorème sont équivalentes à la famille d'inégalités de transport suivante : pour tout $t \in]0, 1[$, pour tous $\mu, \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}^n)$,

$$\frac{1}{2\|\Gamma\|^2} \widetilde{\mathcal{T}}_2(\nu_1|\nu_2) \leq \frac{1}{1-t} H(\nu_1|\mu) + \frac{1}{t} H(\nu_2|\mu),$$

ou encore, de façon équivalente, d'après le Théorème II.3.1, pour toute fonction $g : \mathcal{X}^n \to \mathbb{R}$ bornée inférieurement,

$$\left(\int e^{t\widetilde{Q}g} d\mu \right)^{1/t} \left(\int e^{-(1-t)g} d\mu \right)^{1/(1-t)} \leq 1,$$

où

$$\widetilde{Q}g(x) = \inf_{p \in \mathcal{P}(\mathcal{X}^n)} \left\{ \int g dp + \frac{1}{2\|\Gamma\|^2} \sum_{i=1}^n \left(\int \mathbb{1}_{x_i \neq y_i} dp(y) \right)^2 \right\}.$$

En particulier, en appliquant cette inégalité à la fonction $g = i_A$, $A \subset \mathcal{X}^n$, on obtient la propriété de concentration de Talagrand ([Tal95], “convex hull method”) étendue à n'importe quelle mesure $\mu \in \mathcal{P}(\mathcal{X}^n)$: pour tout sous-ensemble $A \subset X^n$, pour tout $t \in]0,1[$,

$$\int e^{\frac{t}{2\|\Gamma\|^2} D_T^2(x,A)} d\mu(x) \leq \frac{1}{\mu(A)^{t/(1-t)}}, \tag{IV.5}$$

où $D_T(x, A)$ est la distance convexe de Talagrand définie par

$$\begin{aligned} D_T^2(x,A) &= \left(\sup_{\alpha \in \mathbb{R}^n, |\alpha| \leq 1} \inf_{y \in A} \sum_{i=1}^n \alpha_i \mathbb{1}_{x_i \neq y_i} \right)^2 \\ &= \left(\sup_{\alpha \in \mathbb{R}^n, |\alpha| \leq 1} \inf_{p \in \mathcal{P}(A)} \sum_{i=1}^n \alpha_i \int \mathbb{1}_{x_i \neq y_i} dp(y) \right)^2 \\ &= \left(\inf_{p \in \mathcal{P}(A)} \sup_{\alpha \in \mathbb{R}^n, |\alpha| \leq 1} \sum_{i=1}^n \alpha_i \int \mathbb{1}_{x_i \neq y_i} dp(y) \right)^2 \\ &= \inf_{p \in \mathcal{P}(A)} \sum_{i=1}^n \left(\int \mathbb{1}_{x_i \neq y_i} dp(y) \right)^2 = \widetilde{Q} i_A(x). \end{aligned}$$

La deuxième égalité découle de la linéarité de l'expression en p et du fait que les mesures de Dirac sont les points extrémaux de l'ensemble convexe $\mathcal{P}(A)$. La troisième égalité est une conséquence du Théorème minimax de Sion [Sio58, Kom88] qui s'applique ici car l'expression est lineaire en p et α, donc convexe en p et concave en α.

Lorsque μ est la loi d'une suite de variables aléatoires indépendantes, l'inégalité (IV.5) correspond exactement au résultat de concentration de Talagrand puisque $\|\Gamma\| = \|\mathrm{Id}\| = 1$.

Le Théorème IV.2.1 complète celui de Marton pour les chaînes de Markov contractantes [Mar96b, Mar97]. Supposons plus généralement que la suite $(X_k)_{k\geq 1}$ est une chaîne de Markov de noyau K, récurrente au sens de Doeblin ; c'est à dire qu'il existe une probabilité m, un entier r et un réel $\rho \in]0,1[$ tels que pour tout $x \in \mathcal{X}$ et tout sous-ensemble $A \subset \mathcal{X}$

$$K^r(x, A) \geq \rho m(A).$$

Alors le coefficient $\|\Gamma\|$ est borné indépendamment de n,

$$\|\Gamma\| \leq \frac{\sqrt{2}}{1 - \rho^{1/2r}}.$$

Plus généralement, si $(\Phi_k)_{k\geq 1}$ représente la suite des coefficients de Φ-mélange d'une suite de variables aléatoires $X_1, \ldots, X_n$ (voir [1], [Dou94]) alors, on a

$$\|\Gamma\| \leq \sum_{k=1}^n \sqrt{\Phi_k}.$$

Depuis ce résultat, différents auteurs ont montré des inégalités de transport sous des hypothèses de dépendance faible différentes, par exemple sous des conditions de dépendance de type Dobrushin [DGW04, Mar04, Mar10, Wu06, Kon12, Pau12, WW14, Wan14, Pau14].

Parmi ces travaux, le résultat suivant de Paulin [Pau14] concerne exactement la propriété de concentration de Talagrand (IV.5) obtenue sous des conditions de Dobrushin, en appliquant les méthodes de symétrisation de Stein, mises en avant par Chatterjee [Cha07]. Etant donné $1 \leq i \leq n$, on note X_{-i} le vecteur aléatoire défini par

$$X_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n).$$

La matrice $D = (d_{ij})$ des coefficients de dépendance de Dobrushin, est une matrice dont les entrées positives vérifient, pour tout $i \in \{1, \ldots, n\}$, et tous $x, y \in \mathcal{X}^n$,

$$\|\mathcal{L}(X_i|X_{-i} = x_{-i}) - \mathcal{L}(X_i|X_{-i} = y_{-i})\|_{TV} \leq \sum_{j, j\neq i} d_{ij} \mathbb{1}_{x_j \neq y_j}.$$

Théorème IV.2.2. *(Théorème 3.3,[Pau14]) Soit* $\|D\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^n d_{ij}$ *et* $\|D\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^n d_{ij}$. *Si* $\|D\|_1 < 1$ *and* $\|D\|_\infty \leq 1$, *alors pour tout* $A \subset \mathcal{X}^n$, *on a*

$$\int e^{\frac{1-\|D\|_1}{26.1} D_T^2(x,A)} d\mu(x) \leq \frac{1}{\mu(A)}.$$

Des exemples d'applications de ce résultat sont présentés dans [Pau14] (le problème du voyageur de commerce, les arbres de Steiner).

IV.3 Inégalités de type Bernstein pour des suprema de processus empiriques indépendants

Les premières inégalités de déviation des suprema de processus empiriques de type Bernstein ont été démontrées par Talagrand [Tal96b, Tal96a] par la méthode dite "de l'enveloppe convexe". Ces inégalités de concentration sont particulièrement utiles en statistique [Mas00b, Mas07].

Ledoux [Led97] a proposé une "méthode entropique" qui permet de retrouver simplement les résultats de Talagrand. Cette méthode est basée sur la propriété de tensorisation des inégalités de Sobolev et l'argument de Herbst. Elle a été largement développée ensuite notamment pour obtenir les bornes de déviation optimales pour les suprema de processus empiriques indépendants [Mas00a, BLM00, Rio01, BLM03, Rio02, Bou03, KR05, Rio12, Rio13, BLM13].

L'objectif de l'article [3] est de développer une approche alternative, par inégalités de transport, afin d'obtenir les constantes optimales dans les inégalités de déviation des suprema de processus empiriques indépendants, dans le prolongement des travaux de Talagrand, Marton [Mar96a, Mar96b], Dembo [Dem97] et Maurey [Mau91].

Une autre approche a été proposée par Panchenko, basée sur des méthodes de symétrisation [Pan01, Pan02, Pan03]. Enfin, pour compléter ce tableau, les méthodes de Stein ont été mises en avant par Chatterjee pour obtenir des propriétés de concentration semblables à celle de Talagrand [Cha05, Cha07, CD10, Pau14]. Cette dernière méthode présente l'intérêt de s'étendre à des cas de dépendance, sous des conditions de type Dobrushin.

Voici l'un des résultats de concentration pour les suprema de processus empiriques indépendants qui se déduit des inégalités de transport du Théorème IV.1.1 après tensorisation (voir [3]). Soit $\mathcal{F}$ un ensemble dénombrable et $(X_{1,t})_{t\in\mathcal{F}}, \dots, (X_{n,t})_{t\in\mathcal{F}}$, n processus indépendants. On s'intéresse aux déviations de la variable aléatoire

$$Z = \sup_{t\in\mathcal{F}} \sum_{i=1}^{n} X_{i,t}.$$

Notons

$$V = \sup_{t\in\mathcal{F}} \sum_{i=1}^{n} \mathbb{E}\left[[X_{i,t} - X'_{i,t}]_+^2 \middle| X_{i,t} \right],$$

où $(X'_{i,t})_{t\in\mathcal{F}}$ est une copie indépendante de $X_i = (X_{i,t})_{t\in\mathcal{F}}$ et $\mathbb{E}[\,\cdot\,|X_{i,t}]$ est l'espérance conditionnelle étant donné $X_{i,t}$. Dans le théorème suivant, pour tout $t \in \mathcal{F}$ et tout $i \in \{1, \dots, n\}$, $M_{i,t}$ et $m_{i,t}$ sont des constantes qui encadrent les variables aléatoires $X_{i,t}$.

Théorème IV.3.1. *(a) Supposons que $X_{i,t} \leq M_{i,t}$, et $\mathbb{E}\left[(M_{i,t} - X_{i,t})^2\right] \leq 1$, pour tout i et tout t, alors pour tout $u \geq 0$,*

$$\begin{aligned} \mathbb{P}(Z \geq \mathbb{E}[Z] + u) &\leq \exp\left[-\frac{u}{2\left(1 + \varepsilon\left(\frac{u}{\mathbb{E}[V]}\right)\right)} \log\left(1 + \frac{u}{\mathbb{E}[V]}\right)\right] \\ &\leq \exp\left[-\frac{u^2}{2\mathbb{E}[V] + 2u} \right], \end{aligned}$$

avec $\varepsilon(u) = \dfrac{\beta_0(u)}{(1+u)\log(1+u)}$ et $\beta_0(u) := (1+u)\log(1+u) - u$.

(b) Supposons que $m_{i,t} \leq X_{i,t} \leq M_{i,t}$, avec $M_{i,t} - m_{i,t} = 1$ pour tout i et tout t, alors pour tout $u \geq 0$,

$$\mathbb{P}(Z \leq \mathbb{E}[Z] - u) \leq \exp\left[-\mathbb{E}[V]\beta_0\left(\frac{u}{\mathbb{E}[V]}\right)\right] \leq \exp\left[-\frac{u^2}{2\mathbb{E}[V] + \frac{2}{3}u}\right],$$

avec $\beta_0(u) = (1+u)\log(1+u) - u$.

L'optimalité de ces résultats est expliquée dans l'article [3]. Rappelons que par des techniques de symétrisation usuelles ([LT91], Lemme 6.3 et Théorème 4.12), le terme de variance $\mathbb{E}[V]$ s'estime de la manière suivante

$$\mathbb{E}[V] \leq \sup_{t\in\mathcal{F}} \sum_{i=1}^{n} \mathrm{Var}(X_{i,t}) + 16\,\mathbb{E}\left[\sup_{t\in\mathcal{F}} \left| \sum_{i=1}^{n} (X_{i,t} - \mathbb{E}[X_{i,t}]) \right| \right].$$

Dans [3], certains résultats intermédiaires sur les moments exponentiels ne nécessitent pas que les variables aléatoires soient bornées. Pour plus de détails, voir par exemple le Théorème 3.4 et le Corollaire 3.5 de [3]. Dans [Ada08], en utilisant l'inégalité Hoffman-Jørgensen et les résultats de Talagrand, Adamczak a étendu les propriétés de concentration à des suprema de variables aléatoires non-bornées, par des arguments de troncature des variables aléatoires. Ces mêmes arguments pourraient être utilisés avec les résultats intermédiaires donnés dans [3], pour démontrer de nouvelles inégalités de déviation pour des variables non-bornées.

Idée de la preuve. Nous présentons seulement une preuve élémentaire de (b), pour montrer les liens entre l'inégalité de transport avec un coût $\widehat{\mathcal{T}}_{\beta_t}$ et les inégalités de déviation d'une somme empirique autour de sa moyenne.

Soit μ_i la loi du processus $X_i = (X_{i,t})_{t\in\mathcal{F}}$. Le point (b) découle simplement de la forme duale de l'inégalité de transport tensorisée (IV.2) : $\mathcal{X} = \mathbb{R}^{\mathcal{F}}$, pour toute fonction $g : \mathcal{X}^n \to \mathbb{R}$,

$$\int e^{-g} d\mu \leq \exp\Big(-\int \widehat{Q}g\, d\mu\Big), \tag{IV.6}$$

où $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ et pour tout $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$

$$\widehat{Q}g(x) = \inf_{p\in\mathcal{P}(\mathcal{X}^n)} \left\{ \int g(y)\, dp(y) + \sum_{i=1}^n \int \beta_0 \left(\mathbb{1}_{x_i\neq y_i} \frac{dp_i}{d\mu_i}(y_i) \right) d\mu_i(y_i) \right\},$$

les probabilités p_i sont les marginales de p. Choisissons $g = \lambda f$ avec $\lambda \geq 0$ et

$$f(x) = \sup_{t\in\mathcal{F}} \sum_{i=1}^n x_{i,t}, \qquad x = (x_1, \ldots, x_n) \in \mathcal{X}^n.$$

On suppose, pour simplifier, que pour tout $x \in \mathcal{X}^n$, le supremum est atteint en un unique point $\tau(x) \in \mathcal{F}$:

$$\sup_{t\in\mathcal{F}} \sum_{i=1}^n x_{i,t} = \sum_{i=1}^n x_{i,\tau(x)}.$$

Alors pour tout $x, y \in \mathcal{X}^n$

$$f(y) \geq f(x) + \sum_{i=1}^n (y_{i,\tau(x)} - x_{i,\tau(x)}) = f(x) + \sum_{i=1}^n (y_{i,\tau(x)} - x_{i,\tau(x)}) \mathbb{1}_{x_i\neq y_i}.$$

Par suite, pour tout x,

$$\begin{aligned}
&\widehat{Q}g(x)\\
&\geq \lambda f(x) - \sup_p \left\{ \int \sum_{i=1}^n \lambda(x_{i,\tau(x)} - y_{i,\tau(x)}) \mathbb{1}_{x_i \neq y_i} dp(y) \right.\\
&\qquad\qquad\qquad\qquad\qquad \left. - \sum_{i=1}^n \int \beta_0 \left(\mathbb{1}_{x_i \neq y_i} \frac{dp_i}{d\mu_i}(y_i) \right) d\mu_i(y_i) \right\}\\
&= \lambda f(x) - \sum_{i=1}^n \sup_{p_i \in \mathcal{P}(\mathcal{X})} \left\{ \int \lambda(x_{i,\tau(x)} - y_{i,\tau(x)}) \mathbb{1}_{x_i \neq y_i} \frac{dp_i}{d\mu_i}(y_i) d\mu_i(y_i) \right.\\
&\qquad\qquad\qquad\qquad\qquad \left. - \int \beta_0 \left(\mathbb{1}_{x_i \neq y_i} \frac{dp_i}{d\mu_i}(y_i) \right) d\mu_i(y_i) \right\}\\
&\geq \lambda f(x) - \sum_{i=1}^n \int \sup_{h \geq 0} \left\{ \lambda(x_{i,\tau(x)} - y_{i,\tau(x)}) h - \beta_0(h) \right\} d\mu_i(y_i)\\
&= \lambda f(x) - \sum_{i=1}^n \int \beta_0^* \left(\lambda [x_{i,\tau(x)} - y_{i,\tau(x)}]_+ \right) d\mu_i(y_i)\\
&\geq \lambda f(x) - \beta_0^*(\lambda) \sum_{i=1}^n \int [x_{i,\tau(x)} - y_{i,\tau(x)}]_+^2 d\mu_i(y_i)\\
&\geq \lambda f(x) - \beta_0^*(\lambda) \sup_{t \in \mathcal{F}} \sum_{i=1}^n \int [x_{i,t} - y_{i,t}]_+^2 d\mu_i(y_i),
\end{aligned}$$

où $\beta_0^*(s) = e^s - s - 1$, $s \geq 0$. L'avant dernière inégalité est une conséquence du fait que $[x_{i,\tau(x)} - y_{i,\tau(x)}]_+ \leq M_{i,t} - m_{i,t} \leq 1$ et $\beta_0^*(\lambda u) \leq u^2 \beta_0^*(\lambda)$ pour $0 \leq u \leq 1$. En insérant l'estimation précédente de $\widehat{Q}g(x)$ dans l'inégalité de transport (IV.6), on obtient pour tout $\lambda \geq 0$,

$$\mathbb{E}\left[e^{-\lambda Z}\right] \leq e^{-\lambda \mathbb{E}[Z] + \mathbb{E}[V]\beta_0^*(\lambda)}.$$

L'inégalité de déviation de (b) découle immédiatement par l'inégalité de Markov en optimisant sur les valeurs de λ. □

IV.4 Inégalités de transport pour la loi uniforme sur le groupe symétrique.

Dans cette section nous présentons des inégalités de transport pour la loi uniforme sur le groupe symétrique S_n, qui redonnent des résultats de concentration obtenus par Maurey [Mau79] et Talagrand [Tal95].

Soit μ la loi uniforme sur S_n,

$$\mu(\sigma) = \frac{1}{n!}, \qquad \sigma \in S_n.$$

Le résultat de Maurey, démontré par des techniques de martingales est le suivant.

Théorème IV.4.1. *[Mau79] Soit d la distance de Hamming sur le groupe symétrique, pour tout $\sigma, \tau \in S_n$,*

$$d(\sigma, \tau) = \sum_{i=1}^{n} \mathbb{1}_{\sigma(i) \neq \tau(i)}.$$

Alors pour tout sous-ensemble $A \subset S_n$ tel que $\mu(A) \geq 1/2$, et pour tout $t \geq 0$, on a

$$\mu(A_t) \geq 1 - 2e^{-\frac{t^2}{64n}},$$

où $A_t = \{y \in S_n, d(x, A) \leq t\}$.

Ce résultat a été généralisé par V. Milman et Schechtman, à des groupes dont la distance est invariante par translation [MS86]. Talagrand a établi une propriété de concentration plus forte quant à la dépendance en n, obtenue par la méthode de l'enveloppe convexe, dont voici une version légèrement modifiée. Cette propriété implique celle du théorème précédent à des constantes près.

Théorème IV.4.2. *[Tal95] Pour tout sous-ensemble $A \subset S_n$,*

$$\int_{S_n} e^{f(A,\sigma)/16} d\mu(\sigma) \leq \frac{1}{\mu(A)},$$

où la quantité $f(A, \sigma)$ évalue la distance de σ à A de la façon suivante

$$f(A, \sigma) = \inf_{p \in \mathcal{P}(A)} \sum_{i=1}^{n} \left(\int \mathbb{1}_{\sigma(i) \neq \tau(i)} dp(\tau) \right)^2.$$

Ce théorème a été généralisé à des produits de groupes de permutations par McDiarmid [McD02], puis par Luczak et McDiarmid [LM03] à des groupes de permutations plus généraux agissant localement.

Les Théorèmes IV.4.1 et IV.4.2 sont conséquences des inégalités de transport suivantes.

Théorème IV.4.3. *[14] Soit μ la loi uniforme sur le groupe symétrique.*

(a) Pour toutes probabilités ν_1 et ν_2 sur S_n,

$$\frac{1}{2(n-1)} W_1^2(\nu_2, \nu_1) \leq \left(\sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)} \right)^2,$$

où $W_1(\nu_1, \nu_2) = \inf_{\pi \in \Pi(\nu_2, \nu_1)} \iint d(\sigma, \tau) d\pi(\sigma, \tau)$.

(b) Pour toutes probabilités ν_1 et ν_2 sur S_n,

$$\frac{1}{8} \widehat{\mathcal{T}}_2(\nu_1|\nu_2) \leq \left(\sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)} \right)^2,$$

où

$$\widehat{\mathcal{T}}_2(\nu_1|\nu_2) = \inf_{\pi \in \Pi(\nu_2, \nu_1)} \int \sum_{i=1}^{n} \left(\int \mathbb{1}_{\sigma(i) \neq \tau(i)} dp_\sigma(\tau) \right)^2 d\nu_2(\sigma),$$

avec $\pi(\sigma, \tau) = \nu_2(\sigma) p_\sigma(\tau)$.

La preuve de (b), inspirée de celle du théorème de Talagrand, est donnée dans l'article [14] en préparation. Nous présentons à la fin de ce chapitre une preuve similaire plus simple de (a). En fait, l'inégalité de transport de (a) est plus connue sous sa forme duale fonctionnelle : pour toute fonction $f : S_n \to \mathbb{R}$, 1-lipschitzienne pour la distance de Hamming d sur le groupe symétrique,

$$\int e^f d\mu \leq e^{\int f\, d\mu + (n-1)t^2/2}, \qquad \forall t \geq 0.$$

Cette inégalité exponentielle découle des inégalités de Hoeffding pour des martingales bornées. Elle est largement commentée et redémontrée dans l'article [BHT06].

Voici un exemple utile de propriété de concentration issue de l'inégalité de transport faible du point (b).

Théorème IV.4.4. *[14] Soit μ la loi uniforme sur S_n. Soit $g : S_n \to \mathbb{R}$ et $\alpha_k : S_n \to \mathbb{R}^+$, $k \in \{1, \dots, n\}$ des fonctions telles que pour tout $\tau, \sigma \in S_n$,*

$$g(\tau) - g(\sigma) \leq \sum_{k=1}^n \alpha_k(\tau) \mathbb{1}_{\tau(k) \neq \sigma(k)}.$$

Alors, pour tout $t \geq 0$, on a

$$\mu\Big(g \leq \int g\, d\mu - t\Big) \leq \exp\Big(-\frac{t^2}{8\int |\alpha|^2 d\mu}\Big),$$

et

$$\mu\Big(g \geq \int g\, d\mu + t\Big) \leq \exp\Big(-\frac{t^2}{8\, \sup_{\sigma \in S^n} |\alpha(\sigma)|^2}\Big),$$

avec $|\alpha(\sigma)|^2 = \sum_{k=1}^n \alpha_k^2(\sigma)$, $\sigma \in S_n$.

En appliquant ce résultat à la fonction particulière $g(\sigma) = \varphi(x_\sigma)$ où $\varphi : [0,1]^n \to \mathbb{R}$ est lipschitzienne convexe et étant donné $(x_1, \dots, x_n) \in [0,1]^n$, $x_\sigma = (x_{\sigma(1)}, \dots, x_{\sigma(n)})$, on retrouve l'inégalité de déviation de Adamczak, Chafaï et Wolff [ACW14] (Théorème 3.1) issue du Théorème IV.4.2 de Talagrand. Cette propriété de concentration jouent un rôle essentiel dans leur approche pour étudier la convergence de la mesure spectrale empirique des matrices aléatoires à entrées échangeables, lorsque la taille des matrices tend vers l'infini.

Une preuve par transport du point (a) du Théorème IV.4.3. L'objet de cette démonstration est de donner l'idée de la preuve du point (b), plus complexe (voir [14]). Comme la distance W_1 satisfait l'inégalité triangulaire, il suffit de montrer que pour toute probabilité ν_1 sur S_n,

$$\frac{1}{2(n-1)} W_1^2(\nu_1, \mu) \leq H(\nu_1|\mu).$$

D'après la Proposition II.3.1, la formulation duale de cette inégalité de transport est la suivante, pour toute fonction φ sur S_n et tout $\lambda \geq 0$,

$$\int e^{\lambda Q\varphi} d\mu \leq e^{\int \lambda\varphi\, d\mu + (n-1)\lambda^2/2}, \tag{IV.7}$$

avec

$$
\begin{aligned}
Q\varphi(\sigma) &= \inf_{p\in\mathcal{P}(S_n)} \left\{\int \varphi dp + \int d(\sigma,\tau)\, dp(\tau)\right\} \\
&= \inf_{p\in\mathcal{P}(S_n)} \left\{\int \varphi dp + \sum_{k=1}^{n} \int \mathbb{1}_{\sigma(k)\neq\tau(k)}\, dp(\tau)\right\}.
\end{aligned}
$$

Nous allons démontrer l'inégalité (IV.7) par récurrence sur l'entier n.

Lorsque $n = 2$, S_n est l'espace à deux points et

$$Q\varphi(\sigma) = \inf_{p\in\mathcal{P}(S_n)} \left\{\int \varphi dp + 2\int \mathbb{1}_{\sigma\neq\tau}\, dp(\tau)\right\}.$$

L'inégalité (IV.7) correspond alors exactement à la forme duale de l'inégalité de Csizár-Kullback-Pinsker donnée par la Proposition II.3.1 : pour toute probabilité ν sur un espace métrique séparable $\mathcal{X}$, pour toute fonction mesurable $f : \mathcal{X} \to \mathbb{R}$,

$$\int e^{\lambda Rf} d\nu \leq e^{\lambda \int f\, d\nu + \lambda^2/2}, \qquad \forall\lambda \geq 0, \tag{IV.8}$$

avec

$$Rf(x) = \inf_{p\in\mathcal{P}(\mathcal{X})} \left\{\int f dp + 2\int \mathbb{1}_{x\neq y} dp(y)\right\}, \qquad x \in \mathcal{X}.$$

Le principe de récurrence repose lui aussi sur cette forme duale de l'inégalité de Csizár-Kullback-Pinsker. Soit $(H_i)_{1\leq i\leq n}$ la partition de S_n définie par,

$$H_i = \{\sigma \in S_n, \sigma(i) = n\}$$

Si p est une probabilité sur S_n, elle admet une unique décomposition définie par

$$p = \sum_{i=1}^{n} \hat{p}(i) p_i, \qquad \text{avec} \quad p_i \in \mathcal{P}(H_i) \qquad \text{et} \quad \hat{p}(i) = p(H_i).$$

On a défini ainsi une probabilité $\hat{p}$ sur $\{1, \ldots, n\}$. En particulier, pour la loi uniforme μ sur S_n, on a

$$\mu = \frac{1}{n} \sum_{i=1}^{n} \mu_i,$$

où μ_i est la loi uniforme sur H_i, $\mu_i(\sigma) = \frac{1}{(n-1)!}$ pour tout $\sigma \in H_i$. On a donc

$$\int e^{\lambda Q\varphi} d\mu = \frac{1}{n} \sum_{i=1}^{n} \int e^{\lambda Q\varphi(\sigma)} d\mu_i(\sigma).$$

Pour toute fonction $f : H_i \to \mathbb{R}$, notons

$$Q_{H_i} f(\sigma) = \inf_{p\in\mathcal{P}(H_i)} \left\{\int f\, dp + \sum_{k\neq i} \int \mathbb{1}_{\sigma(k)\neq\tau(k)} dp(\tau)\right\}.$$

Soit τ_{ij} la transposition qui échange i en j, $i, j \in \{1, \ldots, m\}$. L'application de H_i dans H_n définie par $\tau \mapsto \tau\tau_{in}$ est bijective, donc par changement d'indice dans la somme,

$$\begin{aligned} Q_{H_i} f(\sigma) &= \inf_{p \in \mathcal{P}(H_i)} \left\{ \int f(\tau)\, dp(\tau) + \sum_{k \neq n} \int \mathbb{1}_{\sigma\tau_{in}(k) \neq \tau\tau_{in}(k)} dp(\tau) \right\} \\ &= \inf_{q \in \mathcal{P}(H_n)} \left\{ \int f(\tau\tau_{in})\, dq(\tau) + \sum_{k \neq n} \int \mathbb{1}_{\sigma\tau_{in}(k) \neq \tau(k)} dq(\tau) \right\} \\ &= Q_{H_n} f^{\tau_{in}}(\sigma\tau_{in}). \end{aligned}$$

où $f^{\tau_{in}}(\tau) = f(\tau\tau_{in})$ pour tout $\tau \in H_n$. Par suite, d'après l'hypothèse de récurrence, pour toute fonction $f : H_i \to \mathbb{R}$, pour tout $\lambda \geq 0$,

$$\begin{aligned} \int e^{\lambda Q_{H_i} f} d\mu_i &= \int e^{\lambda Q_{H_n} f^{\tau_{in}}(\sigma\tau_{in})} d\mu_i(\sigma) = \int e^{\lambda Q_{H_n} f^{\tau_{in}}} d\mu_n \\ &\leq \exp\left[\lambda \int f^{\tau_{in}} d\mu_n + (n-2)\frac{\lambda^2}{2}\right] = \exp\left[\lambda \int f d\mu_i + (n-2)\frac{\lambda^2}{2}\right]. \end{aligned}$$

La preuve s'appuie ensuite sur le lemme suivant.

Lemme IV.4.1. *Pour toute fonction $\varphi : H_i \to \mathbb{R}$ et pour tout $\sigma \in H_i$, on a*

$$Q\varphi(\sigma) \leq \inf_{\hat{p} \in \mathcal{P}(\{1,\ldots,n\})} \left\{ \sum_{l=1}^n Q_{H_i} \varphi^{\tau_{il}}(\sigma)\hat{p}(l) + 2\sum_{l=1}^n \mathbb{1}_{l \neq i}\hat{p}(l) \right\}.$$

La démonstration de ce lemme se fait en décomposant sur les ensembles H_j, les probabilités p qui interviennent dans la définition de $Q\varphi(\sigma)$, pour établir que si $\sigma \in H_i$ alors

$$\begin{aligned} Q\varphi(\sigma) = &\inf_{\hat{p} \in \mathcal{P}(\{1,\ldots,n\})} \inf_{p_1 \in \mathcal{P}(H_1), \ldots, p_n \in \mathcal{P}(H_n)} \\ &\left\{ \sum_{l=1}^n \left[\int \varphi\, dp_l + \sum_{k \notin \{l,i\}} \int \mathbb{1}_{\sigma(k) \neq \tau\tau_{il}(k)} dp_l(\tau) \right] \hat{p}(l) + 2\sum_{l=1}^n \mathbb{1}_{l \neq i}\hat{p}(l) \right\}. \end{aligned}$$

On poursuit la preuve de (a) en appliquant ce lemme, l'inégalité de Hölder, puis l'hypothèse de récurrence, ce qui donne

$$\begin{aligned} &\int e^{\lambda Q\varphi(\sigma)} d\mu_i(\sigma) \leq \inf_{\hat{p} \in \mathcal{P}(\{1,\ldots,n\})} \left\{ \prod_{l=1}^n \left(\int e^{\lambda Q_{H_i} \varphi^{\tau_{il}}} d\mu_i \right)^{\hat{p}(l)} e^{2\lambda \sum_{l=1}^n \mathbb{1}_{l \neq i}\hat{p}(l)} \right\} \\ &\leq \exp\left[\inf_{\hat{p} \in \mathcal{P}(\{1,\ldots,n\})} \left\{ \lambda \sum_{l=1}^n \left(\int \varphi^{\tau_{il}} d\mu_i \right) \hat{p}(l) + (n-2)\frac{\lambda^2}{2} + 2\lambda \sum_{l=1}^n \mathbb{1}_{l \neq i}\hat{p}(l) \right\} \right] \\ &= \exp\left[\lambda \inf_{\hat{p} \in \mathcal{P}(\{1,\ldots,n\})} \left\{ \sum_{l=1}^n \hat{\varphi}(l)\hat{p}(l) + 2\sum_{l=1}^n \mathbb{1}_{l \neq i}\hat{p}(l) \right\} + (n-2)\frac{\lambda^2}{2} \right], \end{aligned}$$

où $\hat{\varphi}(l) = \int \varphi^{\tau_{il}} d\mu_i = \int \varphi d\mu_l$. Considérons sur l'espace $\mathcal{X} = \{1, \ldots, n\}$, l'opérateur d'infimum-convolution $R\hat{\varphi}$ défini précédemment,

$$R\hat{\varphi}(i) = \inf_{\hat{p} \in \mathcal{P}(\{1,\ldots,n\})} \left\{ \sum_{l=1}^n \hat{\varphi}(l)\hat{p}(l) + 2\sum_{l=1}^n \mathbb{1}_{l \neq i}\hat{p}(l) \right\}.$$

Par conséquent, en appliquant l'inégalité de (IV.8) avec la loi uniforme ν sur $\{1,\ldots,n\}$, le résultat précédent donne

$$\begin{aligned}\int e^{\lambda Q\varphi}d\mu &= \frac{1}{n}\sum_{i=1}^{n}\int e^{\lambda Q\varphi(\sigma)}d\mu_i(\sigma)\\ &\leq \left(\frac{1}{n}\sum_{i=1}^{n}e^{\lambda R\hat{\varphi}(i)}\right)e^{(n-2)\lambda^2/2}\\ &\leq \exp\left[\frac{\lambda}{n}\sum_{i=1}^{n}\hat{\varphi}(i)+\frac{\lambda^2}{2}+(n-2)\frac{\lambda^2}{2}\right]\\ &= \exp\left[\lambda\int\varphi\,d\mu+(n-1)\frac{\lambda^2}{2}\right].\end{aligned}$$

$\square$

CHAPITRE V

Inégalité de Poincaré et concentration adimensionnelle

Dans le chapitre II, nous avons vu que les inégalités de transport $\mathbf{T}_c(a_1, a_2)$ sont équivalentes à des propriétés de concentration adimensionnelles (Proposition II.4.2). Dans ce chapitre, nous nous intéressons plus particulièrement à l'inégalité de transport dont le coût "usuel" sur un espace polonais $(\mathcal{X}, d)$ est donné par

$$c(x,p) = \int \theta^a(d(x,y))p(dy),$$

où $\theta^a(h) = \theta(ah), h \geq 0$, $a > 0$ et θ est la fonction de coût quadratique puis linéaire définie par

$$\theta(h) = \begin{cases} h^2 & \text{si } h \in [0,1], \\ 2h-1 & \text{si } h \geq 1. \end{cases}$$

Plus précisément, $\mu \in \mathcal{P}(\mathcal{X})$ satisfait *l'inégalité de transport notée* $\mathbf{T}_{\theta^a}$ si

$$\mathbf{T}_{\theta^a} : \qquad \mathcal{T}_{\theta^a}(\mu, \nu) \leq H(\nu|\mu), \qquad \forall \nu \in \mathcal{X}$$

avec

$$\mathcal{T}_{\theta^a}(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \iint \theta^a(d(x,y))\, d\pi(x,y).$$

Puisque $\sup_{h>0} \frac{h\theta'(h)}{\theta(h)} = 2 = p_\theta$, nous avons vu au chapitre II que cette inégalité de transport est équivalente à la famille d'inégalités de transport donnée par (II.6) : pour tout $t \in]0,1[$,

$$\mathcal{T}_{\theta^a}(\nu_1,\nu_2) \leq \frac{H(\nu_1|\mu)}{t} + \frac{H(\nu_2|\mu)}{(1-t)}, \qquad \forall \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}).$$

Gozlan a montré que $\mathbf{T}_{\theta^a}$ est équivalente à la propriété de concentration adimensionnelle suivante [Goz09, Goz10] : il existe une constante b telle que pour tout entier n et tout borélien $A \in \mathcal{B}(\mathcal{X}^n)$ de mesure $\mu^n(A) \geq 1/2$,

$$\mu^n(A_{r,c^n}) \geq 1 - be^{-r}, \tag{V.1}$$

où l'élargissement de A est donné par

$$A_{r,c^n} = \{x \in \mathcal{X}^n, c^n(x,A) \leq r\}, \qquad c^n(x,A) = \inf_{y\in A} \sum_{i=1}^n \theta^a(d(x_i,y_i)).$$

Ce résultat est aussi un cas particulier de la Proposition II.4.2.

Par linéarisation, l'inégalité de transport $\mathbf{T}_{\theta^a}$ implique l'inégalité de Poincaré suivante (voir [Mau91, OV00]) : pour toute fonction lipschitzienne $f : \mathcal{X} \to \mathbb{R}$, on a

$$\lambda \operatorname{Var}(f) \leq \int |\nabla^- f|^2 d\mu, \tag{V.2}$$

avec $\lambda = 2a^2$ et

$$|\nabla^- f|(x) = \limsup_{y \to x} \frac{[f(y) - f(x)]_-}{d(y,x)}, \qquad x \in \mathcal{X}.$$

En fait, Bobkov, Gentil et Ledoux [BGL01] ont montré l'équivalence, aux constantes près, entre l'inégalité de Poincaré (V.2) et l'inégalité de transport $\mathbf{T}_{\theta^a}$, avec $a = \kappa\sqrt{\lambda}$ (κ est une constante universelle). Ainsi l'inégalité de Poincaré et l'inégalité de transport $\mathbf{T}_{\theta^a}$ sont équivalentes à une même propriété de concentration adimensionnelle (V.1).

Dans [10], nous étudions une version plus faible de propriété de concentration adimensionnelle : μ satisfait la *propriété de concentration* $\mathbf{CI}_2^\infty(\alpha)$ si pour tout entier n et tout borélien $A \subset \mathcal{B}(\mathcal{X}^n)$ de mesure $\mu^n(A) \geq 1/2$,

$$\mu^n(A_r) \geq 1 - \alpha(r), \tag{V.3}$$

où A_r est l'élargissement défini par $A_r = \{x \in \mathcal{X}^n, d_2(x, A) \leq r\}$, avec

$$d_2(x,A) = \inf_{y \in A} d_2(x,y), \qquad d_2(x,y) = \left(\sum_{i=1}^n d^2(x_i, y_i) \right)^{1/2}.$$

La fonction $\alpha : \mathbb{R}^+ \to [0, 1/2]$ est ici un profil de concentration quelconque *non-trivial* indépendant de la dimension n : il existe $r > 0$ tel que $\alpha(r) < 1/2$.

En utilisant le fait que $t \to \theta(\sqrt{t})$ est sous-additive, on montre simplement que

$$\sum_{i=1}^n \theta(ad(x_i, y_i)) \geq \theta(ad_2(x,y)), \qquad x, y \in \mathcal{X}^n,$$

et par conséquent,

$$A_{\theta^a(r), c^n} \subset A_r.$$

Par suite si μ satisfait la propriété de concentration adimensionnelle (V.1), alors μ satisfait la propriété de concentration adimensionnelle (V.3) avec le profil de concentration

$$\alpha(r) = be^{-\theta(ar)} \leq ebe^{-2ar}.$$

Dans [10] (Théorème 1.6.), nous montrons que dès qu'une probabilité μ satisfait la propriété de concentration $\mathbf{CI}_2^\infty(\alpha)$ avec un profil α non-trivial, elle satisfait une inégalité de Poincaré.

Théorème V.0.5. *[10] Si μ satisfait la propriété de concentration* $\mathbf{CI}_2^\infty(\alpha)$*, alors μ satisfait l'inégalité de Poincaré* (V.2) *avec*

$$\sqrt{\lambda} = \sup\left\{\frac{\overline{\Phi}^{-1}(\alpha(r))}{r}\,; r > 0 \textit{ tel que } \alpha(r) \leq 1/2\right\},$$

où $\overline{\Phi}(x) = \frac{1}{\sqrt{2\pi}}\int_x^{+\infty} e^{-u^2/2}du$.

La démonstration repose sur les trois principes suivants : la formulation fonctionnelle du principe de concentration grâce à un opérateur d'infimum-convolution (énoncé en toute généralité dans la section I.2), l'équation d'Hamilton-Jacobi satisfaite par l'opérateur d'infimum convolution (du type Théorème II.5.3), et le théorème de la limite centrale.

Comme nous l'avons vu, l'inégalité de Poincaré est équivalente au principe de concentration adimensionnelle (V.1) avec $a = \kappa\sqrt{\lambda}$. On retrouve ainsi un résultat de Talagrand [Tal91] (Théorème 5.1) qui avait remarqué, sur $\mathcal{X} = \mathbb{R}$, qu'un principe de concentration adimensionnelle $\mathbf{CI}_2^\infty(\alpha)$ non trivial ($\alpha \neq 1/2$) peut toujours être amélioré en un principe de concentration adimensionnelle $\mathbf{CI}_2^\infty(\tilde{\alpha})$ pour lequel le profil de concentration décroît exponentiellement. Le théorème précédent donne plus exactement

$$\tilde{\alpha}(r) = be^{-\kappa\sqrt{\lambda}r}, \quad \text{avec} \quad \sqrt{\lambda} = \sup\left\{\frac{\overline{\Phi}^{-1}(\alpha(r))}{r}\,; r > 0 \text{ tel que } \alpha(r) \leq 1/2\right\},$$

où b et κ sont des constantes positives universelles.

Le résultat du Théorème V.0.5 reste valable si on considère cette fois-ci l'inégalité de Poincaré restreinte aux fonctions $f : \mathcal{X} \to \mathbb{R}$ convexes sur un espace géodésique, et si on remplace la propriété de concentration $\mathbf{CI}_2^\infty(\alpha)$ par la même propriété de concentration adimensionnelle, restreinte aux sous-ensembles A convexes (notée $\mathbf{CCI}_2^\infty(\alpha)$ dans [10]). Pour plus de précision sur ce sujet, voir la section 6 de [10].

Nous avons vu dans le chapitre III que l'inégalité de Poincaré convexe sur $\mathcal{X} = \mathbb{R}$ est équivalente à une inégalité de transport barycentrique (voir Théorème III.5.2), elle-même équivalente à un principe de concentration adimensionnelle restreint aux ensembles convexes sur $\mathbb{R}^n$. On peut ainsi établir, comme précédemment, un résultat identique d'auto-amélioration de la propriété de concentration $\mathbf{CCI}_2^\infty(\alpha)$ restreinte aux ensembles convexes sur $\mathbb{R}^n$.

CHAPITRE VI

Notion de courbure

On distingue sur les variétés riemanniennes plusieurs notions de courbure, entre autres la courbure sectionnelle, et la courbure scalaire ou courbure de Ricci. Grâce à la formule de Bochner, la notion de courbure de Ricci minorée permet de comparer le volume des boules sur la variété au volume des boules sur les variétés de références, la sphère, le plan et l'espace hyperbolique.

La notion de courbure de Ricci s'étend de manière plus abstraite aux espaces mesurés : la mesure de volume v sur la variété riemannienne est remplacée par une mesure de référence absolument continue par rapport à v.

L'un des critères très utilisé en probabilité, généralisant la notion de courbure de Ricci minorée, est le critère courbure-dimension de Bakry-Emery [BÉ85] pour une diffusion ou un semi-groupe de Markov. Ce critère est associé à l'opérateur carré du champ et son itéré, par analogie à la formule de Bochner. Un autre critère est celui de Lott-Sturm-Villani [Vil09, LV09, Stu06a, Stu06b], établi comme propriété de convexité de l'entropie sur l'espace de Wasserstein par des techniques de transport de mesure.

Récemment Ambrosio-Gigli-Savaré ont établi des équivalences entre ces deux critères sur les espaces géodésiques [AGS08, AGS14, AGS15, AGMR15]. Ces critères sont à l'origine de nombreuses inégalités fonctionnelles, Poincaré, log-Sobolev, transport-entropie, isopérimétrie. En outre, ils fournissent des estimations de vitesse de convergence des semi-groupes de Markov (voir par exemple [Led97, ABC$^+$00, Vil03, Vil09]).

Mes recherches ont plus particulièrement porté sur le critère de convexité de l'entropie comme définition de la notion de courbure de Ricci minorée. Dans l'article [5], cette définition permet d'étendre aux espaces géodésiques un résultat de E. Milman [Mil09a, Mil09b, Mil12] sur l'équivalence entre certaines inégalités fonctionnelles et certaines propriétés de concentration lorsque la courbure de Ricci est minorée. Ces résultats sont présentés dans la partie VI.1 ci-après. Dans [9], nous nous sommes attaché à établir une notion de courbure sur certains espaces discrets, en termes de propriété de convexité de l'entropie le long de pseudo-géodésiques. De nombreuses approches se développent actuellement autour de la notion de courbure sur les espaces discrets. Contrairement au cadre riemannien, les rapports entre ces différentes approches sont souvent difficiles à analyser. Dans le paragraphe VI.2, nous expliquerons l'approche élémentaire que nous avons cherché à développer sur les espace discrets.

VI.1 Concentration et inégalité de Sobolev logarithmique sous hypothèse de courbure

Reprenons la définition de la courbure de Ricci minorée sur les espaces géodésiques proposée par Lott-Sturm-Villani [Vil09, LV09, Stu06a, Stu06b]. Soit $(\mathcal{X}, d)$ un espace métrique complet séparable localement compact. On suppose par ailleurs que c'est un *espace de longueur*, ce qui signifie que la distance d a la propriété d'être la longueur minimale d'un chemin entre deux points : pour tous $x, y \in \mathcal{X}$,

$$d(x,y) = \inf\left\{\ell(\gamma); \gamma : [0,1] \to \mathcal{X},\ \text{continue}\ , \gamma(0) = x,\ \gamma(1) = y\right\},$$

où on définit *la longueur d'un chemin* γ par

$$\ell(\gamma) = \sup_{N \geq 1} \sup_{0 = t_0 < t_1 < \cdots < t_N = 1} \left\{\sum_{i=1}^N d(\gamma(t_{i-1}), \gamma(t_i))\right\}.$$

Comme $\mathcal{X}$ est complet et localement compact, $(\mathcal{X}, d)$ est un *espace géodésique* (Théorème de Hopf Rinow, Cohn-Vossen [Coh36], voir Chapitre 1 [Bal12]), c'est à dire que pour tous points x_0, x_1, il existe une géodésique $(x_t)_{t\in[0,1]}$ allant de x_0 à x_1 :

$$d(x_s, x_t) = |t-s| d(x_0, x_1), \qquad \forall s, t \in [0,1].$$

De plus $(\mathcal{X}, d)$ étant géodésique, l'espace de Wasserstein $(\mathcal{P}_2(\mathcal{X}), W_2)$ est lui aussi géodésique (voir [Stu06a]).

Par extension du cas riemannien, d'après la définition de Lott-Sturm-Villani, l'espace probabilisé $(\mathcal{X}, d, \mu)$ a *une courbure de Ricci minorée* par K si pour toutes probabilités $\nu_0, \nu_1 \in \mathcal{P}_2(\mathcal{X})$ absolument continues par rapport à μ, il existe une géodésique $(\nu_t)_{t\in[0,1]}$ dans l'espace de Wasserstein joignant ν_0 à ν_1, telle que pour tout $t \in [0,1]$,

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{K}{2} t(1-t) W_2^2(\nu_0, \nu_1). \qquad \text{(VI.1)}$$

Lorsque t tend vers 0, cette propriété de courbure donne l'inégalité HWI symétrique (voir Proposition 3.36 [LV09]) : pour toutes probabilités ν_1 et ν_0 de densité h_0 par rapport à μ, lipschitzienne,

$$H(\nu_0|\mu) - H(\nu_1|(\mu) \leq W_2(\nu_0, \nu_1)\sqrt{I(\nu_0|\mu)} - \frac{K}{2} W_2^2(\nu_0, \nu_1), \qquad \forall \nu_0, \nu_1 \in \mathcal{P}_2(\mathcal{X}),$$

où $I(\nu_0|\mu)$ est l'information de Fisher relative de ν_0 par rapport à μ,

$$I(\nu_0|\mu) = \int \frac{|\nabla^- h_0|^2}{h_0} d\mu.$$

Lorsqu'on choisit $\nu_1 = \mu$ on obtient l'inégalité HWI :

$$H(\nu_0|\mu) \le W_2(\nu_0,\mu)\sqrt{I(\nu_0|\mu)} - \frac{K}{2}W_2^2(\nu_0,\mu), \qquad \forall \nu_0 \in \mathcal{P}_2(\mathcal{X}).$$

Si le minorant de la courbure satisfait $K > 0$, en optimisant l'inégalité HWI sur toutes les valeurs de $w = W_2(\nu_0,\mu)$, puisque $\sup_{w\ge 0}\{uw - Kw^2/2\} = u_+^2/(2K)$, on obtient que μ satisfait l'inégalité de Sobolev logarithmique

$$H(\nu_0|\mu) \le \frac{1}{2K} I(\nu_0|\mu), \qquad \forall \nu_0 \in \mathcal{P}_2(\mathcal{X}).$$

D'après le Théorème II.5.1 d'Otto-Villani, l'inégalité de Sobolev logarithmique redonne l'inégalité de transport $\mathbf{T}_2$ de Talagrand :

$$W_2^2(\nu_0,\mu) \le \frac{2}{K}\, H(\nu_0|\mu), \qquad \forall \nu_0 \in \mathcal{P}_2(\mathcal{X}).$$

Cette inégalité découle aussi directement de l'hypothèse de courbure donnée par la propriété de convexité (VI.1) (en remarquant que $H(\nu_t|\mu) \ge 0$). Comme expliqué dans le chapitre II.2, cette inégalité de transport $\mathbf{T}_2$ implique une propriété de concentration gaussienne de la mesure μ. Pour résumer, les implications suivantes sont satisfaites

$$\begin{aligned}&\text{Courbure minorée } \Rightarrow \text{HWI } \Rightarrow_{K>0} \text{log-Sobolev}\\ &\qquad\qquad\qquad\qquad \Rightarrow \text{Transport-entropie } \Rightarrow \text{Concentration.}\end{aligned}$$

Pour compléter ce schéma, notons que Ledoux [Led00] a montré que si la courbure de Ricci d'une variété riemannienne est strictement positive, $K > 0$, alors μ satisfait la propriété isopérimétrique gaussienne suivante, pour tout borélien A,

$$\mu^+(\partial A) \ge \sqrt{K}\ \Phi' \circ \Phi^{-1}(\mu(A)),$$

où $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty} e^{-u^2/2}du$, $x \in \mathbb{R}$. Ce résultat s'appuie sur une formulation fonctionnelle des inégalités isopérimétriques due à Bobkov [Bob97]. En fait, sur une variété riemannienne, sans aucune hypothèse de courbure, les implications suivantes sont vérifiées,

$$\begin{aligned}&\text{Isopérimétrie gaussienne } \Rightarrow \text{Log-Sobolev}\\ &\qquad\qquad\qquad\qquad \Rightarrow \text{Transport-entropie } \Rightarrow \text{Concentration.}\end{aligned}$$

La première implication est attribuée à Bechner (voir [Led00]). Dans [Led94], Ledoux a montré que la réciproque de cette première implication est vraie sous hypothèse de courbure nulle (minorée par 0), avec une constante d'isopérimétrie dépendant de la dimension. Comme nous le rappelons dans [5], d'autres implications peuvent être inversées sous certaines hypothèses proposées dans différents travaux. Plus récemment, dans une série de papiers [Mil09b, Mil09a, Mil12], E. Milman a largement amélioré ces résultats en montrant que sous hypothèse de courbure éventuellement négative, si μ satisfait une propriété de concentration gaussienne suffisamment forte, alors μ satisfait la propriété isopérimétrique gaussienne. Un point essentiel des résultats de Milman est que les constantes des inégalités fonctionnelles de la chaîne d'implications inverses ne dépendent pas de la dimension de la variété. Il démontre par ailleurs le même type de résultats sous hypothèse de courbure nulle, en considérant l'inégalité isopérimétrique de Cheeger plutôt que l'inégalité isopérimétrique gaussienne, ainsi qu'une propriété de concentration plus faible. La démonstration des résultats de Milman a été ensuite simplifiée par Ledoux par une approche de semi-groupes [Led11].

Dans [5], nous avons étendu en partie les résultats de Milman, dans le cadre des espaces de longueur, en partant de la définition de Lott-Sturm-Villani de la notion de courbure de Ricci minorée. Sous hypothèse de courbure, éventuellement négative, si μ satisfait une propriété de concentration suffisamment forte, alors μ vérifie une inégalité de Sobolev logarithmique (voir Théorème 1.13, [5]).

Théorème VI.1.1. *Soit $(\mathcal{X}, d)$ un espace de longueur localement compact complet et μ une probabilité dans l'espace de Wasserstein satisfaisant la condition* (VI.1) *de courbure minorée par K avec $K \leq 0$. Supposons que μ satisfait la propriété de concentration gaussienne suivante : pour tout sous-ensemble A tel que $\mu(A) \geq 1/2$,*

$$\mu(A_r) \geq 1 - Me^{-ar^2}, \qquad \forall r \geq 0, \tag{VI.2}$$

où $A_r = \{x \in X; d(x, A) \leq r\}$. Si les constantes a, M et K satisfont la relation

$$\frac{K_-}{2a} < \frac{\log(2)}{\left(2\sqrt{M} + \sqrt{\log(2)}\right)^2},$$

où K_- désigne la partie négative de K, alors μ vérifie l'inégalité de Sololev logarithmique

$$H(\nu|\mu) \leq C\, I(\nu|\mu), \qquad \forall \nu << \mu,$$

où C est une constante ne dépendant que de K, a et M.

Nous avons obtenu le même type de résultats avec l'inégalité de Poincaré : sous l'hypothèse de courbure nulle et de concentration exponentielle, μ satisfait une inégalité de Poincaré (voir Théorème 1.14 [5]).

Idée de la preuve du Théorème VI.1.1. Les étapes de la démonstration sont les suivantes. L'hypothèse de courbure minorée implique l'inégalité HWI qui s'écrit de la manière suivante : pour toute probabilité ν de densité f^2 lipschitzienne par rapport à μ,

$$H(\nu|\mu) \leq W_2(\mu, \nu)\sqrt{I(\nu|\mu)} + \frac{K_-}{2} W_2^2(\mu, \nu),$$

Par suite, en utilisant l'égalité $2uv = \inf_{\lambda > 0}\{\lambda u^2 + v^2/\lambda\}$, pour tout $u, v > 0$, on obtient

$$\mathrm{Ent}_\mu(f^2) \leq \frac{\lambda + K_-}{2} W_2^2(\mu|\nu) + \frac{1}{2\lambda}\int \left|\nabla^- f\right|^2 d\mu \qquad \forall \lambda > 0.$$

A partir de cette inégalité, l'objectif est d'établir une inégalité de Sobolev logarithmique, lorsque μ satisfait une propriété de concentration gaussienne. L'un des arguments clefs est que la propriété de concentration gaussienne est équivalente à une inégalité de transport $\mathbf{T}_2$ non-tendue (voir Corollaire 2.20 [5]).

Lemme VI.1.1. *Si μ vérifie la propriété de concentration* (VI.2)*, alors pour toute probabilité ν, on a*

$$W_2^2(\mu,\nu) \leq \frac{u}{a} H(\nu|\mu) + \frac{4Mau}{u-1}, \qquad \forall u > 1.$$

Cette estimation de $W_2^2(\mu,\nu)$ utilisée avec l'inégalité précédente donne : pour tout $\lambda > 0$ et tout $u > 0$,

$$\mathrm{Ent}_\mu(f^2) \leq \frac{(\lambda+K_-)u}{2a}\mathrm{Ent}_\mu(f^2) + \frac{4Mau(\lambda+K_-)}{2(u-1)} \int f^2\, d\mu + \frac{1}{2\lambda}\int \left|\nabla^- f\right|^2 d\mu.$$

Si les paramètres λ et u vérifient

$$\frac{(\lambda+K_-)u}{2a} < 1,$$

l'inégalité précédente correspond à une inégalité de Sobolev logarithmique non-tendue :

$$\mathrm{Ent}_\mu(f^2) \leq b_1(\lambda,u)\int \left|\nabla^- f\right|^2 d\mu + b_2(\lambda,u)\int f^2\, d\mu.$$

On conclut en utilisant un phénomène d'auto-tension des inégalités de Sobolev logarithmique, observé pour la première fois par Wang [Wan04] (voir Proposition 3.3 [5]).

Proposition VI.1.1. *Si une probabilité μ satisfait l'inégalité de Sobolev logarithmique non-tendue :*

$$\mathrm{Ent}_\mu(f^2) \leq b_1 \int \left|\nabla^- f\right|^2 d\mu + b_2 \int f^2\, d\mu,$$

pour toute fonction bornée f lipschitzienne et si $b_2 < \log 2$, alors μ satisfait l'inégalité de Sobolev logarithmique

$$\mathrm{Ent}_\mu(f^2) \leq b \int \left|\nabla^- f\right|^2 d\mu,$$

pour toute fonction bornée f lipschitzienne avec

$$b = b_1\left(1 + 2\,\frac{2+b_2}{2-e^{b_2}}\right).$$

La preuve du Théorème VI.1.1 consiste donc à choisir les meilleurs paramètres $\lambda > 0$ et $u > 1$ tels que $b_2(\lambda,u) < \log(2)$. Ceci n'est possible que si le paramètre de concentration a est suffisamment grand comparé au paramètre de courbure K_-. □

VI.2 Notion de courbure dans les espaces discrets

Nous allons uniquement présenter la notion de propriété de convexité de l'entropie introduite dans l'article [9] en l'illustrant par les exemples les plus simples. Nous n'abordons pas les développements actuels autour de la notion de courbure sur les espaces discrets, l'approche flot gradient [Maa11, EM12, EM14, EMT15], l'approche par le critère Γ_2 de Bakry-Emery [KKRT15], l'approche par propriété de contraction de la distance de Wasserstein [Oll09, Oll10, JO10, CJ13], ou encore l'approche par propriété de convexité de l'entropie qui s'inspire de la définition de Lott-Sturm-Villani [BS09, OV12, Mie13, Léo13a, Léo13b, Hil14a, Hil14c, Hil14b].

Nous nous plaçons sur un graphe fini $G = (\mathcal{X}, E)$ connexe non-orienté, muni de sa distance naturelle d : pour tout $x, y \in \mathcal{X}$, $d(x, y)$ est le nombre minimal d'arêtes dans E nécessaires pour joindre x à y. Nous proposons la construction naturelle suivante de chemins dans $\mathcal{P}(\mathcal{X})$ joignant deux probabilités ν_0 et ν_1. Cette construction, associée à la distance d, s'inspire de l'interpolation binômiale de Johnson [Joh07].

Soit $\Gamma(x, y)$ l'ensemble des chemins γ joignant x à y de longueur minimale $d(x, y)$, γ peut être modélisé par un sous-ensemble de $d(x, y) + 1$ points z de $\mathcal{X}$ comprenant x et y et ordonné selon leur distance au point x, $d(x, z)$. On considère alors l'interpolation binômiale sur ce chemin γ qui joint la mesure de Dirac en x, δ_x, à la mesure de Dirac en y, δ_y, définie par

$$\nu_t^\gamma(z) = \binom{d(x,y)}{d(x,z)} t^{d(x,z)}(1-t)^{d(y,z)}, \qquad z \in \gamma, \quad \gamma \in \Gamma(x, y), \quad t \in [0, 1],$$

où $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, $0 \le k \le n$. On a bien $\nu_0^\gamma = \delta_x$ et $\nu_0^\gamma = \delta_y$. L'interpolation $(\nu_t^{x,y})_{t\in[0,1]}$ entre δ_x et δ_y est alors obtenue en moyennant uniformément les interpolations $(\nu_t^\gamma)_{t\in[0,1]}$ sur tous les chemins $\gamma \in \Gamma(x, y)$:

$$\nu_t^{x,y}(z) = \sum_{\gamma\in\Gamma(x,y)} \nu_t^\gamma(z) \frac{\mathbb{1}_{z\in\gamma}}{|\Gamma(x, y)|}, \qquad z \in \mathcal{X},$$

où $|\Gamma(x, y)|$ est le cardinal de l'ensemble $\Gamma(x, y)$. Une fois construits les chemins de probabilité entre les mesures de Dirac, et étant donné un couplage π sur l'espace produit $\mathcal{X} \times \mathcal{X}$ ayant pour marginales $\nu_0 \in \mathcal{P}(\mathcal{X})$ et $\nu_1 \in \mathcal{P}(X)$, $\pi \in \Pi(\nu_0, \nu_1)$, nous construisons le chemin de probabilité $(\nu_t^\pi)_{t\in[0,1]}$ associé à π entre $\nu_0 = \nu_0^\pi$ et $\nu_1 = \nu_1^\pi$ en posant

$$\nu_t^\pi(z) = \sum_{x,y\in\mathcal{X}} \nu_t^{x,y}(z)\pi(x, y), \qquad z \in \mathcal{X}.$$

Lorsque π^* est un couplage optimal pour la distance de Wasserstein, c'est à dire

$$W_1(\nu_0, \nu_1) = \inf_{\pi\in\Pi(\nu_0,\nu_1)} \iint d(x, y)d\pi(x, y) = \sum_{x,y\in\mathcal{X}} d(x, y)\pi^*(x, y),$$

nous montrons que le chemin $(\nu_t^{\pi^*})_{t\in[0,1]}$ est une géodésique de l'espace de Wasserstein $(\mathcal{P}(\mathcal{X}), W_1)$ (voir Proposition 2.6, [10]),

$$W_1(\nu_t^{\pi^*}, \nu_s^{\pi^*}) = |t-s| W_1(\nu_0,\nu_1), \qquad \forall t,s\in[0,1].$$

Soit μ une probabilité de référence sur $\mathcal{X}$. On dit que μ satisfait *une propriété de déplacement convexe de constante* K (voir [10]), si pour tout ν_0,ν_1, il existe un couplage $\pi\in\Pi(\nu_0,\nu_1)$ tel que

$$H(\nu_t^\pi|\mu) \le (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - Kt(1-t)\, I(\pi), \tag{VI.3}$$

où la fonction $I : \pi\in\mathcal{P}(\mathcal{X}\times\mathcal{X}) \to \mathbb{R}^+$ est une fonction de coût. Dans [10], nous considérons les fonctions $I = J_2$ et $I = I_2 + \bar{I}_2$ avec

$$J_2(\pi) = \left(\sum_{x,y\in\mathcal{X}} d(x,y)\pi(x,y)\right)^2,$$

et avec la décomposition $\pi(x,y) = \nu_0(x)p(x,y) = \nu_1(y)\bar{p}(y,x)$,

$$I_2(\pi) = \sum_{x\in\mathcal{X}}\left(\sum_{y\in\mathcal{X}} d(x,y)p(x,y)\right)^2 \nu_0(x),\ \bar{I}_2(\pi) = \sum_{y\in\mathcal{X}}\left(\sum_{x\in\mathcal{X}} d(x,y)\bar{p}(y,x)\right)^2 \nu_1(y).$$

Remarquons que si π^* est un couplage optimal pour la distance de Wasserstein W_1, alors

$$J_2(\pi^*) = W_1^2(\nu_0,\nu_1).$$

De même, les fonctions de coût I_2 et $\bar{I}_2$ sont associées aux coûts faibles de type Marton

$$\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) = \inf_{\pi\in\Pi(\nu_0,\nu_1)} I_2(\pi), \qquad \text{et} \qquad \widetilde{\mathcal{T}}_2(\nu_0|\nu_1) = \inf_{\pi\in\Pi(\nu_0,\nu_1)} \bar{I}_2(\pi).$$

La propriété de convexité (VI.3) se tensorise (voir Théorème 1.6, [10]). Considérons par exemple le cas où $I = J_2$, si μ satisfait (VI.3) avec constante $K\ge 0$, alors la mesure produit μ^n sur le graphe produit $\mathcal{X}^n$ satisfait (VI.3) avec la même constante K et le coût

$$J_2^{(n)}(\pi) = \sum_{i=1}^n \left(\sum_{x,y\in\mathcal{X}} d(x_i,y_i)\pi(x,y)\right)^2, \qquad \pi\in\Pi(\nu_0,\nu_1),$$

avec $x=(x_1,\dots,x_n), y=(y_1,\dots,y_n)\in\mathcal{X}^n$ et $\nu_0,\nu_1\in\mathcal{P}(\mathcal{X}^n)$. La distance $d^{(n)}$ sur le graphe produit $\mathcal{X}^n$ est donnée par

$$d^{(n)}(x,y) = \sum_{i=1}^n d(x_i,y_i).$$

L'inégalité de Cauchy-Schwarz donne donc

$$J_2^{(n)}(\pi) \ge \frac{1}{n}\left(\sum_{x,y\in\mathcal{X}} d^{(n)}(x,y)\pi(x,y)\right)^2 \ge \frac{1}{n} W_1^2(\nu_0,\nu_1).$$

Pour illustrer simplement cette propriété de tensorisation, considérons une loi de Bernoulli, de paramètre quelconque, sur $\mathcal{X} = \{0,1\}$. Comme on le verra plus loin dans ce paragraphe, elle satisfait la propriété de convexité (VI.3) avec constante $K = 2$ (pour $I = J_2$). Donc, par tensorisation, toute mesure produit $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ sur le cube discret $\mathcal{X}^n = \{0,1\}^n$ satisfait la propriété de convexité suivante : pour tous $\nu_0, \nu_1 \in \mathcal{P}(\{0,1\}^n)$, il existe $\pi \in \Pi(\nu_0,\nu_1)$ telle que

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{2}{n}t(1-t)\, W_1^2(\nu_0,\nu_1), \qquad \forall t \in [0,1].$$

Puisque $H(\nu_t^\pi|\mu) \geq 0$, en choisissant $\nu_1 = \mu$, cette propriété redonne par exemple l'inégalité de transport connue pour les mesures produit μ sur le cube discret

$$W_1^2(\nu_0,\mu) \leq \frac{n}{2}H(\nu_0|\mu), \qquad \forall \nu_0 \in \mathcal{P}(\{0,1\}^n).$$

La propriété de convexité (VI.3) pour la loi de Bernoulli, avec constante $K = 2$, est une conséquence de la propriété de convexité “universelle” associée à la distance en variation totale suivante (voir Proposition 4.1 [10]) : pour toutes probabilités μ, ν_0 et ν_1 sur $\mathcal{X}$,

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{1}{2}t(1-t)\, \|\nu_0 - \nu_1\|_{TV}^2, \qquad \forall t \in [0,1],$$

où $\nu_t = (1-t)\nu_0 + t\nu_1$. L'article [10] propose une preuve simple de cette inégalité et en remarquant que $H(\nu_t|\mu) \geq 0$, on obtient l'inégalité de Csizár-Kullback-Pinsker (IV.1). Nous démontrons par ailleurs dans [10] une autre propriété de convexité universelle associée à la distance de Marton

$$\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) = \inf_{\pi \in \Pi(\nu_0,\nu_1)} \left\{ \int \left(\int \mathbb{1}_{x\neq y}\, dp_x(y)\right)^2 d\nu_0(x), d\pi(x,y) = d\nu_0(x)dp_x(y) \right\}.$$

Pour toutes probabilités μ, ν_0 et ν_1 sur $\mathcal{X}$, on a, pour tout $t \in [0,1]$

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{1}{2}t(1-t)\left(\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1)\right), \tag{VI.4}$$

avec $\nu_t = (1-t)\nu_0 + t\nu_1$. En remarquant que $H(\nu_t|\mu) \geq 0$, puis en optimisant sur le paramètre $t \in]0,1[$, cette propriété de convexité fournit l'inégalité de transport universelle (IV.4).

Dans [9], nous expliquons par ailleurs comment la propriété de convexité de l'entropie (VI.3) peut s'interpréter en termes d'inégalité de type Prékopa-Leindler. L'intérêt des coûts I_2 et $\bar{I}_2$ est aussi de fournir des inégalités de Sobolev logarithmiques discrètes, lorsque la propriété de convexité (VI.3) est satisfaite avec $K > 0$.

CHAPITRE VII

Versions améliorées quantitatives de l'inégalité de Sobolev logarithmique gaussienne

L'inégalité de Sobolev logarithmique pour la mesure gaussienne standard γ_n sur $\mathbb{R}^n$ est due à Gross [Gro75, Gro93], comme propriété équivalente de la propriété d'hypercontractivité du semi-groupe d'Ornstein-Uhlenbeck (voir [ABC$^+$00]). Nous avons vu dans la partie précédente qu'elle se déduit par ailleurs de l'inégalité HWI, puisque la courbure de l'espace probabilisé $(\mathbb{R}^n, |\cdot|, \gamma_n)$ est minorée par $K = 1$ au sens de Lott-Sturm-Villani.

En théorie de l'information, ces inégalités s'interprêtent comme des relations entre des distances classiques. Soit Z une variable aléatoire de loi gaussienne γ_n. Notons $D(X|Z)$ et $I(X|Z)$ l'entropie relative et l'information de Fisher relative d'une variable aléatoire X par rapport à Z. Par définition, si μ est de densité lipschitzienne f par rapport à γ_n,

$$D(X|Z) = H(\mu|\gamma_n) = \mathrm{Ent}_{\gamma_n}(f), \qquad I(X|Z) = I(\mu|\gamma_n) = \int \frac{|\nabla f|^2}{f} d\gamma_n.$$

Par analogie, notons

$$W_2(X, Z) = W_2(\mu, \gamma_n),$$

la distance de Wasserstein d'ordre 2 entre μ et γ_n. L'inégalité de Sobolev logarithmique de Gross s'écrit alors

$$D(X|Z) \leq \frac{1}{2} I(X|Z),$$

pour tout vecteur aléatoire X de densité lipschitzienne.

L'objet de l'article [11] est d'améliorer l'inégalité de Gross en proposant des bornes inférieures sur le déficit

$$\delta(f) = \frac{1}{2} I(X|Z) - D(X|Z).$$

Un premier résultat dans ce sens est celui de Bakry-Ledoux (Proposition 2, [BL06]) qui peut être formulé de la façon suivante : pour tout vecteur aléatoire X de densité g lipschitzienne,

$$I(X|Z) - 2D(X|Z) \geq n\Delta\left(\frac{I(X)}{n} - 1\right), \tag{VII.1}$$

où Δ est la fonction positive définie par

$$\Delta(t) = t - \log(1+t), \qquad t > -1,$$

et $I(X)$ est l'information de Fisher (non-relative) définie par

$$I(X) = \int \frac{|\nabla g(x)|^2}{g(x)} dx.$$

De manière surprenante, l'inégalité améliorée (VII.1) est une conséquence de l'inégalité de Gross elle-même. L'idée est d'appliquer l'inégalité de Sobolev logarithmique à la variable λX, $\lambda > 0$ puis d'optimiser en λ (voir [11]).

Remarquons que l'inégalité HWI donne une autre estimation du déficit car elle s'écrit encore

$$I(X|Z) - 2D(X|Z) \geq \left(\sqrt{I(X|Z)} - W_2(X,Z)\right)^2,$$

Ainsi le déficit dans l'inégalité $W_2 - I$,

$$W_2^2(X,Z) \leq I(X|Z),$$

contrôle celui de l'inégalité de Sobolev logarithmique. Puisque par ailleurs on a

$$\sqrt{I(X|Z)} - W_2(X,Z) \geq \sqrt{2D(X|Z)} - W_2(X|Z),$$

le contrôle du déficit dans l'inégalité de transport-entropie de Talagrand

$$W_2^2(X,Z) \leq 2D(X|Z),$$

fournit lui-aussi un contrôle du déficit dans l'inégalité de Sobolev logarithmique.

Revenons à l'inégalité HWI. Dans [11] (Théorème 1.1), nous améliorons cette inégalité de la façon suivante.

Théorème VII.0.1. *[11] Pour tout vecteur aléatoire X dont la densité est lipschitzienne et telle que $I(X|Z)$ soit fini, on a*

$$I(X|Z) - 2D(X|Z) \geq \left(\sqrt{I(X|Z)} - W_2(X,Z)\right)^2 + n\Delta\left(\frac{W_2(X,Z)}{\sqrt{I(X|Z)}}\left(\frac{I(X)}{n} - 1\right)\right),$$

Cette version dimensionnelle améliorée de l'inégalité HWI est dans l'esprit des résultats de Wang pour les mesures satisfaisant une condition de courbure dimensionnelle [Wan08]. L'espace gaussien $(\mathbb{R}^n, |\cdot|, \gamma_n)$ n'entre cependant pas directement dans le cadre des résultats de Wang.

L'inégalité du Théorème VII.0.1 ne se compare pas directement à celle de Bakry-Ledoux (VII.1) (voir [11]). Nous avons obtenu ce résultat par des arguments propres à la théorie de l'information. Ceci donne une nouvelle preuve de l'inégalité HWI, différente de celle de Otto-Villani [LV09] qui s'appuie sur des arguments de transport optimal de mesure, ou encore de celle de Bakry-Gentil-Ledoux [BGL01] qui utilise les inégalités de type Harnack sur les semi-groupes.

La preuve du théorème VII.0.1 s'appuie sur les deux lemmes suivants. Le premier Lemme est l'inégalité d'information de Fisher due à Stam [Sta59].

Lemme VII.0.1. *Si X et Y sont des vecteurs aléatoires indépendants ayant des densités lipschitziennes, alors*

$$\frac{1}{I(X+Y)} \geq \frac{1}{I(X)} + \frac{1}{I(Y)}.$$

Pour tout vecteur aléatoire X, notons

$$X_t := X + \sqrt{t}Z,$$

où Z est un vecteur aléatoire gaussien standard indépendant de X. Le second lemme est une inégalité de transport-entropie inverse, due à Wu [Wu11], dont nous rappelons la preuve élémentaire dans [11] (Appendice A).

Lemme VII.0.2. *Soit X et Y des vecteurs aléatoires admettant un moment d'ordre 2, alors pour tout $t > 0$,*

$$D(X_t|Y_t) \leq \frac{1}{2t} W_2^2(X,Y).$$

Les Théorèmes 1.3 et 1.4 de [11] donnent également des estimations du déficit dans l'inégalité de Sobolev logarithmique gaussienne, qui améliorent les résultats d'Indrei et Marcon [IM14]. L'avantage du Théorème 1.3 est de fournir les cas d'égalité dans l'inégalité de Gross, qui correspondent aux lois gaussiennes translatées (voir Appendice C [11]). Ces théorèmes sont établis à partir de versions améliorées de l'inégalité de transport de Talagrand pour la mesure gaussienne sur $\mathbb{R}$, comme celle de Barthe-Kolesnikov [BK08]. Les estimations du déficit dans l'inégalité de Sobolev logarithmique uni-dimensionnelle sont donc obtenues par des estimations du déficit dans l'inégalité de transport de Talagrand. Des arguments de tensorisation permettent ensuite de sortir du cas uni-dimensionnel.

Bibliographie

[ABC$^+$00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.

[ACW14] R. Adamczak, D. Chafaï, and P. Wolff. Circular law for random matrices with exchangeable entries. *ArXiv e-prints*, February 2014.

[Ada05] R. Adamczak. Logarithmic Sobolev inequalities and concentration of measure for convex functions and polynomial chaoses. *Bull. Pol. Acad. Sci. Math.*, 53(2) :221–238, 2005.

[Ada08] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13 :no. 34, 1000–1034, 2008.

[AGMR15] L. Ambrosio, N. Gigli, A. Mondino, and T. Rajala. Riemannian Ricci curvature lower bounds in metric measure spaces with σ-finite measure. *Trans. Amer. Math. Soc.*, 367(7) :4661–4701, 2015.

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[AGS14] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Math. J.*, 163(7) :1405–1490, 2014.

[AGS15] L. Ambrosio, N. Gigli, and G. Savaré. Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds. *Ann. Probab.*, 43(1) :339–404, 2015.

[Bal12] W. Ballmann. *Lectures on Spaces of Nonpositive Curvature.* Oberwolfach Seminars. Birkhäuser Basel, 2012.

[BÉ85] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

[BG99] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1) :1–28, 1999.

[BGL01] S. G. Bobkov, I. Gentil, and M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. *J. Math. Pures Appl. (9)*, 80(7) :669–696, 2001.

[BHT06] S. G. Bobkov, C. Houdré, and P. Tetali. The subgaussian constant and concentration inequalities. *Israel J. Math.*, 156 :255–283, 2006.

[BK08] F. Barthe and A. V. Kolesnikov. Mass transport and variants of the logarithmic Sobolev inequality. *J. Geom. Anal.*, 18(4) :921–979, 2008.

[BL06] D. Bakry and M. Ledoux. A logarithmic Sobolev form of the Li-Yau parabolic inequality. *Rev. Mat. Iberoam.*, 22(2) :683–702, 2006.

[BLM00] S. Boucheron, G. Lugosi, and P. Massart. A sharp concentration inequality with applications. *Random Structures Algorithms*, 16(3) :277–292, 2000.

[BLM03] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3) :1583–1614, 2003.

[BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

[Bob97] S. G. Bobkov. An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space. *Ann. Probab.*, 25(1) :206–214, 1997.

[Bou03] O. Bousquet. Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 213–247. Birkhäuser, Basel, 2003.

[BR03] F. Barthe and C. Roberto. Sobolev inequalities for probability measures on the real line. *Studia Math.*, 159(3) :481–497, 2003.

[BR08] F. Barthe and C. Roberto. Modified logarithmic Sobolev inequalities on $\mathbb{R}$. *Potential Anal.*, 29(2) :167–193, 2008.

[BS09] A. I. Bonciocat and K. T. Sturm. Mass transportation and rough curvature bounds for discrete spaces. *J. Funct. Anal.*, 256(9) :2944–2966, 2009.

[CD10] S. Chatterjee and P.S. Dey. Applications of Stein's method for concentration inequalities. *Ann. Probab.*, 38(6) :2443–2485, 2010.

[CG06] P. Cattiaux and A. Guillin. On quadratic transportation cost inequalities. *J. Math. Pures Appl. (9)*, 86(4) :341–361, 2006.

[Cha05] S. Chatterjee. Concentration inequalities with exchangeable pairs (Ph.D. thesis). *ArXiv Mathematics e-prints*, July 2005.

[Cha07] S. Chatterjee. Stein's method for concentration inequalities. *Probab. Theory Related Fields*, 138(1-2) :305–321, 2007.

[CJ13] D. Chafaï and A. Joulin. Intertwining and commutation relations for birth-death processes. *Bernoulli*, 19(5A) :1855–1879, 2013.

[Coh36] S. Cohn-Vossen. Existenz kürzester Wege. *Compos. Math.*, 3 :441–452, 1936.

[Csi67] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2 :299–318, 1967.

[Dem97] A. Dembo. Information inequalities and concentration of measure. *Ann. Probab.*, 25(2) :927–939, 1997.

[DGW04] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.*, 32(3B) :2702–2732, 2004.

[Dou94] P. Doukhan. *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994. Properties and examples.

[EM12] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.*, 206(3) :997–1038, 2012.

[EM14] M. Erbar and J. Maas. Gradient flow structures for discrete porous medium equations. *Discrete Contin. Dyn. Syst.*, 34(4) :1355–1374, 2014.

[EMT15] M. Erbar, J. Maas, and P. Tetali. Discrete Ricci curvature bounds for Bernoulli-Laplace and random transposition models. *Ann. Fac. Sci. Toulouse Math. (6)*, 24(4) :781–800, 2015.

[FHT03] A. A. Fedotov, P. Harremoës, and F. Topsøe. Refinements of Pinsker's inequality. *IEEE Trans. Inform. Theory*, 49(6) :1491–1498, 2003.

[GGM05] I. Gentil, A. Guillin, and L. Miclo. Modified logarithmic Sobolev inequalities and transportation inequalities. *Probab. Theory Related Fields*, 133(3) :409–436, 2005.

[GGM07] I. Gentil, A. Guillin, and L. Miclo. Modified logarithmic Sobolev inequalities in null curvature. *Rev. Mat. Iberoam.*, 23(1) :235–258, 2007.

[Gil10] G. L. Gilardoni. On Pinsker's and Vajda's type inequalities for Csiszár's f-divergences. *IEEE Trans. Inform. Theory*, 56(11) :5377–5386, 2010.

[GL10] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Process. Related Fields*, 16(4) :635–736, 2010.

[Goz07] N. Gozlan. Characterization of Talagrand's like transportation-cost inequalities on the real line. *J. Funct. Anal.*, 250(2) :400–425, 2007.

[Goz09] N. Gozlan. A characterization of dimension free concentration in terms of transportation inequalities. *Ann. Probab.*, 37(6) :2480–2498, 2009.

[Goz10] N. Gozlan. Poincaré inequalities and dimension free concentration of measure. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(3) :708–739, 2010.

[Goz12] N. Gozlan. Transport-entropy inequalities on the line. *Electron. J. Probab.*, 17 :no. 49, 18, 2012.

[Gro75] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4) :1061–1083, 1975.

[Gro93] L. Gross. Logarithmic Sobolev inequalities and contractivity properties of semigroups. In *Dirichlet forms (Varenna, 1992)*, volume 1563 of *Lecture Notes in Math.*, pages 54–88. Springer, Berlin, 1993.

[Har66] L. H. Harper. Optimal numberings and isoperimetric problems on graphs. *J. Combinatorial Theory*, 1 :385–393, 1966.

[Hil14a] E. Hillion. Contraction of measures on graphs. *Potential Anal.*, 41(3) :679–698, 2014.

[Hil14b] E. Hillion. Entropy along W_{1,+}-geodesics on graphs. *ArXiv e-prints*, June 2014.

[Hil14c] E. Hillion. $W_{1,+}$-interpolation of probability measures on graphs. *ArXiv e-prints*, February 2014.

[HS87] R. Holley and D. Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *J. Statist. Phys.*, 46(5-6) :1159–1194, 1987.

[IM14] E. Indrei and D. Marcon. A quantitative log-Sobolev inequality for a two parameter family of functions. *Int. Math. Res. Not. IMRN*, (20) :5563–5580, 2014.

[JO10] A. Joulin and Y. Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.*, 38(6) :2418–2442, 2010.

[Joh07] O. Johnson. Log-concavity and the maximum entropy property of the Poisson distribution. *Stochastic Process. Appl.*, 117(6) :791–802, 2007.

[KKRT15] B. Klartag, G. Kozma, P. Ralli, and P. Tetali. Discrete curvature and abelian groups. *ArXiv e-prints*, January 2015.

[Kom88] H. Komiya. Elementary proof for Sion's minimax theorem. *Kodai Math. J.*, 11(1) :5–7, 1988.

[Kon12] A. Kontorovich. Obtaining measure concentration from Markov contraction. *Markov Process. Related Fields*, 18(4) :613–638, 2012.

[KR05] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3) :1060–1077, 2005.

[Kul67] S. Kullback. Lower bound for discrimination information in terms of variation. *IEEE Trans. Information Theory*, 4 :126–127, 1967.

[Led93] M. Ledoux. Inégalités isopérimétriques en analyse et probabilités. *Astérisque*, (216) :Exp. No. 773, 5, 343–375, 1993. Séminaire Bourbaki, Vol. 1992/93.

[Led94] M. Ledoux. A simple analytic proof of an inequality by P. Buser. *Proc. Amer. Math. Soc.*, 121(3) :951–959, 1994.

[Led97] M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.*, 1 :63–87 (electronic), 1997.

[Led00] M. Ledoux. The geometry of Markov diffusion generators. *Ann. Fac. Sci. Toulouse Math. (6)*, 9(2) :305–366, 2000. Probability theory.

[Led01] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[Led11] M. Ledoux. From concentration to isoperimetry : semigroup proofs. In *Concentration, functional inequalities and isoperimetry*, volume 545 of *Contemp. Math.*, pages 155–166. Amer. Math. Soc., Providence, RI, 2011.

[Léo13a] C. Léonard. Lazy random walks and optimal transport on graphs. *ArXiv e-prints*, August 2013.

[Léo13b] C. Léonard. On the convexity of the entropy along entropic interpolations. *ArXiv e-prints*, October 2013.

[Lév51] P Lévy. *Problèmes concrets d'analyse fonctionnelle. Avec un complément sur les fonctionnelles analytiques par F. Pellegrino.* Gauthier-Villars, Paris, 1951. 2d ed.

[LM03] M. J. Luczak and C. McDiarmid. Concentration for locally acting permutations. *Discrete Math.*, 265(1-3) :159–171, 2003.

[LT91] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)].* Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

[LV07] J. Lott and C. Villani. Hamilton-Jacobi semigroup on length spaces and applications. *J. Math. Pures Appl. (9)*, 88(3) :219–229, 2007.

[LV09] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3) :903–991, 2009.

[Maa11] J. Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8) :2250–2292, 2011.

[Mar96a] K. Marton. Bounding $\overline{d}$-distance by informational divergence : a method to prove measure concentration. *Ann. Probab.*, 24(2) :857–866, 1996.

[Mar96b] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3) :556–571, 1996.

[Mar97] K. Marton. Erratum to : "A measure concentration inequality for contracting Markov chains" [Geom. Funct. Anal. **6** (1996), no. 3, 556–571 ; MR1392329 (97g :60082)]. *Geom. Funct. Anal.*, 7(3) :609–613, 1997.

[Mar03] K. Marton. Measure concentration and strong mixing. *Studia Sci. Math. Hungar.*, 40(1-2) :95–113, 2003.

[Mar04] K. Marton. Measure concentration for Euclidean distance in the case of dependent random variables. *Ann. Probab.*, 32(3B) :2526–2544, 2004.

[Mar10] K. Marton. Correction : Measure concentration for Euclidean distance in the case of dependent random variables [mr2078549]. *Ann. Probab.*, 38(1) :439–442, 2010.

[Mas00a] P. Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, 28(2) :863–884, 2000.

[Mas00b] P. Massart. Some applications of concentration inequalities to statistics. *Ann. Fac. Sci. Toulouse Math. (6)*, 9(2) :245–303, 2000. Probability theory.

[Mas07] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

[Mau79] B. Maurey. Construction de suites symétriques. *C. R. Acad. Sci. Paris Sér. A-B*, 288(14) :A679–A681, 1979.

[Mau91] B. Maurey. Some deviation inequalities. *Geom. Funct. Anal.*, 1(2) :188–197, 1991.

[McD02] C. McDiarmid. Concentration for independent permutations. *Combin. Probab. Comput.*, 11(2) :163–178, 2002.

[Mie13] A. Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Partial Differential Equations*, 48(1-2) :1–31, 2013.

[Mil09a] E. Milman. Concentration and isoperimetry are equivalent assuming curvature lower bound. *C. R. Math. Acad. Sci. Paris*, 347(1-2) :73–76, 2009.

[Mil09b] E. Milman. On the role of convexity in functional and isoperimetric inequalities. *Proc. Lond. Math. Soc. (3)*, 99(1) :32–66, 2009.

[Mil12] E. Milman. Properties of isoperimetric, functional and transport-entropy inequalities via concentration. *Probab. Theory Related Fields*, 152(3-4) :475–507, 2012.

[MS86] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.

[MSWW11] Y. Ma, S. Shen, X. Wang, and L. Wu. Transportation inequalities : from Poisson to Gibbs measures. *Bernoulli*, 17(1) :155–169, 2011.

[Muc72] B. Muckenhoupt. Hardy's inequality with weights. *Studia Math.*, 44 :31–38, 1972. Collection of articles honoring the completion by Antoni Zygmund of 50 years of scientific activity, I.

[Oll09] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3) :810–864, 2009.

[Oll10] Y. Ollivier. A survey of Ricci curvature for metric spaces and Markov chains. In *Probabilistic approach to geometry*, volume 57 of *Adv. Stud. Pure Math.*, pages 343–381. Math. Soc. Japan, Tokyo, 2010.

[OV00] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2) :361–400, 2000.

[OV12] Y. Ollivier and C. Villani. A curved Brunn-Minkowski inequality on the discrete hypercube, or : what is the Ricci curvature of the discrete hypercube ? *SIAM J. Discrete Math.*, 26(3) :983–996, 2012.

[Pan01] D. Panchenko. A note on Talagrand's concentration inequality. *Electron. Comm. Probab.*, 6 :55–65 (electronic), 2001.

[Pan02] D. Panchenko. Some extensions of an inequality of Vapnik and Chervonenkis. *Electron. Comm. Probab.*, 7 :55–65 (electronic), 2002.

[Pan03] D. Panchenko. Symmetrization approach to concentration inequalities for empirical processes. *Ann. Probab.*, 31(4) :2068–2081, 2003.

[Pau12] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *ArXiv e-prints*, December 2012.

[Pau14] D. Paulin. The convex distance inequality for dependent random variables, with applications to the stochastic travelling salesman and other problems. *Electron. J. Probab.*, 19 :no. 68, 34, 2014.

[Pin64] M. S. Pinsker. *Information and information stability of random variables and processes*. Translated and edited by Amiel Feinstein. Holden-Day Inc., San Francisco, Calif., 1964.

[Rio01] E. Rio. Inégalités de concentration pour les processus empiriques de classes de parties. *Probab. Theory Related Fields*, 119(2) :163–175, 2001.

[Rio02] E. Rio. Une inégalité de Bennett pour les maxima de processus empiriques. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6) :1053–1057, 2002. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

[Rio12] E. Rio. Sur la fonction de taux dans les inégalités de Talagrand pour les processus empiriques. *C. R. Math. Acad. Sci. Paris*, 350(5-6) :303–305, 2012.

[Rio13] E. Rio. On McDiarmid's concentration inequality. *Electron. Commun. Probab.*, 18 :no. 44, 11, 2013.

[RR91] M. M. Rao and Z. D. Ren. *Theory of Orlicz spaces*. Marcel Dekker Inc., 1991.

[Sio58] M. Sion. On general minimax theorems. *Pacific J. Math.*, 8 :171–176, 1958.

[Sta59] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2 :101–112, 1959.

[Str65] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36 :423–439, 1965.

[Stu06a] K.T. Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1) :65–131, 2006.

[Stu06b] K.T. Sturm. On the geometry of metric measure spaces. II. *Acta Math.*, 196(1) :133–177, 2006.

[Tal91] M. Talagrand. A new isoperimetric inequality and the concentration of measure phenomenon. In *Geometric aspects of functional analysis (1989–90)*, volume 1469 of *Lecture Notes in Math.*, pages 94–124. Springer, Berlin, 1991.

[Tal95] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, (81) :73–205, 1995.

[Tal96a] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3) :505–563, 1996.

[Tal96b] M. Talagrand. A new look at independence. *Ann. Probab.*, 24(1) :1–34, 1996.

[Tal96c] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3) :587–600, 1996.

[Vil03] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[Vil09] C. Villani. *Optimal transport : old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

[Wan04] F.-Y. Wang. Spectral gap for hyperbounded operators. *Proc. Amer. Math. Soc.*, 132(9) :2629–2638, 2004.

[Wan08] F.-Y. Wang. Generalized transportation-cost inequalities and applications. *Potential Anal.*, 28(4) :321–334, 2008.

[Wan14] N.-Y. Wang. Concentration inequalities for Gibbs sampling under d_{l_2}-metric. *Electron. Commun. Probab.*, 19 :no. 63, 11, 2014.

[Wu06] L. Wu. Poincaré and transportation inequalities for Gibbs measures under the Dobrushin uniqueness condition. *Ann. Probab.*, 34(5) :1960–1989, 2006.

[Wu11] Y. Wu. A simple transportation-information inequality, with applications to hwi inequalities and predictive density estimation. Preprint, 2011.

[WW77] D. L. Wang and P. Wang. Extremal configurations on a discrete torus and a generalization of the generalized Macaulay theorem. *SIAM. J. Appl. Math.*, 33(1) :55–59, 1977.

[WW14] N.-Y. Wang and L. Wu. Convergence rate and concentration inequalities for Gibbs sampling in high dimension. *Bernoulli*, 20(4) :1698–1716, 2014.