université Paris Ouest Nanterre La Défense

ED 139 : Ecole doctorale Connaissance, langage, modélisation
EA 3454 - Modélisation aléatoire de Paris-X

UNIVERSITÉ PARIS LUMIÈRES

Yan SHU

Opérateurs d'inf-convolution et inégalités de transport sur les graphes

Thèse présentée et soutenue publiquement le 7 juillet 2016
en vue de l'obtention du doctorat de Mathématiques
de l'Université Paris Ouest Nanterre La Défense

sous la direction de M. Cyril ROBERTO et M. Nathaël GOZLAN

Jury :

Président :	M. Christian LEONARD	Université Paris Ouest Nanterre La Défense
Rapporteur:	M. Ivan GENTIL	Université Claude Bernard Lyon 1
Rapporteur:	M. Mokshay MADIMAN	University of Delaware
Examinateur:	M. Frank BARTHE	Université Paul Sabatier
Examinateur:	M. Nathaël GOZLAN	Université Paris Ouest Nanterre La Défense
Examinateur:	M. Cyril ROBERTO	Université Paris-Est Marne-la-Vallée

Résumé

Dans cette thèse, nous nous intéressons à différents opérateurs d'inf-convolutions et à leurs applications à une classe d'inégalités de transport générales, plus spécifiquement sur les graphes. Notre objet de recherche s'inscrit donc dans les théories du transport de mesure et de l'analyse fonctionnelle.

En introduisant une notion de gradient adapté au cadre discret (et plus généralement à tout espace métrique dont les boules sont compactes), nous prouvons que certains opérateurs d'inf-convolution sont solutions d'une inéquation d'Hamilton Jacobi sur les graphes. Ce résultat nous permet d'étendre au cadre discret un théorème classique de Bobkov, Gentil et Ledoux. Plus précisément nous montrons que des inégalités de transport faible (adaptées au cadre discret) sont équivalentes, sur un graphe, à l'hypercontractivité des opérateurs d'inf-convolutions. On en déduit plusieurs résultats concernant différentes inégalités fonctionnelles, dont celle de Sobolev logarithmique et de transport faible.

Nous étudions par ailleurs les propriétés générales de différents opérateurs d'inf-convolutions, incluant le précédent, mais aussi un opérateur relié à un modèle issu de la physique (et au phénomène de grande déviation), toujours sur les graphes (dérivabilités, convexité, points extremum etc.).

Dans un deuxième temps, nous nous intéressons aux liens entre différentes notions de courbure de Ricci sur les graphes – proposées récemment par plusieurs auteurs – et les inégalités fonctionnelles de type transport-entropie, ou transport-information associées à une chaîne de Markov. Nous obtenons également une borne supérieure sur le diamètre d'un graphe dont la courbure, en un certain sens, est minorée, un résultat à la Bonnet-Myers.

Enfin, en nous restreignant au cas de la dimension 1, sur la droite réelle, nous obtenons une caractérisation d'une inégalité de transport faible et de l'inégalité de Sobolev logarithmique restreinte aux fonctions convexes. Ces résultats utilisent des propriétés géométriques liés à l'ordre convexe.

Mots clés : inf-convolution, equation d'Hamilton-Jacobi, inégalité de transport faible, courbure de Ricci, espace discret, ordre convexe

Abstract

In this thesis, we interest in different inf-convolution operators and their applications to a class of general transportation inequalities, more specifically in the graphs. Therefore, our research topic fits in the theories of transportation and functional analysis.

By introducing a gradient notion adapting to a discrete space (more generally to all space in which all closed balls are compact), we prove that some inf-convolution operators are solutions of a Hamilton-Jacobi's inequation. This result allows us to extend a classical theorem from Bobkov, Gentil and Ledoux. More precisely, we prove that, in a graph, some weak transport inequalities are equivalent to the hypercontractivity of inf-convolution operators. Thanks to this result, we deduce some properties concerning different functional inequalities, including Log-Sobolev inequalities and weak-transport inequalities.

Besides, we study some general properties (differentiability, convexity, extreme points etc.) of different inf-convolution operators, including the one before, but also an operator related to a physical model (and to a large deviation phenomenon). We stay always in a graph.

Secondly, we interest in connections between different notions of discrete Ricci curvature on the graphs which are proposed by several authors in the recent years, and functional inequalities of type transport-entropy, or transport-information related to a Markov chain. We also obtain an extension of Bonnet-Myers' result: an upper bound on the diameter of a graph of which the curvature is floored in some ways.

Finally, restricting in the real line, we obtains a characterisation of a weak transport inequality and a log-Sobolev inequality restricted to convex functions. These results are from the geometrical properties related to the convex ordering.

key words: inf-convolution, Hamilton-Jacobi equation, weak transport inequality, Ricci curvature, discrete space, convex ordering

Contents

Chapter 1

Introduction Générale

Dans cette thèse, je me suis intéressé aux opérateurs d'inf-convolution et à leurs applications aux théories des inégalités fonctionnelles, du transport optimal et de la courbure de Ricci sur un espace discret. Dans l'introduction, je me placerai tout d'abord sur un espace continu pour présenter le problème classique du transport optimal et son problème dual, afin d'introduire les opérateurs d'inf-convolution (aussi appelés formules de Hopf-Lax), dans un contexte bien connu. Ces opérateurs sont des outils essentiels dans l'étude des inégalités fonctionelles auxquelles je me suis intéressé : inégalités de *Poincaré*, de *Sobolev logarithmique*, et inégalités dites de *transport*. Je présenterai des résultats dans deux cadres différents : la version classique où l'espace est continu, et une version où l'espace est discret. Je décrirai brièvement le rôle joué par les opérateurs d'inf-convolution dans l'étude de ces inégalités et leurs relations, leurs conséquences et leurs applications. Ensuite, j'introduirai la notion (classique en continu) de courbure de Ricci sur une variété riemannienne avant d'en présenter plusieurs définitions récentes sur les graphes. J'expliquerai en particulier comment une borne inférieure sur la courbure de Ricci entraîne certaines inégalités fonctionnelles avant de décrire les difficultés pour généraliser ces résultats aux espaces discrets. Enfin je discuterai du problème du transport et des inégalités fonctionnelles en dimension un.

Le matériel constituant cette thèse est issu des travaux suivants :

- *Hamilton-Jacobi equations on graphs and applications*, article disponible à l'adresse http ://arxiv.org/abs/1512.02416 ;
- *Curvature and transport inequalities for Markov chains in discrete spaces*, article écrit en collaboration avec Max Fathi et disponible à l'adresse http ://arxiv.org/abs/1509.07160 ;
- *Characterization of a class of weak transport-entropy inequalities on the line*, article écrit en collaboration avec Nathaël Gozlan, Cyril Roberto, Paul-Marie Samson et Prasad Tetali et disponible à l'adresse http ://arxiv.org/abs/1509.04202 ;
- *Characterization of modified log-Sob inequality*, preprint en fin d'écriture écrit en collaboration avec Michal Strzelecki ;

– *Large deviations for the invariant distribution of Markov chains with exponentially small jump rates*, in progress, en collaboration avec Alessandra Faggionato, David Gabrielli, Mauro Mariani.

1.1 Transport optimal

Dans cette section nous introduisons le notion de transport optimal, d'abord dans le cadre classique d'un espace continu, puis dans le cadre des espaces discrets avec la notion dite de transport faible.

1.1.1 Transport optimal classique

Après avoir défini le problème du transport optimal, nous introduirons quelques notions qui lui sont reliées : opérateur d'inf-convolution et équation d'Hamilton-Jacobi. Le lecteur peut par exemple consulter le livre de Cédric Villani [99] pour une introduction plus détaillée et plus complète de ce thème de recherche ainsi que pour une présentation exhaustive de la littérature. On commence par la définition du transport optimal :

Définition 1.1.1. *Sur un espace métrique $(\mathcal{X}, d)$, étant données deux mesures de probablité μ, ν et une fonction c de $\mathcal{X} \times \mathcal{X}$ dans $\mathbb{R}^+$, le problème du transport optimal consiste à minimiser*

$$\mathcal{T}_c(\mu, \nu) := \inf_{\pi} \iint_{\mathcal{X}} c(x, y) d\pi,$$

sur l'ensemble des couplages π de μ et ν.

On rappelle que π est un couplage de μ et ν si π est une mesure de probabilité sur $\mathcal{X} \times \mathcal{X}$ vérifiant $\pi(dx, \mathcal{X}) = \mu(dx)$ et $\pi(\mathcal{X}, dy) = \nu(dy)$ (i.e. π a pour première marginale μ et pour seconde marginale ν).

S'il existe un couplage, souvent noté π^*, qui réalise l'infimum dans la définition du transport optimal, on parle de couplage optimal.

Le problème du transport optimal a été introduit par Gaspard Monge au XVIIIème siècle pour modéliser le déplacement d'une masse d'un endroit à un autre de façon optimale. La quantité $c(x, y)$ représente alors le coût du déplacement d'une unité de masse du point x au point y, et pour cette raison on appelle c la fonction de coût. En termes probabilistes, le problème du transport optimal peut être reformulé comme suit

$$\mathcal{T}_c(\mu, \nu) = \inf_{X \sim \mu, Y \sim \nu} \mathbb{E}[c(X, Y)]$$

où (X, Y) est un couple de variables aléatoires, X de loi μ, Y de loi ν (on note $X \sim \mu$ et $Y \sim \nu$) et où $\mathbb{E}$ désigne l'espérance.

En particulier, dans une variété riemanienne (M, d), lorsque $c(x, y) := d(x, y)^p$ avec $p \geqslant 1$, on retrouve la distance bien connue de Wasserstein :

$$W_p := \mathcal{T}_c^{1/p}.$$

Si $\mathcal{P}(M)$ désigne l'ensemble des mesures de probabilités sur M, il est facile de voir que $(\mathcal{P}(M), W_p)$ est un espace métrique pour tout $p \geqslant 1$. Suite aux travaux fondateurs de Otto et Villani [87] et Cordero-Erausquin, McCann et Schmuckenschläger [24], les propriétés des géodésiques de l'espace $(\mathcal{P}(M), W_p)$ se sont avérées étroitement reliées aux propriétés géométriques de l'espace M lui-même, ce qui a finalement abouti à une définition alternative de la courbure de Ricci et à son extension à des espaces métriques plus généraux à travers la théorie de Sturm-Lott-Villani [69, 96], voir aussi [2] pour de récents développements.

Dans ce document, les fonctions de coût seront toujours de la forme $c(x,y) = \theta(d(x,y))$, avec $\theta\colon \mathbb{R} \to \mathbb{R}^+$ une fonction convexe. On notera $\mathcal{T}_\theta(\mu,\nu)$ au lieu de $\mathcal{T}_c(\mu,\nu)$. En outre, si $\theta(x) = |x|^p$, avec $p \geqslant 1$, on notera encore plus simplement $\mathcal{T}_p$.

Maintenant on introduit le théorème de la dualité donné par Kantorovich. Ce théorème relie le coût de transport à un opérateur appelé opérateur d'inf-convolution. On a l'égalité suivante :

$$\mathcal{T}_c(\mu,\nu) = \inf_{f \in \mathcal{C}_c^b} \int Qf\,d\mu - \int f\,d\nu. \tag{1.1.2}$$

Où $\mathcal{C}_c^b$ désigne l'ensemble des fonctions continues à support compact et

$$Q^c f(x) := \inf_{y \in \mathcal{X}} \{f(y) + c(x,y)\}$$

est l'opérateur d'inf-convolution. Si $c(x,y) := \frac{1}{2} d(x,y)^2$ la famille d'opérateurs

$$Q_t f(x) = \inf\{f(y) + \frac{1}{2t} d(x,y)^2\}, \tag{1.1.3}$$

indexée par $t \geqslant 0$ est un semi-groupe (dans le sens où $Q_t(Q_s) = Q_{t+s}$). On voit bien que pour $t = 1$ on retrouve $Q^c f$. L'équation (1.1.3) s'appelle aussi la formule de Hopf-Lax qui est connue pour être la solution de viscosité de l'équation d'Hamilton Jacobi suivante :

$$\begin{cases} \frac{\partial v(x,t)}{\partial t} + \frac{1}{2} |\nabla_x v(x,t)|_x^2 = 0 & (x,t) \in M \times (0,\infty) \\ v(x,0) = f(x) & x \in M, \end{cases}$$

où $|\cdot|_x$ désigne la norme sur l'espace tangent $T_x M$ correspondant à la métrique g au point x (voir par exemple [33]). L'équation d'Hamilton-Jacobi joue un rôle très important dans la domaine des équations aux dérivées partielles. Elle permet d'appliquer les outils de calcul différentiel aux problèmes de transport cf. [99], on va en discuter à travers tout ce document. Elle est également reliée à la théorie des systèmes dynamiques, à la théorie d'Aubry-Mather etc.

1.1.2 Transport optimal faible

Motivés par certaines formes du transport introduites par Marton [76, 75], Gozlan, Roberto, Samson et Tetali [50] introduisent une définition générale de ce qu'ils appellent le transport faible que nous allons maintenant rappeler. Au lieu de considérer les fonctions de coût de la forme

$$c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+,$$

ces auteurs introduisent les fonctions de coût de la forme

$$c : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}^+.$$

Ainsi, le transport faible, que nous allons encore noter $\mathcal{T}_\theta$, est défini par :

$$\mathcal{T}_\theta(\nu|\mu) := \inf_\pi \left\{ \int c(x, p_x)\mu(dx) \right\},$$

où π est un couplage de μ et ν et p_x est le noyau conditionnel de π par rapport à μ, i.e. $\pi(dxdy) = \mu(dx)p_x(dy)$. Le terme "faible" provient du fait que $\mathcal{T}_\theta(\nu|\mu)$ est toujours plus petit que $\mathcal{T}_\theta(\mu, \nu)$.

Sous certaines hypothèses portant sur l'espace et le coût c, ces auteurs généralisent le théorème de dualité de Kantorovich [50]. Selon le choix de c, on peut définir différents types de transport faibles et même retrouver le transport optimal classique (choisir $c(x, y) = \int \theta(d(x, y)p_x(dy))$. Dans ce document, on s'intéresse plus spécifiquement aux deux cas suivants :

(i) en prenant $c(x, p) = \theta(\int d(x, y)p(dy))$, on retrouve le transport de Marton [76, 75]

$$\widetilde{\mathcal{T}}_\theta(\nu|\mu) := \inf_\pi \left\{ \int \theta \left(\int d(x, y))p_x(dy) \right) \mu(dx) \right\}$$

qui, en terme probabiliste, s'écrit également

$$\widetilde{\mathcal{T}}_\theta(\nu|\mu) := \inf_{X\sim\mu, Y\sim\nu} \mathbb{E}\left[\theta(\mathbb{E}[d(X, Y)|X])\right]$$

où $\mathbb{E}(\cdot|X)$ désigne l'espérance conditionnelle sachant X. Si θ est convexe, la formule de dualité (1.1.2) devient [50]

$$\widetilde{\mathcal{T}}_\theta(\nu|\mu) = \inf_{f\in\mathcal{C}_c^b} \int \widetilde{Q}f d\mu - \int f d\nu$$

avec

$$\widetilde{Q}f(x) := \inf_{p\in\mathcal{P}(\mathcal{X})} \left\{ \int f(y)p(dy) + \theta \left(\int d(x, y)p(dy) \right) \right\}.$$

(ii) Sur un espace vectoriel normé, en prenant $c(x, p) = \theta\left(\|x - \int yp(dy)\|\right)$, Gozlan et al. définissent [50] le transport faible suivant

$$\overline{\mathcal{T}}_\theta(\nu|\mu) := \inf_\pi \left\{ \int \theta \left(\|x - \int yp_x(dy)\| \right) \mu(dx) \right\},$$

(où π décrit l'ensemble des couplages de μ et ν), qui s'écrit encore

$$\overline{\mathcal{T}}_\theta(\nu|\mu) := \inf_{X\sim\mu, Y\sim\nu} \mathbb{E}\left[\theta(\|X - \mathbb{E}(Y|X)\|)\right].$$

Si θ est convexe, la formule de dualité (1.1.2) prend la forme [50]

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf_{f\in\mathcal{C}_c^b} \int \overline{Q} f d\mu - \int f d\nu$$

avec

$$\overline{Q} f(x) := \inf_{p\in\mathcal{P}(\mathcal{X})} \left\{ \int f(y)p(dy) + \theta\left(\|x - \int y p_x(dy)\| \right) \right\}.$$

1.2 Inégalités fonctionnelles

Cette section est dédiée à l'introduction de plusieurs inégalités fonctionnelles, dont celles bien connues de Sobolev logarithmique et de transport. Comme pour la section précédente, nous commencerons par présenter ces inégalités dans le cadre classique d'un espace continu avant d'en donner certaines formes valables dans les espaces discrets. Le lecteur intéressé pourra consulter par exemple [5] pour une introduction plus complète.

1.2.1 Espace continu

Pour simplifier on se place dans le cadre de $\mathbb{R}^n$ muni de la norme euclidienne, notée $|\cdot|$, mais l'essentiel des notions présentées ici restent valables dans un contexte bien plus général.

Définition 1.2.1. *On dit qu'une mesure de probabilité μ sur $\mathbb{R}^n$ vérifie l'inégalité de Sobolev logarithmique, s'il existe une constante C, (on notera alors* LS(C) *ou* LS *dans le cas où on ne précise pas la dépendance par raport à la constante C) telle que pour toute fonction $f : \mathbb{R}^n \to \mathbb{R}$, assez régulière, l'inégalité suivante est satisfaite :*

$$\mathrm{Ent}_\mu(f^2) := \int f^2 \log\left(\frac{f^2}{\int f^2 d\mu} \right) d\mu \leqslant C \int |\nabla f|^2 d\mu.$$

Dans l'expression précédente, $\mathrm{Ent}_\mu(f^2)$ est l'entropie de f^2, calculée par rapport à la mesure de probabilité μ, et $|\nabla f|^2 = \sum_{i=1}^n \partial_i f^2$. Cette inégalité est introduite par Gross [52] en 1975. L'inégalité de Sobolev logarithmique peut être reformulée en utilisant la terminologie de la théorie de l'information. Si $\int f^2 d\mu = 1$, $\nu := f^2\mu$ est une mesure de probabilité de densité f^2 par rapport à μ. L'inégalité de Sobolev logarithmique s'écrit alors

$$H(\nu|\mu) \leqslant C\mathcal{I}(\nu|\mu),$$

pour toute mesure ν absolument continu par rapport à μ, où $H(\nu|\mu) := \mathrm{Ent}_\mu(f^2)$ est l'entropie relative de ν par rapport à μ et $\mathcal{I}(\nu|\mu) := \int |\nabla f|^2 d\mu$ est l'information de Fisher.

Par un changement de fonction $f \to e^f$, et en utilisant la formule de dérivation des fonctions composées, on obtient une autre formulation encore. Plus précisément, l'inégalité de Sobolev logarithmique LS(C) est équivalente à

$$\text{(mLS)} : \quad \mathrm{Ent}_\mu(e^f) \leqslant \frac{C}{4} \int |\nabla f|^2 e^f d\mu.$$

Cette dernière inégalité va jouer un rôle particulier dans le cas où l'espace n'est pas continu. Elle est souvent appelée inégalité de Sobolev logarithmique modifiée (on écrira mLS). Notons dès à présent que dans un cadre discret l'inégalité de Sobolev logarithmique modifiée est (strictement) plus faible que l'inégalité de Sobolev logarithmique.

Si on restreint mLS à la classe des fonctions convexes, alors on dit que μ satisfait à une inégalité de Sobolev logarithmique modifiée convexe et on notera CmLS.

L'une des conséquences classiques de l'inégalité LS est la concentration Gaussienne suivante :

Théorème 1.2.2. *Si μ vérifie* LS(C)*, alors pour toute fonction régulière $f : \mathcal{X} \to \mathbb{R}$ vérifiant $\sup_x |\nabla f|(x) \leqslant 1$ et tout $t > 0$, on a*

$$\mu(f - \int f \geqslant t) \leqslant e^{-\frac{1}{C}t^2}.$$

Pour prouver ce résultat on peut utiliser l'argument de Herbst qui consiste à appliquer l'inégalité LS(C) à la fonction $\exp\{\lambda f/2\}$ et à étudier le comportement de la transformée de Laplace $F(\lambda) := \int \exp\{\lambda f\} d\mu$, voir par exemple [64, 5, 7]. Il existe des extensions à des cadres plus généraux, voir notamment [81, 51].

Parmi les propriétés fondamentales satisfaites par l'inégalité de Sobolev logarithmique, notons en particulier la tensorisation : si μ vérifie LS(C) sur $\mathbb{R}^n$, alors le produit $\mu^{\otimes m}$ vérifie également LS(C) sur $(\mathbb{R}^n)^m$ pour tout $m \in \mathbb{Z}^+$. Par conséquent, $\mu^{\otimes m}$ vérifie la concentration Gaussienne ci-dessus pour tout m. On dit alors que μ vérifie une propriété de concentration Gaussienne adimensionelle.

La deuxième inégalité fonctionnelle qui va nous intéresser est l'inégalité de transport qui compare le transport optimal à l'entropie relative.

Définition 1.2.3. *Etant donnée une fonction de coût $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$, on dit qu'une mesure de probabilité μ vérifie l'inégalité transport-entropie* $\mathrm{T_c}(C)$ *s'il existe une constante C, telle que pour toute mesure de probabilité ν, l'inégalité suivante est satisfaite :*

$$\mathcal{T}_c(\mu, \nu) \leqslant C H(\nu|\mu).$$

Cette inégalité a été introduite par Talagrand pour la mesure gaussienne, avec $c(x,y) = |x-y|^2$, où il prouve que $C = 2$ est la constante optimale [98]. Pour cette raison l'inégalité de transport-entropie introduite ci-dessus est parfois aussi appelée inégalité de Talagrand. On la notera $T_2(C)$ (ou plus simplement T_2). Nous renvoyons à [99] pour une présentation complète de cette inégalité. Donnons néanmoins une caractérisation très utile, due à Bobkov et Götze [13] (pour l'obtenir, il suffit de comparer la formule duale de $\mathcal{T}_c$ (i.e. le théorème de Kantorovich) et la formule duale de l'entropie $\mathrm{Ent}_\mu(f) = \sup_{\int e^g = 1} \int f g d\mu$).

Théorème 1.2.4 (Caractérisation de Bobkov et Götze)**.** *Les assertions suivantes sont équivalentes :*
(i) *μ vérifie* $\mathrm{T_c}(C)$.
(ii) *Pour toute fonction $f : \mathbb{R}^n \to \mathbb{R}$ continue bornée, on a*

$$\int \exp\left\{\frac{Q^c f}{C}\right\} d\mu \leqslant \exp\left\{\frac{1}{C}\int f d\mu\right\}.$$

On peut montrer que l'inégalité de Talagrand (pour $c(x,y) = |x-y|^2$) entraîne le même phénomène de concentration gaussien que l'inégalité de log-Sobolev. En outre, une telle inégalité tensorise entraînant ainsi une concentration Gaussienne adimensionelle.

Ces deux inégalités (log-Sobolev et de Talagrand) donnant le même phénomène de concentration, une question naturelle est de savoir si ces inégalités sont comparables. Une réponse est apportée par Otto et Villani dans [87] :

Théorème 1.2.5. *Si μ vérifie l'inégalité de Sobolev logarithimique, alors μ vérifie l'inégalité de Talagrand avec la même constante.*

Il y a plusieurs preuves de ce résultat (c.f.[87, 11, 42, 40, 45, 46, 47]). La preuve de [42] est la plus générale, valable dans n'importe quel espace métrique. Elle consiste à montrer que T_2 est en fait équivalente à la concentration Gaussienne adimensionnelle. Par conséquent, toutes les inégalités impliquant la concentration gaussienne adimensionelle devraient entraîner T_2. La preuve de Bobkov, Gentil et Ledoux [11] va particulièrement nous intéresser car elle fait pour la première fois le lien entre l'inégalité de transport-entropie et l'opérateur d'inf-convolution, via l'équation d'Hamilton-Jacobi.

Théorème 1.2.6 (Hypercontractivité [11])**.** *Supposons que μ est une mesure de probabilité absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$. Alors si*
(i) *il existe une constante $\rho > 0$ telle que μ vérifie* $\mathrm{LS}(2/\rho)$,
alors
(ii) *pour tout $a > 0$ et tout $t \geqslant 0$,*

$$\|e^{Q_t f}\|_{a+\rho t} \leqslant \|e^f\|_a.$$

Inversement, si (ii) *est valable pour un certain* $a > 0$ *et pour tout* $t \geqslant 0$*, alors* μ *vérifie* (i).

La dernière famille d'inégalités qui va nous intéresser a été introduite en 2009, dans [54], par Guillin, Léonard, Wu et Yao. Ces inégalités, que nous introduisons dans la définition qui suit, comparent le transport optimal à l'information de Fisher, et s'appellent pour cela les inégalités de transport-information.

Définition 1.2.7. *On dit qu'une mesure de probabilité* μ *vérifie l'inégalité transport-information* $\mathrm{T_cI}(C)$*, si pour toute mesure de probabilité* ν*, absolument continue par rapport à* μ*,*

$$\mathcal{T}_c(\mu,\nu) \leqslant C\mathcal{I}(\nu|\mu).$$

Si la fonction de coût est quadratique, on notera $\mathrm{T_2I}$ pour l'inégalité de transport-information. Guillin et *al.* [53] ont montré que $\mathrm{T_2I}$ est entraînée par l'inégalité de log-Sobolev et entraîne celle de Talagrand.

Signalons enfin que Gozlan et *al.* montrent que les inégalités CmLS sont équivalentes à une inégalité de transport faible dans [50].

1.2.2 Inégalités de transport-entropie faibles

Plaçons-nous dans le cadre d'un espace polonais $\mathcal{X}$ dont les boules sont compactes. Il est facile de voir que l'inégalité de Talagrand T_2 n'est pas valable pour des mesures à support non connexe. Ainsi une telle inégalité n'est valable, dans un graphe, que si μ est une mesure Dirac. Gozlan, Roberto, Samson et Tetali [50] ont proposé de s'intéresser aux inégalités de transport-entropie faibles, où l'entropie est comparée au transport faible plutôt qu'au transport classique. Plus précisément, et plus spécifiquement, on s'intéressera aux inégalités suivantes.

Définition 1.2.8. *Soit* μ *une mesure de probabilité sur* $\mathcal{X}$ *(avec* $\mathcal{X} = \mathbb{R}^n$ *pour les trois dernières inégalités). On dit que* μ *vérifie l'inégalité*

- $\widetilde{\mathrm{T}}^+_\theta(C)$ *si pour toute probabilité* ν*,*
$$\widetilde{\mathcal{T}}_\theta(\nu|\mu) \leqslant CH(\nu|\mu);$$
- $\widetilde{\mathrm{T}}^-_\theta(C)$ *si pour toute probabilité* ν*,*
$$\widetilde{\mathcal{T}}_\theta(\mu|\nu) \leqslant CH(\nu|\mu);$$
- $\widetilde{\mathrm{T}}_\theta(C)$ *si* μ *vérifie* $\widetilde{\mathrm{T}}^-_\theta(C)$ *et* $\widetilde{\mathrm{T}}^+_\theta(C)$.
- $\overline{\mathrm{T}}^+_\theta(\mathrm{C})$ *si pour toute probabilité* ν*,*
$$\overline{\mathcal{T}}_\theta(\nu|\mu) \leqslant CH(\nu|\mu);$$
- $\overline{\mathrm{T}}^-_\theta(C)$ *si pour toute probabilité* ν*,*
$$\overline{\mathcal{T}}_\theta(\mu|\nu) \leqslant CH(\nu|\mu);$$

– $\overline{\mathrm{T}}_\theta(C)$ *si* μ *vérifie* $\overline{\mathrm{T}}_\theta^-(C)$ *et* $\overline{\mathrm{T}}_\theta^+(C)$.

Ces inégalités de transport faibles peuvent être définies dans un cadre très général.

Dans $\mathbb{R}^n$, en utilisant le transport faible $\overline{\mathcal{T}}_\theta$, Gozlan, Roberto, Samson et Tetali retrouvent l'hypercontractivité pour les inégalités de Sobolev logarithimiques modifiées restreintes à la classe des fonctions convexes. Ceci est liée à la théorie des ordres convexes que nous présentons à présent.

Ordre convexe et Théorème de Strassen

Gozlan et *al.* montrent que le transport faible $\overline{\mathcal{T}}$ est lié aux anciens résulats de Strassen [95] concernant l'existence de martingales ayant des marginales données.

Définition 1.2.9 (ordre convexe)**.** *Soient* μ, ν *deux mesures de probabilité sur* $\mathbb{R}^m$*, on dit que* μ *est majorée par* ν *pour l'ordre convexe, et noté* $\mu \preceq \nu$ *si*

$$\int f d\mu \leqslant \int f d\nu$$

pour tout $f : \mathbb{R}^m \to \mathbb{R}$ *convexe.*

Théorème 1.2.10 (Strassen)**.** *Soient* $\mu, \nu \in \mathcal{P}(\mathbb{R}^m)$*,* $\mu \preceq \nu$ *si et seulement s'il existe une martingale* (X, Y) *telle que* X *suit la loi* μ *et* Y *suit la loi* ν*.*

Par définition de $\overline{\mathcal{T}}$, pour un coût convexe positif θ, il est facile de voir que $\mu \preceq \nu$ est équivalente à $\overline{\mathcal{T}}_\theta(\mu|\nu) = 0$.

Lorsqu'on regarde l'inégalité de transport faible $\overline{\mathcal{T}}$, $\overline{\mathcal{T}}(\nu|\mu) = 0$ est équivalente à l'existence de $X \sim \mu$ et $Y \sim \nu$ telle que (X, Y) est un couple de martingale.

1.2.3 Caractérisation des inégalités fonctionnelles sur $\mathbb{R}$

Lorsqu'on se restreint à la dimension un (sur $\mathbb{R}$), il est parfois possible de caractériser complètement certaines inégalités fonctionnelles. Dans les travaux de Muckenhoupt [82], concernant l'inégalité de Hardy, est établie une caractérisation de l'inégalité de Poincaré. On en présente ici une version simplifiée.

Théorème 1.2.11 (Muckenhoupt72)**.** *Soient deux mesures de Borel* μ *et* ν *sur* $\mathbb{R}^+$ *et* A *la meilleure constante telle que pour toute fonction* f *mesurable positive,*

$$\int_0^\infty \left(\int_0^x f(t) dt \right)^2 d\mu(x) \leqslant A \int_0^\infty f^2(x) d\nu(x), \tag{1.2.12}$$

alors A *est finie si et seulement si la constante*

$$B = \sup_{x \geqslant 0} \mu([x, \infty[) \left(\int_0^\infty \frac{1}{\nu'(t)} dt \right)$$

est finie. Et dans ce cas, on a $A \leqslant B \leqslant 4B$*. Ici* ν' *désigne la densité de la partie absolument continue de* ν *par rapport à la mesure de Lebesgue.*

L'inégalité (1.2.12) est une forme particulière de l'inégalité de Hardy. La constante B, ou toute autre constante semblable, est souvent appelée constante de Hardy, et l'encadrement du type $A \leqslant B \leqslant 4B$ un encadrement de type Hardy. L'inégalité de Hardy se trouve dans les travaux de Hardy dans les années 1920. Miclo [79] généralise cette caractérisation au cadre discret, sur $\mathbb{N}$, en 1999.

Bobkov et Gotze [13] sont les premiers à utiliser les inégalités de Hardy pour caractériser sur $\mathbb{R}$ les mesures satisfaisant l'inégalité de Sobolev logarithmique. Pour simplifier, on présente ce résultat en supposant que μ est une mesure de probabilité de médiane 0, absolument continue par rapport à la mesure de Lebesgue et symétrique par rapport à 0. Alors μ vérifie LS(C) si et seulement si la constante de Hardy :

$$D := \sup_{x>0} \left\{ \mu([x, \infty[) \log \frac{1}{\mu([x, \infty[)} \int_0^x \frac{1}{\mu'(t)} dt \right\}$$

est finie et on a $\frac{1}{150} D \leqslant C \leqslant 2880D$. Miclo [79] améliore ce résultat dans le cadre des espaces discrets et obtient $\frac{1}{31} D \leqslant C \leqslant 20D$. Voir aussi [36] pour une preuve simplifiée dans le cas de la droite réelle. Enfin, signalons que Roberto [89] généralise ces résultats à certaines inégalités à poids.

Grâce à la caractérisation de Bobkov et Götze (théorème 1.2.4), Gozlan propose une caractérisation de l'inégalité de Talagrand. Pour simplifier, on ne précise pas les constantes des inégalités.

Théorème 1.2.13. *Soit μ est une mesure de probabilité sur $\mathbb{R}$; on note τ la mesure exponentielle de densité $\frac{1}{2}\exp(-|x|)$. Alors μ satisfait l'inégalité de Talagrand T_2 si et seulement si*

- *μ vérifie l'inégalité de Poincaré.*
- *l'application de transport U envoyant τ sur μ (i.e $U\sharp\tau = \mu$) vérifie la propriété de contraction suivante :*

$$\exists a, b > 0, \forall x, h \in \mathbb{R},\ |U(x+h) - U(x)| \leqslant \sqrt{a + b|h|}.$$

Dans notre thèse, nous donnerons des caractérisations d'autres inégalités sur $\mathbb{R}$, notamment certaines inégalités de transport faibles avec différentes fonctions de coût, les inégalités de Sobolev logarithmiques modifiées restreintes aux fonctions convexes etc.

1.2.4 Cadre d'un graphe

Lorsque l'espace est discret, afin de définir l'inégalité de Sobolev logarithmique et les inégalités de transport-information, il faut trouver un remplaçant de l'information de Fisher. Sur les graphes, l'inégalité de Sobolev logarithmique est bien étudiée dans la littérature (voir [41, 17]). On rappelle quelques définitions.

Définition 1.2.14 (Le cadre d'un graphe)**.** *Considérons un ensemble $\mathcal{X}$ fini ou infini dénombrable, K est un noyau de Markov sur $\mathcal{X}$ si pour tout $x, y \in \mathcal{X}$, $K(x,y) \geqslant 0$ et*

$$\sum_{y \in \mathcal{X}} K(x,y) = 1.$$

On définit le graphe avec l'ensemble des sommets $\mathcal{X}$, et l'ensemble des arêtes $\{(x,y) | K(x,y) > 0, x \neq y\}$. Ainsi, l'espace qu'on considère est $\mathcal{X}$, muni de la distance de graphe

$$d(x,y) := \inf\{n \in \mathbb{N}; \exists x_0, .., x_n | x_0 = x,\ x_n = y,\ K(x_i, x_{i+1}) > 0\ \forall 0 \leqslant i \leqslant n-1\}.$$

On suppose qu'il existe une mesure réversible π associée au noyau K, i.e. on a

$$K(x,y)\pi(x) = K(y,x)\pi(y) \quad \forall x, y \in \mathcal{X}.$$

On note L le générateur Markovien,

$$Lf(x) = \sum_y (f(y) - f(x)) K(x,y).$$

On note $P_t = e^{tL}$ le semigroupe associé à L, et on définit l'opérateur Γ, le carré du champ, donné par

$$\Gamma(f,g)(x) := \frac{1}{2} \sum_y (f(y) - f(x))(g(y) - g(x)) K(x,y)$$

et on note $\Gamma(f) := \Gamma(f,f)$.

L'opérateur L est une analogie du Laplacien Δ, et par suite le semigroupe P_t est une analogie du semigroupe de la chaleur. On peut se référer à [7] pour les propriétés de carré du champ Γ, de L et pour différentes applications.

Puisque la règle de dérivation des fonctions composées n'est plus valable en discret, il y a plusieurs façons de définir l'information de Fisher. Par exemple on peut poser

$$\mathcal{I}(\nu|\mu) = \mathcal{I}_\mu(f) := 4 \int \Gamma\left(\sqrt{f}, \sqrt{f}\right) d\mu,$$

ou

$$\bar{\mathcal{I}}(\nu|\mu) = \bar{\mathcal{I}}_\mu(f) := \int \frac{\Gamma(f)}{f} d\mu,$$

où f est la densité de ν par rapport à μ. Dans ce cadre, les relations entre les inégalités fonctionnelles ne sont pas claires, on les développera dans la suite.

1.3 Courbure de Ricci en discret

Nous présentons rapidement dans cette section la notion de courbure de Ricci, sur un espace continu puis sur un espace discret où plusieurs définitions sont possibles.

La courbure de Ricci joue un rôle très important en géométrie Riemannienne. Le lecteur peut se référer à [2, 4, 23, 99] pour une introduction complète. Dans une variété Riemannienne, il y a beaucoup de définitions de la courbure de Ricci bornée inférieurement par un réel positif. On peut se référer à [97] pour un synthèse de toutes ces définitions et la preuve de leur équivalence. À cause de la perte de la règle de dérivation des fonctions composées, ces définitions ne sont plus équivalentes sur un graphe. Dans ce document, je me concentrerai sur trois définitions : la condition de courbure de Bakry et Émery $\mathrm{CD}(\kappa, \infty)$, la courbure exponentielle CDE'(κ_e, ∞) et la courbure d'Ollivier κ_c. Dans la suite j'introduirai ces trois notions sur un graphe et sur une variété riemannienne en même temps : lorsque $\mathcal{X}$ est un graphe, L et Γ se définissent comme dans la définition 1.2.14, lorsque $\mathcal{X}$ est une variété Riemannienne, L sera le laplacien Δ. Je commence par la condition de Bakry et Émery [6].

Définition 1.3.1. *On définit l'opérateur* Γ_2 *comme*

$$\Gamma_2(f) = \frac{1}{2} L\Gamma(f) - \Gamma(f, Lf).$$

On dit que la condition de courbure $\mathrm{CD}(\kappa, \infty)$ *est vérifiée si et seulement si pour toute* f*, on a*

$$\Gamma_2(f) \geqslant \kappa \Gamma(f).$$

Dans une variété Riemannenne, $\mathrm{CD}(\kappa, \infty)$ implique l'ingalité de Sobolev logarithmique, et donc l'inégalité de transport-information et l'inégalité de Talagrand, elle implique aussi inégalité de log-Harnack et l'inégalité de Buser, etc. Dans le cadre d'une variété Riemannienne, Γ_2 associe aussi le tenseur de Ricci, et la condition $\mathrm{CD}(\kappa, \infty)$ est équivalente à dire que la borne inférieure de courbure de Ricci est κ. Dans le cadre d'un graphe, cette condition de courbure a été introduite et étudiée en premier par Schmuckenshlager dans [92], puis étendue par S-T. Yau et ses collaborateurs dans [59]. La condition $\mathrm{CD}(\kappa, \infty)$ sert aussi à prouver une version discrète de l'inégalité de Buser [63].

Une autre adaptation en discret est proposée par Bauer et *al.* dans [9]. Ils considèrent un opérateur Γ_2 modifié au lieu du Γ_2 habituel :

Définition 1.3.2. *On définit l'opérateur* $\widetilde{\Gamma}_2$ *par*

$$\widetilde{\Gamma}_2(f) := \Gamma_2(f) - \Gamma\left(f, \frac{\Gamma(f)}{f}\right).$$

On dit que la condition de courbure exponentielle $\mathrm{CDE}'(\kappa_e, \infty)$ *est vérifiée si et seulement si pour toute fonction* f *positive, on a*

$$\widetilde{\Gamma}_2(f) \geqslant \kappa_e \Gamma(f).$$

Dans le cadre d'un graphe, le terme $\widetilde{\Gamma}_2$ permet de retrouver certaines règles de dérivation. Comme application Bauer et *al.* prouvent une version discrète de l'inégalité de log-Hanack. Dans une variété Riemannienne, ces deux notions de courbures sont équivalentes.

Une autre formulation équivalente de $\mathrm{CD}(\kappa, \infty)$ est la décroissance exponentielle de distance de Wasserstein

$$W_1(P_t^*\mu, P_t^*\nu) \leqslant \exp(-\kappa t) W_1(\mu, \nu),$$

où P_t^* est l'opérateur adjoint de P_t. Cette notion est introduite par Ollivier sur les graphes dans [85].

1.4 Présentation des chapitres

Dans cette section nous présentons très brièvement les résultats de notre thèse, regroupés en chapitre.

Résumé du Chapitre II

Dans ce chapitre, on note $\mathcal{X}, S$ deux espaces et $\mathcal{F}(\mathcal{X}), \mathcal{F}(S)$ les ensembles des fonctions de $\mathcal{X}, S$ à valeur dans $\mathbb{R}$. On va s'intéresser aux opérateurs de la forme

$$Rf(x) = \inf_{s \in S} \{a(f)(s) + b(x, s)\},$$

où $x \in \mathcal{X}$, $a : \mathcal{F}(\mathcal{X}) \to \mathcal{F}(S)$ et $R : \mathcal{F}(\mathcal{X}) \to \mathcal{F}(\mathcal{X})$.

Dans ce chapitre, on rappellera d'abord certains résultats classiques de la formule de Hopf-Lax suivante, avec $\alpha \colon \mathcal{X} \to \mathbb{R}^+$ (on suppose $(\mathcal{X}, d)$ métrique)

$$Q_t f(x) := \inf_{y \in \mathcal{X}} \left\{ f(y) + t\alpha\left(\frac{d(x,y)}{t}\right) \right\}.$$

qui est une simple généralisation de (1.1.3). Dans une variété Riemannienne, $Q_t f$ vérifie la propriété de semi-groupe et les équations d'Hamilton-Jacobi. De plus, elle est reliée à la théorie du transport optimal par la formule de dualité (1.1.2).

Ensuite on se placera dans un espace $\mathcal{X}$ polonais dont les boules sont compactes. On donnera des résultats concernant l'opérateur $\widetilde{Q}_t$ défini par

$$\widetilde{Q}_t f(x) = \inf_{p \in \mathcal{P}(\mathcal{X})} \left\{ \int f(y) p(dy) + t\theta\left(\frac{\int d(x,y) p(dy)}{t}\right) \right\}.$$

Les propriétés de $\widetilde{Q}_t f$ permettent de comprendre le transport faible $\widetilde{\mathcal{T}}$ introduit dans section 1.1.2. Dans la section 2 du chapitre II, on étudiera les propriétés de régularité de $\widetilde{Q}_t f$, l'existence de minimum, la convexité de cette fonction. Dans la section 3, on établira une équation d'Hamilton-Jacobi pour $\widetilde{Q}_t f$ de la forme

$$\frac{\partial}{\partial t}\widetilde{Q}_t f(x) + \theta^* \left(|\widetilde{\nabla} \widetilde{Q}_t f|(x) \right) \leqslant 0,$$

avec

$$|\widetilde{\nabla} f| = \sup_y \frac{[f(x) - f(y)]_-}{d(x,y)}. \tag{1.4.1}$$

On prouvera l'existence de la dérivation par rapport à t à gauche et à droite, et on en donnera une formule explicite. Tous ces résultats sont issus de [93] et seront utilisés dans l'étude des inégalités de transport $\widetilde{\mathcal{T}}$ du chapitre III et IV.

A la fin du chapitre, on regardera le cas d'équilibre. En prenant un graphe fini orienté $G(V,E)$, on se concentrera sur l'équation $\omega(x) = \inf_{y:(y,x)\in E}\{\omega(y) + k(y,x)\}$. Cette équation est motivée par un modèle issu de la théorie des grandes déviations. Lorsque G est un graphe fini, les solutions forment un polycône convexe de $\mathbb{R}^{|V|}$. On caractérisera en particulier les solutions extrêmes (les points extrêmaux de ce polycône convexe). Ce travail est issu de [34].

Résumé du Chapitre III

Dans ce chapitre, l'espace est encore polonais avec les boules compactes. Le but est d'étudier certaines inégalités fonctionnelles sur les espaces discrets. On se concentrera sur les inégalités de transport faible (c.f. Definition 1.2.8) impliquant $\widetilde{\mathcal{T}}$. En prenant $\widetilde{\nabla}$ (voir (1.4.1)) comme un gradient modifié, on introduira les inégalités de Sobolev modifiée et de Poincaré modifiées suivantes :

$$\mathrm{Ent}_\mu(e^f) \leqslant C \int |\widetilde{\nabla} f|^2 e^f d\mu,$$

$$\mathrm{Var}_\mu(f) \leqslant C \int |\widetilde{\nabla} f|^2 d\mu.$$

En premier lieu, on reliera ces inégalités aux inégalités usuelles connues sur les graphes. Ensuite, en utilisant certains outils du chapitre II et en suivant les idées de [11], on étudiera une version discrète des inégalités de Sobolev-logarithmique et leurs propriétés d'hypercontractivité. En particulier on obtient le théorème suivant :

Théorème 1.4.2. *Soient μ une mesure de probabilité sur $\mathcal{X}$ et $C > 0$.*

(i) *Si pour toute fonction $f\colon \mathcal{X} \to \mathbb{R}$ mesurable bornée,*

$$\mathrm{Ent}_\mu(e^f) \leqslant C \int |\widetilde{\nabla} f|^2 e^f \, d\mu, \tag{1.4.3}$$

Alors pour tout $\rho \geqslant 0$, tout $t \geqslant 0$ et toute f mesurable bornée,

$$\|e^{\widetilde{Q}_t f}\|_{\rho + \frac{2t}{C}} \leqslant \|e^f\|_\rho. \tag{1.4.4}$$

Inversement, si on a (1.4.4) pour certain $\rho > 0$ et pour tout $t \geqslant 0$, alors on a (1.4.3).

(*ii*) *Si pour toute fonction* $f\colon X \to \mathbb{R}$ *mesurable bornée,*

$$\mathrm{Ent}_\mu(e^f) \leqslant C \int |\widetilde{\nabla}(-f)|^2 e^f \, d\mu, \tag{1.4.5}$$

alors on a (1.4.4) *pour tout* $\rho \leqslant 0$, *tout* $t \in [0, -\rho C/2]$ *et toute fonction* f *mesurable bornée. Inversement, si on a (1.4.4) pour certain* $\rho < 0$ *et pour tout* $t \in [0, -\rho C/2)$, *alors on a* (1.4.5).

Ce théorème permet d'étabilir une version discrète de la chaîne d'implication :

$$\mathrm{LS} \Rightarrow \mathrm{T}_2 \Rightarrow \text{Poincaré}.$$

Cette partie est issue de [93].

Résumé du Chapitre IV

Dans ce chapitre, on se placera dans le cadre d'un graphe associé à une chaîne de Markov qui est définie dans la section 1.2.4. Le but sera d'obtenir des inégalités fonctionnelles dans un graphe sous les conditions de courbure de Ricci. On s'intéressera aux trois différentes courbures discrètes : la courbure de Bakry et Émery $\mathrm{CD}(\kappa, \infty)$, la courbure exponentielle CDE'(κ_e, ∞) et la courbure d'Ollivier κ_c. Avec Max Fathi, sous l'hypothèse de $\mathrm{CD}(\kappa, \infty)$, en prenant les même notation de la section 1.2.4, nous obtenons une inégalité de transport information

$$W_1(f\pi, \pi)^2 \leqslant \frac{J}{\kappa} \mathcal{I}_\pi(f)$$

et une estimation de diamètre du graphe (théorème du type Bonnet-Myers) :

$$d(x, y)\kappa \leqslant 2\left(J(x) + J(y)\right).$$

Sous l'hypothèse de CDE'(κ_e, ∞), nous obtenons

$$\widetilde{W}_2(f\pi|\pi)^2 \leqslant \frac{2J}{\kappa_e^2} \mathcal{I}_\pi(f) \leqslant \frac{2}{\kappa_e^2} \mathcal{I}_\pi(f).$$

Finalement sous l'hypothèse de la courbure d'Ollivier avec constante κ_c, nous obtenons

$$W_1(f\pi, \pi)^2 \leqslant \frac{1}{\kappa_c^2} \mathcal{I}_\pi(f) \left(J - \frac{1}{8}\mathcal{I}_\pi(f)\right) \leqslant \frac{1}{\kappa_c^2} \mathcal{I}_\pi(f).$$

Cette partie est issue de [35], une collaboration avec Max Fathi.

Résumé du Chapitre V

Dans ce chapitre, nous nous sommes concentrés sur la droite réelle $\mathbb{R}$ et le transport faible $\overline{\mathcal{T}}$. L'objectif est de trouver les conditions nécessaires et suffisantes pour qu'une mesure μ vérifie l'inégalité de transport-entropie faible $\overline{\mathcal{T}}_\theta$ pour un coût θ convexe.

Nos résultats principaux des sept premières sections sont

– En dimension un, lorsque le coût est convexe, le plan optimal de transport faible $\overline{\mathcal{T}}_\theta$ ne dépend pas de θ.

– Si θ est une fonction quadratique-linéaire, l'inégalité de Poincaré restreinte aux fonctions convexes, l'inégalité de transport faible $\overline{\mathrm{T}}_\theta^+$ et l'inégalité de transport faible $\overline{\mathrm{T}}_\theta^-$ sont équivalentes.

– Si θ se comporte comme x^2 au voisinage de 0, alors l'inégalité $\overline{\mathrm{T}}_\theta$ est équivalente à une propriété de contraction :

$$\exists a, b > 0, \forall h > 0, \quad \sup_x \{U(x+h) - U(x)\} \leqslant a\theta^{-1}(h+b).$$

où $U := F_\mu^{-1} \circ F_\tau$ avec F la fonction de distribution et τ la distribution exponentielle de densité $\tau(dx) = \frac{1}{2}\exp(-|x|)dx$.

Ces résultats sont issus de [49] en collaboration avec Nathaël Gozlan, Cyril Roberto, Paul-Marie Samson et Prasad Tetali.

Dans la section 8, sous certaines hypothèses sur θ, nous montrons que l'inégalité CmLS est équivalente à la propriété de contraction suivante :

$$\exists a, b > 0, \forall h > 0, \quad \sup_x \{U(x+h) - U(x)\} \leqslant \sqrt{a + bh}.$$

Ce travail est extrait de [94] en collaboration avec Michal Strzelecki.

Le chapitre se conclut par une comparaison des inégalités de transport-information et de Talagrand en dimension un. Cette section propose une condition nécessaire pour les inégalités $\mathrm{T}_\theta\mathrm{I}$. Grâce à cette condition, nons donnons un contre-exemple d'une mesure probabilité μ qui vérifie l'inégalité T_2 mais pas $\mathrm{T}_\theta\mathrm{I}$ pour tout θ convexe croissant. Les résultats de cette section sont un développement de mon mémoire de Master.

Chapitre 2

Infimum convolution operators

Abstract

In this chapter, we are interested in infimum convolution operators of the forms

$$Rf(x) = \inf_{s\in S}\{a(f)(s) + b(x,s)\}.$$

This kind of formulas are widely used in many areas of mathematics. Here we will focus on the following special cases.

– The classical Hopf-Lax formula

$$Q_t f(x) := \inf_{x\in\mathcal{X}}\{f(y) + t\alpha\left(\frac{d(x,y)}{t}\right)\}, \tag{2.0.1}$$

defined for all (say) bounded functions f on some metric space $(\mathcal{X}, d)$.

– The modified Hopf-Lax formula introduced in [50]

$$\widetilde{Q}_\alpha\varphi(x) = \inf_{p\in\mathcal{P}(\mathcal{X})}\left\{\int \varphi(y)p(dy) + \alpha\left(\int d(x,y)p(dy)\right)\right\}, \tag{2.0.2}$$

where $\mathcal{P}(\mathcal{X})$ denotes the set of probability measures on $\mathcal{X}$.

– The equilibrium case

$$\omega(x) = \inf_{y:(y,x)\in E}\{\omega(y) + k(y,x)\},$$

where k is some cost function defined on some finite set E.

We will first recall briefly some properties of Q_t, then introduce a modified H-L-O formula in a more general space. We will then analyse the regularity properties of this operator $\widetilde{Q}$ and establish Hamilton-Jacobi type equations for this operator. Finally, motivated by a large deviation problem, we shall focus on the equilibrium case on a graph setting.

2.1 Hopf-Lax formula in geodesic spaces

This section is a reminder about the classical Hopf-Lax-Oleinik formula. Let $\alpha : \mathbb{R}^+ \mapsto \mathbb{R}^+$ be an increasing convex function of class C^1 such that $\alpha(0) = 0$. If $f : \mathcal{X} \mapsto \mathbb{R}$ is a semi continuous function bounded from below, one can define for all $t > 0$ the function $Q_t f$ as follows :

$$Q_t f(x) := \inf_{x \in \mathcal{X}} \{ f(y) + t\alpha \left(\frac{d(x,y)}{t} \right) \}. \tag{2.1.1}$$

This formula is usually called the Hopf-Lax-Oleinik formula. The Hopf-Lax-Oleinik formula is known to be the solution of Hamilton Jacobi equations, and is connected to optimal transport theory through the Kantorovich duality formula. We recall these properties below.

First of all, the family of operators $\{Q_t\}_{t>0}$ satisfies the semi-group property

$$Q_{t+s} f = Q_t(Q_s f).$$

For $t = 1$, one has the Kantorovich duality formula

Theorem 2.1.2 (Kantorovich duality)**.** *Let μ, ν be two probability measures on $\mathcal{X}$, then it holds*

$$\mathcal{T}_\alpha(\mu, \nu) = \sup_{f \in \mathcal{C}_c^b} \int Q_1 f d\mu - \int f d\nu$$

where $\mathcal{T}_\alpha(\mu, \nu)$ is the optimal transportation cost with cost function α.

We recall as well the definition of the optimal transportation cost $\mathcal{T}_\alpha(\mu, \nu)$.

Definition 2.1.3 (Optimal transportation cost)**.** *Let $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$, with $\alpha(0) = 0$, be a measurable function referred to as the* cost *function. Then, the usual optimal* transportation cost*, in the sense of Kantorovich, between two probability measures μ and ν on $\mathbb{R}$ is defined by*

$$\mathcal{T}_\alpha(\nu, \mu) := \inf_\pi \iint \alpha(d(x,y))\, \pi(dxdy), \tag{2.1.4}$$

where the infimum runs over the set of couplings π between μ and ν, i.e., probability measures on $\mathcal{X}^2$ such that $\pi(dx \times \mathcal{X}) = \mu(dx)$ and $\pi(\mathcal{X} \times dy) = \nu(dy)$.

Thanks to the Kantorovich duality theorem, the properties of this infimum convolution operator Q lead to various properties of the optimal transport cost. It has been proved to be a very powerful tool for analysing transport inequalities. We refer [11], [99] and [48] for related works.

The Hopf-Lax formula satisfies the following Hamilton-Jacobi equation :

$$\frac{d}{dt_+} Q_t f + \alpha^* \left(|\nabla^- Q_t f| \right) = 0, \quad (x,t) \in \mathcal{X} \times (0, \infty) \tag{2.1.5}$$

Where α^* is the Legendre transformation of α and the gradient $|\nabla^- f|$ is defined by

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

(by convention, we set $|\nabla^- f|(x) = 0$ if x is and isolated point in $\mathcal{X}$). When the space is $(\mathbb{R}^n, \|.\|_2)$, these properties are very classical. Recently, Ambrosio, Gigli and Savaré in [3] and Gozlan, Roberto, Samson in [47] extended this result independently in a general geodesic space.

Here are some more properties of Q_t borrowed from [47] in a geodesic space.

Theorem 2.1.6. *Let $f : \mathcal{X} \mapsto \mathbb{R}$ be a lower semicontinuous function bounded from below. For all $t > 0$ and $x \in \mathcal{X}$, denote by $m(t,x)$ the set of points where the infimum* (2.1.1) *defining $Q_t f(x)$ is reached :*

$$m(t,x) = \Big\{ y \in \mathcal{X}, Q_t f(x) = f(y) + t\alpha\left(\frac{d(x,y)}{t}\right) \Big\}.$$

These sets are always non empty and compact and it holds for all $t > 0$ and $x \in \mathcal{X}$,

$$\frac{d}{dt_+} Q_t f(x) = -\beta\left(\frac{1}{t} \max_{y \in m(t,x)} d(x,y)\right)$$

and

$$\frac{d}{dt_-} Q_t f(x) = -\beta\left(\frac{1}{t} \min_{y \in m(t,x)} d(x,y)\right),$$

where $\beta(h) = h\alpha'(h) - \alpha(h)$, $h \geqslant 0$.

Proposition 2.1.7 (Convexity)**.** *Let $\mathcal{X} = \mathbb{R}^n$ and $d := \|.\|_2$ be the Euclidean norm. Assume that f is a convex function on $\mathcal{X}$, then the function $Q_t f$ is convex.*

We refer [47] for more regularity properties of $Q_t f$.

2.2 Modified Hopf-Lax formula in general space

In this section we only assume that $(\mathcal{X}, d)$ is a complete and separable metric space in which closed balls are compact (for example $(\mathcal{X}, d)$ is a finite graph equipped with graph distance). Observe that the proofs of properties of the latter section rely heavily on the fact that the space is geodesic. If we consider the space is not a geodesic space, although the operator Q is still well defined, the semi-group property, the Hamilton-Jacobi equation and many of the regularity properties mentioned before will fail. Therefore, we should introduce a modified Hopf-Lax formula in order to adapt the results in a non geodesic space. An natural idea is to look at infimum convolution of form

$$Q_t f(x) := \inf_{y \in V} \{ f(y) + D_t(y,x) \}.$$

However, in a discrete space, say graph, there is not hope of finding such a family of mappings $(D_t)_{t>0}$ such that the usual semi-group property $Q_{t+s} = Q_t(Q_s)$ holds.

More precisely, we have the following result.

Proposition 2.2.1 (Semi-group property fails). *Let $G = (V, E)$ be a finite graph. Assume we are given a family of mappings $D_t \colon V \times V \to \mathbb{R}^+$, $t > 0$ that satisfies $D_t(x, x) = 0$ for all $x \in V$ and all $t > 0$. Assume furthermore that for any $f \colon V \to \mathbb{R}$ and any $x \in V$, $Q_t f(x) := \inf_{y\in V}\{f(y) + D_t(y, x)\} \to f(x)$ when $t \to 0$. Then, there exists f, $x \in V$ and $t, s > 0$ such that $Q_{t+s}f(x) \neq Q_t(Q_s f)(x)$.*

proof. By contradiction assume that for all f bounded on V, all $x \in \mathcal{X}$ and $s, t > 0$, it holds $Q_t Q_s f = Q_{t+s}f$. The proof is based on the following claims.

Claim 2.2.2. *For all $x, z \in V$, all $s < r \in (0, \infty)$, it holds $D_r(z, x) = \min_{y\in V} \{D_s(z, y) + D_{r-s}(y, x)\}$.*

Claim 2.2.3. *For all $x, z \in V$, the map $(0, \infty) \ni t \mapsto D_t(z, x)$ is non-increasing and, if $x \neq z$, $D_t(z, x) \to \infty$ as t goes to 0.*

We postpone the proof of the above claims to end the prove of the proposition.

Fix $x, z \in V$, $x \neq z$. Then, by Claim 2.2.2, for all $s \in (0, 1)$, it holds

$$\begin{aligned} D_1(z, x) &= \min_{y\in V}\{D_s(z, y) + D_{1-s}(y, x)\} \\ &= \min\left(D_{1-s}(z, x); \min_{y\neq z}\{D_s(z, y) + D_{1-s}(y, x)\}\right). \end{aligned}$$

By Claim 2.2.3 and since the graph is finite, $\lim_{s\to 0} \min_{y\neq z}\{D_s(z, y)+D_{1-s}(y, x)\} = \infty$. Hence, there exists $s_o \in (0, 1)$ such that, for $s < s_o$, $D_1(z, x) = D_{1-s}(z, x)$ so that $u_o := \sup\{u \in (0, 1) : D_{1-u}(z, x) = D_1(z, x)\}$ is well-defined thanks to Claim 2.2.3. By a similar argument, there exists $s_1 \in (0, 1 - u_o)$ such that $D_{1-u_o-s}(z, x) = D_{1-u_o}(z, x)$ for all $s < s_1$. This contradicts the definition of u_o and ends the proof of the proposition provided that we prove Claim 2.2.3 and Claim 2.2.2.

Proof of Claim 2.2.2. Since $D_t(x, z)$ is non-negative and $D_t(x, x) = 0$, the claim is trivial if $x = z$. Assume that $x \neq z$. Let $s < r$ and consider $f \colon V \to \mathbb{R}$ defined by $f(z) = 0$ and $f(y) = D_r(z, x) + 1$ for all $y \neq z$. Then

$$\begin{aligned} Q_r f(x) &= \min_{y\in V}\{f(y) + D_r(y, x)\} = \min\left(D_r(z, x); \min_{y\neq z}\{f(y) + D_r(y, x)\}\right) \\ &= D_r(z, x). \end{aligned}$$

On the other hand, by the semi-group property, similarly (necessarily $u = z$) it holds

$$\begin{aligned} Q_r f(x) = Q_{r-s}(Q_s f)(x) &= \min_{u,y\in V}\{f(u) + D_s(u, y) + D_{r-s}(y, x)\} \\ &= \min_{y\in V}\{D_s(z, y) + D_{r-s}(y, x)\} \end{aligned}$$

which leads to the conclusion. □

Proof of Claim 2.2.3. If $x = z$, the map $t \mapsto D_t(z,x)$ is constant and so there is nothing to prove. Assume that $x \neq z$. By Claim 2.2.2 we have for $s < r$ (take $y = x$), $D_r(z,x) = \inf_{y \in V}\{D_s(z,y) + D_{r-s}(y,x)\} \leqslant D_s(z,x)$ which proves that $t \mapsto D_t(z,x)$ is non-increasing and that the limit $\lim_{r\to 0} D_r(z,x)$ exists in $[0,\infty]$. For $M > 0$, let $f \colon V \to \mathbb{R}$ be defined by $f(z) = 0$, $f(x) = M$ and $f(y) = M+1$ for all $y \neq z, x$. Then

$$
\begin{aligned}
Q_r f(x) &= \min_{y\in V}\{f(y) + D_r(y,x)\} = \min\left(D_r(z,x); f(x); \min_{y\neq z,x}\{f(y) + D_r(y,x)\}\right) \\
&= \min\left(D_r(z,x); M\right) \leqslant \frac{1}{2}\left(D_r(z,x) + M\right).
\end{aligned}
$$

Now, by assumption $Q_r f(x) \to f(x) = M$ as r goes to 0 so that, taking the limit in the latter guarantees that $\lim_{r\to 0} D_r(z,x) \geqslant M$ which ends the proof of Claim 2.2.3 since M is arbitrarily large. □

The proof of the proposition is complete. □

Another approach is to consider the following modified Hopf-Lax formula introduced by Gozlan-Roberto-Samson-Tetali in [50].

$$
\widetilde{Q}_t^{\alpha}\varphi(x) = \inf_{p\in\mathcal{P}(\mathcal{X})}\left\{\int \varphi(y)p(dy) + t\alpha\left(\frac{\int d(x,y)p(dy)}{t}\right)\right\}, \tag{2.2.4}
$$

With this particular operator, they manage to recover the Kantorovich duality corresponding to the Marton transportation cost [76, 75].

Definition 2.2.5 (Marton's transportation cost)**.** *Let μ, ν be two probability measures on $\mathcal{X}$ and θ be a positive cost function. Define*

$$
\widetilde{\mathcal{T}}_{\alpha}(\nu|\mu) := \inf_{\pi\in\Pi(\mu,\nu)}\left\{\int \alpha\left(\int d(x,y)p_x(dy)\right)\mu(dx)\right\},
$$

where $\Pi(\mu,\nu)$ is the set of all couplings π whose first marginal is μ and second marginal is ν, p_x is the probability kernel such that $\pi(dxdy) = p_x(dy)\mu(dx)$.

Theorem 2.2.6 (GRST)**.** *With the definitions and the notations before, it holds*

$$
\widetilde{\mathcal{T}}_{\alpha}(\nu|\mu) = \sup_{\varphi}\left\{\int \widetilde{Q}_1^{\alpha}\varphi d\mu - \int \varphi d\nu\right\}.
$$

Therefore, it will be very interesting to look at the properties of this operator $\widetilde{Q}$ and connect them with the Marton's transportation cost.

This section presents some technical results about $\widetilde{Q}_t$, the goal is to establish a Hamilton Jacobi type equations in a space non necessary geodesic. We start with a convexity property.

2.2.1 Convexity

Proposition 2.2.7 (Convexity). *Let g be a function defined on $\mathcal{X}$, then for all $x \in \mathcal{X}$, the function $t \mapsto \widetilde{Q}_t g(x)$ is convex.*

proof. Fix $x \in \mathcal{X}$, define $G(t) := \widetilde{Q}_t g(x)$ Observe that for any $\lambda \in [0,1]$, and $p_1, p_2 \in \mathcal{P}(\mathcal{X})$, setting $p := \lambda p_1 + (1-\lambda)p_2 \in \mathcal{P}(\mathcal{X})$, for all $t, s > 0$, denote $\Lambda = \frac{\lambda t}{\lambda t + (1-\lambda)s}$, the convexity of the cost function α implies that :

$$\Lambda \alpha(u_1) + (1-\Lambda)\alpha(u_2) \leqslant \alpha(\Lambda u_1 + (1 - \Lambda u_2)).$$

Now let $u_1 = \int d(x,z)p_1(dz)/t$ and $u_2 = \int d(x,z)p_2(dz)/s$, we deduce that

$$\lambda t \alpha\left(\frac{\int d(x,z)p_1(dz)}{t}\right) + (1-\lambda)s\alpha\left(\frac{\int d(x,z)p_2(dz)}{s}\right) \geqslant (\lambda t + (1-\lambda)s)\alpha\left(\frac{\int d(x,z)p(dz)}{\lambda t + (1-\lambda)s}\right)$$

As a consequence, we get

$$\begin{aligned}\lambda\left(\int g(z)p_1(dz) + t\alpha\left(\frac{\int d(x,z)p_1(dz)}{t}\right)\right) + (1-\lambda)\left(\int g(z)p_2(dz) + s\alpha\left(\frac{\int d(x,z)p_2(dz)}{s}\right)\right) \\ \geqslant \int g(z)p(dz) + (\lambda t + (1-\lambda)s)\alpha\left(\frac{\int d(x,z)p(dz)}{(\lambda t + (1-\lambda)s)}\right) \geqslant G(\lambda t + (1-\lambda)s)\end{aligned}$$

Taking the infimum over all $p_1, p_2 \in \mathcal{P}(\mathcal{X})$ on the left hand side of the inequality, the conclusion follows. □

Remark 2.2.8. *This statement does not hold true for classical Hopf-Lax formula in geodesic space.*

Next, we are interested in the existence and properties of points where the infimum defining $\widetilde{Q}_t$ is reached.

2.2.2 Existence of minimum

Now we will analyse the set of probabilities such that the infimum defined in $\widetilde{Q}_t f(x)$ is reached. We recall here the assumption on the space : $(\mathcal{X}, d)$ is a complete and separable metric space in which closed balls are compact.

In all what follows, as in the classical case, we need to assume that $f : \mathcal{X} \to \mathbb{R}$ is a lower semi-continuous function bounded from below. We let

$$m_f(t,x) := \left\{ p \in \mathcal{P}(\mathcal{X}) : \widetilde{Q}_t f(x) = \int f\, dp + t\alpha\left(\frac{\int d(x,y)\, p(dy)}{t}\right)\right\} \tag{2.2.9}$$

be the set (possibly empty) of probability measures p realizing the infimum in the definition of $\widetilde{Q}_t f(x)$. In fact, this set is always non empty.

Proposition 2.2.10. *If $f : \mathcal{X} \to \mathbb{R}$ is lower semicontinuous and bounded from below, then $m_f(t,x) \neq \emptyset$ for all $t > 0$ and $x \in \mathcal{X}$.*

proof. Consider the function

$$F(p) := \int f\, dp + t\alpha\left(\frac{\int d(x,y)\, p(dy)}{t}\right).$$

The problem reduces to prove that the minimum of F exists. We first show that F is a lower semicontinuous function bounded from below. It is easy to see that F is bounded from below by $m = \inf_X f$. Since f is lower semicontinuous and bounded from below, the function $p \mapsto \int f\, dp$ is lower semicontinuous w.r.t. the weak convergence topology of $\mathcal{P}(\mathcal{X})$. For the same reason $p \mapsto \int d(x,y)\, dp$ is also lower semicontinuous. Moreover, the sub-level sets of F are compact. Indeed, for all $r \geqslant m$, it holds

$$\{F \leqslant r\} \subset \left\{p \in \mathcal{P}(\mathcal{X}) : \int d(x,y)\, p(dy) \leqslant C_{t,r}\right\}, \quad \text{with} \quad C_{t,r} = t\alpha^{-1}\left(\frac{r-m}{t}\right).$$

In particular, if $p \in \{F \leqslant r\}$, then $p(B(x,R)^c) \leqslant C_{t,r} R^{-1}$, for all $R > 0$. Since balls in $\mathcal{X}$ are assumed to be compact, the compactness of $\{F \leqslant r\}$ follows from Prokhorov theorem.

Since F is lower semi-continuous, bounded from below and has compact sub-level sets, F attains its minimum and so $m_f(t,x)$ is not empty. □

2.2.3 Regularity

In this section, we are going to analyse the regularity of the set $m_f(t,x)$, by comparing the operator $\widetilde{Q}$ with the classical Hopf-Lax-Oleinik formula Q which is defined on $\mathbb{R}$. We will show that when $x \in \mathcal{X}$ is given, there exists a real value function $\tilde{f}_x : \mathbb{R}^+ \mapsto \mathbb{R}$ such that $\widetilde{Q}_t f(x) = Q_t \tilde{f}_x(0)$. This helps us understand properties of $\widetilde{Q}$ through properties of Q. Especially, since the function $\tilde{f}_x$ does not depend on t, one can deduce the time derivative of $\widetilde{Q}_t f(x)$ from the time derivative of Q_t.

Now we are going to construct the function $\tilde{f}_x$. We start with the following two definitions.

Definition 2.2.11. *Given $x \in \mathcal{X}$, let $I_x := \{d(x,y), y \in \mathcal{X}\} \subset \mathbb{R}^+$ be the image of the function $y \in \mathcal{X} \mapsto d(x,y)$. Since $(\mathcal{X}, d)$ is a polish space such that all closed balls are compact, I_x is a closed subset of $\mathbb{R}$. Define $f_x\colon [0,+\infty) \to \mathbb{R} \cup \{\infty\}$ as*

$$f_x(u) := \min_{y \in \mathcal{X} : d(x,y) = u} \{f(y)\}.$$

with convention that $\min \emptyset = \infty$. We notice that $f_x(0) = f(x)$ and $f_x(u)$ is finite if and only if $u \in I_x$.

Let $\widetilde{I}_x$ be the convex hull of I_x. Since closed balls are assumed to be compact, $\widetilde{I}_x$ is one of the following intervals $[0, \sup I_x]$ (if I_x is bounded) or $[0, +\infty)$ (if I_x is unbounded). Let $\widetilde{f}_x : \mathbb{R}^+ \to \mathbb{R} \cup \{+\infty\}$ be the convex hull of f_x, that is to say the greatest convex function $g : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ such that $g(u) \leqslant f_x(u)$ for all $u \in I_x$. The function $\widetilde{f}_x$ takes finite values on $\widetilde{I}_x$ and is $+\infty$ outside $\widetilde{I}_x$.

Definition 2.2.12. *Let $\widetilde{f}_x : \mathbb{R}^+ \to \mathbb{R} \cup \{+\infty\}$ be the convex hull of f_x, that is to say the greatest convex function $g : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ such that $g(u) \leqslant f_x(u)$ for all $u \in I_x$.*

Proposition 2.2.13. *Let $f : \mathcal{X} \to \mathbb{R}$ be bounded from below and lower semi-continuous. Then for all $t > 0$, all $x \in \mathcal{X}$, it holds*

$$\widetilde{Q}_t f(x) = Q_t \widetilde{f}_x(0).$$

proof. Fix $f : \mathcal{X} \to \mathbb{R}$ bounded from below and lower semi-continuous, and $x \in \mathcal{X}$. It holds

$$\begin{aligned}\widetilde{Q}_t f(x) &= \inf_{p \in \mathcal{P}(\mathcal{X})} \left\{ \int f\,dp + t\alpha\left(\frac{\int d(x,y)\,p(dy)}{t} \right) \right\} \\ &= \inf_{u \in \mathbb{R}^+} \left\{ g_x(u) + t\alpha\left(\frac{u}{t}\right) \right\},\end{aligned}$$

where

$$g_x(u) = \inf\left\{ \int f\,dp : p \in \mathcal{P}(\mathcal{X}) : \int d(x,y)\,p(dy) = u \right\}, \qquad u \in \mathbb{R}^+.$$

Let us show that $g_x(u) = \widetilde{f}_x(u)$ $u \in \mathbb{R}^+$. If u is outside $\widetilde{I}_x$, then both functions are equal to $+\infty$ and there is nothing to prove. It is enough to show that $g_x = \widetilde{f}_x$ on $\widetilde{I}_x$.

First choosing, in the definition of g_x, $p = \delta_y$ for some $y \in \mathcal{X}$ such that $d(x,y) = u \in I_x$, one gets that $g_x(u) \leqslant f(y)$. Optimizing over all y such that $d(x,y) = u$, one concludes that $g_x(u) \leqslant f_x(u)$ for all $u \in I_x$. Moreover the function g_x is easily seen to be convex. By definition of the convex hull of f_x, it follows that $g_x(u) \leqslant \widetilde{f}_x(u)$ for all $u \in \widetilde{I}_x$. Now let us show that $g_x \geqslant \widetilde{f}_x$. For all $y \in \mathcal{X}$, it holds $f(y) \geqslant f_x(d(x,y))$. Therefore, if p is such that $\int d(x,y)\,p(dy) = u \in \widetilde{I}_x$, then denoting by $\widetilde{p} \in \mathcal{P}_u(I_x)$ the image of p under the map $y \mapsto d(x,y)$, it holds

$$\int f(y)\,p(dy) \geqslant \int f_x(d(x,y))\,p(dy) = \int f_x(v)\,\widetilde{p}(dv) \geqslant \int \widetilde{f}_x(v)\,\widetilde{p}(dv) \geqslant \widetilde{f}_x(u), \tag{2.2.14}$$

where the last inequality follows from Jensen inequality. Optimizing over p, yields to $g_x \geqslant \widetilde{f}_x$ on $\widetilde{I}_x$ and so $g_x = \widetilde{f}_x$ and this completes the proof. □

Here we have another way to define $\widetilde{f}_x$ on $\widetilde{I}_x$, which allows to get more regularity property. With this lemma, we are able to show that there exists a probability measure of form $\lambda\delta_x + (1-\lambda)\delta_y$, who achieves the infimum defined on $\widetilde{Q}_t$, for strictly increasing cost function.

Let $\mathcal{P}_u(I_x)$ be the set of probability measures on I_x with expectation u, *i.e.* $\int_{I_x} y\, p(dy) = u$.

Lemma 2.2.15. *Let $f : \mathcal{X} \to \mathbb{R}$ be a lower semicontinuous function and define f_x and $\widetilde{f}_x$ as above. Then, for all $u \in \widetilde{I}_x$,*

$$\widetilde{f}_x(u) = \inf\left\{\int_{I_x} f_x(w)\, q(dw) : q \in \mathcal{P}_u(I_x) \text{ charging at most two points}\right\}. \quad (2.2.16)$$

Moreover, the function $\widetilde{f}_x$ is continuous on $\widetilde{I}_x$ and lower semicontinuous on $\mathbb{R}$.

Proof of Lemma 2.2.15. Fix $f : \mathcal{X} \to \mathbb{R}$ bounded from below and lower semicontinuous, $x \in \mathcal{X}$ and $u \in \mathbb{R}^+$. According to *e.g.* [56][Proposition B.2.5.1],

$$\widetilde{f}_x(u) = \inf\left\{\int_{I_x} f_x(w)\, q(dw) : q \in \mathcal{P}_u(I_x) \text{ with finite support}\right\}.$$

Applying Caratheodory's Theorem (see *e.g.* [56][Theorem A.1.3.6]), ones sees that one can assume that the infimum is over probability measures q charging at most three points. Let us explain how to reduce to two points.

Fix $\varepsilon > 0$; there exist $w_1, w_2, w_3 \in I_x$, and $\lambda_1, \lambda_2, \lambda_3 \in [0,1]$ with $\sum_i \lambda_i = 1$ such that $u = \lambda_1 w_1 + \lambda_2 w_2 + \lambda_3 w_3$ and

$$\widetilde{f}_x(u) \geqslant \lambda_1 f_x(w_1) + \lambda_2 f_x(w_2) + \lambda_3 f_x(w_3) - \varepsilon$$

Without loss of generality we can assume that $w_1 < w_2 < w_3$, and for example that $w_1 \leqslant u \leqslant w_2$ (the other case is similar). Then there exist $a, b \in [0,1]$ such that

$$u = aw_1 + (1-a)w_2 = bw_1 + (1-b)w_3.$$

Then it is not difficult to check that there is a unique $\lambda \in [0,1]$ such that $\lambda_1 = \lambda a + (1-\lambda)b$, $\lambda_2 = \lambda(1-a)$ and $\lambda_3 = (1-\lambda)(1-b)$. Therefore it holds

$$u = (\lambda a + (1-\lambda)b)w_1 + \lambda(1-a)w_2 + (1-\lambda)(1-b)w_3$$

and

$$\widetilde{f}_x(u) \geqslant (\lambda a + (1-\lambda)b) f_x(w_1) + \lambda(1-a) f_x(w_2) + (1-\lambda)(1-b) f_x(w_3) - \varepsilon.$$

By definition of $\widetilde{f}_x(u)$, necessarily,

$$\widetilde{f}_x(u) \leqslant \min_{s\in[0,1]}\{(sa + (1-s)b) f_x(w_1) + s(1-a) f_x(w_2) + (1-s)(1-b) f_x(w_3)\}.$$

Since, in the right hand side of the latter, the function of s that needs to be minimized is an affine function, the minimum is reached at $s = 0$ or $s = 1$. Therefore

$$\widetilde{f}_x(u) \geqslant \min\{a f_x(w_1) + (1-a) f_x(w_2), b f_x(w_1) + (1-b) f_x(w_3)\} - \varepsilon$$

which proves that, for all $\varepsilon > 0$, there exists $q \in \mathcal{P}_2(I_x)$ such that $\int v\, q(dv) = u$ and $\int f_x(v)\, q(dv) \geqslant \widetilde{f}_x(u) \geqslant \int_{I_x} f_x(v)\, q(dv) - \varepsilon$. Since $\varepsilon > 0$, this completes the proof.

Now let us prove that $\widetilde{f}_x$ is continuous on $\widetilde{I}_x$. By definition, $\widetilde{f}_x$ is a convex function on the closed interval $\widetilde{I}_x$, thus it is continuous on the interior of $\widetilde{I}_x$. Hence it only remains to prove that $\widetilde{f}_x$ is continuous at 0 and, in case I_x is bounded, at $b = \max I_x$. We only give the proof of the continuity at 0, the other case is similar.

Take $x_o \in \mathcal{X} \setminus \{x\}$ and let $u_o = d(x, x_o) \in I_x \setminus \{0\}$. Since $\widetilde{f}_x$ is convex, on $\widetilde{I}_x$, it holds, for all $0 \leqslant u \leqslant u_o$

$$\widetilde{f}_x(u) = \widetilde{f}_x\left(\frac{u}{u_o}.u_o + (1 - \frac{u}{u_o}).0\right) \leqslant \frac{u}{u_o}\widetilde{f}_x(u_o) + \left(1 - \frac{u}{u_o}\right)\widetilde{f}_x(0).$$

Thus letting $u \to 0^+$, one gets that $\limsup_{u\to 0^+} \widetilde{f}_x(u) \leqslant \widetilde{f}_x(0)$. Now, we prove that $\liminf_{u\to 0^+} \widetilde{f}_x(u) \geqslant \widetilde{f}_x(0)$. Thanks to the lower semicontinuity of f, for all $\varepsilon \in (0,1)$, there exists η, for all $y \in B(x,\eta)$, $f(y) \geqslant f(x) - \varepsilon$. Thus, from the definition of f_x, it follows that for all $u \in [0,\eta)$,

$$f_x(u) \geqslant f_x(0) - \varepsilon.$$

On the other hand, if m is a lower bound for f, then $f_x(u) \geqslant m$ for all $u \in [0,\infty)$. Therefore, it holds

$$f_x(u) \geqslant (f_x(0) - \varepsilon)\mathbf{1}_{[0,\eta)}(u) + m\mathbf{1}_{[\eta,\infty)} := g_\varepsilon(u), \qquad \forall u \in [0,\infty),$$

(here we use that by definition $f_x(u) = +\infty$ when $u \notin I_x$). Taking a smaller m if necessary, one can assume that $f_x(0) - \varepsilon > m$ for all $\varepsilon \in (0,1)$. Now consider, the affine function h_ε joining $(0, f_x(0) - \varepsilon)$ to (η, m). It is clear that $g_\varepsilon \geqslant h_\varepsilon$ on $[0,\infty)$. Therefore, by definition of $\widetilde{f}_x$ as the greatest convex function below f_x, it holds $\widetilde{f}_x \geqslant h_\varepsilon$ on $[0,\infty)$. In particular,

$$\liminf_{u\to 0^+} \widetilde{f}_x(u) \geqslant \liminf_{u\to 0^+} h_\varepsilon(u) = f_x(0) - \varepsilon.$$

Since ε is arbitrary, one concludes that $\liminf_{u\to 0^+} \widetilde{f}_x(u) \geqslant f_x(0) \geqslant \widetilde{f}_x(0)$. In conclusion, $\lim_{u\to 0^+} \widetilde{f}_x(u) = \widetilde{f}_x(0) = f_x(0)$, which completes the proof. $\square$

The following lemma illustrates when the latter infimum could be achieved. This lemma seems classical and it might be found in some convex analyses document.

Lemma 2.2.17. *Let f be a lower semi-continuous function bounded from below define on a close set $I \subset \mathbb{R}$. Let g be the largest convex function such that $g \leqslant f$ on I. For all affine function h, define $I_h := [a,b]$ be the maximum interval such that $g - h$ reaches its minimum. Then, if $a \neq \infty$, it holds that $a \in I$ and $f(a) = g(a)$, and the same conclusion holds for b if $b \neq \infty$.*

proof. Without loss of generality, we can suppose that $I_h = [a, b]$ with $a \neq \pm\infty$. It is enough to show that $f(a) = g(a)$, the other cases are similar. The definition of g implies directly that $g(a) \leqslant f(a)$, so we now turn to prove the inverse inequality. Changing h into $h + constant$, we can suppose that $g - h = 0$ on I_h and $g - h > 0$ on $\mathbb{R} \setminus I_h$. Let h_n the affine function such that $h_n(a - 1/n) = g(a - 1/n)$ and $h_n(a + 1/n) = g(a + 1/n)$. By definition of I_h, $h_n(a - 1/n) > h(a - 1/n)$ and $h_n(a + 1/n) \geqslant h(a + 1/n)$. It follows that $h_n(a) > h(a) = g(a)$. Thus, if we define $g_n : x \mapsto \max\{g(x), h_n(x)\}$, then g_n is a convex function greater than g. Thus, the definition of g implies that the existence of $z_n \in I$ such that $f(z_n) < g_n(z_n)$. Notice that $g_n = g$ on $\mathbb{R}\setminus[a-1/n, a+1/n]$, so $z_n \in [a-1/n, a+1/n]$. Hence, $\lim_{n\to\infty} z_n = a$ and it holds

$$\begin{aligned} g(a) = h(a) = \lim_{n\to\infty} h(z_n) &\leqslant \lim_{n\to\infty} f(z_n) \\ &\leqslant \lim_{n\to\infty} g_n(z_n) \leqslant \lim_{n\to\infty} \max\{g(a - 1/n), g(a + 1/n)\} = g(a) \end{aligned}$$

Thus, by lower semi-continuity of f, we have $g(a) = \lim_{n\to\infty} f(z_n) \geqslant f(\lim_{n\to\infty} z_n) = f(a)$. The proof is completed. □

As a consequence of the latter lemma, suppose that the largest affine part contains $(u, \widetilde{f}_x(u))$ is $([a_u, b_u], \widetilde{f}_x([a_u, b_u]))$, if $b_u < \infty$, then we have $a_u, b_u \in I_x$ and $\widetilde{f}_x(a_u) = f_x(a_u)$, $\widetilde{f}_x(b_u) = f_x(b_u)$. Hence,

$$\widetilde{f}_x(u) = \int_{I_x} f_x(w)\, q(dw),$$

where $q = \lambda\delta_{a_u} + (1 - \lambda)\delta_{b_u}$ with λ satisfies $u = \lambda a_u + (1 - \lambda b_u)$.

Finally, let

$$\widetilde{m}_f(t, x) := \left\{u \in \mathbb{R}^+ : Q_t\widetilde{f}_x(0) = \widetilde{f}_x(u) + t\alpha\left(\frac{u}{t}\right)\right\}. \tag{2.2.18}$$

This set is easily seen to be non-empty using the lower semicontinuity of $\widetilde{f}_x$ (see also Item (ii) of the following result.)

2.2.4 Time derivative

Now we are ready to calculate the time derivative :

Theorem 2.2.19. *Set $\beta(x) := x\alpha'(x) - \alpha(x)$, $x \geqslant 0$. Let $f : \mathcal{X} \to \mathbb{R}$ be bounded from below and lower semi-continuous. Then,*

(i) Assume that the cost function α is strictly increasing, then for all $t > 0$ and all $x \in \mathcal{X}$, it holds

$$\left\{\int d(x, y)\, p(dy) : p \in m_f(t, x)\right\} = \widetilde{m}_f(t, x). \tag{2.2.20}$$

(ii) For all $x \in \mathcal{X}$ and all $t > 0$, the function $u \mapsto \beta(u/t)$ is constant on $\widetilde{m}_f(t,x)$. In particular, the function $p \mapsto \beta\left(\int d(x,y)\,p(dy)/t\right)$ is constant on $m_f(t,x)$.

(iv) For all $t > 0$, $x \in \mathcal{X}$ and $p \in m_f(t,x)$, it holds

$$\frac{\partial}{\partial t}\widetilde{Q}_t f(x) = -\beta\left(\frac{\int d(x,y)\,p(dy)}{t}\right); \tag{2.2.21}$$

Proof of Theorem 2.2.19. Item (i). Let $p \in m_f(t,x)$ and $u = \int d(x,y)\,p(dy)$. Then, according to (2.2.14), one has $\widetilde{f}_x(u) \leqslant \int f\,dp$. Hence, using the very definition of $m_f(t,x)$, Item (i) and the definition of $Q_t\widetilde{f}_x(0)$, it holds

$$\begin{aligned}\widetilde{f}_x(u) + t\alpha\left(\frac{u}{t}\right) &\leqslant \int f\,dp + t\alpha\left(\frac{\int d(x,y)\,p(dy)}{t}\right) = \widetilde{Q}_t f(x) = Q_t\widetilde{f}_x(0)\\ &\leqslant \widetilde{f}_x(u) + t\alpha\left(\frac{u}{t}\right)\end{aligned}$$

It follows that $Q_t\widetilde{f}_x(0) = \widetilde{f}_x(u) + t\alpha\left(\frac{u}{t}\right)$ and thus that $u \in \widetilde{m}_f(t,x)$ which, in turn, guarantees that $\{\int d(x,y)\,p(dy) : p \in m_f(t,x)\} \subset \widetilde{m}_f(t,x)$.

Conversely, let $u \in \widetilde{m}_f(t,x)$. Firstly assume that the cost function α is strictly increasing. If $u = 0$, then it suffice to take $p = \delta_0$ and it is easy to see that $p \in m_f(t,x)$. Now suppose that $u > 0$. Let $([a_u,b_u], \widetilde{f}_x([a_u,b_u]))$ be the largest affine part of the graph $\widetilde{f}_x$ which contains $(u, \widetilde{f}_x(u))$. If $b_u < \infty$, then thanks to lemma2.2.17 $f_x(a_u) = \widetilde{f}_x(a_u)$ and $f_x(b_u) = \widetilde{f}_x(b_u)$. As a consequence, there exist y_1 and y_2 such that $f_x(a_u) = f(y_1)$ and $f_x(b_u) = f(y_2)$, $d(x,y_1) = a_u$, $d(x,y_2) = b_u$. It is suffice to define $p := \lambda\delta_{y_1} + (1-\lambda)\delta_{y_2}$ where λ satisfies $\lambda a_u + (1-\lambda)b_u = u$. Moreover, by Item (i) and by definition of $\widetilde{m}_f(t,x)$ we have

$$\widetilde{Q}_t f(x) = Q_t\widetilde{f}_x(0) = \widetilde{f}_x(u) + t\alpha\left(\frac{u}{t}\right) = \int f\,dp + t\alpha\left(\frac{\int d(x,y)\,p(dy)}{t}\right) \geqslant \widetilde{Q} f(x)$$

which proves that $p \in m_f(t,x)$ and thus that $u \in \{\int d(x,y)\,p(dy) : p \in m_f(t,x)\}$.

Now we turn to the case $b_u = \infty$. Let h be the affine function which is coincide with $\widetilde{f}_x$ on $[a_u,\infty)$. Since $\widetilde{f}_x$ is bounded from below, so is h. It follows that $h' \geqslant 0$. Hence, $z \mapsto \widetilde{f}_x(z) + t\alpha(z/t)$ is strictly increasing on $[a_u,\infty)$. On the other hand, $u \in \widetilde{m}_f(t,x)$ implies that u achieves the minimum of function $z \mapsto \widetilde{f}_x(z) + t\alpha(z/t)$. Thus $u = a_u$ and there exists $y \in \mathcal{X}$ such that $d(x,y) = u$ and $f(y) = f_x(u) = \widetilde{f}_x(u)$ by lemma 2.2.17. Again by Item (i) and by definition of $\widetilde{m}_f(t,x)$ we deduce that the probability $p := \delta_y \in m_f(t,x)$ and $u \in \{\int d(x,y)\,p(dy) : p \in m_f(t,x)\}$.

Let us prove Item (ii). By definition, $\widetilde{m}_f(t,x)$ is the set where the convex function $F(v) = \widetilde{f}_x(v) + t\alpha(v/t)$ attains its minimum on $\mathbb{R}^+$. Therefore $\widetilde{m}_f(t,x)$ is an interval. Suppose that $u_1 < u_2$ are in $\widetilde{m}_f(t,x)$, then F is constant on $[u_1,u_2]$. Since both functions $\widetilde{f}_x$ and $t\alpha(\,\cdot\,/t)$ are convex, this easily implies that these functions $\widetilde{f}_x$ and $t\alpha(\,\cdot\,/t)$ are both affine on $[u_1,u_2]$. In particular, $\alpha'(u/t)$ is constant on $[u_1,u_2]$.

It follows that $\beta(u_2/t) = (u_2/t)\alpha'(u_2/t) - \alpha(u_2/t) = (u_2/t)\alpha'(u_1/t) - \alpha(u_1/t) - \alpha'(u_1/t)(u_2 - u_1)/t = \beta(u_1/t)$. This shows that $\beta(\,\cdot\,/t)$ is constant on $\widetilde{m}_f(t,x)$.

Let us turn to the proof of Item (iv). According to [47, Theorem 1.10] (which applies since $\widetilde{f}_x : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is bounded from below and, according to Lemma 2.2.15, lower-semicontinuous), it holds

$$\frac{dQ_t\widetilde{f}_x(0)}{dt_+} = -\beta\left(\frac{\max \widetilde{m}_f(t,x)}{t}\right),$$

and

$$\frac{dQ_t\widetilde{f}_x(0)}{dt_-} = -\beta\left(\frac{\min \widetilde{m}_f(t,x)}{t}\right),$$

where $d/dt_\pm$ stands for the right and left derivatives. According to Item (ii) the function $\beta(\,\cdot\,/t)$ is constant on $\widetilde{m}_f(t,x)$. Therefore, the left and the right derivatives of $t \mapsto Q_t\widetilde{f}_x(0)$ are equal, and so the function is actually differentiable in t. According to Item (i), $\widetilde{Q}_t f(x) = Q_t\widetilde{f}_x(0)$ and, according to Item (ii), $\{\int d(x,y)\,p(dy) : p \in m_f(t,x)\} \subset \widetilde{m}_f(t,x)$ which proves (2.2.21). $\square$

Let us mention an interesting consequence of the proof of Item (ii). Let us denote by $\mathcal{P}_2(X)$ the set of probability measures on X charging at most two points :

$$\mathcal{P}_2(X) := \{(1-s)\delta_x + s\delta_y :\ s \in [0,1],\ x, y \in \mathcal{X}\}\,.$$

Proposition 2.2.22. *Let $f : \mathcal{X} \to \mathbb{R}$ be a lower semicontinuous function bounded from below. Then*

$$\widetilde{Q}_t f(x) = \inf\left\{\int f\,dp + t\alpha\left(\frac{\int d(x,y)\,p(dy)}{t}\right) : p \in \mathcal{P}_2(X)\right\}.$$

Démonstration. It is enough to show that for all $\varepsilon > 0$, $m_f^\varepsilon(t,x) \cap \mathcal{P}_2(X) \neq \emptyset$ (recall the definition of $m^\varepsilon(t,x)$ given in Item (ii) of Theorem 2.2.19). Actually, this follows immediately from the argument given in the proof of Item (ii). Indeed, we showed there that for all $u \in \widetilde{m}_f(t,x)$ there exists $p \in \mathcal{P}_2(X) \cap m_f^\varepsilon(t,x)$ such that $\int d(x,y)\,p(dy) = u$. $\square$

2.2.5 Change of scale

Since the we have always the term $\widetilde{Q}_1$ in the Kantorovich duality formula for the transport cost, I would like to detail a lemma for the change of scale for some special cost. Let the cost function $\alpha(x) = ax^2$, then it is easy to show that $\widetilde{Q}_1(tf)(x) = t\widetilde{Q}_t f(x)$ for all $t > 0$.

Now define a quadratic-linear cost function α_a^h, $a, h > 0$ with

$$\alpha_a^h(x) = \begin{cases} ax^2 & x \leqslant h \\ 2ax - ah^2 & x > h. \end{cases}$$

then we have the following lemma.

Lemma 2.2.23 (Change of scale)**.** *Let f be an l-Lipschitz function and $\widetilde{Q}_t$ be the inf-convolution for a quadratic-linear cost function α_a^h, $a, h > 0$. Then, for all $x \in \mathcal{X}$ and all $t < (ah)/l$, it holds $\widetilde{Q}_1(tf)(x) = t\widetilde{Q}_t f(x)$.*

proof. Fix $t < ah/l$ and $x \in \mathcal{X}$. For all $p \in m_{tf}(1, x)$ (defined in (2.2.9)) we have by Item (i) of Theorem 2.2.19

$$\int tf(y)\, p(dy) + \alpha_a^h \left(\int d(x, y)\, p(dy) \right) = \widetilde{Q}_1(tf)(x) \leqslant tf(x).$$

Hence

$$\begin{aligned}\alpha_a^h \left(\int d(x, y)\, p(dy) \right) &\leqslant t \int f(x) - f(y)\, p(dy) \leqslant t|\widetilde{\nabla} f|(x) \int d(x, y)\, p(dy) \\ &\leqslant tl \int d(x, y)\, p(dy) \leqslant ah \int d(x, y)\, p(dy),\end{aligned}$$

where we used that $f(x) - f(y) \leqslant |\widetilde{\nabla} f|(x) d(x, y)$ and the fact that f is l-Lipschitz. Since for quadratic-linear cost $\alpha_a^h(u) \leqslant ahu$ if and only if $u \leqslant h$, the above inequality implies that $\int d(x, y)\, p(dy) \leqslant h$ and that $\alpha_a^h \left(\int d(x, y)\, p(dy) \right) = a \left(\int d(x, y)\, p(dy) \right)^2$. Therefore

$$\widetilde{Q}_1(tf)(x) = \inf_{p \in \mathcal{P}(\mathcal{X})} \left\{ \int tf\, dp + a \left(\int d(x, y)\, p(dy) \right)^2 \right\}.$$

Similarly for all $q \in m_f(t, x)$ it holds

$$\int f(y)\ q(dy) + t\alpha_a^h \left(\frac{\int d(x, y)\, q(dy)}{t} \right) \leqslant f(x).$$

Therefore

$$\begin{aligned}\alpha_a^h \left(\frac{\int d(x, y)\, q(dy)}{t} \right) &\leqslant \int f(x) - f(y)\, q(dy) \leqslant |\widetilde{\nabla} f|(x) \int d(x, y)\, q(dy) \\ &\leqslant l \int d(x, y)\, p(dy) \leqslant \frac{ah}{t} \int d(x, y)\, q(dy).\end{aligned}$$

This (due to the specific shape of the quadratic-linear cost) leads to $\int d(x, y)\, q(dy)/t \leqslant h$ and $\alpha_a^h \left(\frac{\int d(x,y)\, q(dy)}{t} \right) = a \left(\frac{\int d(x,y)\, q(dy)}{t} \right)^2$. Therefore,

$$\widetilde{Q}_t f(x) = \inf_{p \in \mathcal{P}(\mathcal{X})} \left\{ \int f\, dp + \frac{a}{t} \left(\int d(x, y)\, p(dy) \right)^2 \right\}.$$

As a conclusion,

$$\begin{aligned}t\widetilde{Q}_t f(x) &= t \inf_{q \in \mathcal{P}(\mathcal{X})} \left\{ \int f\, dq + \frac{a}{t} \left(\int d(x, y)\, p(dy) \right)^2 \right\} \\ &= \inf_{p \in \mathcal{P}(\mathcal{X})} \left\{ \int tf\, dp + a \left(\int d(x, y)\, p(dy) \right)^2 \right\} = \widetilde{Q}_1(tf)(x).\end{aligned}$$

□

We will use this lemma to deal with the connection with Poincaré inequality and transport entropy inequality in the next chapter.

2.3 Hamilton-Jacobi equation

In this section, we will show that the modified Hopf-Lax formula (2.2.4) satisfies some Hamilton-Jacobi type equation. In order to illustrate the idea, imagine in the first place that we can do all the manipulations that we want, functions are continuous and bounded, infimums and supremums are all minimums and maximums. Then informally, we can rewrite the operator $\widetilde{Q}$ as follows :

$$\widetilde{Q}_t f(x) = \inf_{p \in \mathcal{P}(\mathcal{X})} \left\{ \langle f, p \rangle + \frac{1}{2t} W_1^2(\delta_x, p) \right\}$$

where we consider the integration of f with respect to measure p as a scalar product $\langle f, p \rangle$, and W_1 denotes the L^1 Wasserstein distance.

Intuitively, by analogy with the classical case, the time derivative of $\widetilde{Q}_t f(x)$ should be

$$-\frac{1}{2t^2} W_1^2(\delta_x, p_x).$$

The core difficulty is to find a proper space derivative.

In a discrete space, an estimation of a space derivative is essentially some kind of estimation of difference. Now take $x, y \in \mathcal{X}$ and let us try to say something about $\widetilde{Q}_t f(x) - \widetilde{Q}_t f(y)$. Denote by p_x and p_y the probability measures such that the infimum defined in $\widetilde{Q}_t f(x)$ and $\widetilde{Q}_t f(x)$ are reached respectively. One can write

$$\widetilde{Q}_t f(x) - \widetilde{Q}_t f(y) = \langle f, p_x - p_y \rangle + \frac{1}{2t} \left(W_1^2(\delta_x, p_x) - W_1^2(\delta_y, p_y) \right). \tag{2.3.1}$$

Now take an interpolation between p_x and p_y :

$$p_\lambda := \lambda p_y + (1 - \lambda) p_x.$$

Then for all $\lambda \in [0, 1]$, one has

$$G(\lambda) := \langle f, p_\lambda \rangle + \frac{1}{2t} W_1^2(\delta_x, p_\lambda) \leqslant \widetilde{Q}_t f(x) = G(0).$$

We deduce that $G'(0) \geqslant 0$, which is equivalent to,

$$\langle f, p_y - p_x \rangle + \frac{1}{2t} 2 W_1(\delta_x, p_x) \left(W_1(\delta_x, p_x) - W_1(\delta_x, p_y) \right) \geqslant 0,$$

plug it in to equation (2.3.1), we get

$$\begin{aligned}
\widetilde{Q}_t f(x) - \widetilde{Q}_t f(y) &\leqslant \frac{1}{2t}\Big(W_1^2(\delta_x,p_x) - W_1^2(\delta_y,p_y) - 2W_1(\delta_x,p_x)\left(W_1(\delta_x,p_x) - W_1(\delta_x,p_y)\right)\Big) \\
&= \frac{1}{2t}\Big(2W_1(\delta_x,p_x)W_1(\delta_x,p_y) - W_1^2(\delta_x,p_x) - W_1^2(\delta_y,p_y)\Big) \\
&\leqslant \frac{1}{2t}\left(2W_1(\delta_x,p_x)W_1(\delta_x,p_y) - 2W_1(\delta_x,p_x)W_1(\delta_y,p_y)\right) \\
&\leqslant \frac{1}{t} W_1(\delta_x,p_x) d(x,y)
\end{aligned}$$

Here the last step we used the triangle inequality

$$W_1(\delta_x,p_y) - W_1(\delta_y,p_y) \leqslant W_1(\delta_x,\delta_y) = d(x,y).$$

As we had mentioned before, the time derivative $\frac{\partial}{\partial t}\widetilde{Q}_t f(x)$ should be $-\frac{1}{2t^2}W_1^2(\delta_x,p_x)$. Thus, the following holds for all x,y :

$$\frac{\partial}{\partial t}\widetilde{Q}_t f(x) + \frac{1}{2}\left(\frac{[\widetilde{Q}_t f(x) - \widetilde{Q}_t f(y)]_+}{d(x,y)}\right)^2 \leqslant 0.$$

Taking the supremum over $y \in \mathcal{X}$, and defining $|\widetilde{\nabla} f|(x) := \sup_y \frac{[f(y)-f(x)]_-}{d(x,y)}$, we have a Hamilton-Jacobi type inequality :

$$\frac{\partial}{\partial t}\widetilde{Q}_t f + \frac{1}{2}|\widetilde{\nabla}\widetilde{Q}_t f|^2 \leqslant 0.$$

Now we are going to prove it rigorously.

Recall that α^* the Fenchel-Legendre transform of α.

Proposition 2.3.2. *Let f be a lower semi-continuous function bounded from below.*

(i) For all $x \in \mathcal{X}$, all $t > 0$ and all $p \in m_f(t,x)$, it holds

$$|\widetilde{\nabla}\widetilde{Q}_t f|(x) \leqslant \alpha'\left(\frac{\int d(x,y)\,p(dy)}{t}\right). \tag{2.3.3}$$

(ii) Assume that f reaches its minimum at a unique point $x_o \in \mathcal{X}$, then for all $x \in \mathcal{X} \setminus \{x_o\}$, it holds

$$|\widetilde{\nabla} f|(x) = |\widetilde{f}_x'(0)|, \tag{2.3.4}$$

and $|\widetilde{\nabla} f|(x_o) = 0$. Moreover, if f reaches its minimum in two or more points, or if f does not reach its minimum, then (2.3.4) holds for all $x \in \mathcal{X}$.

Remark 2.3.5. *Observe that, if f reaches its minimum at a unique point x_o, then it could be that $\widetilde{f}_{x_o}'(0) \neq 0$. For example consider, on $X = \mathbb{R}^+$, $f(x) = x$ that reaches its minimum at $x_o = 0$. Trivially $\widetilde{f}_{x_0}(x) = x$ for all $x \in \mathcal{X}$ so that $\widetilde{f}_{x_o}'(0) = 1$. Hence, there is no hope for (2.3.4) to be true at x_o in general.*

proof. First let us prove item (i). Consider y such that $\widetilde{Q}_t f(y) < \widetilde{Q}_t f(x)$ (if there is no such y, then $|\widetilde{\nabla \widetilde{Q}_t f}|(x) = 0$ and there is nothing to prove). By Lemma 2.2.10, there exist $p_o \in m_f(t,x), p_1 \in m_f(t,y)$ and according to Item (ii) of Theorem 2.2.19, $u_o = \int d(x,z)\, p_0(dz) \in \widetilde{m}_f(t,x)$ and $u_1 = \int d(y,z)\, p_1(dz) \in \widetilde{m}_f(t,y)$ and it holds

$$\widetilde{Q}_t f(x) = \int f\, dp_o + t\alpha(u_o/t) \qquad \text{and} \qquad \widetilde{Q}_t f(y) = \int f\, dp_1 + t\alpha(u_1/t). \tag{2.3.6}$$

Now, set $p_\lambda := (1-\lambda)p_o + \lambda p_1$, $\lambda \in [0,1]$, $u := \int d(x,z)\, p_1(dz)$ and observe that, by definition of $\widetilde{Q}_t$,

$$\widetilde{Q}_t f(x) \leqslant \int f\, dp_\lambda + t\alpha\left(\frac{\int d(x,z)\, p_\lambda(dz)}{t}\right) = \int f dp_\lambda + t\alpha\left(\frac{\lambda u + (1-\lambda)u_o}{t}\right).$$

Since the latter holds for all $\lambda \in [0,1]$ the function

$$g\colon \lambda \mapsto \int f\, dp_\lambda + t\alpha\left(\frac{\lambda u + (1-\lambda)u_o}{t}\right) - \widetilde{Q}_t f(x)$$

is always non-negative. Therefore, since $g(0) = 0$ and

$$g'(0) = (\int f\, dp_1 - \int f\, dp_o) + (u - u_o)\alpha'(u_o/t) \geqslant 0$$

which ensures that

$$\int f\, dp_o - \int f\, dp_1 \leqslant (u - u_o)\alpha'(u_o/t). \tag{2.3.7}$$

On the other hand, since $d(x,z) \leqslant d(x,y) + d(y,z)$, it holds

$$u = \int d(x,z)\, p_1(dz) \leqslant \int (d(x,y) + d(y,z))\, p_1(dz) \leqslant u_1 + d(x,y).$$

As a consequence, it holds

$$u - u_1 \leqslant d(x,y). \tag{2.3.8}$$

Thanks to (2.3.6), (2.3.7) and (2.3.8) together with the fact that $\alpha' \geqslant 0$, for all y such that $\widetilde{Q}_t f(x) > \widetilde{Q}_t f(y)$, it holds

$$\begin{aligned}
&[\widetilde{Q}_t f(y) - \widetilde{Q}_t f(x)]_- = \widetilde{Q}_t f(x) - \widetilde{Q}_t f(y) \\
&= \int f\, dp_o - \int f\, dp_1 + t\left(\alpha\left(\frac{u_o}{t}\right) - \alpha\left(\frac{u_1}{t}\right)\right) \\
&\leqslant (u - u_o)\alpha'(\frac{u_o}{t}) + t\left(\alpha\left(\frac{u_o}{t}\right) - \alpha\left(\frac{u_1}{t}\right)\right) \\
&\leqslant d(x,y)\alpha'\left(\frac{u_o}{t}\right) + (u_1 - u_o)\alpha'\left(\frac{u_o}{t}\right) + t\left(\alpha\left(\frac{u_o}{t}\right) - \alpha\left(\frac{u_1}{t}\right)\right).
\end{aligned}$$

Therefore, by convexity of α, we conclude that

$$(u_1 - u_o)\alpha'\left(\frac{u_o}{t}\right) + t\left(\alpha\left(\frac{u_o}{t}\right) - \alpha\left(\frac{u_1}{t}\right)\right) \geqslant 0$$

and in turn that for all $x, y \in \mathcal{X}$,

$$[\widetilde{Q}_t f(y) - \widetilde{Q}_t f(x)]_- \leqslant d(x,y)\alpha'\left(\frac{u_o}{t}\right)$$

which leads to the expected result by taking the supremum over $y \neq x$.

Now we turn to the proof of Item (ii). Fix $x \in \mathcal{X}$. The proof relies on the existence of a point $y \neq x$ such that $f(y) \leqslant f(x)$. Such an existence is guaranteed for all $x \in \mathcal{X}$ (resp. for all $x \in \mathcal{X} \setminus \{x_o\}$) when f does not reach its minimum or reaches its minimum in more than two points (resp. when f reaches its minimum at a unique point x_o). Given such a point y, by definition of $\widetilde{f}_x$, we have

$$\widetilde{f}_x(0) = f(x) \geqslant f(y) \geqslant \widetilde{f}_x\left(d(x,y)\right).$$

Thanks to the convexity of $\widetilde{f}_x$, the slope function

$$u \mapsto \frac{\widetilde{f}_x(u) - \widetilde{f}_x(0)}{u}$$

is non-decreasing. It follows that

$$\widetilde{f}_x{}'(0) = \lim_{u \to 0^+} \frac{\widetilde{f}_x(u) - \widetilde{f}_x(0)}{u} = \inf_{u>0} \frac{\widetilde{f}_x(u) - \widetilde{f}_x(0)}{u} \leqslant \frac{\widetilde{f}_x\left(d(x,y)\right) - \widetilde{f}_x(0)}{d(x,y)} \leqslant 0.$$

Taking the absolute value, we get

$$|\widetilde{f}_x{}'(0)| = \sup_{u>0} \frac{\widetilde{f}_x(0) - \widetilde{f}_x(u)}{u}.$$

Observe that, according to Lemma 2.2.15, for all $u > 0$, $\widetilde{f}_x(u) = \inf \int f\,dp$ where the infimum is running over all $p \in \mathcal{P}_2(X)$ such that $\int d(x,\cdot)\,dp = u$. Hence, setting $p = \lambda\delta_{y_1} + (1-\lambda)\delta_{y_2}$, $y_1, y_2 \in \mathcal{X}$, $\lambda \in [0,1]$, we have (recall that $\widetilde{f}_x(0) = f(x)$)

$$\begin{aligned}
|\widetilde{f}_x{}'(0)| &= \sup_{u>0} \frac{\widetilde{f}_x(0) - \widetilde{f}_x(u)}{u}. \\
&= \sup_{u>0} \sup_{\substack{y_1,y_2 \in \mathcal{X}, \lambda \in [0,1] s.t \\ \lambda d(x,y_1) + (1-\lambda)d(x,y_2) = u}} \frac{f(x) - (\lambda f(y_1) + (1-\lambda)f(y_2))}{u} \\
&= \sup_{y_1,y_2 \in \mathcal{X}, \lambda \in [0,1]} \frac{\lambda(f(x) - f(y_1)) + (1-\lambda)(f(x) - f(y_2))}{\lambda d(x,y_1) + (1-\lambda)d(x,y_2)} \\
&= \sup_{y \neq x} \frac{f(x) - f(y)}{d(x,y)} = |\widetilde{\nabla} f|(x),
\end{aligned}$$

where the last equality comes from the fact that the function $\lambda \mapsto \frac{\lambda a+(1-\lambda)b}{\lambda c+(1-\lambda)d}$ (with $c, d > 0$ and $a, b \in \mathbb{R}$) is monotone on $[0,1]$. This proves (2.3.4). That $|\widetilde{\nabla} f|(x_o) = 0$ is a direct consequence of the definition of the gradient. □

Now we are in position to state the Hamilton-Jacobi equation which is satisfied by $\widetilde{Q}$.

Theorem 2.3.9. *Let $f\colon \mathcal{X} \to \mathbb{R}$ be a lower semi-continuous function bounded from below. Then, for all $x \in \mathcal{X}$, it holds*

(i) For all $t > 0$, $\frac{\partial}{\partial t}\widetilde{Q}_t f(x) + \alpha^*\left(|\widetilde{\nabla}\widetilde{Q}_t f|(x)\right) \leqslant 0$.

(ii) Assume that α^* is well define on $[0,l)$, (i.e $\forall x \in [0,l)$, $\alpha^*(l) < \infty$.) Then for all x such that $|\widetilde{\nabla} f|(x) \in [0,l)$, $\lim_{t\to 0} \widetilde{Q}_t f = f$ and it holds

$$\frac{\partial}{\partial t}\widetilde{Q}_t f(x)|_{t=0} + \alpha^*\left(|\widetilde{\nabla} f|(x)\right) = 0.$$

Remark 2.3.10. *In Item (ii), if $\lim_{x\to\infty} \alpha(x)/x = \infty$, we can take $l = \infty$, then the latter equation holds for almost every $x \in \mathcal{X}$.*

If f is $l - \varepsilon$-lipschiz then $|\widetilde{\nabla} f|(x) < l$ and the latter equality holds. Moreover, if there exists h such that $\alpha'(h) = l$, then the latter holds for all x such that $|\widetilde{\nabla} f|(x) \in [0,l]$.

proof. We will first prove Item (i). On the one hand, by Theorem 2.2.19, for all $t > 0$, it holds

$$\frac{\partial}{\partial t}\widetilde{Q}_t f(x) = -\beta\left(\frac{u_o}{t}\right), \qquad x \in \mathcal{X}$$

where $u_o \in \widetilde{m}_f(t,x)$. On the other hand, since α^* is non-decreasing, Proposition 2.3.2 ensures that

$$\alpha^*\left(|\widetilde{\nabla}\widetilde{Q}_t f|(x)\right) \leqslant \alpha^*\left(\alpha'\left(\frac{u_o}{t}\right)\right).$$

In order to conclude, it is enough to observe that, the function $G := y \mapsto y\alpha'\left(\frac{u_o}{t}\right) - \alpha(y)$ is a concave function and $G'\left(\frac{u_o}{t}\right) = 0$. Hence,

$$\alpha^*\left(\alpha'\left(\frac{u_o}{t}\right)\right) = \sup_{y\in\mathbb{R}}\left\{y\alpha'\left(\frac{u_o}{t}\right) - \alpha(y)\right\} = \frac{u_o}{t}\alpha'\left(\frac{u_o}{t}\right) - \alpha\left(\frac{u_o}{t}\right) = \beta\left(\frac{u_o}{t}\right).$$

Now we turn to the proof of Item (ii). If $x = x_o$ is a minimum of f (if any), then (observe that $\widetilde{Q}_t f(x_o) = f(x_o)$ for all $t > 0$) it is easy to see that

$$\frac{\partial}{\partial t}\widetilde{Q}_t f(x)|_{t=0} := \lim_{t\to 0}\frac{\widetilde{Q}_t f(x) - f(x)}{t} = \alpha^*\left(|\widetilde{\nabla} f|(x)\right) = 0$$

and the claim follows. For the remaining of the proof we assume that $x \in \mathcal{X}$ is not a minimum of f. Thanks to Theorem 2.2.19, for all $t > 0$, it holds

$$\frac{\widetilde{Q}_t f(x) - f(x)}{t} = \frac{Q_t\widetilde{f}_x(0) - \widetilde{f}_x(0)}{t} = \frac{\widetilde{f}_x(u) - \widetilde{f}_x(0)}{t} + \alpha\left(\frac{u}{t}\right),$$

where $u \in \widetilde{m}_f(t,x)$.

Let us prove that $u > 0$. Since x is not a minimum of f, there exists $y \in \mathcal{X}$ such that $f(y) < f(x)$. Fix $t > 0$, by the very definition of $\widetilde{Q}_t$, for all $\lambda \in [0,1]$, choosing

$$p = (1-\lambda)\delta_x + \lambda\delta_y,$$

it holds that

$$\widetilde{Q}_t f(x) \leqslant (1-\lambda)f(x) + \lambda f(y) + t\alpha\left(\frac{\lambda d(x,y)}{t}\right).$$

Define

$$G : \lambda \in [0,1] \mapsto (1-\lambda)f(x) + \lambda f(y) + t\alpha\left(\frac{\lambda d(x,y)}{t}\right).$$

Then

$$G'(0) = \alpha'(0)d(x,y) + f(y) - f(x) = f(y) - f(x) < 0.$$

Thus, there exist $\lambda \in (0,1)$ such that

$$\widetilde{Q}_t f(x) \leqslant G(\lambda) < G(0) = f(x).$$

Hence

$$\widetilde{f}_x(u) \leqslant \widetilde{Q}_t f(x) < f(x) = \widetilde{f}_x(0)$$

and therefore $u > 0$.

According to Lemma 2.2.15, for all $x \in \mathcal{X}$, $\widetilde{f}_x$ is convex and continuous on $\widetilde{I}_x$. It follows that $\frac{\widetilde{f}_x(u)-\widetilde{f}_x(0)}{u} \geqslant \widetilde{f}_x{}'(0)$. Since $\widetilde{f}_x(u) \leqslant Q_t\widetilde{f}_x(0) \leqslant \widetilde{f}_x(0)$, we have that $\frac{\widetilde{f}_x(u)-\widetilde{f}_x(0)}{u}$ is non-positive and $\frac{\widetilde{f}_x(0)-\widetilde{f}_x(u)}{u} \leqslant |\widetilde{f}_x{}'(0)|$. Hence,

$$\begin{aligned}\frac{f(x)-\widetilde{Q}_t f(x)}{t} &= \frac{\widetilde{f}_x(0)-\widetilde{f}_x(u)}{t} - \alpha\left(\frac{u}{t}\right) = \frac{\widetilde{f}_x(0)-\widetilde{f}_x(u)}{u}\frac{u}{t} - \alpha\left(\frac{u}{t}\right)\\ &\leqslant \alpha^*\left(\frac{\widetilde{f}_x(0)-\widetilde{f}_x(u)}{u}\right) \leqslant \alpha^*\left(|\widetilde{f}_x{}'(0)|\right)\end{aligned}$$

where the last inequality comes from the fact that α^* is non-decreasing. This leads to

$$\liminf_{t\to 0} \frac{\widetilde{Q}_t f(x) - f(x)}{t} \geqslant -\alpha^*\left(|\widetilde{f}_x{}'(0)|\right), \tag{2.3.11}$$

by passing to the limit.

Next, we prove that $\limsup_{t\to 0} \frac{\widetilde{Q}_t f(x)-f(x)}{t} \leqslant -\alpha^*\left(|\widetilde{f}_x{}'(0)|\right)$. By convexity of $\widetilde{f}_x$, for all $h \in (0,u)$, it holds

$$\frac{\widetilde{f}_x(u)-\widetilde{f}_x(0)}{u} \leqslant \frac{\widetilde{f}_x(u)-\widetilde{f}_x(u-h)}{h}. \tag{2.3.12}$$

On the other hand, since (by definition of u)

$$\widetilde{f}_x(u) + t\alpha\left(\frac{u}{t}\right) \leqslant \widetilde{f}_x(u-h) + t\alpha\left(\frac{u-h}{t}\right),$$

we have

$$\frac{\widetilde{f}_x(u) - \widetilde{f}_x(u-h)}{h} \leqslant \frac{t\left(\alpha\left(\frac{u-h}{t}\right) - \alpha\left(\frac{u}{t}\right)\right)}{h}. \tag{2.3.13}$$

According to (2.3.12) and (2.3.13), for all $h \in (0,u)$, it holds :

$$\frac{\widetilde{Q}_t f(x) - f(x)}{t} = \frac{\widetilde{f}_x(u) - \widetilde{f}_x(0)}{t} + \alpha\left(\frac{u}{t}\right) \leqslant \frac{u}{t}\frac{\alpha\left(\frac{u-h}{t}\right) - \alpha\left(\frac{u}{t}\right)}{h/t} + \alpha\left(\frac{u}{t}\right).$$

Let h goes to 0, we get that

$$\frac{\widetilde{Q}_t f(x) - f(x)}{t} \leqslant -\frac{u}{t}\alpha'\left(\frac{u}{t}\right) + \alpha\left(\frac{u}{t}\right) = -\beta\left(\frac{u}{t}\right) = -\alpha^*\left(\alpha'\left(\frac{u}{t}\right)\right) \tag{2.3.14}$$

where we recall that β is defined in Section 2.2. Hence, it is enough to prove that $\lim_{t\to 0}\alpha'\left(\frac{u}{t}\right) = |\widetilde{f}_x'(0)|$. Since $\widetilde{f}_x$ is convex, it is right and left differentiable at every point. Hence taking the left derivative of $v \mapsto \widetilde{f}_x(v) + t\alpha\left(\frac{v}{t}\right)$, for all $t \in \mathbb{R}^+$ and all $u \in \widetilde{m}_f(t,x)$, we have

$$\alpha'\left(\frac{u}{t}\right) \leqslant -\frac{d}{du_-}\widetilde{f}_x(u).$$

Let $l := \lim_{x\to\infty}\alpha'(x)$, it is easy to see that $\alpha^*(x) < \infty$ when $x \leqslant l$ and $= \infty$ when $x > l$. By Item (ii) of Proposition 2.3.2 and convexity of $\widetilde{f}_x$ and Equation (2.3.4), there exists $h_1 < l$ such that the following holds :

$$\alpha'\left(\frac{u}{t}\right) \leqslant -\frac{d}{du_-}\widetilde{f}_x'(u) \leqslant -\widetilde{f}_x'(0) = |\widetilde{\nabla} f|(x) \leqslant \alpha'(h_1).$$

By convexity of α, the latter inequality leads to $\frac{u}{t} \leqslant h_1$ for all $t > 0$. We conclude from the above argument that $u \in \widetilde{m}(t,x)$ goes to 0 as t goes to 0.

Now, taking the right derivative of $v \mapsto \widetilde{f}_x(v) + t\alpha\left(\frac{v}{t}\right)$, for all $t \in \mathbb{R}^+$ and all $u \in \widetilde{m}_f(t,x)$, we have

$$\alpha'\left(\frac{u}{t}\right) \geqslant -\frac{d}{du_+}\widetilde{f}_x(u).$$

Since $\lim_{u\to 0}\frac{d}{du_+}\widetilde{f}_x'(u) = \widetilde{f}_x'(0)$ and using the monotonicity and the (right) continuity of α^* when t goes to 0, we have thanks to 2.3.14

$$\limsup_{t\to 0}\frac{\widetilde{Q}_t f(x) - f(x)}{t} \leqslant -\alpha^*\left(|\widetilde{f}_x'(0)|\right) \tag{2.3.15}$$

This combined with 2.3.11 and Proposition2.3.2 leads to the desired result.

□

2.3.1 Examples

Example of $\mathbb{R}^n$, equality case

The equality of Hamilton-Jacobi system can hold in some specific cases.

Let $\mathcal{X} = \mathbb{R}^n$ equipped with norm euclidian $\|\cdot\|_2$). Assume that the cost function α is convex, of class C^1 and $\alpha'(x)$ goes to ∞ as x goes to ∞. Then for all function f convex and of class C^2 and all $t \geqslant 0$, it holds

$$\frac{\partial}{\partial t}\widetilde{Q}_t f + \alpha^*(|\widetilde{\nabla}\widetilde{Q}_t f|) = 0.$$

We illustrate this with the following proposition.

Proposition 2.3.16. *Assume that $\mathcal{X} = \mathbb{R}^n$ equipped with a distance d coming from a norm $\|\cdot\|$. Then, for all $f \colon \mathbb{R}^n \mapsto \mathbb{R}$ convex and bounded from below, $\widetilde{Q}_t f = Q_t f$.*

proof. By convexity of f and of the norm, Jensen's Inequality and the monotonicity of α imply that, for all $p \in \mathcal{P}(\mathbb{R}^n)$ such that $\int \|x\|\, p(dx)$ is finite, it holds

$$\int f(y)\, p(dy) + t\alpha\left(\frac{\int \|x-y\|\, p(dy)}{t}\right) \geqslant f\left(\int y\, p(dy)\right) + t\alpha\left(\frac{1}{t}\|x - \int y\, p(dy)\|\right).$$

Hence, setting $z := \int y\, p(dy) \in \mathbb{R}^n$ and optimizing we get

$$\begin{aligned}\widetilde{Q}_t f(x) &= \inf_{p \text{ s.t. } \int \|x\|\, p(dx) < \infty} \left\{ \int f\, dp + t\alpha\left(\frac{\int \|x-y\|\, p(dy)}{t}\right)\right\} \\ &\geqslant \inf_{z \in \mathbb{R}^n} \left\{ f(z) + t\alpha\left(\frac{\|x-z\|}{t}\right)\right\} = Q_t f(x)\end{aligned}$$

which leads to the desired result. □

According to the latter Proposition, we have the following equality case :

Let $\alpha(x) = x^2/2$, $x \in \mathbb{R}^+$ and $f \colon \mathbb{R}^n \to \mathbb{R}$ convex. Then for all $t \geqslant 0$,

$$\frac{\partial}{\partial t_+}\widetilde{Q}_t f(x) + \alpha^*(|\widetilde{\nabla}\widetilde{Q}_t f|(x)) = 0 = 0,$$

i.e. there is actually equality in Item (i) of Theorem 2.3.9.

To prove this fact, we observe first that, since $\lim_{h\to\infty} \alpha'(h) = \infty$, the thesis follows from Item (ii) of Theorem 2.3.9 when $t = 0$. For $t > 0$, since f is convex, Proposition 2.3.16 ensures that $\widetilde{Q}_t f = Q_t f$. Moreover, for all convex function f, $Q_t f$ is a convex function which guarantees that $|\widetilde{\nabla} Q_t f| = |\nabla Q_t f|$ (where $|\nabla \cdot|$ is the Euclidean length of the usual gradient). Hence, the claim follows from the classical Hamilton-Jacobi equation that precisely asserts that for $t > 0$, $\frac{\partial}{\partial t} Q_t f(x) + \frac{1}{2}|\nabla Q_t f|^2(x) = 0$.

Example of the two points space $\{0,1\}$, strict inequality case

Let $\alpha(x) = x^2/2$ and $X = \{0,1\}$ (the graph consisting of two points). Consider f such that $f(0) = 1$ and $f(1) = 0$. It is easy to see that for $t \in (0,1)$, $\widetilde{Q}_t f(0) = 1 - \frac{t}{2}$ and $\widetilde{Q}_t f(1) = 0$. It leads to $|\widetilde{\nabla}\widetilde{Q}_t f|(0) = 1 - \frac{t}{2}$ and $\frac{\partial}{\partial t}\widetilde{Q}_t f(0) = -\frac{1}{2}$. Thus, for all $t \in (0,1)$, $\frac{\partial}{\partial t}\widetilde{Q}_t f(0) + \frac{1}{2}|\widetilde{\nabla}\widetilde{Q}_t f|^2(0) < 0$, *i.e.* the inequality in Item (i) of Theorem 2.3.9 is strict. We observe that, more generally, the same conclusion holds as soon as $\mathcal{X}$ has at least one isolated point x_o (take f with $f(x_o) = 0$ and $f(y) = 1$ for all $y \neq x_o$).

2.4 Equilibrium case

In this section, we focus on the equilibrium case : the equation $\omega(x) = \inf_{y:(y,x)\in E}\{\omega(y) + k(y,x)\}$. The motivation is from a consideration of the following large deviation problem.

2.4.1 Motivation and assumptions

We consider an ε–parametrized discrete–time Markov chain $X^{(\varepsilon)} = (X_n^{(\varepsilon)})_{n\geqslant 0}$ with finite state space V. We write $p_\varepsilon(\cdot,\cdot)$ for the transition probabilities, i.e. $p_\varepsilon(x,y) = \mathbb{P}(X_{n+1}^{(\varepsilon)} = y | X_n^{(\varepsilon)} = x)$.

We assume the following properties :

(A1) There exists an ε–independent subset $E \subset V \times V$ such that $p_\varepsilon(x,y) > 0$ if and only if $(x,y) \in E$;

(A2) For any $\varepsilon > 0$ the Markov chain $X^{(\varepsilon)}$ is irreducible;

(A3) There exists a function $k : e = (x,y) \in E \to \mathbb{R}_+$ such that

$$\lim_{\varepsilon\downarrow 0} -\varepsilon \log p_\varepsilon(x,y) = k(x,y) \qquad \forall (x,y) \in E\,. \tag{2.4.1}$$

We introduce the oriented graph $G = (V,E)$ having vertex set V and edge set E. Then, due to assumption (A1), assumption (A2) is equivalent to the property that the graph G is strongly connected, i.e. for any $x,y \in V$ there is an oriented path with edges in E from x to y.

Since G is strongly connected and V is finite, the Markov chain has a unique invariant distribution, that we denote by $\pi^{(\varepsilon)}$. Recall that $\pi^{(\varepsilon)}(x) > 0$ for any x.

The following static large deviation principle holds (cf. [39]) : For any $x \in V$ the limit $\lim_{\varepsilon\downarrow} -\varepsilon \log \pi^{(\varepsilon)}(x) =: W(x)$ exists, moreover $W(x) \geqslant 0$ for any $x \in V$ and $\min W(x) = 0$.

It is simple to check that the rate function W satisfies the following equation :

$$W(x) = \inf_{y:(y,x)\in E}\{W(y) + k(y,x)\}. \tag{2.4.2}$$

Indeed, since $\pi^{(\varepsilon)}$ is invariant, it holds $\pi^{(\varepsilon)}(x) = \sum_{y\in V} \pi^{(\varepsilon)}(y)p_\varepsilon(y,x)$, hence

$$\max_{y\in V} \pi^{(\varepsilon)}(y)p_\varepsilon(y,x) \leqslant \pi^{(\varepsilon)}(x) \leqslant |V| \max_{y\in V} \pi^{(\varepsilon)}(y)p_\varepsilon(y,x) \,.$$

By taking $-\varepsilon\log(\cdot)$ and the limit $\varepsilon \to 0$, one gets (2.4.2).

We now recall a graphical construction of the rate function W presented in [39]. Given a vertex x in V, an oriented graph g is said to belong to the family $\mathcal{T}_x$ if it is a directed tree inside G having vertices V, rooted at x and pointing towards the root. This means that

– g is an oriented graphs, with edges in E ;
– the vertices of g are all the states in V ;
– every vertex y in $V\setminus\{x\}$ is the initial point of exactly one oriented edge ;
– for any $y\in V\setminus\{x\}$ there exists a (unique) oriented path in g from y to x ;
– no oriented edge exits from x.

Then (cf. [39])

$$W(x) = \min_{g\in\mathcal{T}_x} \sum_{e\in g} k(e) - \min_z \min_{g\in\mathcal{T}_z} \sum_{e\in g} k(e) \,, \tag{2.4.3}$$

where the sum among $e \in g$ means the sum among the oriented edges e of the directed tree g.

2.4.2 The equation $\omega(x) = \inf_{y:(y,x)\in E}\{\omega(y) + k(y,x)\}$

We focus now on the equation

$$\omega(x) = \inf_{y:(y,x)\in E}\{\omega(y) + k(y,x)\} \,, \qquad \forall x \in V \,. \tag{2.4.4}$$

with an assumption on k :

Assumption 1. *For all x, there exists y such that $(x,y)\in E$ and $k(x,y) = 0$.*

First we observe that for any x we are taking the infimum on a finite non–empty set (and therefore the infimum is indeed a minimum). In fact, given $x\in V$ we can fix some $z\in V\setminus\{x\}$ and, since G is strongly connected, we know that the is a path from z to x. Considering the last edge of this path, we conclude that there exists $y\in V$ such that $(y,x)\in E$.

Trivially, for any $c\in\mathbb{R}$, if ω is a solution of (2.4.4) then also $\omega+c$ is a solution of (2.4.4). We call *canonical solutions* the solution ω such that $\min\omega = 0$. An example of canonical solution is $\omega = W$.

Definition 2.4.5. *(Pseudo-distance function) We define $D : V\times V \mapsto \mathbb{R}_+$ as*

$$D(x,y) = \min\{\sum_{e\in\gamma} k(e) \,:\, \gamma \text{ is a path from } x \text{ to } y \text{ in } G\}. \tag{2.4.6}$$

We denote γ_o the beginning vertex of the path γ and γ_f the final vertex of γ. By convention that $\gamma = (x)$ is a path from x to x, we set that $D(x,x) = 0$ for all $x \in V$. Now we can extend the domain of the Pseudo-distance function : define the pseudo-distance between two subset of V, or between a subset and an element of V as following :

$$\forall V_1, V_2 \subset V, D(V_1, V_2) = \min\{D(v_1, v_2), (v_1, v_2) \in V_1 \times V_2\}$$

and

$$\forall x \in V, U \subset V, D(U,x) = \min_{y \in U} D(y,x),\ \ D(x,U) = \min_{y \in U} D(x,y).$$

Remark 2.4.7. *Assume that ω is a solution of* (2.4.4). *By iterating equation* (2.4.4), *it holds for all $(x,y) \in V \times V$:*

$$\omega(x) \leqslant \omega(y) + D(y,x). \tag{2.4.8}$$

Definition 2.4.9. *We introduce another oriented graph $G' = (V, E')$ having vertex set V and edge set $E' := \{(x,y) \in E\ :\ k(x,y) = 0\}$. Since $\sum_{y \in V} p_\varepsilon(x,y) = 1$ and since V is finite, we conclude that for any vertex $x \in V$ there is at least one outgoing edge $(x,y) \in E'$. Note that it could be $(x,x) \in E'$.*

Observe that, if ω satisfies (2.4.4)*, then $\omega(x) \leqslant \omega(y)$ if $(y,x) \in E'$ and $\omega(x) < \omega(y)$ if $(y,x) \in E \setminus E'$: along the edges of $E \setminus E'$ the solution ω strictly decreases and along the edges of E' decreases.*

Definition 2.4.10. *We consider the following partition of V :*

$$V = \sqcup_{i \in I} C_i \sqcup_{j \in J} \{x_j\},$$

where C_i is a strongly connected component of G' with at least 2 points and $\{x_j\}$ is a strongly connected component of G' with exactly 1 point. Moreover, we partitioned the set I as $I = A \sqcup B$ where $i \in A$ if and only if C_i has no outgoing edges. We call the strongly connected component with more than two point a large component. The set of large components of G' is $\{C_i, i \in I\}$.

Remark 2.4.11. *One can easily deduce that A is always non empty from the fact that every vertex of G' has at least one outgoing edge.*

Theorem 2.4.12. *[Characterization of the set of 0] Let ω be a canonical solution of* (2.4.4)*, denote $U := \{x \in V, \omega(x) = 0\}$ the maximal set of vertex where ω vanished. Let $A' \subset A, B' \subset B$ such that for all $i \in A' \cup B'$, $C_i \cap U \neq \emptyset$, where the set A, B, C_i are defined as in Definition2.4.10. Denote $C_o := \cup_{i \in A' \cup B'} C_i$ and $C_f := \cup_{i \in A'} C_i$. Then*

$$U = \cup_{\gamma_o \in C_o, \gamma_f \in C_f}\ \gamma. \tag{2.4.13}$$

proof. We prove at first $U \supset \cup_{\gamma_o \in C_o, \gamma_f \in C_f} \gamma$.. Let $x \in C_i \cap U$. Since ω is constant on C_i, for all $y \in C_i$, $\omega(y) = \omega(x) = 0$. It follows that $C_o \cap C_f \subset U$. As we mentioned in definition2.4.9, solution ω decreases and along the edges of E'. Thus for all $x \in \cup_{\gamma_o \in C_o, \gamma_f \in C_f} \gamma$, $\omega(x) = 0$.

Now we turn to prove $U \subset \cup_{\gamma_o \in C_o, \gamma_f \in C_f} \gamma$.

Let $x \in U$. Since ω is a solution of (2.4.4), there exists y such that $\omega(x) = \omega(y) + k(y,x)$. It follows that $y \in U$ and $k(y,x) = 0$. It means that for all $x \in U$ there exists $y \in U$ such that $(y,x) \in E'$. By iterating this process, since the graph is finite, one can find a connected component $C_i \subset U$ and a path γ' such that $\gamma'_o \in C_i \subset C_o$ and $\gamma'_f = x$.

On the other hand, consider the a path γ'' beginning with x, following the edges of E' and ending up in a connected component without out going edges C_i, where $i \in A'$. By Remark2.4.9, for any $y \in \gamma''$,$\omega(y) \geqslant \omega(x) = 0$. Considering the fact that ω is a canonical solution, one deduce that $\gamma'' \in U$. Now consider the path $\gamma' \cup \gamma''$, it is a path containing x, beginning with $\gamma_o = \gamma'_o \in C_o$ and $\gamma_f = \gamma''_f \in C_f$. As a consequence, it holds

$$U \subset \cup_{\gamma_o \in C_o, \gamma_f \in C_f} \gamma.$$

The conclusion follows. □

Proposition 2.4.14. *Take $A' \subset A, B' \subset B$. Consider $C_o := \cup_{i \in A' \cup B'} C_i$ and $C_f := \cup_{i \in A'} C_i$. Define the set $U \subset V$ as following :*

$$U := \cup_{\gamma_o \in C_o, \gamma_f \in C_f} \gamma.$$

Then the function $D(U,.)$ is a canonical solution of (2.4.4).

proof. First we shaw prove that (2.4.4) holds for all $x \in U$.

Aware of the fact that $D(U,x) = 0$ for all $x \in U$, it is enough to prove that for all $x \in U$, there exists $y \in U$ such that $k(y,x) = 0$, or equivalently, $(y,x) \in E'$. Assume that $x \in C_o$, there exists $i \in A' \cup B'$ such that $x \in C_i$. Since C_i is a strongly connected component which contains at least 2 points, thus there exists $y \in C_i \subset U$ such that $(y,x) \in E'$.

Now assume that $x \in U \backslash C_o$, then by definition of the set U, there exists a path γ contains x and of the form $(z_o \in C_o, z_1, ..., z_i, z_{i+1}, ..., z_n \in C_f)$ with $z_{i+1} = x, i \geqslant 0$. One deduce that $z_i \in U$ and $(z_i, x) \in E'$.

Now we turn to the case when $x \in V \setminus U$. On one hand, for all y such that $k(y,x) \in E$, it holds

$$\begin{aligned} D(U,x) + k(y,x) &= \min_{\gamma_o \in U, \gamma_f = y} \sum_{e \in \gamma} k(e) + k(y,x) \\ &\geqslant \min_{\gamma_o \in U, \gamma_f = x} \sum_{e \in \gamma} k(e) = D(U,x). \end{aligned}$$

On the other hand, assume that $D(U,x) = \sum_{e\in\gamma} k(e)$, where γ is a path from U to x in G. Let us write $\gamma = (z_0 \in U, z_1, \dots, z_n = x)$ with $n \geqslant 1$. Denote γ_1 the path $z_0 \in U, z_1, \dots, z_{n-1})$. Then it holds

$$D(U, z_{n-1}) + k(z_{n-1}, x) \leqslant \sum_{e\in\gamma_1} k(e) + k(z_{n-1}) = D(U,x).$$

Combing the last two inequalities, the conclusion holds. □

Remark 2.4.15. *According to* (2.4.8)*, one deduce the following fact. Let ω be a solution and U be the maximal set where ω is vanished. Then $\omega \leqslant D(U,.)$.*

Proposition 2.4.16. *System* (2.4.4) *has a unique canonic solution if and only if $|A| = 1$ and $|B| = 0$. More precisely, system* (2.4.4) *has a unique canonic solution if and only if the graph G' has exactly one connected component C ; that component does not have outgoing edges ; and the solution is $D(C,.)$.*

proof. We assume at first that $|A| = 1$ and $|B| = 0$. Due to Proposition2.4.14, $D(C,.)$ is a canonical solution. Let ω be a canonical solution, it is enough to prove that $\omega = D(C.)$. Theorem2.4.12 ensures that C is the maximum set where ω is vanished. By Remark2.4.15, one has $\omega \leqslant D(C,.)$. On the other hand, given $x \in V \setminus C$, one can construct a path $(x_0, x_1, \dots, x_i, \dots)$ with $x_0 = x$, $\omega(x_i) = \omega(x_{i+1}) + k(x_{i+1}, x_i)$. Since V is finite, there exists $i \neq k$ such that $x_i = x_k$. Thus $(x_i, \dots, x_k)$ is a circle in G' and we deduce that $x_i \in C$. It follows that

$$\begin{aligned}\omega(x) &= \omega(x_i) + \sum_{j=0}^{i-1} k(x_{j+1}, x_j)\\ &\geqslant \min_{\gamma, \gamma_o \in C, \gamma_f = x} \{\sum_{e\in\gamma} k(e)\} = D(C,.)\end{aligned}$$

Now suppose that the solution (2.4.4) is unique. Let C be a strongly connected component without outgoing edges of G' and define U_1 as in Proposition2.4.14 by taking $C_o = C_f = C$. It is easy to see that $U_1 = C$. Thus $D(C,.)$ is a canonical solution. Now consider U_2 as Proposition2.4.14 by taking $C_o = C_f = \cup_{i\in I} C_i$ the union of all connected components. By uniqueness of the canonical solution, one has $D(C,.) = D(U_2,.)$. Otherwise, since C does not have outgoing edges of G', for all $x \in V \setminus C$, $D(C,x) > 0$. One deduce that $U_2 = C$ and the conclusion follows. □

Theorem 2.4.17. *Define I as in definition2.4.10 and suppose that k satisfies Assumption1. Let S_1 be the solution set of the following system*

$$\forall i \in I,\ u(i) \leqslant u(j) + \tilde{D}(j,i),\ \min_I u = 0. \tag{2.4.18}$$

Where the function $\tilde{D}$ is defined as $\tilde{D}(j,i) := D(C_j, C_i)$ for all $i,j \in I$.

Let S be the canonical solution set of system(2.4.4)*. Then there is a bijection $\varphi : S_1 \mapsto S$ defined as following. For $u \in S_1$,*

$$\varphi(u)(x) := \min_{i \in I}\{u(i) + D(C_i, x)\},$$

and for $\omega \in S$, let $x_i \in C_i$,

$$\varphi^{-1}(\omega)(i) := \omega(x_i).$$

proof. It is easy to see that φ is well define. Meanwhile, since ω is constant on strongly connected component C_i for all $i \in I$, one can this constant by $\omega(x_i)$, where $x_i \in C_i$. Then the application $\psi : \omega \in S \mapsto \psi(\omega)$,$\psi(\omega)(i) := \omega(x_i)$ is well define. Now we prove the theorem by 3 steps following :

(i) $\forall u \in S_1, \varphi(u) \in S$.

$(ii)\forall \omega \in S, \psi(\omega) \in S_1$.

(iii) $\varphi \circ \psi$ and $\psi \circ \varphi$ are identity mapping on S and S_1 respectively.

(i) : Given $u \in S_1$. It is easy to see $\min \varphi(u) = 0$ from the fact that $\min u = 0$. Given $x \in V$, For all y such that $(y, x) \in E$, there exists $j \in I$ such that the following holds :

$$\varphi(u)(y)+k(y,x) = u(j)+D(C_j,y)+k(y,x) \geqslant u(j)+D(C_j,x) \geqslant \varphi(u)(x). \quad (2.4.19)$$

One the other hand, there exists $i \in I$ such that $\varphi(u)(x) = u(i) + D(C_i, x)$. If $x \in C_i$, then one can find $y \in C_i$ such that $(y, x) \in E'$, and it holds

$$\varphi(u)(x) = u(i) = \varphi(u)(y) = \varphi(u)(y) + k(y, x).$$

Otherwise, if $x \notin C_i$, there is y such that $(y, x) \in E$ and

$$D(C_i, x) = D(C_i, y) + k(y, x).$$

We deduce that

$$\varphi(u)(x) = u(i) + D(C_i, y) + k(y, x) \geqslant \varphi(u)(y) + k(y, x). \quad (2.4.20)$$

Together with (2.4.19), we conclude that $\varphi(u)(x) = \min \varphi(u)(y) + k(y, x)$, $\min \varphi(u) = 0$ and (i) follows.

(ii) : Now take any $\omega \in S$, according to Proposition 2.4.14, there exists $i \in I$ such that for all $x \in C_i$, $\omega(x) = 0$. It follows that $\min \psi(\omega) = \psi(\omega)(i) = 0$. On the other hand, for all $x_i \in C_i, x_j \in C_j, i, j \in I$, $\omega(x_i) \leqslant \omega(x_j) + D(x_j, x_i)$. We deduce that

$$\psi(\omega)(i) \leqslant \psi(\omega)(j) + D(C_j, C_i),$$

and (ii) follows.

(iii) : For any $u \in S_1$ and any $i \in I$, it holds for all $x_i \in C_i$

$$\psi \circ \varphi(u)(i) = \varphi(u)(x_i) = u(i).$$

On the other hand, for any $\omega \in S$, any $x \in V$, it holds

$$\varphi \circ \psi(\omega)(x) = \min_{i \in I} \psi(\omega)(i) + D(C_i, x).$$

It is enough to prove that

$$\omega(x) = \min_{i \in I} \psi(\omega)(i) + D(C_i, x). \tag{2.4.21}$$

Firstly, given $i \in I$, take $y \in C_i$, one deduce that

$$\psi(\omega)(i) + D(C_i, x) = \omega(y) + D(y, x) \geqslant \omega(x). \tag{2.4.22}$$

Now we argue as in Proposition If there exists i such that $x \in C_i$, it is easy to see that (2.4.21) holds. Otherwise, assume that given $x \in V \setminus \cup_i C_i$, one can construct a path $(x_0, x_1, \dots, x_j, \dots)$ with $x_0 = x$, $\omega(x_j) = \omega(x_{j+1}) + k(x_{j+1}, x_j)$. Since V is finite, there exists $j \neq l$ such that $x_j = x_l$. Thus $(x_j, \dots, x_l)$ is a circle in G' and this circle is a subgraph of a strongly connected component C_i, $i \in I$. It follows that

$$\omega(x) = \omega(x_j) + \sum_{s=1}^{j} k(x_{s+1}, x_s) \geqslant \omega(x_j) + D(C_i, x) = \psi(\omega)(i) + D(C_i, x).$$

Together with (2.4.22), one deduce that (2.4.21) holds. Thus one has

$$\varphi \circ \psi(\omega)(x) = \min_{i \in I} \psi(\omega)(i) + D(C_i, x) = \omega(x)$$

for all $x \in V$ and all $\omega \in S$. The proof is completed.

□

Extreme solutions

As we can see in the latter theorem, one can determine the solution of system (2.4.4) by solving the system (2.4.18). In system (2.4.18), the function u is defined on a space of $|I|$ points. If we forget the normalizing constraint $\min u = 0$, then each constraints of (2.4.18) defines a half-space of $\mathbb{R}^{|I|}$. Here we identify a function u on I with a vector on $\mathbb{R}^{|I|}$ as $\overrightarrow{u} := (u(x_1), ..., u(x_{|I|})) \in \mathbb{R}^{|I|}$. Thus, the solution set is intersection of all those half-space, which gives a convex set S_c of $\mathbb{R}^{|I|}$. In order to describe a convex set, it is enough to determine all the extreme points. Recall that an extreme of a convex set on $\mathbb{R}^n$ is defined as following.

Definition 2.4.23. *Let $C \subset \mathbb{R}^n$ convex. $u \in C$ is an extreme point if and only if for all $u_1, u_2 \in C$, $\lambda \in (0, 1)$,*

$$\overrightarrow{u} = \lambda \overrightarrow{u_1} + (1 - \lambda) \overrightarrow{u_2} \Rightarrow \overrightarrow{u} = \overrightarrow{u_1} = \overrightarrow{u_2}.$$

Observing the fact that The convex S_c is invariant from translation by vector $c\mathbf{1}$, one should define the extreme point of on the sub space $\mathbf{1}^{\perp}$. Denote P the orthogonal projection from $\mathbb{R}^{|I|}$ onto $\mathbf{1}^{\perp}$, where $P : \overrightarrow{u} \mapsto \overrightarrow{u} - \sum_{i=1}^{|V|} u(x_i)$ We define Extreme point of solution set S_c as following :

Definition 2.4.24 (Extreme solution)**.** *One says that $\overrightarrow{u} \in S_c$ is an extreme point of S_c if and only if $P(\overrightarrow{u})$ is an extreme point of $P(S_c)$. It means that for all $\overrightarrow{u_1}, \overrightarrow{u_2} \in S_c$, $\lambda \in (0,1)$,*

$$P(\overrightarrow{u}) = \lambda P(\overrightarrow{u_1}) + (1-\lambda) P(\overrightarrow{u_2}) \Rightarrow P(\overrightarrow{u}) = P(\overrightarrow{u_1}) = P(\overrightarrow{u_2}).$$

One says that u is an extreme solution of S_c if $\overrightarrow{u}$ is an extreme point of S_c.

Remark 2.4.25. *One can always take $\lambda = 1/2$ in the latter definition.*

Theorem 2.4.26 (Characterisation of extreme solution)**.** *Given $u \in S_c$. Consider the reduce directed graph corresponding to u, defined as $G_u(I, E_u)$, with $(i,j) \in E_u$ if and only if $u(j) = u(i) + \tilde{D}(i,j)$. Then u is an extreme solution if and only if G_u is weakly connected. Where weakly connected stands for the correspond non directed graph is connected.*

In order to prove the Theorem, we need the following lemma :

Lemma 2.4.27. *Assume that $u, v \in S_c$ satisfy $E_u \subset E_v$, then for all weakly connected component $C \in G_u$ there exists $c \in \mathbb{R}$, such that for all $i \in C$, $u(i) = v(i) + c$. In particular, if G_u is weakly connected, then there exists $c \in \mathbb{R}$ such that $u = v + c$, and $G_u = G_v$.*

proof of lemma2.4.27. Let $i \in I$, for all $j \in I$ weakly connected to i, there exists a (general) path $\gamma := (i_1, \ldots, i_n)$ such that $i_1 = i, i_n = j$ and for all $1 \leqslant k \leqslant n-1$, E_u contains (i_k, i_{k+1}) or (i_{k+1}, i_k). Thus it holds

$$u(j) = u(i) + \sum_{k \in A^+} \tilde{D}(i_k, i_{k+1}) - \sum_{k \in A^-} \tilde{D}(i_{k+1}, i_k), \tag{2.4.28}$$

where $A^+ := \{k, (i_k, i_{k+1}) \in E_u\}$ and $A^- := \{k, (i_{k+1}, i_k) \in E_u\}$. Since $E_u \in E_v$, (2.4.28) holds for v as well. Now set $c = u(i) - v(i)$, one can deduce that for all j connected to i, $v(j) - u(j) = v(i) - u(i) = c$. The lemma follows. □

proof of theorem. Necessary condition :

Suppose that $u \in S_c$ and G_u is weakly connected. Assume that $u_1, u_2 \in S_c$ satisfy $u = (u_1 + u_2)/2$. In order to prove that u is an extreme solution, according to Lemma2.4.27, it is enough to show that $G_{u_1} = G_{u_2} = G_u$, or equivalently, since G_u is connected, it is enough to show that $E_u \subset E_{u_1} \cap E_{u_2}$.

If $(i,j) \in E_u$, then it holds

$$u_1(j) + u_2(j) = 2u(j) = 2u(i) + 2\tilde{D}(i,j) = u_1(i) + u_2(i) + 2\tilde{D}(i,j) \tag{2.4.29}$$

Combining with the fact that $u_1(j) \leqslant u_1(i)+\tilde{D}(i,j)$ and $u_2(j) \leqslant u_2(i)+\tilde{D}(i,j)$ (since $u_1, u_2 \in S_c$), one deduces that $u_1(j) = u_1(i) + \tilde{D}(i,j)$ and $u_2(j) = u_2(i) + \tilde{D}(i,j)$. As a consequence, $(i,j) \in E_{u_1} \cap E_{u_2}$. Thus $E_u \subset E_{u_1} \cap E_{u_2}$.

Sufficient condition :

We will prove it by contradiction : Assume that u is an extreme solution and there exists $I_1 \subset I$ such that $I_2 := I \setminus I_1 \neq \emptyset$ and I_1, I_2 are disconnected. For all $i \in I_1$ and $j \in I_2$, $(i,j),(j,i) \notin E_u$, it holds that $u(j) < u(i) + \tilde{D}(i,j)$ and $u(i) < u(j) + \tilde{D}(j,i)$. Denote $\varepsilon_{ij} = \min\{u(i) + \tilde{D}(i,j) - u(j), u(j) + \tilde{D}(j,i) - u(i)\}$. One deduces that $\varepsilon_{ij} > 0$ for all $i \in I_1, j \in I_2$. Let $\varepsilon = \min_{ij} \varepsilon_{ij}$>0, define two function u_1, u_2 on I as following :

$$u_1(i) := \begin{cases} u(i) & \forall i \in I_1 \\ u(i) + \varepsilon & \forall i \in I_2 \end{cases}$$

and

$$u_2(i) := \begin{cases} u(i) & \forall i \in I_1 \\ u(i) - \varepsilon & \forall i \in I_2 \end{cases}.$$

It is easy to check that u_1,u_2 satisfy the system (2.4.18) and $u = (u_1 + u_2)/2$, but $P(u) \neq P(u_1)$. This is a contradiction with the fact that u is an extreme solution. $\square$

Chapitre 3

Functional inequalities in discrete space

Abstract

In this chapter, we are interested in functional inequalities in general space. We will assume that the space $(\mathcal{X}, d)$ is a polish space, such that closed balls are compact, the goal is to deal with functional inequalities in a discrete space or a non geodesic space. We will mainly focus on Log-Sob inequality, transport inequalities and Poincaré inequality. Those inequalities are essential tools in the study of concentration of measure, in the estimation of the relaxation time of various ergodic systems. We refer the readers to to [5],[64] for a more general introduction and applications of functional inequalities.

3.1 Introduction

Firstly, we will recall some classical results in the continuous settings. In order to avoid technical assumptions and proofs, we assume that the space is $\mathbb{R}^n$. Denote $\mathcal{P}(\mathbb{R}^n)$ the set of all probability measure of $\mathbb{R}^n$, in this case, we consider the distance d is the Euclidean norm. Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$, one can define the related entropy, the Fisher information and the optimal transportation cost between them, each of those quantities can somehow measure the "differences" between μ and ν.

Definition 3.1.1. *(Entropy) Let f be a real valued function defined on $\mathbb{R}^n$, the entropy of f with respect to μ is defined as :*

$$\mathrm{Ent}_\mu(f) := \int f \log f d\mu - \left(\int f d\mu\right) \log \left(\int f d\mu\right),$$

If f is a density function of ν with respect to μ, one can define the related entropy of ν with respect to μ :

$$H(\nu|\mu) = \int \left(\log \frac{d\nu}{d\mu}\right) d\nu = \mathrm{Ent}_\mu(f),$$

if ν is not absolutely continuous with respect to μ, then $H(\nu|\mu) = \infty$.

Definition 3.1.2. *(Fisher information) Let f be a positive function defined on $\mathcal{X}$, assume that f is differentiable almost everywhere, the fisher information of f with respect to μ is*

$$\mathcal{I}_\mu(f) = 4 \int |\nabla \sqrt{f}|^2 d\mu = \int \frac{|\nabla f|^2}{f} d\mu$$

If f is a density function of ν with respect to μ. one says that the Fisher information of ν with respect to μ is

$$\mathcal{I}(\nu|\mu) = \mathcal{I}_\mu(f).$$

Here $|.|$ is the Euclidean norm. In more general case, one could also consider more general Dirichlet forms than $|\nabla f|^2$.

Definition 3.1.3. *(Transport cost, Wasserstain distance) Let $\theta : \mathbb{R}^+ \to \mathbb{R}^+$, with $\theta(0) = 0$, be a measurable function referred to as the* cost *function. Then, the usual optimal* transport cost*, in the sense of Kantorovich, between two probability measures μ and ν on $\mathbb{R}^n$ is defined by*

$$\mathcal{T}_\theta(\nu, \mu) := \inf_\pi \iint \theta(d(x, y))\, \pi(dxdy), \tag{3.1.4}$$

where the infimum runs over the set of couplings π between μ and ν, i.e., probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ such that $\pi(dx \times \mathbb{R}^n) = \mu(dx)$ and $\pi(\mathbb{R}^n \times dy) = \nu(dy)$.

When the cost function θ is $x \mapsto x^p$, $p > 1$, the optimal transport cost is correspond to the power p of L^p-Wasserstain distance. Namely, we have

Definition 3.1.5 (L^p-Kantorovich-Wasserstein distances)**.** *Let $p \geqslant 1$. The L^p-Wasserstein distance W_p between two probability measures μ and ν on a metric space $(\mathcal{X}, d)$ is defined as*

$$W_p(\mu,\nu) := \Big(\inf_{\pi} \int d(x,y)^p \pi(dxdy)\Big)^{1/p}$$

where the infimum runs over all couplings π of μ and ν.

Recall that a probability measure μ on $\mathbb{R}^n$ satisfies a Poincaé (or spectral gap) inequality $\mathrm{P}(c)$ if there exists a constant $c > 0$ such that for every smooth function $f : \mathbb{R}^n \mapsto \mathbb{R}$, μ satisfies

$$\mathrm{Var}_\mu(f) \leqslant c\mathcal{I}_\mu(f^2).$$

The variance is defined by

$$\mathrm{Var}_\mu(f) := \int f^2 d\mu - \Big(\int f d\mu\Big)^2.$$

We say that μ satisfies the logarithmic Sobolev inequality $\mathrm{LS}(c)$ if there exists a constant $c > 0$ such that for all smooth positive valued function f one has

$$\mathrm{Ent}_\mu(f) \leqslant c\mathcal{I}_\mu(f),$$

which is equivalence to say that for all probability measure ν, it holds

$$H(\nu|\mu) \leqslant c\mathcal{I}(\nu|\mu).$$

We say that μ satisfies transport-entropy inequality $\mathrm{T}_\theta(\mathrm{c})$ if there exists $c > 0$ such that for all $\nu \in \mathcal{P}(\mathbb{R})$, it holds

$$\mathcal{T}_\theta(\nu,\mu) \leqslant cH(\nu|\mu).$$

We say that μ satisfies transport-information inequality $\mathrm{T}_\theta\mathrm{I}(\mathrm{c})$ if there exists $c > 0$ such that for all $\nu \in \mathcal{P}(\mathbb{R})$, it holds

$$\mathcal{T}_\theta(\nu,\mu) \leqslant c\mathcal{I}(\nu|\mu).$$

When the cost function θ is quadratic, say $x \mapsto x^2$, we denote T_2 and $\mathrm{T}_2\mathrm{I}$ for the transport-entropy inequality and transport information inequality. T_2 is also referred to Talagrand inequality, since it was introduced by Talagrand in [98], 1996. Assume that μ satisfies one of those inequalities : T_2, $\mathrm{T}_2\mathrm{I}$ or LS, we can deduce that the measure μ satisfies a Gaussian concentration : there exists $a > 0$, for all f $1 - Lip$, it holds for all $t \geqslant 0$.

$$\mu(f \geqslant \int f d\mu + t) \leqslant e^{-at^2}.$$

Under an assumption of Poincaré inequality P, we deduce that the reference measure μ satisfies an exponential concentration : there exists $a > 0$, for all f $1 - Lip$, it holds for all $t \geqslant 0$.

$$\mu(f \geqslant \int f d\mu + t) \leqslant e^{-at}.$$

One can easily see that the concentration property deduced by P is weaker than which deduced by T_2,LS and T_2I. In fact, up to some explicit constant, the following implication chain holds

$$\mathrm{LS} \Rightarrow \mathrm{T}_2\mathrm{I} \Rightarrow \mathrm{T}_2 \Rightarrow \mathrm{P}. \tag{3.1.6}$$

The implication $\mathrm{T}_2 \Rightarrow \mathrm{P}$ followed from a simple linearization argument. The implication LS $\Rightarrow \mathrm{T}_2$ was obtained by Otto and Villani in [87], 2000, sometimes also called Otto-Villani's theorem. The proof of this theorem was recovered by Bobkov, Gentil and Ledoux [11] with a different method, via the Hopf-Lax formula and Hamilton-Jacobi equation. Using a similar argument, Guillin, Léonard and Wu [53] proved the implication $\mathrm{T}_2\mathrm{I} \Rightarrow \mathrm{T}_2$. Later, Gozlan, Roberto and Samsons[48] ; Ambrosio, Gigli and Savaré [3] generalized independently the argument of [11] in a general geodesic space, and recovered the implication chain.

In this chapter, we will discuss discrete version of those functional inequalities, and establish a similar implication chain.

3.2 Weak transport inequalities in discrete setting

Space. Let $(\mathcal{X}, d)$ stand for a polish space (*i.e.* complete and separable), such that closed balls are compact. Denote $\mathcal{P}(\mathcal{X})$ for the set of all probability measure on $\mathcal{X}$.

For example, the space can be a graph : $G = (V, E)$, which denotes a (simple) connected graph with vertex set V and edge set E (given $(x, y) \in E$, we may write $x \sim y$), we assume also that all vertices have finite degree. In this case, the space is the vertex set V equipped with the graph distance d.

We are hoping to establish the implication chain 3.1.6 in a discrete space. However, the Talagrand inequality almost never holds true. More precisely, when the support of μ is disconnected, $\mathrm{T}_2(c)$ never holds, for any $c > 0$. To our knowledge, this was first proved in [47] in graph settings. We give here a different proof, as a consequence of a more general result :

Lemma 3.2.1. *Let $(\mathcal{X}, d)$ be a metric space and μ a probability measure defined on $\mathcal{X}$. Assume that there exist $C_1, C_2 \subset \mathcal{X}$ such that*

(i) $\inf_{x \in C_1, y \in C_2} d(x, y) > 0$,

(ii) $\mathrm{supp}(\mu) \subset C_1 \cup C_2$,

(iii) $\mu(C_1) > 0$, $\mu(C_2) > 0$.

Then μ does not satisfies $\mathrm{T}_2(c)$ for any $c > 0$.

proof. For $h < \min\{\mu(C_1), \mu(C_2)\}$, define

$$\nu_h(dx) := \begin{cases} \mu(dx)(1 + \frac{h}{\mu(C_1)}), & x \in C_1 \\ \mu(dx)(1 - \frac{h}{\mu(C_2)}), & x \in C_2 \end{cases}$$

Let $d := \inf_{x\in C_1, y\in C_2} d(x,y) > 0$. Then we have $W_2(\mu,\nu)^2 \geqslant d^2 h$, and the entropy is

$$(\mu(C_1) + h)\log(1 + h/\mu(C_1)) + (\mu(C_2) - h)\log(1 - h/\mu(C_2)).$$

When h goes to 0, the entropy is $\mathcal{O}(h^2)$. The conclusion follows since W_2^2 have order $\mathcal{O}(h)$. □

Thus, in the case that the space is a simple graph, then all points are isolated. Then T_2 holds if and only if the reference measure μ is a Dirac mass, which is not interesting.

To recover a discrete version of T_2, we therefore have to redefine the transport cost. Erbar and Maas recovered some of those functional inequality results with the notion of entropic Ricci curvature on graphs, and we refer the reader to [32, 70] for more details. Another way to deal with it is to take the weak transport cost introduced by Marton [76] :

Definition 3.2.2. *Let μ, ν be two probabilities measures on $\mathcal{X}$ and θ be a positive cost function. Define*

$$\widetilde{\mathcal{T}}_2(\nu|\mu) := \inf_{\pi\in\Pi(\mu,\nu)} \left\{ \int \theta\left(\int d(x,y)p_x(dy)\right)\mu(dx)\right\}.$$

Where $\Pi(\mu,\nu)$ is the set of all couplings π whose first marginal is μ and second marginal is ν, p_x is the probability kernel such that $\pi(dxdy) = p_x(dy)\mu(dx)$. Using probabilistic notations, on has

$$\widetilde{\mathcal{T}}_\theta(\nu|\mu) = \inf_{X\sim\mu, Y\sim\nu} \mathbb{E}\left[\theta\left(\mathbb{E}(d(X,Y)|X)\right)\right].$$

Without the conditional expectation, we reduce to the classical transportation cost :

$$\mathcal{T}_\theta(\nu|\mu) = \inf_{X\sim\mu, Y\sim\nu} \mathbb{E}\left[\theta\left(d(X,Y)\right)\right].$$

Note that for a quadratic cost, the weak transport cost could be also seen as a weak Wasserstein-like distance. In order to agree with the notations of Wassertein distance, we note $\widetilde{W}_2(\nu|\mu)^2 := \widetilde{\mathcal{T}}_2(\nu|\mu)$. However, it is *not* a distance, since it is not symmetric. Note that by Jensen's inequality, it holds

$$W_1 \leqslant \widetilde{W}_2(\nu|\mu) \leqslant W_2(\mu,\nu).$$

In [50], Gozlan and al. studied this weak transport cost and establish a counter part of Kantorovich duality ; using the infimum-operator $\widetilde{Q}$ we studied in the first chapter :

$$\widetilde{Q}_\theta\varphi(x) = \inf_{p\in\mathcal{P}(\mathcal{X})} \left\{ \int \varphi(y)p(dy) + \theta\left(\int d(x,y)p(dy)\right)\right\}$$

Theorem 3.2.3 (Gozlan-Roberto-Samson-Tetali). *Adapting the settings of definitions before, it holds*

$$\widetilde{\mathcal{T}}_\theta(\nu|\mu) = \sup_\varphi \left\{ \int \widetilde{Q}_\theta \varphi d\mu - \int \varphi d\nu \right\}.$$

They also defined weak-transport entropy inequality and established a discrete version of criteria of Bobkov and Gotze :

Definition 3.2.4. *Adapting the settings of definitions before, we say that μ satisfies (i) the weak transport-entropy inequality $\widetilde{\mathrm{T}}_\theta^+(c)$ if for all probability measures ν, we have*

$$\widetilde{\mathcal{T}}_\theta(\nu|\mu)^2 \leqslant cH(\nu|\mu);$$

(ii) the weak transport-entropy inequality $\widetilde{\mathrm{T}}_\theta^-(c)$ if for all probability measures ν, we have

$$\widetilde{\mathcal{T}}_\theta(\mu|\nu)^2 \leqslant cH(\nu|\mu);$$

(iii)the weak transport-entropy inequality $\widetilde{\mathrm{T}}_\theta(c_1, c_2)$ if for all probability measures ν_1, ν_2, we have

$$\widetilde{\mathcal{T}}_\theta(\nu_1|\nu_2) \leqslant c_1 H(\nu_1|\mu) + c_2 H(\nu_2|\mu).$$

Theorem 3.2.5 (Gozlan-Roberto-Samson-Tetali). *Inequality $\widetilde{\mathrm{T}}_\theta^+(c)$ holds for probability measure μ if and only if for all φ, it holds*

$$\exp\left(\int \widetilde{Q}_\theta \varphi d\mu \right) \left(\int \exp\left(\frac{-\varphi}{c} \right) d\mu \right)^c \leqslant 1.$$

Inequality $\widetilde{\mathrm{T}}_\theta^-(c)$ holds if and only if for all φ, it holds

$$\left(\int \exp\left(\frac{\widetilde{Q}_\theta \varphi}{c} \right) d\mu \right)^c \left(\exp \int -\varphi d\mu \right) \leqslant 1.$$

Inequality $\widetilde{\mathrm{T}}_\theta(c_1, c_2)$ holds if and only if for all φ, it holds

$$\left(\int \exp\left(\frac{\widetilde{Q}_\theta \varphi}{c_1} \right) d\mu \right)^{c_1} \left(\int \exp\left(\frac{-\varphi}{c_2} \right) d\mu \right)^{c_2} \leqslant 1.$$

We introduce now the notion of gradient. A usual notion of gradient is

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

However, when x is an isolated point, this is not clear. In the case that there are few isolated points, one can set 0 by convention if x is isolated, see for example [47]. When the space is a set of isolated points, such as the graph settings, a set of isolated points equipped with graph structure and graph distance, one can consider the gradient length $|\nabla f|$ (see [17]) :

$$|\nabla f(x)| = \sup_{y \sim x} [f(y) - f(x)]_-,$$

or something defined by Γ calculus which we will introduce in the next chapter.

Now let $(\mathcal{X}, d)$ be a complete, separable metric space such that balls are compact. We define the (length of the) gradient as

$$|\widetilde{\nabla} f|(x) := \sup_{y \in \mathcal{X}} \frac{[f(y) - f(x)]_-}{d(x, y)}$$

where $[a]_- = \max(0, -a)$ is the negative part of $a \in \mathbb{R}$ (by convention $0/0 = 0$). The main reason for this gradient is its connection with the infimum-convolution operator $\widetilde{Q}_t$ and the related Hamilton-Jacobi equation we described in Chapter 2. We observe that, this gradient length $|\widetilde{\nabla} f|(x)$ is usually larger than the both two notions of gradient $|\nabla f|(x)$ we described before. In $\mathbb{R}^n$ equipped with the usual Euclidean distance, if f is a smooth convex function, $|\widetilde{\nabla} f|$ coincides with the usual length of the gradient $|\nabla f|(x) = \sqrt{\sum_i \partial_i f^2}(x)$. Thus, the functional inequalities involving $\widetilde{\nabla}$ are often comparable to those who involving other gradient, we will analyse the graph case in Chapter 4.

Unlike the usual notions of gradient, this special gradient is a global notion.

Now we define the modified log sobolev inequalities and the modified Poincaré inequality.

Definition 3.2.6. *One says that $\mu \in \mathcal{P}(\mathcal{X})$ satisfies the Poincaré inequality, respectively the modified log-Sobolev inequality*[1] *of type I and type II, respectively the weak transport-entropy inequality of type I and type II, if there exists a constant $C \in (0, \infty)$ such that for all $f \colon \mathcal{X} \to \mathbb{R}$ bounded it holds*

$$\mathrm{Var}_\mu(f) \leqslant C \int |\widetilde{\nabla} f|^2 \, d\mu \qquad (\textit{Poincaré Inequality}), \tag{3.2.7}$$

respectively

$$\mathrm{Ent}_\mu(e^f) \leqslant C \int \theta^* \left(|\widetilde{\nabla} f| \right) e^f \, d\mu \quad (\textit{Modified log-Sob Ineq. of type I}), \tag{3.2.8}$$

$$\mathrm{Ent}_\mu(e^f) \leqslant C \int \theta^* \left(|\widetilde{\nabla} (-f)| \right) e^f \, d\mu \quad (\textit{Modified log-Sob Ineq. of type II}). \tag{3.2.9}$$

1. We observe that the terminology here is not optimal since there already exist, in the literature, many different inequalities called modified log-Sobolev inequality that have *a priori* no relation between them, and no relation with our definition.

Assume that the cost function is quadratic : $\theta : x \mapsto x^2$. The log-Sobolev-type inequality (3.2.8) is implied by the usual Gross' inequality [52] in the continuous setting (since $|\widetilde{\nabla} f| \geqslant |\nabla f|$). In discrete, there exist a lot of different versions of the log-Sobolev inequality – that are all equivalent in the continuous, thanks to the chain rule formula – each of them having some nice property (connection to the decay to equilibrium of Markov processes, concentration phenomenon etc.). We refer the reader to the paper by Bobkov and Tetali [17] for an introduction to many of these inequalities and related properties. In particular, in [17], the log-Sobolev type inequality (3.2.8) is studied, with some local gradient in place of $\widetilde{\nabla}$. As we shall prove below, the usual log-Sobolev inequality in discrete, with transitions given by a Markovian matrix, implies (3.2.8). In turn, since such an inequality is very well studied in many situations (see *e.g.* the monographs [41, 5] and [72, 55] for results on general graphs and examples coming from physics) this provides a lot of examples of non trivial measures (on graphs) that satisfy (3.2.8).

3.2.1 Connection with some classical inequalities, on graphs

Given a (simple connected) graph $G = (V, E)$, recall that $K = (K(x, y))_{x,y\in V}$ is a matrix with positive entries if $K(x, y) \geqslant 0$ for all $x, y \in V$, and that it is a Markovian matrix if in addition $\sum_{y\in V} K(x, y) = 1$ for all $x \in V$. Then, the couple (μ, K) satisfies the (say) classical modified log-Sobolev inequality if there exists a constant $C \in (0, \infty)$ such that for all $f : V \to \mathbb{R}$ bounded it holds

$$\mathrm{Ent}_\mu(e^f) \leqslant C \sum_{x,y\in V} (e^{f(y)} - e^{f(x)})(f(y) - f(x))\mu(x)K(x, y). \tag{3.2.10}$$

The latter is known to be a consequence of Gross' Inequality that asserts that

$$\mathrm{Ent}_\mu(f) \leqslant C' \sum_{x,y\in V} (f(y) - f(x))^2\mu(x)K(x, y) \qquad \forall f : V \to \mathbb{R} \quad \text{bounded}. \tag{3.2.11}$$

More precisely Gross' Inequality (3.2.11) with constant C' implies the classical modified log-Sobolev inequality (3.2.10) with constant $C \leqslant C'/4$, see [17, Theorem 3.6].

Proposition 3.2.12. *Let μ be a probability measure on a (simple connected) graph $G = (V, E)$ and K be a matrix with positive entries. Assume that there exists a constant L such that $\sum_{y\in V} d^2(x, y)K(x, y) \leqslant L$ for all $x \in V$ and that for all $x, y \in V$, $\mu(x)K(x, y) = \mu(y)K(y, x)$. Finally, assume that (μ, K) satisfies the classical modified log-Sobolev inequality* (3.2.10) *with constant C, respectively Gross' Inequality* (3.2.11) *with constant C'. Then, μ satisfies the modified log-Sobolev inequality* (3.2.8) *with $\alpha(x) = \alpha^*(x) = x^2/2$ and constant $4LC$, respectively LC'.*

Remark 3.2.13. *The condition $\mu(x)K(x, y) = \mu(y)K(y, x)$, $x, y \in V$, is known as the detailed balance condition in the physics literature and means that the operator K, acting on functions, is symmetric in $\mathbb{L}^2(\mu)$. Most commonly one deals with a Markovian matrix with nearest neighbor jumps (meaning that $K(x, y) = 0$ unless*

$d(x,y) = 1$), which guarantees that $L = 1$. In particular the hypotheses of the proposition are very commonly used and correspond to a lot of practical situations [41].

proof. The result involving the Gross' inequality is an immediate consequence of the result involving the classical modified log-Sobolev inequality since the former implies the latter with $C' \leqslant C/4$.

Hence, we only need to show that

$$\sum_{x,y\in V}(e^{f(y)} - e^{f(x)})(f(y)-f(x))\mu(x)K(x,y) \leqslant 2L\sum_{x\in V}|\widetilde{\nabla} f|^2(x)e^{f(x)}\mu(x).$$

Since $(a-b)(e^a - e^b) \leqslant (a-b)^2\max\{e^a, e^b\}$, we have

$$\begin{aligned}&\sum_{x,y\in V}(e^{f(y)} - e^{f(x)})(f(y)-f(x))\mu(x)K(x,y)\\ &\leqslant \sum_{\substack{x,y\in V:\\ f(x)\geqslant f(y)}}(f(y)-f(x))^2 e^{f(x)}\mu(x)K(x,y) + \sum_{\substack{x,y\in V:\\ f(y)\geqslant f(x)}}(f(x)-f(y))^2 e^{f(y)}\mu(x)K(x,y).\end{aligned}$$

Using the detailed balance condition ensures that

$$\sum_{\substack{x,y\in V:\\ f(y)\geqslant f(x)}}(f(x)-f(y))^2 e^{f(y)}\mu(x)K(x,y) = \sum_{\substack{x,y\in V:\\ f(y)\geqslant f(x)}}(f(x)-f(y))^2 e^{f(y)}\mu(y)K(y,x)$$

which, after a change of variable, implies that

$$\sum_{x,y\in V}(e^{f(y)} - e^{f(x)})(f(y)-f(x))\mu(x)K(x,y) = 2\sum_{\substack{x,y\in V:\\ f(x)\geqslant f(y)}}(f(y)-f(x))^2 e^{f(x)}\mu(x)K(x,y).$$

Now, we observe that

$$\begin{aligned}&\sum_{\substack{x,y\in V:\\ f(x)\geqslant f(y)}}(f(y)-f(x))^2 e^{f(x)}\mu(x)K(x,y)\\ &\qquad = \sum_{\substack{x,y\in V:\\ f(x)\geqslant f(y)}}\left(\frac{[f(y)-f(x)]_-}{d(x,y)}\right)^2 e^{f(x)}\mu(x)K(x,y)d(x,y)^2\\ &\qquad \leqslant \sum_{x\in V}|\widetilde{\nabla} f|^2(x)e^{f(x)}\mu(x)\sum_{y\in V}K(x,y)d(x,y)^2\end{aligned}$$

which leads to the desired result since $\sum_{y\in V}K(x,y)d(x,y)^2 \leqslant L$. The proof is complete. $\square$

The next two section will show that the modified log Sobolev inequalities are equivalence to some hypercontractivity property, which imply some Talagrand type transport entropy inequality. Then we will show that the Poincaré inequality is equivalence to some kind of modified log-sob inequality and transport entropy inequality with certain cost function. Before going into this technical proof, we would need some properties of this special gradient length $\widetilde{\nabla}$.

3.2.2 Properties of the gradient $\widetilde{\nabla}$

In this section we collect some useful facts on the gradient $\widetilde{\nabla}$. Our first result is some sort of chain rule formula for $\widetilde{\nabla}$.

Proposition 3.2.14. *Let $f\colon \mathcal{X} \mapsto \mathbb{R}$ and $G\colon f(X) \mapsto \mathbb{R}$.*

(i) If G is non-decreasing then $|\widetilde{\nabla} G \circ f|(x) \leqslant |\widetilde{\nabla} f|(x) |\widetilde{\nabla} G|\left(f(x)\right)$, $x \in \mathcal{X}$.

(ii) If G is non-increasing then $|\widetilde{\nabla} G \circ f|(x) \leqslant |\widetilde{\nabla}(-f)|(x) |\widetilde{\nabla} G|\left(f(x)\right)$, $x \in \mathcal{X}$.

Here, $|\widetilde{\nabla} G|(u) := \sup_{v\in\mathbb{R}} \frac{[G(v)-G(u)]_-}{|v-u|}$, $u \in \mathbb{R}$, with $|\cdot|$ being the absolute value.

proof. Fix $x \in \mathcal{X}$ and assume that G is non-decreasing. Let $y \in \mathcal{X}$ be such that $f(x) > f(y)$ (if $\{y \in \mathcal{X} : f(x) > f(y)\} = \emptyset$ then $|\widetilde{\nabla} G \circ f|(x) = |\widetilde{\nabla} f|(x) = 0$ and there is nothing to prove). Since G is non-decreasing $G(f(x)) \geqslant G(f(y))$ so that

$$\frac{G\left(f(x)\right) - G\left(f(y)\right)}{d(x,y)} \leqslant \frac{f(x)-f(y)}{d(x,y)} \frac{G\left(f(x)\right) - G\left(f(y)\right)}{f(x)-f(y)} \leqslant |\widetilde{\nabla} f|(x) |\widetilde{\nabla} G|\left(f(x)\right).$$

Taking the supremum over all y such that $f(x) > f(y)$ leads to the desired conclusion of Item (i).

The proof of Item (ii) is similar. Let $y \in \mathcal{X}$ be such that $f(y) > f(x)$, then $G(f(y)) \leqslant G(f(x))$ (since G is non-increasing) so that

$$\begin{aligned}\frac{G\left(f(x)\right) - G\left(f(y)\right)}{d(x,y)} &= \frac{(-f)(x)-(-f)(y)}{d(x,y)} \frac{G\left(f(x)\right) - G\left(f(y)\right)}{|f(y)-f(x)|} \\ &\leqslant |\widetilde{\nabla}(-f)|(x) |\widetilde{\nabla} G|\left(f(x)\right).\end{aligned}$$

The result follows by taking the supremum over all $y \in \mathcal{X}$ such that $f(y) > f(x)$. □

Remark 3.2.15. *Observe that $|\widetilde{\nabla}(Cf)|(x) = C|\widetilde{\nabla} f|(x)$ for $C > 0$, while $|\widetilde{\nabla}(Cf)|(x) = -C|\widetilde{\nabla}(-f)|(x)$ for $C < 0$. Because of the negative part entering in its definition, in general $|\widetilde{\nabla}(-f)| \neq |\widetilde{\nabla} f|$.*

3.3 Hypercontractivity

In this section we follow the work by Bobkov, Gentil and Ledoux [11] to prove a result analogous to the celebrated Otto and Villani Theorem [87]. Namely we shall prove that some log-Sobolev type inequality is equivalent to an hypercontractivity property of the semi-group $\widetilde{Q}_t$, which in turn, by a duality argument due to Gozlan et al. [50], implies some Talagrand type transport-entropy inequality. To state this result one needs to introduce some additional notations.

Consider the usual q-norm of a function g on $\mathcal{X}$ defined by $\|g\|_q = (\int |g|^q \, d\mu)^{1/q}$, $q \in \mathbb{R}$, with, when this makes sense, $\|g\|_0 := \lim_{q\to 0} \|g\|_q = \exp\{\int \log g \, d\mu\}$, and when $g \geqslant 0$.

Theorem 3.3.1. *Let μ be a probability measure on $\mathcal{X}$ and $C > 0$. Then*

(i) *If for all bounded measurable function $f\colon \mathcal{X} \to \mathbb{R}$,*

$$\mathrm{Ent}_\mu(e^f) \leqslant C \int |\widetilde{\nabla} f|^2 e^f \, d\mu, \tag{3.3.2}$$

then for every $\rho \geqslant 0$, every $t \geqslant 0$ and every bounded measurable function f,

$$\|e^{\widetilde{Q}_t f}\|_{\rho + \frac{2t}{C}} \leqslant \|e^f\|_\rho. \tag{3.3.3}$$

Conversely, if (3.3.3) holds for some $\rho > 0$ and for all $t \geqslant 0$, then (3.3.2) *holds.*

(ii) *If for all bounded measurable function $f\colon \mathcal{X} \to \mathbb{R}$,*

$$\mathrm{Ent}_\mu(e^f) \leqslant C \int |\widetilde{\nabla}(-f)|^2 e^f \, d\mu, \tag{3.3.4}$$

then (3.3.3) *holds for every $\rho \leqslant 0$, every $t \in [0, -\rho C/2]$ and every bounded measurable function f. Conversely, if (3.3.3) holds for some $\rho < 0$ for all $t \in [0, -\rho C/2)$, then* (3.3.4) *holds.*

Theorem 3.3.5. *Let μ be a probability measure on $\mathcal{X}$ and $C > 0$. Then the following conditions are equivalent*

(i) *μ satisfies the modified log-sob inequality* (3.3.2) *with constant $C_1 > 0$.*

(ii) *There exists $C_2 > 0$ for all ν probability measure on $\mathcal{X}$,*

$$\widetilde{\mathcal{T}_2}(\mu|\nu) \leqslant C_2 H(\nu|\mu). \tag{3.3.6}$$

where $H(\nu|\mu)$ is the relative entropy of ν with respect to μ, i.e. $H(\nu|\mu) = \mathrm{Ent}_\mu(g)$ if $\nu \ll \mu$ and $g := d\nu/d\mu$, and $H(\nu|\mu) = +\infty$ otherwise. Moreover, $(i) \Rightarrow (ii)$ with $C_2 = C_1/2$, $(ii) \Rightarrow (i)$ with $C_1 = 2C_2$.

Inequality (3.3.6) is related to the concentration phenomenon and was studied by the authors listed above (Dembo, Gozlan, Marton, Roberto, Samson, Tetali, Wintenberger). However, proving directly (3.3.6) for non-trivial measures is not an easy task and, to the best of our knowledge, there exist very few examples of measures satisfying (3.3.6). In fact, Theorem3.3.1 above, together with the important literature on the log-Sobolev inequality provide at once new examples.

That (3.3.2) implies (3.3.6)(with $|\nabla|$ in place of $\widetilde{\nabla}$) is known, in the continuous setting, as Otto-Villani's Theorem [87]. Such a theorem was proved using Otto calculus in the original paper [87] in the Riemannian setting. Soon after, Bobkov, Gentil and Ledoux [11] gave an alternative proof based on Hamilton-Jacobi equation. Then, it was generalized to compact measured geodesic spaces by Lott and Villani [68, 69] (see also [8]), and to general metric spaces by Gozlan [42], see also Gozlan, Roberto and Samson [47] and for an approach based on the Hamilton-Jacobi Semigroup. Later on, the original ingredients of Otto-Villani's paper were successfully adapted to the general metric space framework by Gigli and Ledoux [40]. Our proof follows the Hamilton-Jacobi approach of [11]. We point out that (3.3.6) implies (3.3.2)(with $|\nabla|$ in place of $\widetilde{\nabla}$) is not true in the continuous setting.

Proof of Theorem 3.3.1. We shall show that the modified log-Sobolev inequality (3.3.2) implies the hypercontractivity property (3.3.3) for positive ρ and the modified log-Sobolev inequality (3.3.4) implies the hypercontractivity property (3.3.3) for negative ρ at the same time. To that purpose, fix $\rho \in \mathbb{R}$ and, following [11], define

$$F(t) := \frac{1}{k(t)} \log \left(\int e^{k(t)\widetilde{Q}_t f} \, d\mu \right), \qquad t \geqslant 0$$

with $k(t) := \rho + (t/2C)$. By Theorem 2.2.19, F is differentiable at every point $t > 0$ when $\rho \geqslant 0$ and every $t \in (0, -\rho C/2)$ when $\rho \leqslant 0$. For such points, it holds

$$F'(t) = \frac{k'(t)}{k(t)^2} \frac{1}{\int e^{k(t)\widetilde{Q}_t f} \, d\mu} \left(\mathrm{Ent}_\mu \left(e^{k(t)\widetilde{Q}_t f} \right) + \frac{k(t)^2}{k'(t)} \int e^{k(t)\widetilde{Q}_t f} \frac{\partial}{\partial t} \widetilde{Q}_t f \, d\mu \right).$$

According to Theorem 2.3.9, we have

$$\begin{aligned} \mathrm{Ent}_\mu \left(e^{k(t)\widetilde{Q}_t f} \right) &+ \frac{k(t)^2}{k'(t)} \int e^{k(t)\widetilde{Q}_t f} \frac{\partial}{\partial t} \widetilde{Q}_t f \, d\mu \\ &\leqslant \mathrm{Ent}_\mu \left(e^{k(t)\widetilde{Q}_t f} \right) - \frac{k(t)^2}{2k'(t)} \int |\widetilde{\nabla} \widetilde{Q}_t f|^2 e^{k(t)\widetilde{Q}_t f} \, d\mu \\ &= \mathrm{Ent}_\mu \left(e^{k(t)\widetilde{Q}_t f} \right) - \frac{1}{2k'(t)} \int \left| \widetilde{\nabla} \left[|k(t)| \widetilde{Q}_t f \right] \right|^2 e^{k(t)\widetilde{Q}_t f} \, d\mu \end{aligned}$$

where the last equality follows from Remark 3.2.15. Now we have two cases to deal with : (a) If $\rho \geqslant 0$ and μ satisfies (3.3.2), then $|k(t)| = k(t)$. Hence, applying the modified log-Sobolev inequality (3.3.2) leads to $F'(t) \leqslant 0$. (b) If $\rho \leqslant 0$ and μ satisfies (3.3.4), then $|k(t)| = -k(t)$. Hence applying the modified log-Sobolev inequality (3.3.4) leads also to $F'(t) \leqslant 0$. In both cases $F'(t) \leqslant 0$ implies $F(t) \leqslant F(0)$ which amounts to (3.3.3).

Conversely, suppose that (3.3.3) holds for every $t \geqslant 0$ when $\rho > 0$ (respectively every $t \in [0, -\rho C/2)$ when $\rho < 0$) . Then, in the limit, (3.3.3) implies that $F'(0) \leqslant 0$ and thus (recall that $k'(t) = 1/(2C) > 0$)

$$\mathrm{Ent}_\mu \left(e^{k(0)\widetilde{Q}_0 f} \right) + \frac{k(0)^2}{k'(0)} \int e^{k(0)\widetilde{Q}_0 f} \frac{\partial}{\partial t} \widetilde{Q}_t f |_{t=0} \, d\mu \leqslant 0$$

where we set $\widetilde{Q}_0 f := \lim_{t \to 0} \widetilde{Q}_t f$. By Theorem 2.3.9, since $\alpha(x) = x^2/2$, $\widetilde{Q}_0 f = f$ so that the latter is equivalent to

$$\mathrm{Ent}_\mu \left(e^{\rho f} \right) + 2\rho^2 C \int e^{\rho f} \frac{\partial}{\partial t} \widetilde{Q}_t f |_{t=0} \, d\mu \leqslant 0.$$

Now, according to Theorem 2.3.9, $\frac{\partial}{\partial t}\widetilde{Q}_t f(x)|_{t=0} = -\frac{1}{2}|\widetilde{\nabla} f|^2(x)$, $x \in \mathcal{X}$ so that

$$\mathrm{Ent}_\mu\left(e^{\rho f}\right) - C\int e^{\rho f}|\widetilde{\nabla}(|\rho|f)|^2\, d\mu \leqslant 0.$$

This precisely amounts to proving (3.3.2) (respectively (3.3.4)) when $\rho \geqslant 0$ (resp. $\rho \leqslant 0$). The proof of Theorem 3.3.1 is complete. □

proof of Theorem3.3.5. $(i) \Rightarrow (ii)$, recall the following generalization of Bobkov-Gotze dual characterization borrowed from [50, Theorem 5.5] :

Inequality (3.3.6) holds if and only if for all bounded continuous function $\varphi\colon \mathcal{X} \to \mathbb{R}$ it holds

$$\int \exp\left\{\frac{2}{C}\widetilde{Q}_1\varphi\right\} d\mu \leqslant \exp\left\{\frac{2}{C}\int \varphi\, d\mu\right\}. \tag{3.3.7}$$

Now, (3.3.3) applied to $\rho = 0$ and $t = 1$ precisely amounts to (3.3.7), since by definition $\|g\|_0 := \exp\{\int \log g\, d\mu\}$ for $g \geqslant 0$. Hence the result, thanks to the dual characterization of [50].

Now we turn to prove $(ii) \Rightarrow (i)$. According to [50, Proposition 8.3], (ii) implies that for all $\lambda \in (0, 1/C_2)$, the following inequality holds for all bounded lower semi continuous function f :

$$\mathrm{Ent}_\mu(e^f) \leqslant \frac{1}{1-\lambda C}\int (f - R_c^\lambda f)e^f\, d\mu.$$

Here in our settings, $R_c^\lambda f(x) := \inf_{p\in\mathcal{P}(\mathcal{X})}\{\int f dp + \frac{\lambda}{2}(\int d(x,.)dp)^2\} = \widetilde{Q}_{1/\lambda}f(x)$. According to Proposition 2.2.7, $t \mapsto \widetilde{Q}_t f$ is convex. Thus,

$$R_c^\lambda f - f = \widetilde{Q}_{1/\lambda} f - f \geqslant \frac{1}{\lambda}\frac{\partial}{\partial t}\widetilde{Q}_t f|_{t=0} = -\frac{1}{2\lambda}|\widetilde{\nabla} f|^2.$$

We deduce that

$$\begin{aligned}\mathrm{Ent}_\mu(e^f) &\leqslant \frac{1}{1-\lambda C}\int (f - R_c^\lambda f)e^f\, d\mu \\ &\leqslant \frac{1}{2\lambda(1-\lambda C)}\int |\widetilde{\nabla} f|^2 e^f\, d\mu.\end{aligned}$$

Optimizing λ with $\lambda = \frac{1}{2C}$ yields the result. □

Remark 3.3.8. *Since $|\widetilde{\nabla} f|^2(x) \leqslant 1$ for any 1-Lipschitz function, the usual Herbst argument (see e.g. [5, Chapter 7], [17]) applies and leads to the following concentration result : if μ satisfies the modified log-Sobolev inequality* (3.2.8)*, then any* 1*-Lipschitz function $f\colon \mathcal{X} \to \mathbb{R}$ with $\int f\, d\mu = 0$ satisfies $\mu(f \geqslant h) \leqslant e^{-h^2/(4C)}$ for all $h \geqslant 0$.*

3.4 Modified Poincaré inequality

In this section, we prove that the Poincaré inequality (3.2.7), the transport-entropy inequality (3.2.8) with a quadratic-linear cost, and a modified log-sob type inequality are equivalent. This will extend to our setting similar results known in the continuous, see [11]. We start with the definition of quadratic-linear cost function.

Definition 3.4.1 (Quadratic-linear cost function)**.** *A quadratic-linear cost function* $\alpha_a^h : \mathbb{R}^+ \to \mathbb{R}$*,* $a, h > 0$ *is such that*

$$\alpha_a^h(x) = \begin{cases} ax^2 & x \leqslant h \\ 2ax - ah^2 & x > h. \end{cases}$$

The first main theorem of this section is a characterization of the Poincaré Inequality (3.2.7) in term of a modified log-Sobolev inequality with quadratic-linear cost. Such a characterization is an extension of a well known result of Bobkov and Ledoux [15].

Theorem 3.4.2. *A probability measure* μ *on* $\mathcal{X}$ *satisfies the Poincaré Inequality* (3.4.3) *with constant* C *if and only if* μ *satisfies the modified log-Sobolev inequality* (3.2.8) *with constant* C' *and cost* α_a^h*. More precisely*

- (3.4.3) *implies* (3.2.8) *with* $C' = K(c)$*,* $a = \frac{1}{4K(c)}$ *and* $h = 2cK(c)$ *for any* $c < 2/\sqrt{C}$ *with* $K(c)$ *defined in theorem 3.4.3 ;*
- (3.2.8) *implies* (3.4.3) *with* $C = C'$*.*

We observe that, with respect to [15] there is a loss in the constant $K(c)$. This is technical. Indeed, the proof essentially follows [15], but it cannot be extended directly and one has to be careful in many points, for technical reasons coming from the gradient $\widetilde{\nabla}$. We postpone the proof at the end of this section. Now we state the second main theorem of the section.

Theorem 3.4.3. *Let* μ *be a probability measure on* $\mathcal{X}$*. The following propositions are equivalent.*

(i) *There exists a constant* $C_1 > 0$ *such that* μ *satisfies the Poincaré inequality* (3.2.7) *with constant* C_1*.*

(ii) *There exist constants* $C_2, a, h > 0$ *such that* μ *satisfies the weak transport-entropy inequality* (3.2.8) *with constant* C_2 *and cost* α_a^h*.*

More precisely,

- (ii) *implies* (i) *with* $C_1 = aC_2$ *;*
- (i) *implies* (ii) *with* $C_2 = K(c)/2$*,* $a = \frac{1}{4K(c)}$ *and* $h = 2cK(c)$ *for any* $c < 2/\sqrt{C_1}$ *and*

$$K(c) := \frac{C_1}{2}\left(\frac{2 + 2e^2 + c\sqrt{C_1}}{2 - c\sqrt{C_1}}\right)^2 e^{c\sqrt{5C_1}}.$$

Remark 3.4.4. *As a direct consequence of the above theorem, we observe that the weak transport-entropy inequality $\widetilde{T}_\alpha$ with cost function $\alpha(x) := \frac{x^2}{2}$ and constant C implies the Poincaré inequality* (3.2.7) *with constant $C/2$. Indeed, since $\alpha(x) = \frac{x^2}{2} \geqslant \alpha_{1/2}^2(x)$, the weak transport-entropy inequality $\widetilde{T}_2(C)$ implies $\widetilde{T}_{\alpha_{1/2}^2}(C)$ and the conclusion follows from Item* (*ii*) *of Theorem 3.4.3.*

The proof of Theorem 3.4.3 relies on Theorem 3.4.2.

Proof of Theorem 3.4.3. We will first prove that (i) implies (ii). Fix $c < 2/\sqrt{C}$ and set $C = C_1$, $a = \frac{1}{4K(c)}$ and $h = 2cK(c)$. Thanks to Theorem 3.4.2 for all $f\colon \mathcal{X} \to \mathbb{R}$ bounded, it holds

$$\mathrm{Ent}_\mu(e^f) \leqslant K(c) \int (\alpha_a^h)^*(|\widetilde{\nabla} f|) e^f \, d\mu.$$

Arguing as in the proof of Theorem 3.3.1 with $k(t) = 2t/K(c)$, and using the fact [2] that $(\alpha_a^h)^*(\lambda u) \leqslant \lambda^2 (\alpha_a^h)^*(u)$ as soon as $u \leqslant 2ah$, we obtain (details are left to the reader) that the family of operators $(\exp\{\widetilde{Q}_t\})_{t\geqslant 0}$, with $\widetilde{Q}$ defined with the cost α_a^h, is hypercontractive which in turn guarantees that

$$\int \exp\left\{\frac{2}{K(c)} \widetilde{Q}_1 f\right\} d\mu \leqslant \exp\left\{\frac{2}{K(c)} \int f \, d\mu\right\}$$

for all bounded function f. The conclusion follows from the dual characterization of [50] (that we recalled in (3.3.7)).

Next we prove that $(ii) \Rightarrow (i)$. By an easy argument it is enough to prove (3.2.7) for all bounded Lipschitz function f on $\mathcal{X}$. According to [50] (see (3.3.7)), the transport-entropy inequality (3.2.8), with cost $(\alpha_a^h)^*$, is equivalent to say that for all continuous bounded function φ on $\mathcal{X}$ it holds

$$\int \exp\left\{\frac{2}{C_2} \widetilde{Q}_1 \varphi\right\} d\mu \leqslant \exp\left\{\int \frac{2}{C_2} \varphi \, d\mu\right\}$$

where $\widetilde{Q}$ is defined with the cost α_a^h. Fix $l > 0$, let f be a l-Lipschitz function and set $\varphi := tf$. The latter inequality reduces to $\int \exp\left\{\frac{2}{C_2} \widetilde{Q}_1 tf\right\} d\mu \leqslant \exp\left\{\int \frac{2}{C_2} tf \, d\mu\right\}$. Hence, for $t < (ah)/l$, by Lemma 2.2.23 below, we get

$$\int \exp\left\{\frac{2}{C_2} t \widetilde{Q}_t f\right\} d\mu \leqslant \exp\left\{\int \frac{2}{C_2} tf \, d\mu\right\}.$$

An expansion around $t = 0$ yields that

$$\int \left(1 + \frac{2}{C_2} tf + \frac{1}{2} t^2 \left(\frac{4}{C_2^2} f^2 + \frac{4}{C_2} \frac{\partial}{\partial t} \widetilde{Q}_t f_{|t=0}\right) + o(t^2)\right) d\mu \leqslant 1 + t \frac{2}{C_2} \int f \, d\mu + \frac{1}{2} t^2 \frac{4}{C_2^2} \int f \, d\mu + o(t^2). \quad (3.4.5)$$

2. For the reader convenience we observe that $(\alpha_a^h)^*(x) = K(c)x^2$ if $|x| \leqslant c$ and $(\alpha_a^h)^*(x) = +\infty$ otherwise.

Therefore (comparing the coefficients of t^2), it holds $\mathrm{Var}_\mu(f) \leqslant -C_2 \int \frac{\partial}{\partial t}\widetilde{Q}_t f_{|t=0} d\mu$. Applying Theorem 2.3.9 we arrive at $\mathrm{Var}_\mu(f) \leqslant C_2 \int \alpha_a^{h*}\left(|\widetilde{\nabla} f|\right) d\mu$, which in turn, since $\alpha_a^{h*}\left(|\widetilde{\nabla} f|(x)\right) = a\left(|\widetilde{\nabla} f|(x)\right)^2$ for $l \leqslant ah$, implies that for all ah-Lipschitz function f, it holds

$$\mathrm{Var}_\mu(f) \leqslant aC_2 \int |\widetilde{\nabla} f|^2 \, d\mu.$$

Replacing f by λf with $\lambda \in \mathbb{R}^+$, we conclude that the above inequality holds for all Lipschitz function f and thus μ satisfies the Poincaré inequality with constant aC_2. This ends the proof of the theorem. □

Now we prove Theorem3.4.2.

The proof relies on the following three propositions.

Proposition 3.4.6. *If μ satisfies the Poincaré inequality* (3.2.7) *with constant $C > 0$, then for all $f : \mathcal{X} \mapsto \mathbb{R}$,*

$$\mathrm{Var}_\mu(fe^{f/2}) \leqslant C \int |\widetilde{\nabla} f|^2 \left(1 + e^4 + f + \frac{f^2}{4}\right) e^f \, d\mu.$$

Proposition 3.4.7. *If μ satisfies the Poincaré inequality* (3.2.7) *with constant $C > 0$, then for any bounded c-Lipschitz function f on $\mathcal{X}$ with $c < 2/\sqrt{C}$ and $\int f \, d\mu = 0$,*

$$\int f^2 e^f \, d\mu \leqslant C \left(\frac{2 + 2e^2 + c\sqrt{C}}{2 - c\sqrt{C}}\right)^2 \int |\widetilde{\nabla} f|^2 e^f \, d\mu.$$

Proposition 3.4.8. *If μ satisfies the Poincaré inequality* (3.2.7) *with constant $C > 0$, then for any bounded function f on $\mathcal{X}$ with $\|f\|_{\mathrm{Lip}} \leqslant c$ and $\int f \, d\mu = 0$, we have*

$$\int f^2 \, d\mu \leqslant e^{c\sqrt{5C}} \int f^2 e^{-|f|} \, d\mu.$$

We postpone the proof of the above propositions to prove Theorem 3.4.2.

Proof of Theorem 3.4.2. Changing f into $f+$constant we may assume that $\int f \, d\mu = 0$. Since $u \log u \geqslant u - 1$ for all $u \geqslant 0$, we have

$$\mathrm{Ent}_\mu(e^f) \leqslant \int (fe^f - e^f + 1) \, d\mu = \int \left(\int_0^1 tf^2 e^{tf} \, dt\right) d\mu.$$

Let $\varphi(t) := \int f^2 e^{tf} \, d\mu$, $t \in [0,1]$. By convexity, φ attains its maximum at either $t = 0$ or $t = 1$. By Proposition 3.4.8, and since $e^{-|f|} \leqslant e^f$, $\varphi(0) \leqslant e^{c\sqrt{5C}}\varphi(1)$. Thus, for every $t \in [0,1]$, $\varphi(t) \leqslant e^{c\sqrt{5/C}}\varphi(1)$. It follows that

$$\mathrm{Ent}_\mu(e^f) \leqslant \int_0^1 t\varphi(t) \, dt \leqslant \int_0^1 te^{c\sqrt{5C}}\varphi(1) \, dt = \frac{1}{2} e^{c\sqrt{5C}} \int f^2 e^f \, d\mu.$$

Together with Proposition 3.4.7, Theorem 3.4.2 is established. □

Now let us prove Propositions 3.4.6, 3.4.7 and 3.4.8.

Proof of Proposition 3.4.6. Let $G := u \mapsto ue^{u/2}$ and observe that it is decreasing on $(-\infty, -2]$, increasing on $(-2, \infty)$ and its minimum is $G(-2) = -2e^{-1}$. Now, starting from G, define an increasing function H as G when G is increasing and as the symmetric of G with respect to $y = G(-2)$ when G is non-increasing. More precisely,

$$H(x) := \begin{cases} -xe^{x/2} - 4e^{-1} & \text{if } x \leqslant -2 \\ xe^{x/2} & \text{if } x > -2. \end{cases}$$

Observe that $|H(x) + 2e^{-1}| = xe^{x/2} + 2e^{-1}$, $x \in \mathbb{R}$. Hence, using that $\mathrm{Var}_\mu(|g|) \leqslant \mathrm{Var}_\mu(g)$, it holds

$$\mathrm{Var}_\mu(fe^{f/2}) = \mathrm{Var}_\mu(fe^{f/2} + 2e^{-1}) \leqslant \mathrm{Var}_\mu(H \circ f + 2e^{-1}) = \mathrm{Var}_\mu(H \circ f).$$

Now applying the Poincaré Inequality (3.2.7) and Proposition 3.2.14 we have

$$\mathrm{Var}_\mu(H \circ f) \leqslant C \int |\widetilde{\nabla}(H \circ f)|^2 \, d\mu \leqslant C \int |\widetilde{\nabla} f|^2 |\widetilde{\nabla} H|^2 (f) \, d\mu. \tag{3.4.9}$$

Since H is increasing, we have

$$0 \leqslant |\widetilde{\nabla} H|(u) = \sup_{v<u} \frac{H(u) - H(v)}{u - v} = \sup_{v<u} \left\{ \frac{1}{u-v} \int_{(v,u)} H'(t) \, dt \right\} \leqslant \sup_{t<u} H'(t).$$

After some basic analysis, we have the following facts

- if $u < -4$, $\sup_{t<u} H'(t) = |(1 + u/2)e^{u/2}|$ since $H'(t) = |(1 + t/2)e^{t/2}|$ is increasing on $(-\infty, -4]$;
- if $u \in [-4, 0]$, $\sup_{t<u} H'(t) \leqslant 1 \leqslant e^2 e^{u/2}$;
- if $u > 0$, $\sup_{t<u} H'(t) = |(1 + u/2)e^{u/2}|$ since H' is increasing on $[0, \infty)$ and $H'(u) > H(0) = 1 \geqslant \sup_{t \leqslant 0} H'(t)$.

As a consequence, we have $|\widetilde{\nabla} H|^2(u) \leqslant ((1 + u/2)^2 + e^4)\, e^u$. Therefore

$$\mathrm{Var}_\mu(H \circ f) \leqslant C \int |\widetilde{\nabla} f|^2 |\widetilde{\nabla} H|^2 (f) \; d\mu \leqslant C \int |\widetilde{\nabla} f|^2 \left(1 + e^4 + f + \frac{f^2}{4}\right) e^f \, d\mu.$$

This ends the proof of the proposition. □

Proof of Proposition 3.4.7. Set $a^2 = \int f^2 e^f \, d\mu$ and $b^2 = \int |\widetilde{\nabla} f|^2 e^f \, d\mu$. By the Poincaré inequality (3.2.7), for any two bounded functions g and h on $\mathcal{X}$ with $\int g \, d\mu = 0$,

$$\left(\int gh \, d\mu \right)^2 \leqslant \left(\int g^2 \, d\mu \right) \left(\int h^2 \, d\mu \right) \leqslant \left(C \int |\widetilde{\nabla} g|^2 \, d\mu \right) \left(C \int |\widetilde{\nabla} h|^2 \right) d\mu.$$

Therefore, since $\int f \, d\mu = 0$,

$$\left(\int fe^{f/2} \, d\mu \right)^2 \leqslant C^2 \left(\int |\widetilde{\nabla} f|^2 \, d\mu \right) \left(\int |\widetilde{\nabla} e^{f/2}|^2 \, d\mu \right).$$

Set $G(u) = e^{u/2}$, $u \in \mathbb{R}$. The convexity of G guarantees that $|\widetilde{\nabla} G| = |G'|$. Thus by Proposition 3.2.14, it holds $|\widetilde{\nabla} e^{f/2}|^2 \leqslant \frac{1}{4}|\widetilde{\nabla} f|^2 e^f$. Hence

$$\left(\int f e^{f/2}\, d\mu\right)^2 \leqslant \frac{1}{4} C^2 c^2 b^2.$$

On the other hand, according to Proposition 3.4.6,

$$\begin{aligned}\operatorname{Var}_\mu(f e^{f/2}) &\leqslant C \int |\widetilde{\nabla} f|^2 \left(1 + e^4 + f + \frac{f^2}{4}\right) e^f \, d\mu \\ &\leqslant C\left((1+e^4)b^2 + \int |\widetilde{\nabla} f|^2 f e^f d\mu + \frac{c^2 a^2}{4}\right).\end{aligned}$$

By Cauchy-Schwarz inequality,

$$\int |\widetilde{\nabla} f|^2 f e^f \, d\mu \leqslant \left(\int |\widetilde{\nabla} f|^2 f^2 e^f \, d\mu\right)^{1/2} \left(\int |\widetilde{\nabla} f|^2 e^f \, d\mu\right)^{1/2} \leqslant cab,$$

so that

$$\operatorname{Var}_\mu(f e^{f/2}) \leqslant C\left(\left(b + \frac{ca}{2}\right)^2 + e^4 b^2\right).$$

Then we get that

$$a^2 = \left(\int f e^{f/2}\, d\mu\right)^2 + \operatorname{Var}_\mu(f e^{f/2}) \leqslant \frac{1}{4} C^2 c^2 b^2 + C\left(b + \frac{ca}{2}\right)^2 + C e^4 b^2.$$

Simplifying this inequality, we end up with

$$\frac{a}{b} \leqslant \sqrt{C}\left(\frac{2 + 2e^2 + c\sqrt{C}}{2 - c\sqrt{C}}\right),$$

and the conclusion follows. □

Proof of Proposition 3.4.8. For all $u > 0$ and all $v \in \mathbb{R}$, we have $2|v| \leqslant u + (1/u)v^2$. Hence $2|v|^3 \leqslant uv^2 + (1/u)v^4$ and therefore,

$$2\int |f|^3 \, d\mu \leqslant u \int f^2 \, d\mu + \frac{1}{u}\int f^4 \, d\mu. \tag{3.4.10}$$

By the Poincaré inequality (3.2.7) it holds

$$\int f^2 \, d\mu \leqslant C \int |\widetilde{\nabla} f|^2 \, \mu(dx) \leqslant c^2 C,$$

so that $\left(\int f^2 \, d\mu\right)^2 \leqslant c^2 C \int f^2 \, d\mu$.

On the other hand, set $G(t) = t^2$, $t \geqslant 0$. The convexity of G guarantees that for all $t \geqslant 0$, $|\widetilde{\nabla} G|(t) = |G'|(t)$. Hence, according to Proposition 3.2.14, it holds

$$\operatorname{Var}_\mu(f^2) = \operatorname{Var}_\mu(|f|^2) \leqslant C \int |\widetilde{\nabla}(|f|^2)|^2 d\mu \leqslant 4C \int f^2 |\widetilde{\nabla}|f||^2 d\mu \leqslant 4c^2 C \int f^2 d\mu$$

where in the last inequality we used that $|f|$ is c-Lipschitz. It follows that $\int f^4\,d\mu = (\int f^2\,d\mu)^2 + \mathrm{Var}_\mu(f^2) \leqslant 5c^2C \int f^2\,d\mu$. Hence, from (3.4.10), we obtain that for every $u > 0$,

$$2\int |f|^3\,d\mu \leqslant \left(u + \frac{5c^2C}{u}\right)\int f^2\,d\mu.$$

Minimizing over $u > 0$, we get

$$\int |f|^3\,d\mu \leqslant c\sqrt{5C}\int f^2\,d\mu. \tag{3.4.11}$$

Consider now the probability measure $\tau(dx) = f(x)^2\,\mu(dx)/(\int f^2\,d\mu)$. By Jensen's inequality,

$$\int f^2 e^{-|f|}\,d\mu = \int e^{-|f|}\,d\tau \int |f|^2\,d\mu \geqslant e^{-\int f\,d\tau}\int |f|^2\,d\mu.$$

By (3.4.11) we conclude that

$$\int |f|\,d\tau = \frac{\int |f|^3\,d\mu}{\int f^2\,d\mu} \leqslant c\sqrt{5C},$$

from which the result follows. $\square$

Proof of Theorem 3.4.2. Changing f into f+constant we may assume that $\int f\,d\mu = 0$. Since $u\log u \geqslant u - 1$ for all $u \geqslant 0$, we have

$$\mathrm{Ent}_\mu(e^f) \leqslant \int (fe^f - e^f + 1)\,d\mu = \int \left(\int_0^1 tf^2e^{tf}\,dt\right)d\mu.$$

Let $\varphi(t) := \int f^2 e^{tf}\,d\mu$, $t \in [0,1]$. By convexity, φ attains its maximum at either $t = 0$ or $t = 1$. By Proposition 3.4.8, and since $e^{-|f|} \leqslant e^f$, $\varphi(0) \leqslant e^{c\sqrt{5C}}\varphi(1)$. Thus, for every $t \in [0,1]$, $\varphi(t) \leqslant e^{c\sqrt{5/C}}\varphi(1)$. It follows that

$$\mathrm{Ent}_\mu(e^f) \leqslant \int_0^1 t\varphi(t)\,dt \leqslant \int_0^1 te^{c\sqrt{5C}}\varphi(1)\,dt = \frac{1}{2}e^{c\sqrt{5C}}\int f^2e^f\,d\mu.$$

Together with Proposition 3.4.7, Theorem 3.4.2 is established. $\square$

3.5 Synthesis and Examples

In continuous space, we have the following relations

$$\begin{array}{ccccccccc}
\mathrm{LS} & \Longrightarrow & \mathrm{T_2I} & \Longrightarrow & \mathrm{T_2} & \Longrightarrow & \mathrm{T_\alpha} & \Longleftrightarrow & \mathrm{P} \\
\Updownarrow & & & & \Downarrow & & & & \\
\textit{Hypercon.} & & & & \textit{Gaussian concentration} & & & &
\end{array}$$

where α is a quadratic-linear cost function.

In the discrete settings, with the gradient $\widetilde{\nabla}$, up to some constants, we have proved the relationship following :

$$\begin{array}{lllll} \mathrm{mLS(I)} & \Leftrightarrow H^+ & \Leftrightarrow \widetilde{\mathrm{T}}_2^- \Rightarrow & \widetilde{\mathrm{T}}_\alpha^- \Leftrightarrow \widetilde{\mathrm{P}} \\ \mathrm{mLS(II)} & \Leftrightarrow H^- & \Rightarrow \widetilde{\mathrm{T}}_2^+ \Rightarrow & \widetilde{\mathrm{T}}_\alpha^+ \Rightarrow \widetilde{P} \end{array}$$

Where $mLSI(I)$ and $mLSI(II)$ stand for the modified log-Sobolev inequality of type I and type II respectively, H^+ and H^- stand for the two hypercontrativity properties (3.3.3) with respect to ρ positive and negative respectively. $\widetilde{P}$ denotes the Poincaré inequality with the gradient $\widetilde{\nabla}$.

Next we give examples of measures satisfying log-Sobolev/Poincaré/transport-entropy type inequalities.

Measures satisfying the log-Sobolev inequality (3.3.6) **and the transport-entropy** (3.3.6)

As already mentioned, the classical log-Sobolev inequality (3.2.11) implies the (say) classical modified log-Sobolev inequality (3.2.10) which, thanks to Proposition 3.2.12 implies under mild assumptions the modified log-Sobolev inequality (3.2.8), which finally, thanks to Theorem 3.3.1, implies the transport-entropy inequality (3.3.6). The latter is usually hard to obtain directly. The above chain of implication applies to a lot of different situations, including highly non-trivial examples. Let us mention random walks on the hypercube, on the symmetric group or the complete graph (see [17] where optimal (or almost optimal) bounds are given for (3.2.10)) the optimal bound in (3.2.11) for the lamplighter graph can be found in [1], and in [73] for the Ising model at high temperature, on the lattice or on trees. Many other examples can be found in [27]... Bound on the constant in the tranport-entropy inequality (3.3.6) are new for all examples listed above, to the best of our knowledge.

As an illustration, consider the uniform measure $\mu \equiv 1/2^n$ on the hypercube $\{0, 1\}$ associated to the Markov chain that jumps from x to anyone of its nearest neighbors (*i.e.* any string x' that differs from x in exactly one coordinate) with equal probability $(1/n)$. Then μ satisfies Gross' Inequality (3.2.11) with constant $n/2$ [52], the classical modified log-Sobolev inequality (3.2.10) with constant $n/8$ [17], and thus, by Proposition 3.2.12 (note that $L = 1$), the modified log-Sobolev inequality (3.3.6) with constant $n/4$, and in turn, thanks to Theorem3.3.1, the transport-entropy inequality (3.3.6) holds with constant $n/8$.

In the case of the symmetric group S_n, consisting of $n!$ permutation (of n elements), equipped with the transposition distance (*i.e.* two permutations are at distance 1 if one is the other composed with a transposition). Each permutation has $n(n-1)/2$ neighbors and the Markov chain that jumps uniformly at random to any neighbor is reversible with respect to the uniform measure $\mu \equiv 1/n!$. Gross' Inequality is known to hold with a constant of order $n^3 \log n$ [65], while the classical modified log-Sobolev inequality (3.2.10) holds with constant $C \leqslant n(n-1)^2/2$

[17]. Therefore, by Proposition 3.2.12 (again note that $L=1$), μ satisfies the modified log-Sobolev inequality (3.3.6) with constant $n(n-1)^2$ and in turn, thanks to Theorem 3.3.1, the transport-entropy inequality (3.3.6) with constant $n(n-1)^2/2$.

Poincaré inequality

The next proposition extends a well-known result that asserts that the Poincaré inequality holds on bounded domains. We will then give examples of measures satisfying the Poincaré inequality (3.2.7) but not the one with the usual gradient.

Proposition 3.5.1. *Assume that the support of the probability measure μ has a finite diameter and let $D = \sup_{x,y\in \mathrm{Supp}(\mu)}\{d(x,y)\}$. Then μ satisfies the Poincaré Inequality* (3.2.7) *with constant at most $D^2/2$.*

proof. For all $x,y \in \mathrm{Supp}(\mu)$, $f(x)-f(y) \leqslant d(x,y)|\widetilde{\nabla} f|(x) \leqslant D|\widetilde{\nabla} f|(x)$. Thus, for all continuous function f on $\mathcal{X}$, it holds

$$\mathrm{Var}_\mu(f) = \frac{1}{2}\iint_{\mathrm{Supp}(\mu)^2} \left(f(x)-f(y)\right)^2 \mu(dx)\mu(dy) \leqslant \frac{D^2}{2}\int |\widetilde{\nabla} f|^2 \, d\mu.$$

□

Now, on $X=\mathbb{R}$ consider the following probability measure $\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. We claim that μ satisfies the Poincaré inequality (3.2.7), but not the (classical) Poincaré inequality with the Euclidean gradient.

Indeed, Proposition 3.5.1 applies and leads to the Poincaré inequality (3.2.7) with constant at most 1/2. On the other hand, the mapping $f\colon \mathbb{R} \ni x \mapsto 2x^3 - 3x^2+1$ satisfies $f(0)=1$, $f(1)=0$ and $f'(0)=f'(1)=0$ so that $\mathrm{Var}_\mu(f) = \frac{1}{4}\left(f(0)-f(1)\right)^2 = \frac{1}{4}$ and $\int f'^2\,d\mu = 0$ which proves the claim.

Let us prove now that μ also satisfies the modified log-Sobolev inequality (3.3.4). Given $f\colon \mathbb{R}\to\mathbb{R}$ with $f(0)\geqslant f(1)$ (the other direction is similar), we have $f(0) - f(1) \leqslant |\widetilde{\nabla} f|(0)$ so that $\left(f(0)-f(1)\right)^2 e^{f(0)} \leqslant \int_{\mathbb{R}} |\widetilde{\nabla} f|^2 e^f \, d\mu$. Thus, to prove that the modified log-Sobolev inequality (3.3.4) holds, it is enough to prove the existence of a constant C such that

$$\mathrm{Ent}_\mu(f) \leqslant C\left(f(0)-f(1)\right)^2 e^{f(0)}$$

or equivalently

$$f(0)e^{f(0)} + f(1)e^{f(1)} - \left(e^{f(0)}+e^{f(1)}\right)\log\left(\frac{e^{f(0)}+e^{f(1)}}{2}\right) \leqslant 2C\left(f(0)-f(1)\right)^2 e^{f(0)}.$$

Setting $u := f(0)-f(1) \geqslant 0$, the latter is equivalent to prove that

$$ue^u - (e^u+1)\log\left(\frac{e^u+1}{2}\right) \leqslant Cu^2e^u \qquad \forall u \geqslant 0$$

which is an easy exercise.

Chapitre 4

Ricci curvature on graphs

Abstract

In this chapter, we study various transport-information inequalities under three different notions of Ricci curvature in the discrete setting : the curvature-dimension condition of Bakry and Émery [6], the exponential curvature-dimension condition of Bauer *et al.* [9] and the coarse Ricci curvature of Ollivier [85]. We prove that under a curvature-dimension condition or coarse Ricci curvature condition, an L_1 transport-information inequality holds ; while under an exponential curvature-dimension condition, some weak-transport information inequalities hold. As an application, we establish a Bonnet-Myers theorem under the curvature-dimension condition $\mathrm{CD}(\kappa, \infty)$ of Bakry and Émery [6].

4.1 Introduction

In the analysis of the geometry of Riemannian manifolds, Ricci curvature plays an important role. In particular, Ricci curvature lower bounds immediately yield powerful functional inequalities we mentioned in the last chapter, such as the logarithmic Sobolev inequality, which in turn implies transport-entropy and transport-information inequalities.

However, the space we considered here is a graph. Let $\mathcal{X}$ be a finite (or countably infinite) discrete space and K be an irreducible Markov kernel on $\mathcal{X}$. Assume that for any $x \in \mathcal{X}$, we have

$$\sum_y K(x,y) = 1. \tag{4.1.1}$$

This condition is a normalization of the time scale, enforcing that jump attempts occur at rate 1. We also define $J(x) := 1 - K(x,x)$ and $J := \sup_{x\in\mathcal{X}} J(x)$. J is a measure of the laziness of the chain, estimating how often jump attempts end with the particle not moving. Since we assume the kernel is irreducible, $0 < J \leqslant 1$.
We shall always assume there exists a reversible invariant probability measure π, satisfying the detailed balance relation

$$K(x,y)\pi(x) = K(y,x)\pi(y) \quad \forall x,y \in \mathcal{X}.$$

We denote by L the generator of the continuous-time Markov chain associated to the kernel K, which is given by

$$Lf(x) = \sum_y (f(y) - f(x))K(x,y).$$

Let $P_t = e^{tL}$ be the associated semigroup, acting on functions, and P_t^* its adjoint, acting on measures. We also define the Γ operator, given by

$$\Gamma(f,g)(x) := \frac{1}{2}\sum_y (f(y) - f(x))(g(y) - g(x))K(x,y)$$

and write $\Gamma(f) := \Gamma(f,f)$.

The distance on $\mathcal{X}$ we shall use is the *graph distance* associated to the Markov kernel. If we consider $\mathcal{X}$ as the set of vertices of a graph, with edges between all pairs of vertices (x,y) such that $K(x,y) > 0$, d shall be the usual graph distance. More formally, it is defined as

$$d(x,y) := \inf\{n \in \mathbb{N}; \exists x_0,..,x_n | x_0 = x,\ x_n = y,\ K(x_i, x_{i+1}) > 0\ \forall 0 \leqslant i \leqslant n-1\}.$$

In this setting, the transport cost $\mathcal{T}$, the weak-transport cost $\overline{\mathcal{T}}$ and $\tilde{\mathcal{T}}$ are well define, as well as the related entropy $H(\nu|\pi)$ for all probability measure ν. Thus, the definitions in last chapter of the transport entropy inequality and the weak-transport entropy inequality are still available. In order to define the inequalities involving the fisher information, we need to introduce the notion of fisher information in graphs.

Definition 4.1.2. *(Fisher information) Let f be a nonnegative function defined on $\mathcal{X}$. Define the Fisher information I_π of f with respect to π as*

$$\mathcal{I}_\pi(f) := 4\int \Gamma(\sqrt{f})d\pi = 2\sum_{x\in cX}\sum_{y\in\mathcal{X}}(\sqrt{f(y)}-\sqrt{f(x)})^2K(x,y)\pi(x).$$

The factor 4 in this definition comes from the analogy with the continuous setting, where

$$4\int|\nabla\sqrt{f}|^2d\pi = \int|\nabla\log f|^2fd\mu.$$

In the continuous setting, the Fisher information can be written as $\int\nabla\log f\cdot\nabla fd\pi$, so we can define a modified Fisher information as

$$\tilde{\mathcal{I}}_\pi(f) := \int\Gamma(f,\log f)d\pi, \tag{4.1.3}$$

which corresponds to the entropy production functional of the Markov chain.

There is a third way to rewrite the Fisher information for the continuous settings as $\int\frac{|\nabla f|^2}{f}d\mu$, and one can also define another modified Fisher information as

$$\overline{\mathcal{I}}_\pi(f) := \int\frac{\Gamma(f)}{f}d\pi.$$

Of course, there are many other ways to re-write the Fisher information in the continuous setting, each leading to a different definition in the discrete setting. We only stated here the three versions we shall use in this work.

In the discrete setting, $\mathcal{I}_\pi(f)$, $\tilde{\mathcal{I}}_\pi(f)$ and $\overline{\mathcal{I}}_\pi(f)$ are not equal in general. It is easy to see that $\mathcal{I}_\pi(f)\leqslant\tilde{\mathcal{I}}_\pi(f)$ and $\mathcal{I}_\pi(f)\leqslant\overline{\mathcal{I}}_\pi(f)$ If f is the density function of a probability measure ν with respect to π, since $(\sqrt{f(y)}-\sqrt{f(x)})^2\leqslant f(x)+f(y)$, and since π is reversible, one can deduce that

$$\mathcal{I}_\pi(f)\leqslant 2\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}}\left(f(x)+f(y)\right)K(x,y)\pi(x)\leqslant 4J.$$

Here we can see that in discrete settings, the Fisher information is in fact bounded from above, which is not true in continuous settings.

Now the next question one would want to answer is how to define Ricci curvature lower bounds in discrete settings. The natural approach would be to define it as a discrete analogue of a definition valid in the continuous setting. There are several equivalent definitions one can try to use (see [4] for those definitions in the continuous settings and for the equivalences between them). However, in discrete spaces, we lose the chain rule, and these definitions are no longer equivalent.

Several notions of curvature have been proposed in the last few years. Here we shall consider three of them : the curvature-dimension condition of Bakry and Émery [6], the exponential curvature-dimension condition of Bauer *et al.* [9] and the coarse Ricci curvature of Ollivier [85]. Other notions that have been developed (and which we shall not discuss further here) include the entropic Ricci curvature defined by Erbar and Maas in [32] and Mielke in [80], which is based on the Lott-Sturm-Villani definition of curvature [69, 96], geodesic convexity along interpolations in [50] and [66], rough curvature bounds in [19]. It is still an open problem to compare these various notions of curvature. We refer readers to the forthcoming survey [25] for a more general introduction.

In this chapter, we will obtain transport inequalities under the above three notions of curvature conditions and give some applications.

Since we shall deal with three different types of Ricci curvature lower bounds, in order to avoid confusions, we always denote κ for the Bakry-Émery curvature condition, κ_e for the exponential curvature dimension condition and κ_c for a lower bound on the Coarse Ricci curvature. Throughout the paper, $\kappa, \kappa_e, \kappa_c$ will always be positive numbers.

Let us begin with the curvature-dimension condition of Bakry and Émery $\mathrm{CD}(\kappa, \infty)$.

The third notion of curvature we shall now introduce is the coarse Ricci curvature. In order to define it, we first need to introduce Wasserstein distances.

4.2 Preliminary

In this section, we present some technical lemma and some general results of functional inequalities in the graph setting, without assuming any curvature condition. We shall apply the tools of chapter 2 and chapter 3. Since the fisher information is defined by the operator Γ, we will compare it with the gradient length $\widetilde{\nabla}$, in order to connect it with the weak transport inequalities.

4.2.1 Comparison of the gradient $\widetilde{\nabla}$ and Γ-operator

Lemma 4.2.1. *Let π be the reversible probability measure for the Markov kernel k. For any bounded function f and g on $\mathcal{X}$, the following inequalities hold :*

(i) $\int \Gamma(f,g) d\pi \leqslant \sqrt{2}J \int |\widetilde{\nabla} g||\widetilde{\nabla} f| d\pi$,

(ii) $|\int \Gamma(f,g) d\pi| \leqslant \sqrt{2J} \int |\widetilde{\nabla} g| \sqrt{\Gamma(f)} d\pi$.

Moreover, if we suppose that f is non negative, then we have

(iii) $\int \Gamma(f,g) d\pi \leqslant 2\sqrt{2J} \int |\widetilde{\nabla} g| \sqrt{f \Gamma(\sqrt{f})} d\pi$,

(iv) $\int \Gamma(\sqrt{f}) d\pi \leqslant \frac{J}{4} \int |\widetilde{\nabla} \log f|^2 f d\pi$.

proof. The proofs of these four inequalities all follow similar arguments. Denote the positive part and negative part of a function u as u_+ and u_- respectively.

(i) : Using the relation $(uv)_+ \leqslant u_+v_+ + u_-v_-$, we have

$$
\begin{aligned}
\int \Gamma(f,g)_+ d\pi &= \frac{1}{2}\sum_x \left[\sum_{y\sim x}(f(y)-f(x))(g(y)-g(x))\right]_+ K(x,y)\pi(x) \\
&\leqslant \frac{1}{2}\sum_x\sum_{y\sim x}(g(y)-g(x))_+(f(y)-f(x))_+K(x,y)\pi(x) \\
&+ \frac{1}{2}\sum_x\sum_{y\sim x}(g(y)-g(x))_-(f(y)-f(x))_-K(x,y)\pi(x)
\end{aligned}
$$

Now by reversibility of the measure π, it holds

$$
\begin{aligned}
&\sum_x\sum_x(g(y)-g(x))_+(f(y)-f(x))_+K(x,y)\pi(x) \\
&= \sum_x\sum_{y\sim x}(g(y)-g(x))_-(f(y)-f(x))_-K(x,y)\pi(x) \\
&\leqslant \sum_x |\widetilde{\nabla} g|(x)\sum_{y\sim x}(f(x)-f(y))_-K(x,y)\pi(x),
\end{aligned}
$$

Where the latter inequality follows from $|\widetilde{\nabla} g|(x) \geqslant (g(y)-g(x))_-$ for all $y \sim x$. Therefore, we get

$$
\int \Gamma(f,g)_+ d\pi \leqslant \sum_x |\widetilde{\nabla} g| \sum_{y\sim x}(f(y)-f(x))_-K(x,y)\pi(x). \tag{4.2.2}
$$

In (4.2.2), using $\Gamma(f,g) \leqslant \Gamma(f,g)_+$ and $\sum_{y\sim x}(f(y)-f(x))_-K(x,y) \leqslant |\widetilde{\nabla} f|(x)J(x)$, we get (i).

(ii) : By the Cauchy-Schwarz inequality, it holds

$$
\left(\sum_{y\sim x}(f(y)-f(x))_-K(x,y)\right)^2 \leqslant J(x)\sum_{y\sim x}(f(y)-f(x))_-^2K(x,y) \leqslant 2J\Gamma(f) \tag{4.2.3}
$$

Combining (4.2.2) and (4.2.3) leads to

$$
\int \Gamma(f,g)_+ d\pi \leqslant \int |\widetilde{\nabla} g|\sqrt{2J\Gamma(f)}d\pi. \tag{4.2.4}
$$

Following a similar argument, we have

$$
\int \Gamma(f,g)_- d\pi \leqslant \int |\widetilde{\nabla} g|\sqrt{2J\Gamma(f)}d\pi. \tag{4.2.5}
$$

and (ii) follows by (4.2.4), (4.2.5) and the inequality

$$
\left|\int \Gamma(f,g)d\pi\right| \leqslant \max\left\{\int \Gamma(f,g)_+ d\pi, \int \Gamma(f,g)_- d\pi\right\}.
$$

(iii) : Since f is nonnegative, $\sqrt{f}$ is well defined. Then it holds

$$\begin{aligned}\int \Gamma(f,g)d\pi &= \frac{1}{2}\sum_x \sum_{y\sim x}(f(y)-f(x))(g(y)-g(x))K(x,y)\pi(x)\\ &= \frac{1}{2}\sum_x \sum_{y\sim x}(g(y)-g(x))(\sqrt{f}(y)-\sqrt{f}(x))(\sqrt{f}(y)+\sqrt{f}(x))K(x,y)\pi(x)\end{aligned}$$

Now arguing as in (i) and (ii), by reversibility of π, we get

$$\int \Gamma(f,g)d\pi = \sum_x \sum_{y\sim x}(g(y)-g(x))_-(\sqrt{f}(y)-\sqrt{f}(x))_-(\sqrt{f}(y)+\sqrt{f}(x))K(x,y)\pi(x).$$

Notice that $(\sqrt{f}(y)-\sqrt{f}(x))_-(\sqrt{f}(y)+\sqrt{f}(x)) \leqslant (\sqrt{f}(y)-\sqrt{f}(x))_- 2\sqrt{f}(x)$, we have

$$\begin{aligned}\int \Gamma(f,g)d\pi &\leqslant 2\sum_x \sum_{y\sim x}(g(y)-g(x))_-(\sqrt{f}(y)-\sqrt{f}(x))_-\sqrt{f}(x)K(x,y)\pi(x)\\ &\leqslant 2\sqrt{2J}\int |\widetilde{\nabla} g|\sqrt{f\Gamma(\sqrt{f})}d\pi\end{aligned}$$

where the last step we have used (4.2.3) with $u := \sqrt{f}$.

(iv) : If f is the null function, there is nothing to say. Otherwise, if there exist $x, y \in \mathcal{X}$ such that $f(x) = 0, f(y) > 0$, it is easy to see that $|\widetilde{\nabla} \log f(y)|^2 f(y)\pi(y) = \infty$. So we only need to prove the case $f(x) > 0$ for all $x \in \mathcal{X}$.

Since f is a positive function, one can rewrite $f = e^g$, it is enough to prove that

$$\int \Gamma(e^{g/2})d\pi \leqslant \frac{J}{4}\int |\widetilde{\nabla} g|^2 e^g d\pi$$

holds for all function g. In fact, by convexity of function $x \mapsto e^x$, we have for all $a > b$, $(a-b)e^a \geqslant e^a - e^b$. Thus

$$\begin{aligned}J\int |\widetilde{\nabla} g|^2 e^g &\geqslant \sum_{x\sim y; g(y)\leqslant g(x)} (g(x)-g(y))^2 e^{g(x)}K(x,y)\pi(x)\\ &\geqslant 4\sum_{x\sim y; g(y)\leqslant g(x)} \left(e^{\frac{g(x)}{2}} - e^{\frac{g(y)}{2}}\right)^2 K(x,y)\pi(x)\\ &= 4\int \Gamma(e^{g/2})d\pi\end{aligned}$$

□

4.2.2 Transport-information inequalities implies transport-entropy inequalities

We prove here the discrete version of Theorem 2.1 in [53]. The proof is essentially unchanged, we give it to justify the validity of the theorem in the discrete setting.

Theorem 4.2.6. *Assume that the transport-information inequality*

$$W_1(f\pi,\pi)^2 \leqslant \frac{1}{C^2}\mathcal{I}_\pi(f)$$

holds. Then we have the transport-entropy inequality

$$W_1(f\pi,\pi)^2 \leqslant \frac{2}{C}\,\mathrm{Ent}_\pi(f).$$

proof. The transport-entropy inequality $T_1(C)$ is equivalent to the estimate

$$\int e^{\lambda f} d\pi \leqslant \exp\left(\frac{\lambda^2}{2C}\right)$$

for all 1-Lipschitz function f with $\int f d\pi = 0$ and all $\lambda \geqslant 0$. Let f be such a function. Let $Z(\lambda) := \int e^{\lambda f} d\pi$ and $\mu_\lambda := e^{\lambda f}\pi/Z(\lambda)$. We have

$$\frac{d}{d\lambda}\log Z(\lambda) = \frac{1}{Z(\lambda)}\int f e^{\lambda f} d\pi \leqslant W_{1,d}(\mu_\lambda,\pi) \leqslant \sqrt{\frac{4}{C^2}\int \frac{\Gamma(e^{\lambda f/2})}{Z(\lambda)} d\pi}.$$

Using the inequality $\int \Gamma(f) d\pi \leqslant \int f^2 \Gamma(\log f) d\pi$, we deduce

$$\frac{d}{d\lambda}\log Z(\lambda) \leqslant \frac{\lambda}{C}$$

which integrates into $\log Z(\lambda) \leqslant \lambda^2/(2C)$, and this is the bound we were looking for. $\square$

We shall now show that the weak transport-information inequality $\widetilde{\mathrm{T}}_2^+ I$ implies the weak-transport-entropy inequality $\widetilde{\mathrm{T}}_2^+\mathrm{H}$. The proof is an adaptation of the one for the T_2 and $\mathrm{T}_2\mathrm{I}$ inequalities in the continuous setting from [53].

Theorem 4.2.7. *Assume that π satisfies the modified weak-transport information inequality $\widetilde{\mathrm{T}}_2^+\mathrm{I}(C)$, then π satisfies the weak-transport inequality $\widetilde{\mathrm{T}}_2^+\mathrm{H}(C)$.*

proof. According to [50], the transport-entropy inequality $\widetilde{T}_2$ is equivalent to

$$\int \exp\left(\frac{2}{C}\widetilde{Q}_1 f\right) d\pi \leqslant \exp\left(\frac{2}{C}\int f d\pi\right) \quad \forall f : \mathcal{X} \longrightarrow \mathbb{R} \text{ bounded.}$$

Usually, the class of functions f we must use is the class of bounded continuous functions, but here, since we work on a discrete space endowed with a graph distance, we only have to work with bounded functions.

We write $F(t) := \log \int \exp(k(t)\widetilde{Q}_t f) d\pi - k(t)\int f d\pi$ with $k(t) := Ct$. Let μ_t be the probability measure with density with respect to π proportional to $\exp(k(t)\widetilde{Q}_t f)$.

According to part (iv) of Lemma 4.2.1, we have

$$\int \Gamma(e^{\frac{1}{2}(k(t)\widetilde{Q}_t f)}) d\pi \leqslant \int |\widetilde{\nabla} k(t)\widetilde{Q}_t f|^2 d\mu_t$$

Hence, for $t > 0$, we have

$$
\begin{aligned}
F'(t) &\leqslant \frac{1}{\int \exp(k(t)\widetilde{Q}_t f)d\pi}\left(\int k'(t)\widetilde{Q}_t f e^{k(t)\widetilde{Q}_t f}d\pi - \int k(t)|\widetilde{\nabla}\widetilde{Q}_t f|^2 e^{k(t)\widetilde{Q}_t f}d\pi\right) - k'(t)\int f d\pi \\
&\leqslant \frac{k'(t)}{t}\left(\int \widetilde{Q}_1(tf)d\mu_t - \int tf d\pi\right) - k(t)\int |\widetilde{\nabla}\widetilde{Q}_t f|^2 d\mu_t \\
&\leqslant \frac{k'(t)}{t}\widetilde{W}_2(\mu_t,\pi)^2 - k(t)\int |\widetilde{\nabla}\widetilde{Q}_t f|^2 d\mu_t \\
&\leqslant \frac{k'(t)}{tC^2}\int \Gamma(e^{\frac{1}{2}(k(t)\widetilde{Q}_t f)})d\pi - k(t)\int |\widetilde{\nabla}\widetilde{Q}_t f|^2 d\mu_t \\
&\leqslant \left(\frac{k'(t)}{tC^2} - \frac{1}{k(t)}\right)\int |\widetilde{\nabla}k(t)\widetilde{Q}_t f|^2 d\mu_t \\
&= 0
\end{aligned}
$$

□

4.2.3 Transport-information inequalities imply diameter bounds

We now show that transport-information inequalities imply diameter bounds, in the spirit of the L^1 Bonnet-Myers theorem of [85]. The main idea is the observation following :

Proposition 4.2.8. *If $\nu = f\pi$ is a dirac measure, say δ_z, then it holds*

$$\mathcal{I}_\pi(f) = 4J(z).$$

proof. Since f is the density function correspond to the Dirac mass $\nu = \delta_z$, $f := \frac{1_z}{\pi(z)}$.

$$
\begin{aligned}
\int \Gamma(\sqrt{f})d\pi &= \frac{1}{2}\left(\sum_{y\sim z}(\sqrt{f(z)})^2 k(z,y)\pi(z) + \sum_{y\sim z}(\sqrt{f(z)})^2 k(y,z)\pi(y)\right) \\
&= \sum_{y\sim z}(\sqrt{f(z)})^2 k(z,y)\pi(z) = J(z)
\end{aligned}
$$

□

With this proposition, one deduce immediately

Theorem 4.2.9 (Diameter estimate)**.** *Assume that the transport-information inequality*

$$W_{1,d}(f\pi,\pi)^2 \leqslant \frac{1}{C^2}\mathcal{I}_\pi(f)$$

holds, for some distance d. Then

$$\sup_{x,y\in\mathcal{X}} d(x,y) \leqslant \frac{2}{C}\left(\sqrt{J(x)} + \sqrt{J(y)}\right)$$

Remark 4.2.1. *Here d can be any distance defined on $\mathcal{X}$.*

proof. Let f_x and f_y be the density functions of δ_x and δ_y with respect to π. According to Proposition 4.2.8, it holds for all $x, y \in \mathcal{X}$,

$$d(x,y) = W_1(\delta_x, \delta_y) \leqslant W_1(f_x\pi, \pi) + W_1(f_y\pi, \pi) \leqslant \frac{1}{C}\left(\sqrt{4J(x)} + \sqrt{4J(y)}\right).$$

□

4.3 The Bakry-Émery curvature condition CD(κ, ∞)

In the graph setting, this curvature condition was first studied by Schmuckenshlager in [92] and then by S.-T. Yau and his collaborators in [9, 59, 23, 67]. It has also been used in [63], where a discrete version of Buser's inequality was obtained, as well as curvature bounds for various graphs, such as abelian Cayley graphs and slices of the hypercube. Note that most of these works are set in the framework of graphs rather than Markov chains, which generally makes our definitions and theirs differ by a normalization constant, since we enforce the condition (4.1.1).

Definition 4.3.1. *We define the iterated Γ operator Γ_2 as*

$$\Gamma_2(f) = \frac{1}{2}L\Gamma(f) - \Gamma(f, Lf).$$

We say that the curvature condition CD(κ, ∞) is satisfied if, for all functions f, we have

$$\Gamma_2(f) \geqslant \kappa\Gamma(f).$$

The Γ_2 operator and the curvature condition were first introduced in [6], and used to prove functional inequalities, such as logarithmic Sobolev inequalities and Poincaré inequalities, for measures on Riemannian spaces satisfying CD(κ, ∞) for $\kappa > 0$. In the Riemannian setting, the Γ_2 operator involves the Ricci tensor of the manifold, and the condition CD(κ, ∞) is equivalent to asking for lower bounds on the Ricci curvature, and more generally to the Lott-Sturm-Villani definition of lower bounds on Ricci curvature (see [69] and [96] for the definition, and [4] for the equivalence between the two notions). Hence CD(κ, ∞) can be used as a *definition* of lower bounds on the Ricci curvature for non-smooth spaces, and even discrete spaces.

In this section, we assume that the Markov chain satisfies the curvature condition CD(κ, ∞) for some $\kappa > 0$. One of the main tools we shall use is the following subcommutation relation between Γ and the semigroup P_t, which was obtained in [63] :

Lemma 4.3.2. *Assume that CD(κ, ∞) holds. Then for any $f : \mathcal{X} \longrightarrow \mathbb{R}$, we have*

$$\Gamma(P_t f) \leqslant e^{-2\kappa t} P_t \Gamma(f).$$

Remark 4.3.1. *This property implies that if f is 1-Lipschitz for d_Γ, then $P_t f$ is $e^{-\kappa t}$-Lipschitz. Therefore, the condition $CD(\kappa,\infty)$ implies that the coarse Ricci curvature of the Markov chain, using the distance d_Γ, is bounded from below by κ.*

4.3.1 L^1-transport inequalities

Here we shall prove two inequalities connected to the L^1 Wasserstein distance under $\mathrm{CD}(\kappa,\infty)$.

Theorem 4.3.3. *Let K be an irreducible Markov kernel on $\mathcal{X}$ and π the reversible invariant probability measure associated to K. Assume that $\mathrm{CD}(\kappa,\infty)$ holds with $\kappa > 0$. Then π satisfies the transport-information inequality $T_1 I$ with constant κ. More precisely, for all probability measure $\nu := f\pi$ on $\mathcal{X}$, it holds*

$$W_1(f\pi,\pi)^2 \leqslant \frac{2J}{\kappa^2}\mathcal{I}_\pi(f).$$

With such a result in hand, we can then follow the work by Guillin, Leonard, Wang and Wu [53], to prove a transport-entropy inequality T_1 holds, so that the Gaussian concentration property follows as well.

Proof of Theorem 4.3.3. The proof relies on the Kantorovitch-Rubinstein duality formula

$$W_1(\pi,\nu) = \sup_{g\ 1-lip} \int g d\pi - \int g d\nu.$$

Let g be a 1-Lipschitz function. It implies that $\Gamma(g) \leqslant J$.

First, using the Cauchy-Schwartz inequality, it holds

$$\begin{aligned}
-\int \Gamma(P_t g, f) d\pi &= -\frac{1}{2}\sum_{x,y} (P_t g(y) - P_t g(x))(f(y)-f(x))K(x,y)\pi(x) \\
&= \frac{1}{2}\sum_{x,y} |(P_t g(y) - P_t g(x))(\sqrt{f}(y) - \sqrt{f}(x))(\sqrt{f}(y)+\sqrt{f}(x))|K(x,y)\pi(x) \\
&\leqslant \sum_x \pi(x)\Gamma(\sqrt{f})(x)^{\frac{1}{2}} \left(\sum_y (P_t g(y) - P_t g(x))^2(\sqrt{f}(y)+\sqrt{f}(x))^2 K(x,y)\right)^{\frac{1}{2}},
\end{aligned}$$

Now applying the Cauchy-Schwartz inequality again, the latter quantity is less than

$$\left(\int \Gamma(\sqrt{f})d\pi\right)^{\frac{1}{2}} \left(\sum_{x,y} (P_t g(y) - P_t g(x))^2(\sqrt{f}(y)+\sqrt{f}(x))^2 K(x,y)\pi(x)\right)^{\frac{1}{2}}.$$

Therefore, we have

$$\begin{aligned}
-\int \Gamma(P_t g, f) d\pi &\leqslant \left(\int \Gamma(\sqrt{f})d\pi\right)^{\frac{1}{2}} \left(\sum_{x,y} (P_t g(y) - P_t g(x))^2(\sqrt{f}(y)+\sqrt{f}(x))^2 K(x,y)\pi(x)\right)^{\frac{1}{2}} \\
&\leqslant \sqrt{2}\sqrt{\mathcal{I}_\pi(f)}\sqrt{\int \Gamma(P_t g) f d\pi},
\end{aligned}$$

where the last step we have used the reversibility of the measure π and the fact that

$$(\sqrt{f}(y)+\sqrt{f}(x))^2 \leqslant 2(f(x)+f(y)) \tag{4.3.4}$$

for any nonnegative function f.

Therefore, according to Lemma 4.3.2, we have

$$\begin{aligned}\int g d\pi - \int g f d\pi &= \int_0^{+\infty} \frac{d}{dt}\int (P_t g) f d\pi dt \\ &= -\int_0^{+\infty}\int \Gamma(P_t g, f) d\pi dt \\ &\leqslant \int_0^{+\infty} \sqrt{\mathcal{I}_\pi(f)}\sqrt{\int \Gamma(P_t g) f d\pi} dt \\ &\leqslant \sqrt{2}\sqrt{\mathcal{I}_\pi(f)}\int_0^{+\infty} e^{-\kappa t}\sqrt{\int P_t(\Gamma(g)) f d\pi} dt \\ &\leqslant \frac{\sqrt{2J}}{\kappa}\sqrt{\mathcal{I}_\pi(f)}.\end{aligned}$$

The result immediately follows by taking the supremum over all 1-Lipschitz functions g. □

Using, similar arguments, we can also prove the following Cheeger-type inequality :

Proposition 4.3.5. *Assume that CD(κ, ∞) holds. Then for any probability density f with respect to π, we have*

$$W_1(f\pi, \pi) \leqslant \frac{\sqrt{J}}{\kappa}\int \sqrt{\Gamma(f)} d\pi.$$

We call this a Cheeger-type inequality, by analogy with the classical Cheeger inequality

$$||f\pi - \pi||_{TV} \leqslant C \int |\nabla f| d\pi.$$

Here $\int \sqrt{\Gamma(f)} d\pi$ is an L^1 estimate on the gradient of f, while both $W_{1,d_\Gamma}(f\pi,\pi)$ and $||f\pi - \pi||_{TV}$ are distances of L^1 nature.

proof. Once more, by Kantorovitch duality for W_1, and since 1-Lipschitz functions

g satisfy $\Gamma(g) \leqslant J$, we have

$$\begin{aligned}
W_1(f\pi, \pi) &\leqslant \sup_{g;\Gamma(g)\leqslant 1} \int gf d\pi - \int g d\pi \\
&= \sup_{g;\Gamma(g)\leqslant J} - \int_0^{+\infty} \int \Gamma(P_t g, f) d\pi dt \\
&\leqslant \sup_{g;\Gamma(g)\leqslant J} \int_0^{+\infty} \int \sqrt{\Gamma(P_t g)}\sqrt{\Gamma(f)} d\pi dt \\
&\leqslant \sup_{g;\Gamma(g)\leqslant J} \int_0^{+\infty} e^{-\kappa t} \int \sqrt{P_t\Gamma(g)}\sqrt{\Gamma(f)} d\pi dt \\
&\leqslant \frac{\sqrt{J}}{\kappa} \int \sqrt{\Gamma(f)} d\pi
\end{aligned}$$

□

Since $\mathrm{CD}(\kappa, \infty)$ implies the L^1 transport information inequality, combining with the latter theorem, one can obtain a diameter estimate which is weaker than Corollary 4.3.8. In order to obtain Corollary 4.3.8, we need to revisit the proof of theorem 4.3.3 and prove the following lemma.

Proposition 4.3.6. *Assume that* $\mathrm{CD}(\kappa, \infty)$ *holds, then* $W_1(\pi, \delta_z) \leqslant \frac{2J(z)}{\kappa}$

The proof is essentially the same as the proof of theorem 4.3.3, the fact that supposing $\nu = \delta_z$ leads to some better constants.

proof. Denote f the density function of δ_z with respect to π. Observe that f satisfies

$$(\sqrt{f}(y) + \sqrt{f}(x))^2 \leqslant (f(x) + f(y)) \tag{4.3.7}$$

Then following the lines of the proof of theorem 4.3.3, we have for all 1-lipschitz function g,

$$\begin{aligned}
-\int \Gamma(P_t g, f) d\pi &\leqslant \left(\int \Gamma(\sqrt{f}) d\pi\right)^{\frac{1}{2}} \left(\sum_{x,y} (P_t g(y) - P_t g(x))^2 (\sqrt{f}(y) + \sqrt{f}(x))^2 K(x,y)\pi(x)\right)^{\frac{1}{2}} \\
&\leqslant \sqrt{\mathcal{I}_\pi(f)} \sqrt{\int \Gamma(P_t g) f d\pi} \\
&= \sqrt{\mathcal{I}_\pi(f)} \sqrt{\Gamma(P_t g)(z)}.
\end{aligned}$$

Applying 4.2.8 and inequality $\Gamma(P_t g) \leqslant e^{-2\kappa t}\Gamma(g)$, one deduce that

$$-\int \Gamma(P_t g, f) d\pi \leqslant 2\sqrt{J(z)\Gamma(P_t g)(z)} \leqslant 2e^{-\kappa t} J(z).$$

Therefore, according to Lemma 4.3.2, we have

$$\begin{aligned}\int g d\pi - \int g f d\pi &= \int_0^{+\infty} \frac{d}{dt} \int (P_t g) f d\pi dt \\ &= -\int_0^{+\infty} \int \Gamma(P_t g, f) d\pi dt \\ &\leqslant \frac{2J(z)}{\kappa}.\end{aligned}$$

The proof is completed. □

Corollary 4.3.8. *Assume that* $\mathrm{CD}(\kappa, \infty)$ *holds, then*

$$d(x, y)\kappa \leqslant 2\left(J(x) + J(y)\right).$$

Recall that under coarse Ricci curvature condition (see section 4.5), Ollivier has proved in [85] the same type inequality : $\kappa_c d(x, y) \leqslant J(x) + J(y)$.

proof of Corollary 4.3.8. According to the latter proposition, we have

$$d(x, y) = W_1(\delta_x, \delta_y) \leqslant W_1(\delta_x, \pi) + W_1(\pi, \delta_y) \leqslant \frac{2}{\kappa}(J(x) + J(y)).$$

□

4.3.2 L^2-transport inequalities

Under condition $\mathrm{CD}(\kappa, \infty)$, we have not been able to obtain a transport-entropy inequality involving a weak transport cost. However, we can still obtain a bound on $\widetilde{W}_2(\pi|f\pi)^2$ with the Dirichlet energy.

Proposition 4.3.9. *Assume that CD(κ, ∞) holds. Then for any probability density f with respect to π, we have*

$$\widetilde{W}_2(\pi|f\pi)^2 \leqslant \frac{\sqrt{2}J}{\kappa^2} \int \Gamma(f) d\pi.$$

Unlike the transport-information inequality, this inequality does not seem to imply a transport-entropy inequality, and does not seem to be directly related to concentration inequalities.

proof. First, for any bounded continuous function g on $\mathcal{X}$, we have :

$$\begin{aligned}\int \widetilde{Q} g f d\pi - \int \widetilde{Q} g d\pi &= -\int_0^{+\infty} \frac{d}{dt} \int P_t(\widetilde{Q} g) f d\pi dt \\ &= -\int_0^{+\infty} \int \Gamma(P_t(\widetilde{Q} g), f) d\pi dt = \int_0^{+\infty} \int \sqrt{\Gamma(P_t(\widetilde{Q} g))}\sqrt{\Gamma(f)} d\pi dt \quad (4.3.10)\end{aligned}$$

According to Lemma 4.3.2, $\mathrm{CD}(\kappa,\infty)$ implies that $\Gamma(P_t(g)) \leqslant e^{-2\kappa t}P_t(\Gamma(g))$ holds for all g. Hence,

$$\int_0^{+\infty}\int \sqrt{\Gamma(P_t(\widetilde{Q}g))}\sqrt{\Gamma(f)}d\pi dt \leqslant \int_0^{+\infty} e^{-\kappa t}\int \sqrt{P_t\Gamma((\widetilde{Q}g))}\sqrt{\Gamma(f)}d\pi dt \tag{4.3.11}$$

On the other hand, by Lemma 2.2.7 and (2.1.5), it holds

$$\begin{aligned}\int \widetilde{Q}gd\pi - \int gd\pi &= \int_0^1\int \frac{d}{dt}\widetilde{Q}_tgd\pi dt \\ &\leqslant \int_0^1\int \frac{d}{dt}\widetilde{Q}_tg|_{t=1}d\pi dt \leqslant -\frac{1}{4}\int |\widetilde{\nabla}\widetilde{Q}g|^2 d\pi \end{aligned}\tag{4.3.12}$$

Now applying part (i) of Proposition 4.2.1, we get

$$\begin{aligned}-\frac{1}{4}\int |\widetilde{\nabla}\widetilde{Q}g|^2 d\pi &\leqslant -\frac{1}{4\sqrt{2}J}\int \Gamma(\widetilde{Q}g)d\pi \\ &= -\frac{1}{4\sqrt{2}J}\int P_t\Gamma(\widetilde{Q}g)d\pi = \int_0^{+\infty} e^{-\kappa t}\int -\frac{\kappa}{4\sqrt{2}J}P_t\Gamma(\widetilde{Q}g)d\pi dt\end{aligned}\tag{4.3.13}$$

Combining (4.3.10), (4.3.11), (4.3.12) and (4.3.13), we have

$$\begin{aligned}\int \widetilde{Q}gfd\pi - \int gd\pi &= \int \widetilde{Q}gfd\pi - \int \widetilde{Q}gd\pi + \int \widetilde{Q}gd\pi - \int gd\pi \\ &\leqslant \int_0^{+\infty} e^{-\kappa t}\left(\int \sqrt{P_t\Gamma((\widetilde{Q}g))}\sqrt{\Gamma(f)} - \frac{\kappa}{4\sqrt{2}J}P_t\Gamma(\widetilde{Q}g)d\pi\right)dt \\ &\leqslant \int_0^{+\infty} e^{-\kappa t}\int \frac{\sqrt{2}J}{\kappa}\Gamma(f)d\pi dt \\ &= \frac{\sqrt{2}J}{\kappa^2}\int \Gamma(f)d\pi\end{aligned}$$

The conclusion follows by the duality fomula (2.1.2) while taking the supremum when g runs over all bounded continuous function on the left hand side of the last inequality. □

4.4 The exponential curvature condition

Bauer et al. introduced the following curvature condition in [9], which is a modification of Bakry and Émery's curvature condition. Under this condition, they obtain various Li-Yau inequalities on graphs, and then deduce heat kernel estimates and a Buser inequality for graphs.

As we have mentioned before, the main differences between the continuous and discrete settings is the when the operator L is not a diffusion operator. We lose the chain rule. This leads to additional difficulties, and some results, such as certain forms of the logarithmic Sobolev inequality, do not seem to hold anymore. On the other hand, one of the main difficulties in the continuous setting is to exhibit an algebra of smooth functions satisfying certain conditions, while this property immediately holds in the discrete setting.

The main idea in [9] is the following observation.

The key chain rule used in the continuous setting is the identity

$$L(\Phi(f)) = \Phi'(f)Lf + \Phi''(f)\Gamma(f)$$

which characterizes diffusion operators in the continuous setting, and does not hold in discrete settings. However, a key observation of [9] is that when $\Phi(x) = \sqrt{x}$, the identity

$$2\sqrt{f}L\sqrt{f} = Lf - 2\Gamma(\sqrt{f})$$

holds, even in the discrete setting. This observation motivated the introduction of a modified version of the curvature-dimension condition, designed to exploit this identity :

Definition 4.4.1. *We define the modified Γ_2 operator $\tilde{\Gamma}_2$ as*

$$\tilde{\Gamma}_2(f,f) := \Gamma_2(f) - \Gamma\left(f, \frac{\Gamma(f)}{f}\right).$$

We say that the exponential curvature condition CDE'(κ_e, ∞) is satisfied if, for all nonnegative functions f and all $x \in \mathcal{X}$, we have

$$\tilde{\Gamma}_2(f)(x) \geqslant \kappa_e \Gamma(f)(x).$$

Remark 4.4.1. *We use the notation CDE'(κ_e, ∞) to agree with the notations of [9], where they also consider the case when the condition is only satisfied at points x where $Lf(x) < 0$.*

In [83], it is shown that CDE'(κ_e, ∞) implies CD(κ, ∞) with $\kappa = \kappa_e$. When the operator L is a diffusion, the conditions CDE'(κ_e, ∞) and CD(κ, ∞) are equivalent.

In [59], it was shown that the CDE'(κ_e, ∞) condition tensorizes, and that the associated heat kernel satisfies some Gaussian bounds.

In this section, we assume that the exponential curvature condition CDE'(κ_e, ∞) holds. We will prove some transport inequalities. Since CDE'(κ_e, ∞) implies CD(κ, ∞), the results in the last section are still available. We are going to show some stronger inequalities.

First, we shall study some properties of the CDE'(κ_e, ∞) condition.

4.4.1 Properties of CDE'(κ_e, ∞)

Recall that in classical Bakry-Émery theory, the commutation formula is the following : for all f(smooth enough), $\sqrt{\Gamma(P_t f)} \leqslant e^{-\kappa t} P_t(\sqrt{\Gamma(f)})$. We have not been able to recover this formula under $\mathrm{CDE}(\kappa_e, \infty)$ in graphs settings. However, assuming f is positive, we get the similar commutation formula, which associate the fisher information.

Lemma 4.4.2. *Assue that CDE'(κ, ∞) holds. Then for any nonnegative function $f : \mathcal{X} \longrightarrow \mathbb{R}$ and any $t \geqslant 0$, we have*
(i) $\Gamma(\sqrt{P_t f}) \leqslant e^{-2\kappa_e t} P_t \Gamma(\sqrt{f})$.
(ii) $\frac{\Gamma(P_t f)}{P_t f} \leqslant e^{-2\kappa_e t} P_t \left(\frac{\Gamma(f)}{f}\right)$.

proof. The proof follows a standard interpolation argument. We begin with (i). Let $g := P_{t-s} f$ and define $\varphi(s) := e^{-2\kappa_e s} P_s \left(\Gamma(\sqrt{g})\right)$. To obtain the result, it is enough to show that $\varphi' \geqslant 0$. In fact,

$$\begin{aligned}\varphi'(s) &= e^{-2\kappa_e s} P_s \left[L(\Gamma(\sqrt{g})) - \Gamma\left(\sqrt{g}, \frac{Lg}{\sqrt{g}}\right) - 2\kappa \Gamma(\sqrt{g})\right] \\ &\geqslant 0,\end{aligned}$$

where we have used the assumption on the curvature, which is equivalent to

$$\frac{1}{2} L\Gamma(f) - \Gamma\left(f, \frac{L(f^2)}{2f}\right) \geqslant \kappa_e \Gamma(f)$$

(see (3.11) in [9]).

Similarly, let $\psi(s) := e^{-2\kappa_e s} P_s \left(\frac{\Gamma(g)}{g}\right)$. Again, it is enough to show that $\psi' \geqslant 0$. We have

$$\psi'(s) = e^{-\kappa_e s} P_s \left(L\left(\frac{\Gamma(g)}{g}\right) + \frac{1}{g^2}\left(-2g\Gamma(g, Lg) + \Gamma(g) Lg\right) - 2\kappa \frac{\Gamma(g)}{g}\right). \tag{4.4.3}$$

Since g is positive, we only need to show that

$$g\left(L\left(\frac{\Gamma(g)}{g}\right) + \frac{1}{g^2}\left(-2g\Gamma(g, Lg) + \Gamma(g) Lg\right) - 2\kappa_e \frac{\Gamma(g)}{g}\right) \geqslant 0. \tag{4.4.4}$$

Notice that (4.4.4) is equivalent to

$$gL\left(\frac{\Gamma(g)}{g}\right) - 2\Gamma(g, Lg) + \frac{1}{g}\Gamma(g) Lg \geqslant 2\kappa_e \Gamma(g),$$

and we conclude by writing

$$2\kappa\Gamma(g) \leqslant 2\tilde{\Gamma}_2(g) = 2\left(\Gamma_2(g) - \Gamma(g, \frac{\Gamma(g)}{g})\right) = gL\left(\frac{\Gamma(g)}{g}\right) - 2\Gamma(g, Lg) + \frac{1}{g}\Gamma(g) Lg.$$

□

This lemma immediately yields that under CDE'(κ_e, ∞), the Fisher information and the modified Fisher information have exponential decay with respect to the heat flow :

Theorem 4.4.5. *Assume that CDE'(κ_e, ∞) holds, then we have*

$$\mathcal{I}_\pi(P_t f) \leqslant e^{-2\kappa_e t}\mathcal{I}_\pi(f),$$

and

$$\overline{\mathcal{I}}_\pi(P_t f) \leqslant e^{-2\kappa_e t}\overline{\mathcal{I}}_\pi(f).$$

With this theorem, consider the transportation cost between $\nu_t := P_t f \pi$, we can see that the transport-information inequality in fact tells that the transportation cost has an upper bound which decays exponentially.

Another consequence of Lemma4.4.2 is that under CDE'(κ_e, ∞), a modified log-Sobolev inequality holds.

Theorem 4.4.6. *Suppose π satisfies $CDE'(\kappa_e, \infty)$, then it satisfies modified Log-Sob following : $\forall \nu = f\pi$ probability on $\mathcal{X}$,*

$$Ent_\pi(f) \leqslant \frac{1}{2\kappa_e}\int \frac{\Gamma(f)}{f} d\pi$$

proof.

$$Ent_\pi(f) = Ent(P_0 f) - Ent(P_\infty f) = \int_0^\infty \int \Gamma(\log P_t f, P_t f) d\pi dt.$$

now using

$$\int \Gamma(\log P_t f, P_t f) d\pi \leqslant \int \frac{\Gamma(P_t f)}{P_t f} d\pi$$

we obtain

$$Ent_\pi(f) \leqslant \int\int \frac{\Gamma(P_t f)}{P_t f} d\pi dt$$

Now using item (ii) of Lemma4.4.2, we get an upper bound for entropy :

$$Ent_\pi(f) \leqslant \int\int e^{-2\kappa_e t} P_t\left(\frac{\Gamma(f)}{f}\right) d\pi dt \leqslant \frac{1}{2\kappa}\int \frac{\Gamma(f)}{f} d\pi$$

□

4.4.2 Weak transport-information inequalities under CDE'(κ, ∞)

Using Lemma 4.4.2, we can prove some weak transport-information inequalities under CDE'(κ_e, ∞).

Theorem 4.4.7. *Let $(\mathcal{X}, d)$ be a connected graph equipped with graph distance d. Let K be a irreducible Markov kernel on $\mathcal{X}$ and π the reversible invariant probability measure associated to K. Assume that $CDE'(\kappa_e, \infty)$ holds with $\kappa_e > 0$. Then π satisfies the transport-information inequality $\mathrm{T}_2^+\mathrm{I}$ with constant $\kappa_e/\sqrt{2}$. More precisely, for all probability measure $\nu := f\pi$ on $\mathcal{X}$, it holds*

$$\widetilde{W}_2(f\pi|\pi)^2 \leqslant \frac{2J}{\kappa_e^2}\mathcal{I}_\pi(f) \leqslant \frac{2}{\kappa_e^2}\mathcal{I}_\pi(f).$$

Again, following the ideas of [53], one can deduce that the weak-transport entropy inequality $\widetilde{\mathrm{T}}_2^+\mathrm{H}$ holds. On the other hand, since the weak-transport cost is stronger than the L^1-Wasserstein distance, it yields immediately $\mathrm{T}_1\mathrm{I}$ holds, which implies $\mathrm{T}_1\mathrm{H}$ and Gaussian concentration results.

Proof of Theorem 4.4.7. Let $\alpha(t) = e^{-\kappa_e t}$, for any probability density f with respect to π and any $t > 0$, applying Theorem 2.3.9, it holds :

$$\int_0^\infty \frac{d}{dt}\int \widetilde{Q}_{\alpha(t)}gP_tf d\pi dt \leqslant \int_0^\infty \int -\frac{\kappa_e}{4}e^{-\kappa_e t}|\widetilde{\nabla}\widetilde{Q}_{\alpha(t)}g|^2P_tf + \Gamma(\widetilde{Q}_{\alpha(t)}g, P_tf)d\pi dt \tag{4.4.8}$$

According to part (iii) of Lemma 4.2.1, we have

$$\begin{aligned}
\int \widetilde{Q}gf d\pi - \int g d\pi &= \int_0^\infty \frac{d}{dt}\int \widetilde{Q}_{\alpha(t)}gP_tf d\pi dt \\
&\leqslant \int_0^\infty \int -\frac{\kappa_e}{4}e^{-\kappa_e t}|\widetilde{\nabla}\widetilde{Q}_{\alpha(t)}g|^2P_tf + 2\sqrt{2J}|\widetilde{\nabla}\widetilde{Q}_{\alpha(t)}g|\sqrt{P_tf\Gamma(\sqrt{P_tf})}d\pi dt \\
&\leqslant \int_0^\infty \frac{8Je^{\kappa_e t}}{\kappa_e}\int \Gamma(\sqrt{P_tf})d\pi dt.
\end{aligned}$$

Now we apply Lemma 4.4.2, and we get

$$\begin{aligned}
\int \widetilde{Q}gf d\pi - \int g d\pi &\leqslant \int_0^\infty \frac{8e^{-\kappa_e t}}{\kappa_e}\int P_t\left(\Gamma(\sqrt{f})\right)d\pi dt \\
&= \frac{8J}{\kappa_e^2}\int \Gamma(\sqrt{f})d\pi = \frac{2J}{\kappa_e^2}\mathcal{I}_\pi(f).
\end{aligned}$$

The conclusion then follows from the duality formula (2.1.2) by taking $\mu = \pi$ and $\nu = f\pi$. □

It is easy to see that $\mathcal{I}_\pi(f) \leqslant \overline{\mathcal{I}}_\pi(f) := \int \frac{\Gamma(f)}{f}d\pi$, thus we have the following corollary :

Corollary 4.4.9. *Assume that the exponential curvature condition CDE'(κ_e, ∞) holds, then π satisfies the following weak-transport information inequalities :*

$$\widetilde{W}_2(f\pi|\pi)^2 \leqslant \frac{2J}{\kappa_e^2}\overline{\mathcal{I}}_\pi(f).$$

Unfortunately, we have not been able to establish the relation of $\widetilde{W}_2(\pi|f\pi)^2$ and $\int \Gamma(\sqrt{f})d\pi$. However, as in Corollary 4.4.9, we get a weaker inequality as follows :

Theorem 4.4.10. *Assume that the exponential curvature condition CDE'*(κ_e,∞) *holds, then* π *satisfies the following weak-transport information inequalities :*

$$\widetilde{W}_2(\pi|f\pi)^2 \leqslant \frac{2J}{\kappa_e^2}\overline{\mathcal{I}}_\pi(f).$$

proof. We prove this theorem in a similar way as the previous one.

Let $\alpha(t) := 1 - e^{-\kappa t}$ Arguing as in the latter theorem, we get

$$\begin{aligned}\int \widetilde{Q}g d\pi - \int gf d\pi &= \int_0^\infty \frac{d}{dt}\int \widetilde{Q}_{\alpha(t)}gP_tf d\pi dt\\ &\leqslant \int_0^\infty \int -\frac{\kappa}{4}e^{-\kappa_e t}|\widetilde{\nabla}\widetilde{Q}_{\alpha(t)}g|^2P_tf - \Gamma(\widetilde{Q}_{\alpha(t)}g, P_tf)d\pi dt\end{aligned}$$

Now applying part (ii) of Lemma 4.2.1, it follows that

$$\begin{aligned}\int \widetilde{Q}g d\pi - \int gf d\pi &\leqslant \int_0^\infty \int -\frac{\kappa_e}{4}e^{-\kappa_e t}|\widetilde{\nabla}\widetilde{Q}_{\alpha(t)}g|^2P_tf + \int |\widetilde{\nabla}\widetilde{Q}_{\alpha(t)}g|\sqrt{2J\Gamma(P_tf)}d\pi dt\\ &\leqslant \int_0^\infty \frac{2Je^{\kappa_e t}}{\kappa_e}\int \frac{\Gamma(P_tf)}{P_tf}d\pi dt \leqslant \int_0^\infty \frac{2Je^{-\kappa_e t}}{\kappa_e}\int P_t\left(\frac{\Gamma(f)}{f}\right)d\pi dt\\ &= \frac{2J}{\kappa_e^2}\int\frac{\Gamma(f)}{f}d\pi\end{aligned}$$

□

Remark 4.4.1. (i) *One can get Corollary 4.4.9 by a similar argument : let* $\alpha(t) := e^{-\kappa t}$

$$\begin{aligned}\int \widetilde{Q}gf d\pi - \int g d\pi &= \int_0^\infty \frac{d}{dt}\int \widetilde{Q}_{\alpha(t)}gP_tf d\pi dt\\ &\leqslant \int_0^\infty \int -\frac{\kappa}{2}e^{-\kappa_e t}|\widetilde{\nabla}\widetilde{Q}_{\alpha(t)}g|^2P_tf + \sqrt{2J}|\widetilde{\nabla}\widetilde{Q}_{\alpha(t)}g|\sqrt{|\Gamma(P_tf)|}d\pi dt\\ &\leqslant \int_0^\infty \frac{Je^{\kappa t}}{\kappa_e}\int\frac{\Gamma(P_tf)}{P_tf}d\pi dt \leqslant \frac{J}{\kappa_e^2}\int\frac{\Gamma(f)}{f}d\pi dt.\end{aligned}$$

(ii) *Using the notations of [50], define* $\widetilde{W}_2(f\pi,\pi)^2 = \frac{1}{2}(\widetilde{W}_2^2(f\pi|\pi) + \widetilde{W}^2(\pi|f\pi))$*, denote* $\mathcal{P}_2(\mathcal{X})$ *as the set of the probability measure on* $\mathcal{X}$ *which has a finite second moment. Then* $(\mathcal{P}_2(\mathcal{X}), \widetilde{W}_2(.,.))$ *is a metric space, and if* $\mathcal{X}$ *satisfies the exponential curvature condition, we have an upper bound for* $\widetilde{W}_2(.,\pi)$ *in terms of modified Fisher information. Of course, when we work on a finite space, any probability measure has finite second moment.*

4.5 Coarse Ricci curvature

Finally, we recall the definition of Coarse Ricci curvature, which has been introduced by Ollivier in [85] for discrete-time Markov chains. Since we shall work in continuous time, we shall give the appropriate variant, introduced in [61]. Previous works considering contraction rates in transport distance include [29, 90, 84]. Applications to error estimates for Markov Chain Monte Carlo were studied in [62]. The continuous-time version we use here was introduced in [61]. The particular case of curvature on graphs has been studied in [60].

Definition 4.5.1 (Coarse Ricci curvature)**.** *The coarse Ricci curvature of the Markov chain is said to be bounded from below by κ_c if, for all probability measures μ and ν and any time $t \geqslant 0$, we have*

$$W_1(P_t^*\mu, P_t^*\nu) \leqslant \exp(-\kappa_c t) W_1(\mu, \nu),$$

i.e. if it is a contraction in W_1 distance, with rate κ_c.

Note that unlike the $\mathrm{CD}(\kappa, \infty)$ condition, this property does not only depend on the Markov chain, but also on the choice of the distance d.

In this section, we assume the Markov chain has coarse Ricci curvature bounded from below by κ_c, with respect to the graph distance d_g.

Under coarse Ricci curvature condition, the inequality $T_1 I(\kappa_c)$ holds

Theorem 4.5.2. *Let $\mathcal{X}, \pi, K$ define as before. If the global coarse Ricci curvature is bounded from below by $\kappa_c > 0$, then the following transport inequality holds for all density function f :*

$$W_1(f\pi, \pi)^2 \leqslant \frac{1}{\kappa_c^2} \mathcal{I}_\pi(f) \left(J - \frac{1}{8} \mathcal{I}_\pi(f) \right) \leqslant \frac{1}{\kappa_c^2} \mathcal{I}_\pi(f).$$

As a corollary, this last result implies a T_1 inequality for such Markov chain, which has been previously obtained by Eldan, Lehec and Lee [31].

As a consequence of the bound on the curvature, note that for any 1-Lipschitz function g, $P_t g$ is $e^{-\kappa_c t}$-Lipschitz.

The problem of proving a transport-entropy inequality for Markov chains with positive coarse Ricci curvature was raised in Problem J in [86]. It was proved by Eldan, Lee and Lehec [31]. The transport-information inequality is a slight improvement of this result. Note that T_1 cannot hold in the full generality of the setting of [85], since it implies Gaussian concentration, which does not hold for some examples with positive curvature, such as Poisson distributions on $\mathbb{N}$.

The proof of this result will make use of the following lemma :

Lemma 4.5.3. *If the coarse Ricci curvature is bounded from below by $\kappa_c > 0$, then*

$$W_1(f\pi, \pi) \leqslant \frac{1}{\kappa_c} \sum_{x \neq y} |f(x) - f(y)| K(x, y) \pi(x).$$

proof. By Kantorovitch duality for W_1, we have

$$
\begin{aligned}
W_1(f\pi,\pi) = \sup_{g1-lip} \int gf d\pi - \int g d\pi &= \sup_{g1-lip} \; -\int_0^{+\infty} \frac{d}{dt}\int P_t g f d\pi dt \\
&= -\int_0^{+\infty} \sum_{x,y} (P_t f(y) - P_t g(x))(f(y)-f(x))K(x,y)\pi(x)dt \\
&\leqslant \int_0^{+\infty} ||P_t g||_{lip} \sum_{x,y} |f(y)-f(x)|K(x,y)\pi(x)dt \\
&\leqslant \frac{1}{\kappa_c}\sum_{x\neq y} |f(x)-f(y)|K(x,y)\pi(x).
\end{aligned}
$$

□

We can now prove Theorem 4.5.2 :

Proof of Theorem 4.5.2. Observe that

$$
\begin{aligned}
&\sum_{x\neq y}\left(\sqrt{f}(x)+\sqrt{f}(y)\right)^2 K(x,y)\pi(x) \\
&= \sum_{x\neq y}\left(2f(x)+2f(y)-\left(\sqrt{f}(x)-\sqrt{f}(y)\right)^2\right)K(x,y)\pi(x) \\
&\leqslant \sum_{x\neq y}(2f(x)+2f(y))K(x,y)\pi(x) - \sum_{x\neq y}\left(\sqrt{f}(x)-\sqrt{f}(y)\right)^2 K(x,y)\pi(x) \\
&\leqslant 4J - \frac{1}{2}\mathcal{I}_\pi(f)
\end{aligned}
$$

Now using Lemma 4.5.3, we have

$$
\begin{aligned}
W_1(f\pi,\pi) &\leqslant \frac{1}{\kappa_c}\sum_{x\neq y}|f(x)-f(y)|K(x,y)\pi(x) \\
&= \frac{1}{\kappa_c}\sum_{x\neq y}|\sqrt{f}(x)-\sqrt{f}(y)|\left(\sqrt{f}(x)+\sqrt{f}(y)\right)K(x,y)\pi(x) \\
&\leqslant \frac{1}{\kappa_c}\sqrt{\mathcal{I}_\pi(f)}\sqrt{\frac{1}{4}\sum_{x\neq y}\left(\sqrt{f}(x)+\sqrt{f}(y)\right)^2 K(x,y)\pi(x)} \\
&\leqslant \frac{1}{\kappa_c}\sqrt{\mathcal{I}_\pi(f)}\sqrt{J-\frac{1}{8}\mathcal{I}_\pi(f)}.
\end{aligned}
$$

□

4.6 An example : the discrete hypercube

As an example of Markov chain satisfying CDE'(κ, ∞), we study the example of the symmetric random walk on the discrete hypercube. It is a Markov chain on $\{0,1\}^N$, which at rate 1 selects a coordinate uniformly at random, and flips it with probability 1/2. The transition rates are $K(x,y) = 1/(2N)$ for x, y such that $d_g(x,y) = 1$, and else it is 0.

Theorem 4.6.1. *The symmetric random walk on the discrete hypercube satisfies CDE'*$(1/N, \infty)$

proof. We start with the case $N = 1$. Since then we only have to consider a Markov chain on a two-points space, we can easily do explicit computations. Fix $f : \{0,1\} \longrightarrow \mathbb{R}$. We have

$$\Gamma(f)(0) = \Gamma(f)(1) = \frac{1}{4}(f(0) - f(1))^2$$

and hence $\Gamma\left(f, \frac{\Gamma(f)}{f}\right) = 0$ and

$$\tilde{G}_2(f) = \Gamma_2(f) = -\Gamma(f, Lf) = \Gamma(f).$$

Therefore when $N = 1$, the Markov chain satisfies CDE'(1, ∞).

The general case follows, using a tensorization argument. In the unnormalized case, using Proposition 3.3 of [59], the graph satisfies CDE'(1, ∞) independently of N. Since we consider the case of a Markov chain and enforce (4.1.1), we rescale the generator by a factor $1/N$ (so that there is on average one jump by unit of time), and therefore it satisfies CDE'($1/N$, ∞). $\square$

Remark 4.6.1. *We have shown that for the two-point space, the exponential curvature and the curvature are the same, and equal to* 1*. In [63], it is stated that the curvature is* 2*. The difference is because, since we enforced the normalization condition* (4.1.1)*, the definitions of* L *in the two frameworks differ by a factor* 2*.*

Chapitre 5

Transport inequalities on the line

Abstract

In this chapter, we begin with studying the optimality in a class of weak transport costs. Then we characterize the following inequalities on $\mathbb{R}$:

- The weak transport entropy inequality $\overline{\mathrm{T}}(\theta)$ with θ convex and vanishing on a neighborhood of 0.
- The weak transport entropy inequality $\overline{\mathrm{T}}(\theta)$ with θ quadratic-linear.
- The weak transport entropy inequality $\overline{\mathrm{T}}(\theta)$ with θ a general convex cost.
- The convex modified Sobolev logarithmic inequality CmLS.

We finish by proposing a necessary condition for transport information inequality TI and examples of real probabilities which satisfies T_2 but not TI.

5.1 Introduction

In this chapter, we restrict to a one dimension space, the real line. Denote $\mathcal{P}(\mathbb{R})$ the set of probability measures in $\mathbb{R}$ and $\mathcal{P}_1(\mathbb{R})$ the subset of probability measures having finite first moment.

Let $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a measurable function ; the usual optimal transport cost in the sense of Kantorovich between two probability measures μ and ν on $\mathbb{R}$ is defined by

$$\mathcal{T}_\theta(\nu, \mu) = \inf_\pi \iint \theta(|x-y|)\, \pi(dxdy),$$

where the infimum runs over the set of couplings π between μ and ν, i.e probability measures on $\mathbb{R}^2$ such that $\pi(dx \times \mathbb{R}) = \mu(dx)$ and $\pi(\mathbb{R} \times dy) = \nu(dy)$.

In the last years, since the works by Marton [74, 75, 76], transport-entropy inequalities have been extensively studied as a tool to reach concentration properties for measures on product spaces. We refer to the books or survey [64, 20, 44] for bibliographics references on this field. More precisely, given a measure on a product space, and assuming that each of its conditional one-dimensional marginals satisfies a transport-entropy inequality, many authors have obtained transport-entropy inequalities for the whole measure under weak dependence assumptions (see for instance [28, 77, 100]). Then, the transport-entropy inequality for the whole measure leads to concentration properties by classical arguments as the so-called Marton's argument.

The best known example of transport-entropy inequality is the following Talagrand's transport inequality

$$\mathrm{T}(\theta) : \qquad \mathcal{T}_\theta(\nu, \mu) \leqslant \mathrm{H}(\nu|\mu), \qquad \forall \nu \in \mathcal{P}(\mathbb{R}),$$

for which θ is a quadratic cost function $\theta(x) = Cx^2$ with constant $C \in \mathbb{R}^+$; $\mathrm{H}(\nu|\mu)$ denotes the relative entropy (also called Kullback-Leibler distance) of ν with respect to μ, defined by

$$\mathrm{H}(\nu|\mu) = \int \log\left(\frac{d\nu}{d\mu}\right) d\nu,$$

if ν is absolutely continuous with respect to μ, and $\mathrm{H}(\nu|\mu) = \infty$ otherwise. It follows that the product measure $\mu \otimes \cdots \otimes \mu$ also satisfies the Talagrand's transport inequality with the same constant C, by using the well known tensorisation property of $\mathrm{T}(\theta)$ (see [44]). When the measure on the product space is not product, different non-independent tensorisation's strategies have been proposed that reduce the problem to verify one dimensional transport-entropy inequalities [28, 77, 100].

Therefore, it is of real interest to characterize the probability measures μ on $\mathbb{R}$ satisfying $\mathrm{T}(\theta)$ for any general cost function θ.

Let F_μ denote the cumulative distribution function of a probability measure μ defined by

$$F_\mu(x) := \mu(-\infty, x], \qquad \forall x \in \mathbb{R}.$$

and F_μ^{-1} its general inverse defined by

$$F_\mu^{-1}(u) := \inf\{x \in \mathbb{R}, F_\mu(x) \geqslant u\} \in \mathbb{R} \cup \{\pm\infty\}, \qquad \forall u \in [0,1].$$

In [43], N. Gozlan has obtained necessary and sufficient conditions for the transport-entropy $\mathrm{T}(\theta)$ to hold, when the cost function γ is in particular even, continuous, convex with $\gamma(0) = 0$. These conditions are expressed in terms of the behavior of the modulus of continuity of the increasing map U_μ defined by

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

where τ is the symmetric exponential distribution on $\mathbb{R}$

$$\tau(dx) = \frac{1}{2} e^{-|x|}\, dx.$$

This map can also be expressed as follows

$$U_\mu(x) = \begin{cases} F_\mu^{-1}\left(1 - \frac{1}{2} e^{-|x|}\right) & \text{if } x \geqslant 0 \\ F_\mu^{-1}\left(e^{-|x|}\right) & \text{if } x \leqslant 0 \end{cases}.$$

Following [43], here we focus on the study of a new weak transport-entropy inequality introduced in [50], for which a tensorisation property also holds (see Theorem 5.11 of [50]). More precisely, in dimension one, we consider the optimal weak transport cost of ν with respect to μ defined by

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf_\pi \int \theta\left(\left|x - \int y\, p(x,dy)\right|\right) \mu(dx)$$

where the infimum runs over all couplings $\pi(dxdy) = p(x,dy)\mu(dx)$ of μ and ν, and where $p(x,\cdot)$ denotes the disintegration kernel of π with respect to its first marginal.

In terms of random variables, one has the following interpretation

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf \mathbb{E}\left(\theta(|X - \mathbb{E}(Y|X)|)\right).$$

whereas

$$\mathcal{T}_\theta(\nu,\mu) = \inf \mathbb{E}\left(\theta(|X - Y|)\right),$$

where in both cases the infimum runs over all random variables X, Y such that X follows the law μ and Y the law ν. As a consequence, when θ is convex, by Jensen's inequality, one has

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

Therefore, if a measure μ satisfies $\mathrm{T}(\theta)$ then it also satisfies the following weaker transport-entropy inequalities.

Definition 5.1.1. *Let $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function. A probability measure μ on $\mathbb{R}$ is said to satisfy the transport-entropy inequality*
$\overline{\mathrm{T}}^+(\theta)$ *if for all $\nu \in \mathcal{P}_1(\mathbb{R})$, it holds*

$$\overline{\mathcal{T}}_\theta(\nu|\mu) \leqslant \mathrm{H}(\nu|\mu);$$

$\overline{\mathrm{T}}^-(\theta)$ *if for all $\nu \in \mathcal{P}_1(\mathbb{R})$, it holds*

$$\overline{\mathcal{T}}_\theta(\mu|\nu) \leqslant \mathrm{H}(\nu|\mu);$$

$\overline{\mathrm{T}}(\theta)$ *if μ satisfies $\overline{\mathrm{T}}^+(\theta)$ and $\overline{\mathrm{T}}^-(\theta)$.*

In section 5.2, we recall a dual formulation of these weak transport inequalities in terms of infimum convolution operators. This formulations are related to the so called convex (τ)-property introduced by Maurey [78] and developed in [91].

This new weak transport-entropy inequalities are of particular interest since the class of measures satisfying such inequalities also includes discrete measures on $\mathbb{R}$, for examples, Bernoulli, binomial and Poisson measures [50, 91]. We know that the classical Talagrand's transport inequality can not be satisfied for discrete measures. Indeed it is well known that the Poincaré inequality is a consequence of Talagrand's transport inequality that forces the support of μ to be connected.

However, we will see below that when the cost function θ is quadratic near 0, the above weak transport inequalities are strongly related to the Poincaré inequality, but restricted to the class of convex functions.

5.2 Dual formulation for weak transport-entropy inequalities.

In this short section we recall the Bobkov and Götze dual formulation of the transport-entropy inequality and its extensions, borrowed from [50], related to the transport-entropy inequalities of Definition 5.1.1, in terms of infimum convolution inequalities. The results are stated in dimension one to fit our framework but hold in more general settings (see [50]). They will be used in the next sections.

Lemma 5.2.1. *Let $\mu \in \mathcal{P}_1(\mathbb{R})$ and $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function and, for all functions $g : \mathbb{R} \to \mathbb{R}$ bounded from below, set*

$$Q_t g(x) := \inf_{y \in \mathbb{R}} \left\{ f(y) + t\theta\left(\frac{|x-y|}{t}\right)\right\}, \qquad t > 0, x \in \mathbb{R}.$$

Then the following holds.

(i) μ satisfies $\mathrm{T}(\theta)$ if and only if for all $g : \mathbb{R} \to \mathbb{R}$ bounded from below it holds

$$\exp\left(\int Q_1 g \, d\mu\right) \exp\left(-\int g \, d\mu\right) \leqslant 1.$$

(ii) *μ satisfies $\overline{\mathrm{T}}^+(\theta)$ if and only if for all convex $g : \mathbb{R} \to \mathbb{R}$ bounded from below it holds*

$$\exp\left(\int Q_1 g\, d\mu\right) \int \exp(-g)\, d\mu \leqslant 1.$$

(iii) *μ satisfies $\overline{\mathrm{T}}^-(\theta)$ if and only if for all convex $g : \mathbb{R} \to \mathbb{R}$ bounded from below it holds*

$$\int \exp(Q_1 g)\, d\mu \exp\left(-\int g\, d\mu\right) \leqslant 1.$$

(iv) *If μ satisfies $\overline{\mathrm{T}}(\theta)$, then for all convex $g : \mathbb{R} \to \mathbb{R}$ bounded from below it holds*

$$\int \exp(Q_t g)\, d\mu \int \exp(-g)\, d\mu \leqslant 1, \tag{5.2.2}$$

with $t = 2$. Conversely, if μ satisfies (5.2.2) for some $t > 0$, then it satisfies $\overline{\mathrm{T}}(t\theta(\,\cdot\,/t))$.

Démonstration. The first item is due to Bobkov and Götze [13] and is based on a combination of the well-known duality formulas for the relative entropy and for the transport cost $\mathcal{T}_\theta$. Items (*ii*) and (*iii*) generalize the first item to the framework of weak transport-entropy inequalities. We refer to [50, Proposition 4.5] for a more general statement and for a proof (based on an extension of duality for weak transport costs).

Finally we sketch the proof of Item (*iv*) (which already appeared in a slightly different form in [44, Propositions 8.2 and 8.3]). By the very definition, if μ satisfies $\overline{\mathrm{T}}(\theta)$ then it satisfies $\overline{\mathrm{T}}^\pm(\theta)$ and therefore, it satisfies the exponential inequalities given in Items (*ii*) and (*iii*). Note that if g is convex and bounded from below then $Q_1 g$ is also convex and bounded from below. Therefore it holds

$$\exp\left(\int Q_1 g\, d\mu\right) \int \exp(-g)\, d\mu \leqslant 1$$

and

$$\int \exp(Q_1(Q_1 g))\, d\mu \exp\left(-\int Q_1 g\, d\mu\right) \leqslant 1.$$

Multiplying these two inequalities and noticing that $Q_1(Q_1 g) = Q_2 g$ (for a proof of this well-known semi-group property, see *e.g* [99, Theorem 22.46]) gives (5.2.2) with $t = 2$. The converse implication simply follows from Jensen's inequality. □

5.3 Convex ordering and a majorization theorem

This section is devoted to the study of the convex ordering. The notion of convex ordering, characterized by Strassen's Theorem [95], is crucial for the comprehension of the weak transport costs $\overline{\mathcal{T}_\theta}(\nu|\mu)$. In section 5.3, we recall the definition of the convex order and its geometrical meaning in discrete setting given by Rado's theorem [88] (see Theorem 5.3.10). From this geometrical interpretation, we obtain an intermediate outcome, Theorem 5.3.11, that can be interpreted as the discrete version of Theorem 5.4.2. Then, the proof of Theorem 5.4.2, given in section 5.4, follows by discrete approximation arguments. After recalling some classical definitions and results, we shall prove a majorization theorem which will be a key ingredient in the proof of Theorem 5.4.2.

5.3.1 A reminder on convex ordering and the Strassen's Theorem

We collect here some basic facts about convex ordering of probability measures. We refer the interested reader to [71] and [57] for further results and bibliographic references. All the proofs are well-known, we state some of them for completeness.

We start with the definition of the convex order.

Definition 5.3.1 (Convex order)**.** *Given $\nu_1, \nu_2 \in \mathcal{P}_1(\mathbb{R})$, we say that ν_2 dominates ν_1 in the convex order, and write $\nu_1 \preceq \nu_2$, if for all convex functions f on $\mathbb{R}$, $\int_{\mathbb{R}} f\, d\nu_1 \leqslant \int_{\mathbb{R}} f\, d\nu_2$.*

Remark 5.3.2. *Observe that for any probability measure belonging to $\mathcal{P}_1(\mathbb{R})$ the integral of any convex function always makes sense in $\mathbb{R} \cup \{+\infty\}$.*

The convex ordering of probability measures can be determined by testing only some restricted classes of convex functions as the following proposition indicates.

Proposition 5.3.3. *Let $\nu_1, \nu_2 \in \mathcal{P}_1(\mathbb{R})$; the following are equivalent*

(i) $\nu_1 \preceq \nu_2$,

(ii) $\int x\, \nu_1(dx) = \int x\, \nu_2(dx)$ *and for all Lipschitz and non-decreasing and non-negative convex function* $f \colon \mathbb{R} \to \mathbb{R}^+$, $\int f(x)\, \nu_1(dx) \leqslant \int f(x)\, \nu_2(dx)$.

(iii) $\int x\, \nu_1(dx) = \int x\, \nu_2(dx)$ *and for all* $t \in \mathbb{R}$, $\int [x-t]_+\, \nu_1(dx) \leqslant \int [x-t]_+\, \nu_2(dx)$.

For the reader's convenience and for the sake of completeness, we sketch the proof of this classical result. We refer to [71] for more details.

Sketch of the proof. Let us show that (i) is equivalent to (ii). First, since the functions $x \mapsto x$ and $x \mapsto -x$ are both convex, it is clear that $\nu_1 \preceq \nu_2$ implies $\int x\, \nu_1(dx) = \int x\, \nu_2(dx)$ so that (i) implies (ii). Conversely, since the graph of a convex function always lies above its tangent, subtracting an affine function if necessary, one can restrict to non-negative convex functions. Moreover, if $f \colon \mathbb{R} \to \mathbb{R}^+$ is a convex function, then $f_n \colon \mathbb{R} \to \mathbb{R}^+$ defined by $f_n = f$ on $[-n, n]$, $f_n(x) = f_n(n) + f_n'(n)(x-n)$ if $x \geqslant n$ and $f_n(x) = f_n(-n) + f_n'(-n)(x+n)$ if $x \leqslant -n$ (where f_n' denotes the right derivative of f) is Lipschitz and converges monotonically to f as n goes to infinity. The monotone convergence Theorem then shows that one can further restrict to *Lipschitz* convex functions. Finally, up to the subtraction of an affine map, any Lipschitz convex function is non-decreasing, proving that (ii) implies (i).

Now it is not difficult to check that any convex, non-decreasing Lipschitz function $f : \mathbb{R} \to \mathbb{R}^+$ can be approached by a non-increasing sequence of functions of the form $\alpha_0 + \sum_{i=1}^n \alpha_i [x - t_i]_+$, with $\alpha_i \geqslant 0$ and $t_i \in \mathbb{R}$. This shows that *(ii)* and *(iii)* are equivalent. □

The next classical result, due to Strassen [95], characterizes the convex ordering in terms of martingales.

Theorem 5.3.4 (Strassen)**.** *Let $\nu_1, \nu_2 \in \mathcal{P}_1(\mathbb{R})$; the following are equivalent :*

(i) $\nu_1 \preceq \nu_2$,

(ii) there exists a martingale (X, Y) such that X has law ν_1 and Y has law ν_2.

We refer to [50] for a (two-line) proof of Theorem 5.3.4 involving Kantorovich duality for transport costs of the form $\overline{\mathcal{T}}$.

5.3.2 Majorization of vectors and the Rado Theorem

The convex ordering is closely related to the notion of majorization of vectors that we recall in the following definition. As for the previous subsection, all the proofs are well-known and we state them for completeness.

Definition 5.3.5 (Majorization of vectors)**.** *Let $a, b \in \mathbb{R}^n$; one says that a is* majorized *by b, if the sum of the largest j components of a is less than or equal to the corresponding sum of b, for every j, and if the total sum of the components of both vectors are equal.*

Assuming that the components of $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ are in non-decreasing order (*i.e.* $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$ and $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_n$), a is majorized by b, if

$$a_n + a_{n-1} + \cdots + a_{n-j+1} \leq b_n + b_{n-1} + \cdots + b_{n-j+1}, \qquad \text{for } j = 1, \ldots, n-1,$$

and $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$.

The next proposition recalls the link between majorization of vectors and convex ordering.

Proposition 5.3.6. *Let $a, b \in \mathbb{R}^n$ and set $\nu_1 = \frac{1}{n} \sum_{i=1}^n \delta_{a_i}$ and $\nu_2 = \frac{1}{n} \sum_{i=1}^n \delta_{b_i}$. The following are equivalent*

(i) a is majorized by b,

(ii) ν_1 is dominated by ν_2 for the convex order. In other words, for every convex $f : \mathbb{R} \to \mathbb{R}$, it holds that $\sum_{i=1}^n f(a_i) \leq \sum_{i=1}^n f(b_i)$.

Thanks to the above proposition and with a slight abuse of notation, in the sequel we will also write $a \preceq b$ when a is majorized by b.

proof. Assume without loss of generality that the components of a and b are sorted in increasing order. We observe first that, by construction, the equality $\int x\nu_1(dx) = \int x\nu_2(x)$ is equivalent to $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$.

We will first prove that (i) implies (ii). By Item (iii) of Proposition (5.3.3) we only need to prove that $a \preceq b$ implies

$$\sum_{k=1}^n [a_k - t]_+ \leqslant \sum_{k=1}^n [b_k - t]_+, \qquad \forall t \in \mathbb{R}. \tag{5.3.7}$$

Assume that $t \leqslant \max a_k$ (otherwise (5.3.7) obviously holds). Then, let k_o be the smallest k such that $a_k \geqslant t$ so that $\sum_{k=1}^n [a_k - t]_+ = \sum_{k=k_o}^n (a_k - t)$. Therefore, by the majorization assumption (which guarantees that $\sum_{k=k_o}^n a_k \leqslant \sum_{k=k_o}^n b_k$), we get

$$\sum_{k=1}^n [a_k - t]_+ = \sum_{k=k_o}^n (a_k - t) \leqslant \sum_{k=k_o}^n b_k - t \leqslant \sum_{k=1}^n [b_k - t]_+.$$

Conversely, let us prove that (ii) implies (i). Fix $k \in \{1, \dots, n\}$ and set $f_k(x) := [x - b_k]_+$, $x \in \mathbb{R}$. Plugging f_k into Item (ii) of Proposition (5.3.3) leads to

$$\sum_{i=k}^n a_i - b_k \leqslant \sum_{i=1}^n [a_i - b_k]_+ = n \int f(x)\nu_1(dx) \leqslant n \int f(x)\nu_2(dx) = \sum_{i=1}^n [b_i - b_k]_+ = \sum_{i=k}^n b_i - b_k,$$

so that $\sum_{i=k}^n a_i \leqslant \sum_{i=k}^n b_i$, which proves that a is majorized by b. □

Next we recall a simple classical consequence of Proposition 5.3.6 in terms of discrete optimal transport on the line.

Proposition 5.3.8. *Let $x, y \in \mathbb{R}^n$ be two vectors whose coordinates are listed in non-decreasing order (i.e. $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$, $y_1 \leqslant y_2 \leqslant \cdots \leqslant y_n$). Then for all permutation σ of $\{1, \dots, n\}$ and all convex $\theta : \mathbb{R} \to \mathbb{R}$, it holds*

$$\sum_{i=1}^n \theta(x_i - y_i) \leqslant \sum_{i=1}^n \theta(x_i - y_{\sigma(i)}).$$

proof. Since, for all k, $\sum_{i=k}^n y_i \geqslant \sum_{i=k}^n y_{\sigma(i)}$, it holds for $\sum_{i=k}^n (x_i - y_i) \leqslant \sum_{i=k}^n (x_i - y_{\sigma(i)})$ (with equality for $k = 1$). Therefore, denoting $y_\sigma = (y_{\sigma(1)}, \dots, y_{\sigma(n)})$, it holds $x - y \preceq x - y_\sigma$. Applying Proposition 5.3.6 completes the proof. □

Remark 5.3.9. *In particular, let μ, ν are two discrete probability measures on $\mathbb{R}$ of the form*

$$\mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \qquad \text{and} \qquad \nu = \frac{1}{n} \sum_{i=1}^n \delta_{y_i},$$

where the x_i's and the y_i's are in increasing order, and assume for simplicity that the x_i's are distinct. Then the map T sending x_i on y_i for all i realizes the optimal transport of μ onto ν for every cost function θ.

We end this section with a characterization of the convex ordering (or equivalently of the majorization of vectors, thanks to Proposition 5.3.6), due to Rado [88]. We may give a proof based on Strassen's Theorem. For simplicity, we denote by $\mathcal{S}_n$ the set of all permutations of $\{1,2,\dots,n\}$ and, given $\sigma \in \mathcal{S}_n$ and $x=(x_1,\dots,x_n)\in\mathbb{R}^n$, we set $x_\sigma := (x_{\sigma(1)},\dots,x_{\sigma(n)})$.

Theorem 5.3.10 (Rado). *Let $a,b\in\mathbb{R}^n$; the following are equivalent*

(i) *the vector a is majorized by b ;*

(ii) *there exists a doubly stochastic matrix P such that $a=bP$;*

(iii) *there exists a collection of non-negative numbers $(\lambda_\sigma)_{\sigma\in\mathcal{S}_n}$ with $\sum_{\sigma\in\mathcal{S}_n}\lambda_\sigma=1$ such that $a=\sum_{\sigma\in\mathcal{S}_n}\lambda_\sigma b_\sigma$ (in other words a lies in the convex hull of the permutations of b).*

Démonstration. First we will prove that (i) implies (ii). According to Proposition 5.3.6, $a\preceq b$ is equivalent to saying that $\nu_1=\frac{1}{n}\sum_{i=1}^n\delta_{a_i}$ is dominated by $\nu_2=\frac{1}{n}\sum_{i=1}^n\delta_{b_i}$ in the convex order. Set $\mathcal{X}:=\{a_1,\dots,a_n\}$, $\mathcal{Y}:=\{b_1,\dots,b_n\}$, $k_x:=\#\{i\in\{1,\dots,n\}:a_i=x\}$, $x\in\mathcal{X}$ and $\ell_y=\#\{i\in\{1,\dots,n\}:b_i=y\}$, $y\in\mathcal{Y}$ (where $\#$ denotes the cardinality) ; observe that $\nu_1=\frac{1}{n}\sum_{x\in\mathcal{X}}k_x\delta_x$ and $\nu_2=\frac{1}{n}\sum_{y\in\mathcal{Y}}\ell_y\delta_y$. According to the Strassen Theorem (Theorem 5.3.4), there exists a couple of random variables (X,Y) on some probability space $(\Omega,\mathcal{A},\mathbb{P})$ such that X is distributed according to ν_1, and Y according to ν_2 and $X=\mathbb{E}[Y|X]$. Since X is a discrete random variable,

$$\mathbb{E}[Y|X]=\sum_{x\in\mathcal{X}}\frac{\mathbb{E}[Y\mathbf{1}_{X=x}]}{\mathbb{P}(X=x)}\mathbf{1}_{X=x},\qquad \text{a.s.}$$

Therefore, for all $x\in\mathcal{X}$,

$$x=\frac{\mathbb{E}[Y\mathbf{1}_{X=x}]}{\mathbb{P}(X=x)}=\sum_{y\in\mathcal{Y}}\ell_y y K_{y,x},$$

where $K_{y,x}:=n\frac{\mathbb{P}(X=x,Y=y)}{k_x\ell_y}$. Hence $a=bP$ with $P_{j,i}:=K_{b_j,a_i}$, $i,j=1,\dots,n$. This proves Item (ii), since P is doubly stochastic by construction.

If $a=bP$ with a doubly stochastic matrix P, then it is easily checked that $\sum_{i=1}^n f(a_i)\leqslant\sum_{i=1}^n f(b_i)$ for any convex function f on $\mathbb{R}$ so that (ii) implies (i).

Finally, according to Birkhoff's theorem, the extremes points of the set of doubly stochastic matrices are *permutation matrices*. Therefore every doubly stochastic matrix can be written as a convex combination of permutation matrices showing that (ii) and (iii) are equivalent. □

5.3.3 Geometric aspects of convex ordering and a majorization theorem

Contrary to the previous subsections, the results presented here are new. Fix some vector $b=(b_1,b_2,\dots,b_n)$ of $\mathbb{R}^n$ with *distinct* components (for simplicity). We will be working with the convex hull of the permutations of b, a polytope we denote by $\mathrm{Perm}(b)$ and defined as

$$\mathrm{Perm}(b) := \left\{ \sum_{\sigma \in \mathcal{S}_n} \lambda_\sigma b_\sigma, \text{with } \lambda_\sigma \geqslant 0 \text{ and } \sum_{\sigma \in S_n} \lambda_\sigma = 1 \right\}.$$

Such a polytope is often refered to as the *Permutahedron* generated by b. According to Rado's Theorem 5.3.10, $\mathrm{Perm}(b) = \{a \in \mathbb{R}^n : a \preceq b\}$. Hence, $\mathrm{Perm}(b)$ is a subset of the following affine hyperplane

$$\mathcal{E}_b := \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n x_i = \sum_{i=1}^n b_i \right\} = b + \mathcal{E}_0,$$

with $\mathcal{E}_0 := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.

We will be interested in the faces, facets containing a given face, and normal vectors to such facets of $\mathrm{Perm}(b)$. We need to introduce some notations.

Denote by $[n]$ the set of integers from 1 to n. For $S \subset [n]$, let $v_S(b)$ denote the vector with the $|S|$ largest components of b in the positions indexed by S (in decreasing order, say), and the remaining $n - |S|$ lowest components of b in the other positions indexed by $[n] \setminus S$ (also in a decreasing order). Also, when $S \neq \emptyset$, we denote by $P_S(b)$ the set that contains the vector $v_S(b)$ along with all vectors obtained by permuting any subset of coordinates of $v_S(b)$, as long as the subset is contained in S or in $[n] \setminus S$. (That is, the only permutations that are *not* allowed are those that involve elements from *both* S and $[n] \setminus S$). More precisely

$$P_S(b) := \{(v_S(b))_\sigma, \sigma \in \mathcal{S}_n \text{ such that } \sigma(S) = S\}$$

where $\sigma(\mathcal{S}) := \{\sigma(i), i \in \mathcal{S}\}$ denotes the image of $\mathcal{S}$ by σ.

More generally, given a partition $\mathcal{S} = (S_1, S_2, \ldots, S_k)$ of $[n]$, let $v_{\mathcal{S}}(b)$ denote the vector with the largest $|S_1|$ coordinates of b in the positions indexed by S_1 (in decreasing order), then the next largest $|S_2|$ coordinates in the positions indexed by S_2 and so on (as an illustration, for $b = (1, 4, 5, -2, 3, 9, 6, -5) \in \mathbb{R}^8$ and $\mathcal{S} = (S_1, S_2, S_3)$ with $S_1 = \{1, 2\}$, $S_2 = \{3, 6, 7\}$ and $S_3 = \{4, 5, 8\}$, we get $v_{\mathcal{S}}(b) = (\mathit{9}, \mathit{6}, \mathbf{5}, 1, -2, \mathbf{4}, \mathbf{3}, -5)$ where the italic positions refer to the set S_1, the bold positions to the set S_2 and the remaining positions to S_3). Also, we denote by $P_{\mathcal{S}}(b)$ the set containing the vector $v_{\mathcal{S}}(b)$ along with all vectors obtained by permuting the coordinates of $v_{\mathcal{S}}(b)$ that belong to the same S_i :

$$P_{\mathcal{S}}(b) := \{(v_{\mathcal{S}}(b))_\sigma, \sigma \in \mathcal{S}_n \text{ such that for all } i, \sigma(S_i) = S_i\}.$$

Now we recall two geometric definitions/facts from [10].

Fact 1 : A *facet* of $\mathrm{Perm}(b)$ is the convex hull of $P_S(b)$, for some $S \neq \emptyset, [n]$.

Fact 2 : A *face* of $\mathrm{Perm}(b)$ is the convex hull of $P_{\mathcal{S}}(b)$, for some partition $\mathcal{S} = (S_1, S_2, \ldots, S_k)$ of $[n]$ with $k \geqslant 3$. Furthermore, given a face $F = \mathrm{Conv}(P_{\mathcal{S}}(b))$, there exist exactly $k-1$ facets containing F that are obtained by coalescing the first and last several S_i's in $\mathcal{S}$: that is, for each $1 \leq j \leq k-1$, the facet F_j containing F can be described by taking the partition $[n] = T_1 \cup T_2$ with $T_1 = S_1 \cup \cdots \cup S_j$, and $T_2 = S_{j+1} \cup \cdots \cup S_k$.

The next theorem, which we may call the Majorization Theorem, is a key ingredient in the proof of Theorem 5.4.2. It provides a geometric interpretation of majorization in terms of projection.

Theorem 5.3.11 (Majorization Theorem)**.** *Let $a, b \in \mathbb{R}^n$, assume that b has distinct coordinates and that $a \notin \mathrm{Perm}(b)$. Then the following are equivalent :*

(i) $\hat{c} \in \mathrm{Perm}(b)$ *satisfies*

$$a - \hat{c} \preceq a - c, \quad \forall c \in \mathrm{Perm}(b)\,;$$

(ii) $\hat{c}$ *is the closest point of* $\mathrm{Perm}(b)$ *to* a *; that is,*

$$\hat{c} := \arg \min_{c \in \mathrm{Perm}(b)} (\|a - c\|_2)\,.$$

Moreover the vector $\hat{c}$ is sorted as a : $(a_i \leqslant a_j) \Rightarrow (\hat{c}_i \leqslant \hat{c}_j)$, for all i, j.

Let us recall that the orthogonal projection of a point a on the polytope $\mathrm{Perm}(b)$ is the unique $\bar{c} \in \mathrm{Perm}(b)$ such that

$$\langle a - \bar{c}, c - \bar{c} \rangle \leqslant 0, \qquad \forall c \in \mathrm{Perm}(b). \tag{5.3.12}$$

proof. Observe that if $\sum_{i=1}^n a_i \neq \sum_{i=1}^n b_i$, then letting $\tilde{a} := a - \frac{k}{n}(1, 1, \ldots, 1)$ with $k := \sum_{i=1}^n a_i - \sum_{i=1}^n b_i$, we see (using (5.3.12)) that the orthogonal projection of a and $\tilde{a}$ on $\mathrm{Perm}(b)$ are equal (to some point we denote by $\hat{c}$, say), and that $a - \hat{c} \preceq a - c$ if and only if $\tilde{a} - \hat{c} \preceq \tilde{a} - c$. Therefore we can assume without loss that a and b are such that $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$.

We will first prove that (i) implies (ii) which is the easy part of the proof.

$(i) \implies (ii)$. Let $\bar{c}$ be the closest point of $\mathrm{Perm}(b)$ to a (*i.e.* $\bar{c} := \arg\min_{c \in \mathrm{Perm}(b)}(\|a - c\|_2)$). Then by (i), $a - \hat{c} \preceq a - \bar{c}$, which, by Proposition 5.3.6 (applied to $f(x) = x^2$) implies that $\sum_{i=1}^n (a - \hat{c})_i^2 \leq \sum_{i=1}^n (a - \bar{c})_i^2$. By definition of $\bar{c}$, this is possible only if $\hat{c} = \bar{c}$.

Next we will prove that (ii) implies (i). For the sake of clarity, we first deal with the simple case when $\hat{c}$ lies on a *facet* of $\mathrm{Perm}(b)$, before dealing with the general case of $\hat{c}$ being on a face.

$(ii) \implies (i)$. Let $\hat{c}$ be the closest point of $\mathrm{Perm}(b)$ to a. Since $\mathrm{Perm}(b)$ is invariant by permutation, it easily follows from Proposition 5.3.8 that the coordinates of $\hat{c}$ are in the same order as the coordinates of a. Hence, we are left with the proof that $a - \hat{c} \preceq a - c$ for all $c \in \mathrm{Perm}(b)$.

(a) A simple case : $\hat{c} \in F$ for some facet F. Since $\hat{c}$ is chosen from $\mathrm{Perm}(b)$, and since we assumed that $\sum_i b_i = \sum_i a_i$, we have $\sum_i (a - \hat{c})_i = 0$. Writing $\alpha := a - \hat{c} \in \mathcal{E}_0$, suppose that α is perpendicular to the affine subspace $\mathcal{H} := \mathcal{H}_F$ containing a facet F, defined by some nonempty subset S of $[n]$. For all $x, y \in F$, we thus have $\langle \alpha, x-y \rangle = 0$. Choosing $x = v_S(b)$ and $y = x_{\tau_{ij}}$ obtained by permuting two coordinates of x whose indices are both in S or both in S^c (*i.e.* $\tau_{ij} = (ij)$ is the transposition that permutes i and j, with $i, j \in S$, or $i, j \in S^c$), one sees that the coordinates of α are constant on S and S^c. We denote by α_S and α_{S^c} the values of α on these sets, which verify $k\alpha_S + (n-k)\alpha_{S^c} = 0$ since $\alpha \in \mathcal{E}_0$.

Now (recalling that $\alpha = a - \hat{c}$) our task is to show that

$$\alpha \preceq \alpha - (c' - \hat{c}), \quad \text{for every } c' \in \operatorname{Perm}(b)\,.$$

This amounts to showing that

$$\alpha \preceq \alpha - c, \quad \text{for every } c \text{ such that } \langle \alpha, c \rangle \leqslant 0, \quad \text{and } \sum_i c_i = 0\,.$$

Indeed, on the one hand the choice of $\hat{c}$ implies, by (5.3.12), that for every $c' \in \operatorname{Perm}(b)$ $\langle \alpha, c' - \hat{c} \rangle \ \leq 0$, and on the other hand, since $\hat{c}, c' \in \operatorname{Perm}(b)$, necessarily $\sum_i c_i = \sum_i c'_i - \sum_i \hat{c}_i = 0$.

Now $\langle \alpha, c \rangle \leqslant 0$ and $\sum_i c_i = 0$ together imply (recall that α is constant on S and S^c) that

$$(\alpha_S - \alpha_{S^c}) \sum_{i \in S} c_i \leqslant 0\,.$$

Let us assume that $\alpha_S > \alpha_{S^c}$. Then denoting by $c_S = \sum_{i \in S} c_i$ and by $c_{S^c} = \sum_{i \in S^c} c_i$, one has $c_S \leqslant 0$ and $c_{S^c} \geqslant 0$. Therefore, for any convex function f on $\mathbb{R}$, according to Jensen's inequality and by convexity, we get

$$\begin{aligned}
\sum_{i=1}^n f(\alpha_i - c_i) &= k \frac{\sum_{i \in S} f(\alpha_S - c_i)}{k} + (n-k) \frac{\sum_{i \in S^c} f(\alpha_{S^c} - c_i)}{n-k} \\
&\geqslant k f\left(\alpha_S - \frac{c_S}{k}\right) + (n-k) f\left(\alpha_{S^c} - \frac{c_{S^c}}{n-k}\right) \\
&\geqslant k f(\alpha_S) + (n-k) f(\alpha_{S^c}) - f'(\alpha_S) c_S - f'(\alpha_{S^c}) c_{S^c} \\
&\geqslant \sum_{i=1}^n f(\alpha_i),
\end{aligned}$$

where the last inequality comes from the fact that $f'(\alpha_S)c_S + f'(\alpha_S^c)c_{S^c} = c_S(f'(\alpha_S) - f'(\alpha_{S^c})) \leqslant 0$. According to Proposition 5.3.6, we conclude that $\alpha \preceq \alpha - c$ which is the expected result.

(b) The general case. Suppose that $\hat{c}$ lies in a face F of the polytope. This face is related to a partition $\mathcal{S} = (S_1, \ldots, S_k)$ of $[n]$, with $k \geqslant 3$. Then $\alpha := a - \hat{c} \in N(F)$, where $N(F)$ denotes the normal cone of F. Recall that the extreme rays of $N(F)$ are given by the facet directions for the facets containing F. For all $i \in \{1, \ldots, n-1\}$, let us denote by F_i the facet containing F associated to the partition $\mathcal{T}_i = \{S_1 \cup \ldots \cup S_i\,;\, S_{i+1} \cup \ldots \cup S_k\}$, $1 \leqslant i \leqslant k-1$. Consider the vectors $p_1, p_2, \ldots, p_{k-1} \in \mathcal{E}_0$ defined by

$$p_i = \mathbf{1}_{S_1 \cup S_2 \cup \cdots \cup S_i} - \frac{k_i}{n} \mathbf{1}_{[n]}$$

where $\mathbf{1}_T$ denotes the $0-1$ indicator vector of T, for $T \subseteq [n]$, and $k_i = |S_1| + \cdots + |S_i|$. For each i, the vector p_i is orthogonal to the facet F_i. Moreover, for all $c \in \mathrm{Perm}(b)$ one may check that $\langle c, p_i \rangle \leqslant \langle v_{\mathcal{T}_i}, p_i \rangle$, with equality on F_i. This shows that p_i is an outward normal vector to F_i. Therefore $N(F)$ is the conical hull of the p_i's, and so we may express α, for a suitable choice of $\lambda_i \geq 0$, as :

$$\alpha = \sum_i \lambda_i \mathbf{1}_{S_1 \cup S_2 \cup \cdots \cup S_i} - \sigma \mathbf{1}_{[n]}\,,$$

where $\sigma = (1/n)[\sum_{i=1}^{k-1} \lambda_i |S_1| + \sum_{i=2}^{k-1} \lambda_i |S_2| + \cdots + \lambda_{k-1} |S_{k-1}|]$. In particular, α is constant on each S_j : for all $i \in S_j$, $\alpha_i = \left(\sum_{p=j}^{k-1} \lambda_p \right) - \sigma := A_j$.

In order to establish (i), we need to show that

$$\alpha \preceq \alpha - (c - \hat{c}), \quad \forall c \in \mathrm{Perm}(b)\,,$$

or in other words, we need to show that

$$\alpha \preceq \alpha - c', \quad \forall c' \in \mathrm{Perm}(b) - \hat{c}\,.$$

We now use again the fact that our choice of $\hat{c}$ implies that, for all $1 \leq i \leq k-1$,

$$\langle p_i, \hat{c} \rangle \geq \langle p_i, c \rangle\,, \quad \forall c \in \mathrm{Perm}(b)\,.$$

This in turn gives the following :

$$\mathrm{Perm}(b) - \hat{c} \subseteq \{c' : \langle c', p_i \rangle \leq 0,\ \forall i\}\,.$$

Thus using $N(F)^0 := \{d \in \mathcal{E}_0; \langle d, p_i \rangle \leq 0,\ \forall i\}$ to denote the *polar cone*, it then suffices to show that for α (as above),

$$\alpha \preceq \alpha - d, \quad \forall d \in N(F)^0\,.$$

Now, $d \in N(F)^0$ implies that

$$\langle d, \mathbf{1}_{S_1 \cup S_2 \cup \cdots \cup S_j} \rangle \leq 0 \ \text{ and } \ \sum_i d_i = 0\,,$$

therefore denoting $E_j = \sum_{i \in S_1 \cup \ldots \cup S_j} d_i$, for all $j \in \{0, 1, \ldots, k\}$, one has $E_j \leqslant 0$ and $E_0 = E_k = 0$.

Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function; denoting by f' its right derivative, the convexity of f implies that

$$\sum_{i=1}^n f(\alpha_i - d_i) = \sum_{j=1}^k \sum_{i \in S_j} f(A_j - d_i) \geqslant \sum_{j=1}^k |S_j| f(A_j) - \sum_{j=1}^k f'(A_j) D_j,$$

where $D_j = \sum_{i \in S_j} d_i$. Now, using an Abel transform (and the fact that $E_0 = E_k = 0$), one gets

$$\sum_{j=1}^{k} f'(A_j)D_j = \sum_{j=1}^{k} f'(A_j)(E_j - E_{j-1}) = \sum_{j=1}^{k-1}(f'(A_j) - f'(A_{j+1})E_j \leqslant 0,$$

where the inequality comes from $E_j \leqslant 0$, $A_j \geqslant A_{j+1}$ and the monotonicity of f'. Therefore, one gets

$$\sum_{i=1}^{n} f(\alpha_i - d_i) \geqslant \sum_{j=1}^{k} |S_j| f(A_j) = \sum_{i=1}^{n} f(a_i),$$

which proves that $a \preceq a - d$, thanks to Proposition 5.3.6, as expected. This completes the proof. □

5.4 Optimal coupling for weak transport costs

In dimension one, it is well known that

$$\mathcal{T}_{\theta_1(a\,\cdot)+\theta_2(a\,\cdot)}(\nu|\mu) = \mathcal{T}_{\theta_1(a\,\cdot)}(\nu|\mu) + \mathcal{T}_{\theta_2(a\,\cdot)}(\nu|\mu), \tag{5.4.1}$$

We obtain this equality by showing that the two optimal weak transport costs $\mathcal{T}_{\theta_1(a\,\cdot)}(\nu|\mu)$ and $\mathcal{T}_{\theta_2(a\,\cdot)}(\nu|\mu)$ are achieved by a same coupling. This result is well known for classical transport cost with convex cost function θ in dimension one. Namely, in the case where ν has no atom, the map

$$T_{\nu,\mu} = F_\mu^{-1} \circ F_\nu.$$

is the only one non-decreasing and left-continuous function that pushes forward ν onto μ, that is to say

$$\int f\, d\mu = \int f \circ T_{\nu,\mu}\, d\nu.$$

Moreover, from the works by Hoeffding, Fréchet and Dall'Aglio [26, 38, 58], we know that this map achieves the optimal transport of ν onto μ independently of the convex cost functions θ (see also [21]). In other words, it holds

$$\mathcal{T}_\theta(\mu,\nu) = \int \theta\left(|x - T_{\nu,\mu}(x)|\right)\, \nu(dx).$$

Actually, the expected equality (5.4.1) follows by combining this statement in dimension one with our following main result of this section :

Theorem 5.4.2. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$; there exists a probability measure $\hat{\gamma}$ dominated by ν in the convex order, $\hat{\gamma} \preceq \nu$, such that* for all *convex cost function θ it holds*

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

In particular, for any two convex cost functions θ_1, θ_2, it holds

$$\overline{\mathcal{T}}_{\theta_1+\theta_2}(\mu,\nu) = \overline{\mathcal{T}}_{\theta_1}(\mu,\nu) + \overline{\mathcal{T}}_{\theta_2}(\mu,\nu). \tag{5.4.3}$$

We will establish first a preliminary result which gives some connection between $\mathcal{T}$ and $\overline{\mathcal{T}}$. In the sequel, we denote by $\mathrm{Im}(\mu)$, respectively $\mathrm{Im}^{\uparrow}(\mu)$, the set of probability measures on $\mathbb{R}$ which are images of μ under some map $S:\mathbb{R}\to\mathbb{R}$, respectively some non-decreasing map S, *i.e.,*

$$\mathrm{Im}(\mu)=\{\gamma\in\mathcal{P}(\mathbb{R}):\exists S:\mathbb{R}\to\mathbb{R}\text{ measurable such that }\gamma=S_{\#}\mu\},$$

and

$$\mathrm{Im}^{\uparrow}(\mu)=\{\gamma\in\mathcal{P}(\mathbb{R}):\exists S:\mathbb{R}\to\mathbb{R}\text{ measurable, non-decreasing, such that }\gamma=S_{\#}\mu\}.$$

Proposition 5.4.4. *For all probability measures μ,ν on $\mathbb{R}$, it holds*

$$\inf_{\gamma\preceq\nu,\gamma\in\mathrm{Im}^{\uparrow}(\mu)}\mathcal{T}_\theta(\gamma,\mu)\geqslant\overline{\mathcal{T}}_\theta(\nu|\mu)\geqslant\inf_{\gamma\preceq\nu,\gamma\in\mathrm{Im}(\mu)}\mathcal{T}_\theta(\gamma,\mu).$$

Remark 5.4.5. *Note that when μ has no atoms, then $\mathrm{Im}^{\uparrow}(\mu)=\mathrm{Im}(\mu)$. If μ is a discrete probability measure, then the two sets may be different. For instance, if $\mu=\frac{1}{3}\delta_0+\frac{2}{3}\delta_1$, then $\gamma=\frac{2}{3}\delta_0+\frac{1}{3}\delta_1$ is in $\mathrm{Im}(\mu)$ but not in $\mathrm{Im}^{\uparrow}(\mu)$. In the proof of Theorem 5.4.2 below, we will use Proposition 5.4.4 with μ being the uniform distribution on n distinct points for which it is clear that $\mathrm{Im}^{\uparrow}(\mu)=\mathrm{Im}(\mu)$.*

proof. First we will prove that $\overline{\mathcal{T}}_\theta(\nu|\mu)\geqslant\inf_{\gamma\preceq\nu,\gamma\in\mathrm{Im}(\mu)}\mathcal{T}_\theta(\gamma,\mu)$. To that aim, denote by $\pi(dxdy)=p(x,dy)\mu(dx)$ some coupling between μ and ν and set $S(x):=\int y\,p(x,dy)$, $x\in\mathbb{R}$. Clearly $S_{\#}\mu\in\mathrm{Im}(\mu)$. Moreover if $f:\mathbb{R}\to\mathbb{R}$ is some convex function, by Jensen's inequality, it holds

$$\int f(x)\,S_{\#}\mu(dx)=\int f\left(\int y\,p(x,dy)\right)\mu(dx)\leqslant\iint f(y)\,p(x,dy)\mu(dx)=\int f(y)\,\nu(dy)$$

so that $S_{\#}\mu\preceq\nu$. Therefore,

$$\int\theta\left(x-\int y\,p(x,dy)\right)\mu(dx)=\int\theta(x-S(x))\,\mu(dx)\geqslant\mathcal{T}_\theta(S_{\#}\mu,\mu)\quad\geqslant\inf_{\gamma\preceq\nu,\gamma\in\mathrm{Im}(\mu)}\mathcal{T}_\theta(\gamma,\mu)$$

from which the claim follows by taking the infimum over p.

Now we turn to the proof of the inequality $\overline{\mathcal{T}}_\theta(\nu|\mu)\leqslant\inf_{\gamma\preceq\nu,\gamma\in\mathrm{Im}^{\uparrow}(\mu)}\mathcal{T}_\theta(\gamma,\mu)$. Assume that $\gamma\preceq\nu$ and that $\gamma=S_{\#}\mu$ for some non-decreasing map S. According to Strassen's theorem, there exists a coupling π_1 with first marginal γ and second marginal ν such that $\pi_1(dxdy)=p_1(x,dy)\gamma(dx)$ and $x=\int_{\mathbb{R}}yp_1(x,dy)$, γ almost everywhere. For all $x\in\mathbb{R}$, define the following probability measure $p(x,dy):=p_1(S(x),dy)$. Then for all bounded continuous function f, it holds

$$\begin{aligned}\iint f(y)p(x,dy)\,\mu(dx)&=\iint f(y)p_1(S(x),dy)\,\mu(dx)\\&=\iint f(y)p_1(x,dy)\,\gamma(dx)=\int f(y)\,\nu(dy).\end{aligned}$$

Thus the coupling $\pi(dxdy) = p(x,dy)\mu(dx)$ has μ as first marginal and ν as second marginal. Moreover, by definition of p_1 and p, μ almost everywhere, it holds

$$\int yp(x,dy) = \int yp_1(S(x),dy) = S(x).$$

Since S is non-decreasing, it realizes the optimal transport between μ and ν for the classical transport cost $\mathcal{T}_\theta$ and so it follows that

$$\mathcal{T}_\theta(\gamma,\mu) = \int \theta(|x - S(x)|)\mu(dx) = \int \theta(|x - \int yp(x,dy)|)\mu(dx) \geqslant \overline{\mathcal{T}}_\theta(\nu|\mu)$$

which achieves the proof by taking the infimum over γ. □

We are now in a position to prove Theorem 5.4.2.

Proof of Theorem 5.4.2. The proof of the first part of Theorem 5.4.2 is divided into two steps. In the first step we will deal with uniform discrete measures on n points, while in the second step we will use an approximation argument in order to reach any measures.

Step 1. We first deal with

$$\mu := \frac{1}{n}\sum_{i=1}^n \delta_{a_i} \qquad \text{and} \qquad \nu := \frac{1}{n}\sum_{i=1}^n \delta_{b_i},$$

with $a_1 < a_2 < \ldots < a_n$ and $b_1 < b_2 < \ldots < b_n$. Set $a := (a_1,\ldots,a_n)$ and $b := (b_1,\ldots,b_n)$. According to Theorem 5.3.11, there exists some $\hat{c} \in \mathrm{Perm}(b)$ such that $a - \hat{c} \preceq a - c$, for all $c \in \mathrm{Perm}(b)$. Moreover the coordinates of $\hat{c}$ satisfy $\hat{c}_i \leqslant \hat{c}_{i+1}$. Set $\hat{\gamma} := \frac{1}{n}\sum_{i=1}^n \delta_{\hat{c}_i}$ and observe that ν dominates $\hat{\gamma}$ for the convex order and $\hat{\gamma} \in \mathrm{Im}^\uparrow(\mu)$. (Recall the definition from the beginning of this section.)

Now for any $\gamma := \frac{1}{n}\sum_{i=1}^n \delta_{c_i} \in \mathrm{Im}^\uparrow(\mu)$ with $c_i \leqslant c_{i+1}$ and for any convex cost function θ, it holds (since the coordinates are non-decreasing)

$$\mathcal{T}_\theta(\gamma,\mu) = \frac{1}{n}\sum_{i=1}^n \theta(|a_i - c_i|).$$

In particular

$$\mathcal{T}_\theta(\hat{\gamma},\mu) = \frac{1}{n}\sum_{i=1}^n \theta(|a_i - \hat{c}_i|) \leqslant \inf_{c\in\mathrm{Perm}(b)} \frac{1}{n}\sum_{i=1}^n \theta(|a_i - c_i|). \tag{5.4.6}$$

A probability γ such that $\gamma \preceq \nu, \gamma \in \mathrm{Im}^\uparrow(\mu)$ is of the form $\gamma = \frac{1}{n}\sum_{i=1}^n \delta_{c_i}$ with $c_i \leqslant c_{i+1}$ and $c = (c_1,\ldots,c_n) \in \mathrm{Perm}(b)$, and for such a c, it holds $\frac{1}{n}\sum_{i=1}^n \theta(|a_i - c_i|) = \mathcal{T}_\theta(\gamma,\mu)$. Therefore, the latter implies

$$\mathcal{T}_\theta(\hat{\gamma},\mu) \leqslant \inf_{\gamma\preceq\nu,\gamma\in\mathrm{Im}^\uparrow(\mu)} \mathcal{T}_\theta(\gamma,\mu) = \overline{\mathcal{T}}_\theta(\nu|\mu)$$

where the last equality follows from Proposition 5.4.4 and the fact that for such a distribution μ, it holds $\mathrm{Im}(\mu) = \mathrm{Im}^{\uparrow}(\mu)$ (see Remark 5.4.5). Since obviously $\overline{\mathcal{T}}_\theta(\nu|\mu) \leqslant \mathcal{T}_\theta(\hat\gamma, \mu)$, we conclude that $\mathcal{T}_\theta(\hat\gamma,\mu) = \overline{\mathcal{T}}_\theta(\nu|\mu)$ as expected.

Step 2. In the second step we deal with the general case using an approximation argument.

Let μ and ν be two elements of $\mathcal{P}_1(\mathbb{R})$. By assumption, $\int |x|\,\mu(dx) < \infty$ and $\int |x|\,\nu(dx) < \infty$, hence, according to the de la Vallée-Poussin Theorem (see *e.g.* [18, Theorem 4.5.9]), there exists an increasing convex function $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\beta(t)/t \to \infty$ as $t \to \infty$ and such that $\int \beta(|x|)\,\mu(dx) < \infty$ and $\int \beta(|x|)\,\nu(dx) < \infty$.

Next we will construct discrete approximations of μ and ν. According to Varadarajan's theorem (see *e.g.* [30, Theorem 11.4.11]), if X_i is an i.i.d sequence of law μ, then, with probability 1, the empirical measure $L_n^X := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ converges weakly to μ. On the other hand, according to the strong law of large numbers, with probability 1, $\frac{1}{n}\sum_{i=1}^n |X_i| \to \int |x|\,\mu(dx)$ as $n \to \infty$. Let us take $(x_i)_{i\geqslant 1}$, a positive realization of these events and set $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^{(n)}}$, where $x_1^{(n)} \leqslant x_2^{(n)} \leqslant \ldots \leqslant x_n^{(n)}$ denotes the increasing re-ordering of the vector $(x_1, x_2, \ldots, x_n)$. Then the sequence μ_n converges weakly to μ and $\int |x|\,\mu_n(dx) \to \int |x|\,\mu(dx)$. According to [99, Theorem 6.9], this is equivalent to the convergence of the W_1 distance : $W_1(\mu_n,\mu) \to 0$ as $n \to \infty$. Note that one can assume that the points $x_i^{(n)}$ are distinct. Indeed, if this is not the case, then letting $\tilde{x}_i^{(n)} = x_i^{(n)} + i/n^2$ one obtains distinct points and it is not difficult to check that $\tilde\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{\tilde{x}_i^{(n)}}$ still weakly converges to μ (for instance the W_1 distance between μ_n and $\tilde\mu_n$ is easily bounded from above by $(n+1)/(2n^2)$). The same argument yields a sequence $\nu_n = \frac{1}{n}\sum_{i=1}^n \delta_{y_i^{(n)}}$ with $y_i^{(n)} < y_{i+1}^{(n)}$ converging to ν in the W_1 sense. It is not difficult to check (invoking the strong law of large numbers again) that one can further impose that $\int \beta(|x|)\,\nu_n(dx) \to \int \beta(|x|)\,\nu(dx)$, as $n \to \infty$.

For all $n \geqslant 1$, one applies the result proved in the first step : there exists a unique probability measure $\hat\gamma_n \preceq \nu_n$ such that

$$\overline{\mathcal{T}}_\theta(\nu_n|\mu_n) = \mathcal{T}_\theta(\hat\gamma_n, \mu_n),$$

for all convex cost functions θ. Let us show that one can extract from $\hat\gamma_n$ a subsequence converging to some $\hat\gamma$ in $\mathcal{P}_1(\mathbb{R})$ for the W_1 distance. By construction $\int \beta(|x|)\,\nu_n(dx) \to \int \beta(|x|)\,\nu(dx)$ and so $M = \sup_{n\geqslant 1} \int \beta(|x|)\,\nu_n(dx)$ is finite. Since $\hat\gamma_n \preceq \nu_n$ and since the function $x \mapsto \beta(|x|)$ is convex, it thus holds $\int \beta(|x|)\,\gamma_n(dx) \leqslant \int \beta(|x|)\,\nu_n(dx) \leqslant M$. In particular, setting $c(R) = \inf_{t\geqslant R} \beta(t)/t$, $R > 0$, Markov's inequality easily implies that

$$\int_{[-R,R]^c} |x|\,\hat\gamma_n(dx) \leqslant \frac{\int \beta(|x|)\,\nu_n(dx)}{c(R)} \leqslant \frac{M}{c(R)}.$$

Consider $\tilde{\gamma}_n$ defined by $\frac{d\tilde{\gamma}_n}{d\hat{\gamma}_n}(x) = \frac{1+|x|}{\int 1+|x|\,\hat{\gamma}_n(dx)}$. Then it holds,

$$\sup_{n\geqslant 1} \tilde{\gamma}_n([-R,R]^c) \leqslant \frac{2M}{c(R)}, \qquad \forall R \geqslant 1\,,$$

and so the sequence $\tilde{\gamma}_n$ is tight. Therefore, according to the Prokhorov Theorem, extracting a subsequence if necessary, one can assume that $\tilde{\gamma}_n$ converges to some $\tilde{\gamma}$ for the weak topology. Extracting yet another subsequence if necessary, one can also assume that $\int(1+|x|)\,\gamma_n(dx)$ converges to some number $Z > 0$. The weak convergence of $\tilde{\gamma}_n$ to $\tilde{\gamma}$ means that $\int \varphi\, d\tilde{\gamma}_n \to \int \varphi\, d\gamma$ for all bounded continuous φ, which means that

$$\int (1+|x|)\varphi(x)\,\hat{\gamma}_n(dx) \to \int (1+|x|)\varphi(x)\,\hat{\gamma}(dx),$$

where $\hat{\gamma}(dx) = \frac{Z}{1+|x|}\,\tilde{\gamma}(dx) \in \mathcal{P}_1(\mathbb{R})$. Invoking again [99, Theorem 6.9], this implies $\hat{\gamma}_n \to \hat{\gamma}$ as $n \to \infty$ for the W_1 distance.

Now we will check that $\hat{\gamma}$ is such that $\overline{\mathcal{T}_\theta}(\nu|\mu) = \mathcal{T}_\theta(\hat{\gamma},\mu)$ for all convex cost functions $\theta : \mathbb{R}^+ \to \mathbb{R}^+$. First assume that θ is Lipschitz, and denote by L_θ its Lipschitz constant. According to [50, Theorem 2.11], the following Kantorovich duality formula holds

$$\overline{\mathcal{T}}_\theta(\nu_n|\mu_n) = \sup_\varphi \left\{ \int Q_\theta\varphi(x)\,\nu_n(dx) - \int \varphi(y)\,\mu_n(dy) \right\},$$

where the supremum is taken over the set of convex functions φ bounded from below, with $Q_\theta\varphi(x) := \inf_{y\in\mathbb{R}}\{\varphi(y) + \theta(|x-y|)\}$, $x \in \mathbb{R}$. Define $\bar{\varphi}(y) := \sup_{x\in\mathbb{R}}\{Q_\theta\varphi(x) - \theta(|x-y|)\}$. Then it is easily checked that $\bar{\varphi} \leqslant \varphi$, $\bar{\varphi}$ is bounded from below and $Q_\theta\bar{\varphi} = Q_\theta\varphi$. Moreover, being a supremum of convex and L_θ-Lipschitz functions, the function $\bar{\varphi}$ is also convex and L_θ-Lipschitz. Therefore, the supremum in the duality formula above can be further restricted to the class of convex functions which are L_θ-Lipschitz and bounded from below. Using the fact that $W_1(\nu_n,\nu) = \sup\{\int f\,d\nu_n - \int f\,d\nu\}$ where the supremum runs over 1-Lipschitz function and the fact that $Q_\theta\varphi$ is L_θ-Lipschitz (being an infimum of such functions), we easily get the following inequality

$$|\overline{\mathcal{T}}_\theta(\nu_n|\mu_n) - \overline{\mathcal{T}}_\theta(\nu|\mu)| \leqslant L_\theta W_1(\nu_n,\nu) + L_\theta W_1(\mu_n,\mu).$$

A similar (but simpler reasoning) based on the usual Kantorovich duality for $\mathcal{T}_\theta$ yields the inequality

$$|\mathcal{T}_\theta(\hat{\gamma}_n,\mu_n) - \mathcal{T}_\theta(\hat{\gamma},\mu)| \leqslant L_\theta W_1(\hat{\gamma}_n,\hat{\gamma}) + L_\theta W_1(\mu_n,\mu).$$

Passing to the limit as $n \to \infty$ in the identity $\overline{\mathcal{T}}_\theta(\nu_n|\mu_n) = \mathcal{T}_\theta(\hat{\gamma}_n,\mu_n)$, we end up with $\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\hat{\gamma},\mu)$.

Now it remains to extend this identity to general convex functions θ not necessarily Lipschitz. Let $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function (such that $\theta(0) = 0$)

and for all $n \geqslant 1$, let θ_n be the convex cost function defined by $\theta_n(x) = \theta(x)$, if $x \in [0, n]$ and $\theta_n(x) = \theta(n) + \theta'(n)(x - n)$, if $x \geqslant n$, where θ' denotes the right derivative of θ. It is easily seen that θ_n is Lipschitz and that $Q_{\theta_n}\varphi$ converges to $Q_\theta\varphi$ monotonically as $n \to \infty$, for any function φ bounded from below. Therefore, the monotone convergence theorem implies that for any probability measure γ, it holds $\int Q_\theta \varphi \, d\gamma = \sup_{n\geqslant 1} \int Q_{\theta_n}\varphi \, d\gamma$. We deduce from this that $\overline{\mathcal{T}}_\theta(\nu|\mu) = \sup_{n\geqslant 1} \overline{\mathcal{T}}_{\theta_n}(\nu|\mu)$ and $\mathcal{T}_\theta(\hat\gamma|\mu) = \sup_{n\geqslant 1} \mathcal{T}_{\theta_n}(\hat\gamma, \mu)$. Since $\overline{\mathcal{T}}_{\theta_n}(\nu|\mu) = \mathcal{T}_{\theta_n}(\hat\gamma, \mu)$ for all $n \geqslant 1$, this ends the proof of the first part of the theorem *i.e.* that $\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\hat\gamma|\mu)$).

From the first part of the theorem we conclude that there exists some $\hat\gamma \in \mathcal{P}_1(\mathbb{R})$ such that $\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\hat\gamma, \mu)$ for the three cost functions $\theta = \theta_1, \theta_2, \theta_1+\theta_2$. The result then follows from the well-known additivity of $\mathcal{T}_\theta$ in dimension one : $\mathcal{T}_{\theta_1+\theta_2}(\hat\gamma, \mu) = \mathcal{T}_{\theta_1}(\hat\gamma, \mu) + \mathcal{T}_{\theta_2}(\hat\gamma, \mu)$. This ends the proof of the theorem. □

5.5 Cost function vanishing on a neighorhood of the origin

In this section, we treat the particular case where the cost function vanishes on a neighborhood of 0. N.Gozlan has obtained a characterization of classical transport-entropy inequalities for cost functions vanishing at zero. This section completes this result by extending it to weak transport-entropy inequalities for the same type of cost functions.

Theorem 5.5.1. *Let $\mu \in \mathcal{P}_1(\mathbb{R})$ and $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function such that $\{t \in \mathbb{R}^+ : \beta(t) = 0\} = [0, t_o]$, where $t_o > 0$ is some positive constant. The following propositions are equivalent :*

1. *There is $a > 0$ such that μ satisfies the transport-entropy inequality $\mathrm{T}(\beta(a\,\cdot))$.*
2. *There is $a' > 0$ such that μ satisfies the weak transport-entropy inequality $\overline{\mathrm{T}}(\beta(a'\,\cdot))$.*
3. *There are $b > 0$ and $K > 0$ such that $\max(K^+(b), K^-(b)) \leqslant K$, where*

$$K^+(b) = \sup_{x\geqslant m} \frac{1}{\mu(x,\infty)} \int_x^\infty e^{\beta(b(u-x))}\mu(du),$$

and

$$K^-(b) = \sup_{x\leqslant m} \frac{1}{\mu(-\infty,x)} \int_{-\infty}^x e^{\beta(b(x-u))}\mu(du),$$

where m is a median of μ. (Here we use the convention $0/0 = 0$.)

4. *There is $d > 0$ such that*

$$|U_\mu(u) - U_\mu(v)| \leqslant \frac{1}{d}\beta^{-1}(|u - v|), \qquad \forall u \neq v \in \mathbb{R}.$$

(Note that β^{-1} is well defined on $(0, \infty)$.)

In particular, 2 *implies* 4 *with* $d = a'\frac{t_o}{8\beta^{-1}(\log 3)}$ *and* 4 *implies* 2 *with* $a' = d\frac{t_o}{9\beta^{-1}(2)}$.

Proof of Theorem 5.5.1. The equivalence between the assertions 1, 3 and 4 was first proved in [43] (see Theorem 2.2). Let us complete the proof of Theorem 5.5.1 by showing that $1 \Rightarrow 2 \Rightarrow 3$.

First of all, it follows easily from Jensen inequality that

$$\mathcal{T}_{\beta(a\,\cdot)}(\mu,\nu) \geqslant \max\left(\mathcal{T}_{\beta(a\,\cdot)}(\nu|\mu); \mathcal{T}_{\beta(a\,\cdot)}(\mu|\nu)\right).$$

Therefore 1 implies 2 with $a' = a$.

Now let us show that 2 implies 3. Suppose that μ satisfies $\overline{\mathrm{T}}(\beta(a\,\cdot))$ for some $a > 0$. According to Point 4 of Lemma 5.2.1, for all convex function $g : \mathbb{R} \to \mathbb{R}$ bounded from below, it holds

$$\int \exp(Qf)\,d\mu \int e^{-f}\,d\mu \leqslant 1,$$

where

$$Qf(x) = \inf_{y\in\mathbb{R}}\{f(y) + 2\beta(a|y-x|/2)\}.$$

Consider the convex function f_x which equals to 0 on $(-\infty, x]$ and ∞ otherwise, then $Qf(y) = 0$ on $(-\infty, x]$ and $Qf(y) = 2\beta(a(y-x)/2)$ on (x,∞). Applying the inequality above to f_x thus yields

$$\left(\mu(-\infty,x] + \int_{(x,\infty)} e^{2\beta(a(y-x)/2)}\mu(dy)\right)\mu(-\infty,x] \leqslant 1.$$

Considering $x \geqslant m$ yields that $K^+(a/2) \leqslant 3$. One proves similarly that $K^-(a/2) \leqslant 3$. This shows that 2 implies 3 with $b = a/2$ and $K = 3$. □

5.6 A transport form of the convex Poincaré inequality

This section is devoted to the proof of following theorem.

Theorem 5.6.1. *Let μ be a probability measure on $\mathbb{R}$, then the following assertions are equivalent :*

(a) *There exists $h > 0$ such that*

$$\sup_{x\in\mathbb{R}}[U_\mu(x+1) - U_\mu(x)] \leqslant h.$$

(b) *There exists $C > 0$ such that for all convex function f on $\mathbb{R}$, μ satisfies*

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

(c) *There exist $D, l_o > 0$ such that the probability μ satisfies the transport inequalities*

$$\overline{\mathcal{T}}_\alpha(\mu|\nu) \leqslant H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

and

$$\overline{\mathcal{T}}_\alpha(\nu|\mu) \leqslant H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

for the function α defined by

$$\alpha(u) = \begin{cases} \frac{u^2}{2D} & \text{if } |u| \leqslant l_o D \\ l_o|u| - l_o^2 D/2 & \text{if } |u| > l_o D \end{cases}.$$

Moreover the constants are related as follows :

- $(a) \Rightarrow (c)$ *with* $D = 2Kh^2$ *and* $l_o = c/h$,
- $(a) \Rightarrow (b)$ *with* $C = K'h^2$,
- $(b) \Rightarrow (a)$ *with* $h = K''\sqrt{h}$,
- $(c) \Rightarrow (b)$ *with* $C = D$,

where c, K, K' and K'' are absolute constants.

The equivalence $(a) \Leftrightarrow (b)$ goes back to Bobkov and Götze [12]. Hence, the proof of theorem completes the picture by showing that $(a)/(b)$ also characterize the measures satisfying a weak transport-entropy inequality with a cost function which is quadratic near zero and then linear (like θ_1). The dependence between the constants in the implication $(b) \Rightarrow (a)$ is not given for technical reasons. Indeed, the proof relies on an argument from [12] that uses a non trivial proof from [14] where one loses the explicit dependence on the constants.

We indicate that during the preparation of this work, we learned that this characterization of convex Poincaré inequality in terms of transport-entropy inequality has also been obtained by Feldheim, Marsiglietti, Nayar and Wang in their recent paper [37].

The proof of Theorem 5.6.1 is given in section 5.6. It uses independent results like a new discrete logarithmic Sobolev inequality for the exponential measure τ (see Lemma 5.6.9). By transportation technics, this logarithmic-Sobolev inequality provides logarithmic-Sobolev inequalities restricted to the class of convex or concave functions for measures satisfying the condition (a) (see Corollary 5.6.3).

Then the weak transport-entropy inequalities of item (c) are obtained in their dual forms, involving infimum convolution operators (see Lemma 5.2.1). The method to derive transport-entropy inequalities from logarithmic-Sobolev inequalities is based on classical arguments involving the infimum convolution operator as a solution of the Hamilton-Jacobi equation. This approach is due to [11] and has been also generalized in [47, 50].

Our strategy is to prove a modified logarithmic Sobolev inequality for the exponential probability measure τ and then, using a transport argument, a modified logarithmic Sobolev inequality for general μ (satisfying the assumption of Item (a)) restricted to convex or concave Lipschitz functions. Finally, following the well-known

Hamilton-Jacobi interpolation technique of [11], the desired transport inequalities will follow in their dual forms.

We need some notations. Given a convex or concave function $g\colon \mathbb{R} \to \mathbb{R}$, we set

$$|\nabla g|(x) := \min\{|\theta g'_-(x) + (1-\theta) g'_+(x)|; \theta \in [0,1]\}, \tag{5.6.2}$$

where g'_- and g'_+ denote the left and right derivatives of g (which are well-defined everywhere). In particular, if g is convex

$$|\nabla g|(x) = \begin{cases} |g'_+(x)| & \text{if } g'_+(x) \leqslant 0 \\ 0 & \text{if } g'_-(x) \leqslant 0 \leqslant g'_+(x) \\ g'_-(x) & \text{if } g'_-(x) \geqslant 0\,, \end{cases}$$

and if g is concave

$$|\nabla g|(x) = \begin{cases} |g'_-(x)| & \text{if } g'_-(x) \leqslant 0 \\ 0 & \text{if } g'_+(x) \leqslant 0 \leqslant g'_-(x) \\ g'_+(x) & \text{if } g'_+(x) \geqslant 0. \end{cases}$$

The following result is one of the key ingredients in the proof of Theorem 5.6.1. Recall the definition of U_μ from the introduction.

Proposition 5.6.3. *Let μ be a probability measure on $\mathbb{R}$. Assume that $\sup_{x\in\mathbb{R}}[U_\mu(x+1) - U_\mu(x)] \leqslant h$ for some $h > 0$. Set $K := 2740$ and $c := 1/(10\sqrt{2})$. Then, for all convex or concave and l-Lipschitz functions g with $l \leqslant c/h$, it holds*

$$\mathrm{Ent}_\mu(e^g) \leqslant Kh^2 \int_{\mathbb{R}} |\nabla g(x)|^2 e^{g(x)} \mu(dx). \tag{5.6.4}$$

The proof of Proposition 5.6.3 is postponed to the end of this section.

Proof of Theorem 5.6.1. As already mentioned above, from [12] we conclude that Item (i) is equivalent to Item (ii). In order to make the dependency of the constants explicit in the implication $(i) \Rightarrow (ii)$, one can use a well-known expansion argument : apply (5.6.4) to εf and take the limit $\varepsilon \to 0$, see *e.g.* [5]. On the other hand, using a similar expansion argument, it is easy to prove that Item (iii) implies Item (ii) with $C = 2D$: apply (5.2.2) to $g = \varepsilon f$ and take the limit $\varepsilon \to 0$, see *e.g.* [44, 51]. Hence, we are left with the proof of (i) implies (iii) which closely follows the Hamilton-Jacobi semi-group approach introduced in [11].

Let μ be a probability measure on the line and assume that Item (i) of Theorem 5.6.1 holds. According to Proposition 5.6.3, for any convex or concave differentiable function g which is l-Lipschitz with $l \leqslant c/h := l_o$, it holds

$$\mathrm{Ent}_\mu(e^g) \leqslant Kh^2 \int_{\mathbb{R}} |\nabla g(x)|^2 e^{g(x)} \mu(dx)\,, \tag{5.6.5}$$

with $K = 2740$ and $c = 1/(10\sqrt{2})$. It is easy to check that the latter is equivalent to

$$\mathrm{Ent}_\mu(e^g) \leqslant \int \alpha^*(|\nabla g|) e^g \, d\mu, \tag{5.6.6}$$

for all convex or concave $g\colon \mathbb{R} \to \mathbb{R}$ with

$$\alpha^*(v) := \sup_u \{uv - \alpha(u)\} = \begin{cases} Kh^2v^2 & \text{if } |v| \leqslant l_o \\ +\infty & \text{if } |v| > l_o \end{cases}$$

the convex conjugate of $\alpha(u) := \begin{cases} \frac{u^2}{4Kh^2} & \text{if } |u| \leqslant 2l_oKh^2 \\ l_o|u| - l_o^2 Kh^2 & \text{if } |u| > 2l_oKh^2 \end{cases}$.

Now, introduce the inf-convolution operators Q_t, for $t \in (0,1]$, defined by

$$Q_tf(x) := \inf_{y\in\mathbb{R}} \left\{ f(y) + t\alpha\left(\frac{x-y}{t}\right)\right\}, \qquad x \in \mathbb{R}, \qquad t \in (0,1],$$

which makes sense for instance for any Lipschitz function f or for any function f bounded from below. For simplicity denote by $\mathcal{F}$ the set of functions $f\colon \mathbb{R} \to \mathbb{R}$ that are l-Lipschitz and concave, $l \leqslant l_o$, or convex and bounded below. Then, Q_t satisfies the following technical properties :

(a) If f is convex, then Q_tf is convex.

(b) If f is concave and Lipschitz, then Q_tf is concave.

(c) If $f \in \mathcal{F}$, then Q_tf is l_o-Lipschitz.

(d) If $f \in \mathcal{F}$, then the function $u(t,x) := Q_tf(x)$ satisfies the following Hamilton-Jacobi equation

$$\frac{d}{dt_+}u(t,x) + \alpha^*(|\nabla^- u|)(t,x) = 0, \qquad \forall t \in (0,1], \qquad \forall x \in \mathbb{R}, \tag{5.6.7}$$

where $\frac{d}{dt_+}$ is the right time-derivative and $|\nabla^- u(t,x)| = \limsup_{y\to x} \frac{[u(t,y)-u(t,x)]_-}{|y-x|}$ (where as usual $[X]_- := \max(-X,0)$ denotes the negative part).

Item (a) is easy to check and is a general fact about infimum convolution of two convex functions (f and α). Item (b) follows from the fact that, after change of variables, $Q_tf(x) = \inf_u \left\{ f(x-u) + t\alpha\left(\frac{u}{t}\right)\right\}$ so that Q_tf is an infimum of concave functions and is therefore also concave. As for Item (c) we observe that $x \mapsto t\alpha\left(\frac{x-y}{t}\right)$ is l_o-Lipschitz for any y so that Q_tf is also l_o-Lipschitz as an infimum of l_o-Lipschitz functions. A proof of Item (d) can be found in [47] or [2]. We observe that the conclusions of Item (c) – (d) hold in much more general settings.

With these properties and definitions in hand, let $f \in \mathcal{F}$ and (following [11]) define

$$F(t) := \frac{1}{t}\log\left(\int_{\mathbb{R}} e^{tQ_tf}\,d\mu\right), \qquad t \in (0,1].$$

The function F is right differentiable at every point $t > 0$ (thanks to the above technical properties of Q_t, see *e.g.* [47] for details) and it holds

$$\begin{aligned}\frac{d}{dt_+}F(t) &= \frac{1}{t^2}\frac{1}{\int_{\mathbb{R}} e^{tQ_tf}\,d\mu}\left(\mathrm{Ent}_\mu\left(e^{tQ_tf}\right) + t^2\int_{\mathbb{R}}\left(\frac{d}{dt_+}Q_tf\right)e^{tQ_tf}\,d\mu\right)\\ &= \frac{1}{t^2}\frac{1}{\int_{\mathbb{R}} e^{tQ_tf}\,d\mu}\left(\mathrm{Ent}_\mu\left(e^{tQ_tf}\right) - Kh^2t^2\int_{\mathbb{R}}|\nabla^-Q_tf|^2e^{tQ_tf}\,d\mu\right)\\ &\leqslant \frac{Kh^2}{\int_{\mathbb{R}} e^{tQ_tf}\,d\mu}\left(\int|\nabla Q_tf|^2e^{tQ_tf}\,d\mu - \int_{\mathbb{R}}|\nabla^-Q_tf|^2e^{tQ_tf}\,d\mu\right) \leqslant 0,\end{aligned}$$

where the second equality follows from (5.6.7), the first inequality from (5.6.5) applied to the function $g = tQ_tf$ (which is convex or concave and tl_o-Lipschitz) and the last inequality from the fact that for a convex or concave function g, $|\nabla g| \leqslant |\nabla^- g|$ (we recall that $|\nabla g|$ is defined in (5.6.2)).

Thus the function F is non-increasing and satisfies $F(1) \leqslant \lim_{t\to 0} F(t) = \int f d\mu$. In other words,

$$\int e^{Q_1f}\,d\mu \leqslant e^{\int f\,d\mu} \qquad \forall f \in \mathcal{F}. \tag{5.6.8}$$

Now according to Item (*iii*) of Lemma 5.2.1 one concludes (on the one hand) that μ satisfies the transport-entropy inequality $\overline{\mathrm{T}}^-(\alpha) : \overline{\mathcal{T}}_\alpha(\mu|\nu) \leqslant H(\nu|\mu)$, for all $\nu \in \mathcal{P}_1(\mathbb{R})$.

On the other hand, applying (5.6.8) to $f = -Q_1g$ with g convex and bounded from below (so that f is concave and l_o-Lipschitz) yields to $e^{\int Q_1g\,d\mu}\int e^{Q_1(-Q_1g)}\,d\mu \leqslant 1$. Since $Q_1(-Q_1g) \geqslant -g$ we end up with

$$e^{\int Q_1g\,d\mu}\int e^{-g}\,d\mu \leqslant 1,$$

for all g convex and bounded from below. According to Item (*ii*) of Lemma 5.2.1, this implies that μ satisfies the transport-entropy inequality $\overline{\mathrm{T}}^+(\alpha) : \overline{\mathcal{T}}_\alpha(\nu|\mu) \leqslant H(\nu|\mu)$, for all $\nu \in \mathcal{P}_1(\mathbb{R})$, which completes the proof. □

The end of the section is dedicated to the proof of Proposition 5.6.3.

Proof fo Proposition 5.6.3. Let K and c be defined by Lemma 5.6.9 below. We may deal first with convex functions g and divide the proof into three different (sub-)cases : g monotone (non-decreasing and then non-increasing), and g arbitrary.

Assume first that g is convex *non-decreasing* and l-Lipschitz with $l \leqslant c/h$. Set $f = g \circ U_\mu$ (recall that U_μ is defined in the introduction). Then, since g is non-decreasing, and since $U_\mu(x-1) \leqslant U_\mu(x) - h$ by assumption, for all $x \in \mathbb{R}$, it holds

$$f(x) - f(x-1) \leqslant g(U_\mu(x)) - g(U_\mu(x) - h) \leqslant lh \leqslant c, \qquad \forall x \in \mathbb{R}.$$

Therefore, since μ is the image of τ under the map U_μ, Lemma 5.6.9 (apply (5.6.10) to f) and the latter guarantee that

$$\begin{aligned}\mathrm{Ent}_\mu(e^g) = \mathrm{Ent}_\tau(e^f) &\leqslant K\int_{\mathbb{R}} (f(x)-f(x-1))^2 e^{f(x)}\tau(dx)\\ &\leqslant K\int_{\mathbb{R}} (g(U_\mu(x)) - g(U_\mu(x)-h))^2 e^{g(U_\mu(x))}\tau(dx)\\ &= K\int_{\mathbb{R}} (g(x)-g(x-h))^2 e^{g(x)}\mu(dx) \leqslant Kh^2\int_{\mathbb{R}} |\nabla g(x)|^2 e^{g(x)}\mu(dx),\end{aligned}$$

where the last inequality is due to the fact that g is convex and non-decreasing and therefore satisfies $0 \leqslant g(x)-g(x-h) \leqslant g'_-(x)h = |\nabla g(x)|h$. As a conclusion we proved (5.6.4) for all convex *non-decreasing* and l-Lipschitz functions g with $l \leqslant c/h$.

Now suppose that g is convex, *non-increasing* and l-Lipschitz with $l \leqslant c/h$ and set $f(x) = g(U_\mu(-x))$. The function f is non-decreasing and, since $U_\mu(-x+1) \geqslant U_\mu(-x)+h$ by assumption, satisfies

$$f(x)-f(x-1) = g(U_\mu(-x)) - g(U_\mu(-x+1)) \leqslant g(U_\mu(-x)) - g(U_\mu(-x)+h) \leqslant c.$$

Similarly to the previous lines, Lemma 5.6.9 implies that

$$\begin{aligned}\mathrm{Ent}_\mu(e^g) = \mathrm{Ent}_\tau(e^f) &\leqslant K\int_{\mathbb{R}} (g(U_\mu(-x)) - g(U_\mu(-x+1)))^2 e^{g(U_\mu(-x))}\tau(dx)\\ &\leqslant K\int_{\mathbb{R}} (g(U_\mu(-x)) - g(U_\mu(-x)+h))^2 e^{g(U_\mu(-x))}\tau(dx)\\ &= K\int_{\mathbb{R}} (g(x)-g(x+h))^2 e^{g(x)}\mu(dx) \leqslant Kh^2\int_{\mathbb{R}} |\nabla g(x)|^2 e^{g(x)}\mu(dx),\end{aligned}$$

where we used the symmetry of τ and that $0 \leqslant g(x)-g(x+h) \leqslant g'_+(x)(-h) = |\nabla g(x)|h$. Therefore we proved (5.6.4) for all convex *non-increasing* and l-Lipschitz functions g with $l \leqslant c/h$.

Finally, consider an arbitrary convex and l-Lipschitz function g with $l \leqslant c/h$ and assume without loss of generality that g is not monotone. Being convex, there exists some $a \in \mathbb{R}$ such that g restricted to $(-\infty, a]$ is non-increasing and g restricted to $[a,\infty)$ is non-decreasing. Subtracting $g(a)$ if necessary, one can further assume that $g(a) = 0$ since (5.6.4) is invariant by the change of function $g \to g + C$ (for any constant C). Set $g_1 = g\mathbf{1}_{(-\infty,a]}$ and $g_2 = g\mathbf{1}_{(a,\infty)}$. The functions g_1 and g_2 are convex, monotone and l-Lipschitz. Therefore, according to the two previous sub-cases, it holds

$$\mathrm{Ent}_\mu(e^{g_1}) \leqslant Kh^2\int_{-\infty}^a |\nabla g(x)|^2 e^{g(x)}\mu(dx) \text{ and } \mathrm{Ent}_\mu(e^{g_2}) \leqslant Kh^2\int_a^{+\infty} |\nabla g(x)|^2 e^{g(x)}\mu(dx).$$

So what remains to prove is the following sub-additivity property of the entropy functional

$$\mathrm{Ent}_\mu(e^{g_1+g_2}) \leqslant \mathrm{Ent}_\mu(e^{g_1}) + \mathrm{Ent}_\mu(e^{g_2}),$$

which, since $\int ge^g d\mu = \int g_1 e^{g_1} d\mu + \int g_2 e^{g_2} d\mu$, amounts to proving that

$$\int e^{g_1} d\mu \log\left(\int e^{g_1} d\mu\right) + \int e^{g_2} d\mu \log\left(\int e^{g_2} d\mu\right) \leqslant \int e^{g}\, d\mu \log\left(\int e^{g} d\mu\right).$$

Setting $A = \int e^{g_1} d\mu - 1$, $B = \int e^{g_2} d\mu - 1$ and $X = \int e^g d\mu$ and observing that $A + B + 1 = X$ the latter is equivalent to proving that

$$(A+1)\log(A+1) + (B+1)\log(B+1) \leqslant X \log X\,,$$

which follows from the sub-additivity property of the convex function $\Phi\colon x \mapsto (x+1)\log(x+1)$ on $[0,\infty)$, that satisfies $\Phi(0) = 0$. This completes the proof when g is convex.

The case g concave follows the same lines (use (5.6.11) instead of (5.6.10)). Details are left to the reader. □

In the proof of Proposition 5.6.3 we used the following lemma which is a (discrete) variant of a result by Bobkov and Ledoux [16] and an entropic counterpart of a result (involving the variance) by Bobkov and Götze (see [12, Lemma 4.8]).

Lemma 5.6.9. *For all non-decreasing function $f\colon \mathbb{R} \to \mathbb{R}$ with $f(x) - f(x-1) \leqslant 1/(10\sqrt{2})$, $x \in \mathbb{R}$, it holds*

$$\mathrm{Ent}_\tau(e^f) \leqslant 2740 \int_{\mathbb{R}} \left(f(x) - f(x-1)\right)^2 e^f\, d\tau. \tag{5.6.10}$$

and

$$\mathrm{Ent}_\tau(e^f) \leqslant 2740 \int_{\mathbb{R}} \left(f(x+1) - f(x)\right)^2 e^f\, d\tau. \tag{5.6.11}$$

proof. Let τ^+ be the exponential probability measure on $\mathbb{R}^+ : \tau^+(dx) = e^{-x}\mathbf{1}_{[0,\infty)}dx$. We shall use the following fact, borrowed from [12, Lemma 4.7] (with $a = 0$ and $h = 1$ so that the constant $c(a,h)$ appearing in [12] can be explicitly bounded by $1/200$) : for all $f\colon [-1,\infty) \to \mathbb{R}$ non-decreasing and satisfying $f(0) = 0$, it holds

$$\int f^2 d\tau^+ \leqslant 200 \int (f(x) - f(x-1))^2 d\tau^+(x). \tag{5.6.12}$$

We will first prove (5.6.10). Since (5.6.10) is invariant by the change of function $f \to f + C$ for any constant C, we may assume without loss of generality that $f(0) = 0$. Set $\tilde{f}(y) := -f(-y)$, $y \in \mathbb{R}$ and observe that f is non-decreasing. Since $u \log u \geqslant u - 1$ for all $u \geqslant 0$, one has

$$\begin{aligned}
\mathrm{Ent}_\tau(e^f) &\leqslant \int (fe^f - e^f + 1)\, d\tau = \int \left(\int_0^1 t f^2 e^{tf} dt\right) d\tau \\
&= \frac{1}{2}\int_0^\infty f^2 \left(\int_0^1 te^{tf} dt\right) d\tau^+ + \frac{1}{2}\int_0^\infty \tilde{f}^2 \left(\int_0^1 te^{-t\tilde{f}} dt\right) d\tau^+ \\
&\leqslant \frac{1}{4}\int f^2 e^f d\tau^+ + \frac{1}{4}\int \tilde{f}^2 d\tau^+,
\end{aligned} \tag{5.6.13}$$

where the last inequality comes from the fact both f and $\tilde f$ are non-negative on $\mathbb{R}^+$. Now suppose that the function f is such that $f(y)-f(y-1)\leqslant c$ for all $y\in\mathbb{R}$ and some $c\in(0,1)$. Our aim is to bound each term in the right hand side of the latter.

By (5.6.12) applied to the function $\tilde f$, one has

$$\begin{aligned}\int \tilde f^2 d\tau^+ &\leqslant 200\int (\tilde f(y)-\tilde f(y-1))^2 d\tau^+(y) = \frac{200}{\int e^{-\tilde f} d\nu}\int (\tilde f(y)-\tilde f(y-1))^2 e^{-\tilde f(y)}\, d\tau^+(y)\\ &\leqslant 200\exp\left(\frac{\int \tilde f(y)(\tilde f(y)-\tilde f(y-1))^2\, d\tau^+(y)}{\int(\tilde f(y)-\tilde f(y-1))^2\, d\tau^+(y)}\right)\int (\tilde f(y)-\tilde f(y-1))^2 e^{-\tilde f(y)}\, d\tau^+(y).\end{aligned}$$

where we set $\frac{d\nu}{d\tau^+}(y)=\frac{(\tilde f(y)-\tilde f(y-1))^2}{\int(\tilde f(y)-\tilde f(y-1))^2\, d\tau^+(y)}$ and we used Jensen's inequality to guarantee that $1/\int e^{-\tilde f}d\nu\leqslant e^{\int \tilde f d\nu}$. By Cauchy-Schwarz' inequality and using (5.6.12) again, we get

$$\begin{aligned}\int \tilde f(y)(\tilde f(y)-\tilde f(y-1))^2 d\tau^+(y) &\leqslant \left(\int (\tilde f(y)-\tilde f(y-1))^4 d\tau^+(y)\right)^{1/2}\left(\int \tilde f^2 d\tau^+\right)^{1/2}\\ &\leqslant \sqrt{200}c\int (\tilde f(y)-\tilde f(y-1))^2 d\tau^+(y).\end{aligned}$$

It finally follows that

$$\begin{aligned}\int \tilde f^2 d\tau^+ &\leqslant 200 e^{\sqrt{200}c}\int (\tilde f(y)-\tilde f(y-1))^2 e^{-\tilde f(y)}\, d\tau^+(y)\\ &= 200 e^{\sqrt{200}c}\int_{-\infty}^0 (f(y+1)-f(y))^2 e^{f(y)} e^y\, dy\\ &= 200 e^{\sqrt{200}c-1}\int_{-\infty}^1 (f(y)-f(y-1))^2 e^{f(y-1)} e^y\, dy\\ &\leqslant 400 e^{\sqrt{200}c+1}\int_{-\infty}^1 (f(y)-f(y-1))^2 e^{f(y)}\, d\tau(y)\,,\end{aligned}$$

where in the last line we used that $e^y/(e^{-|y|}/2)\leqslant 2e^2$ for all $y\leqslant 1$.

Next we deal with the first term in the right hand side of (5.6.13). Our aim is to apply (5.6.12) to $g=fe^{f/2}$. Observe that, since f is non-decreasing, $f(x)\geqslant f(-1)\geqslant -c+f(0)=-c\geqslant -1$ so that, since $x\mapsto xe^{x/2}$ is non-increasing on $[-2,\infty)$ we are guaranteed that g is non-decreasing on $[-1,\infty)$ and therefore that we can apply (5.6.12) to g. Applying (5.6.12) to $g=fe^{f/2}$ and using the inequality

$$0\leqslant be^{b/2}-ae^{a/2}\leqslant (b-a)e^{b/2}+\frac{b}{2}(b-a)e^{b/2},\quad -2\leqslant a\leqslant b,$$

we get

$$\begin{aligned}&B:=\int f^2 e^f d\tau^+\leqslant 200\int \left(f(y)e^{f(y)/2}-f(y-1)e^{f(y-1)}\right)^2\, d\tau^+(y)\\ &\leqslant 400\int (f(y)-f(y-1))^2 e^{f(y)}\, d\tau^+(y)+100\int f^2(y)(f(y)-f(y-1))^2 e^{f(y)}\, d\tau^+(y)\\ &\leqslant 400\int (f(y)-f(y-1))^2 e^{f(y)}\, d\tau^+(y)+100c^2 B.\end{aligned}$$

Therefore, provided $c < 1/10$ we end up with $B \leqslant 400/(1-100c^2) \int (f(y)-f(y-1))^2 e^{f(y)}\, d\tau^+(y)$. Hence, plugging the previous two bounds into (5.6.13) and choosing $c = 1/\sqrt{200}$, Inequality (5.6.10) follows with the better constant 939 in factor of the right hand side.

To obtain (5.6.11) from (5.6.10), it suffices to observe that, by a simple change of variables

$$\begin{aligned}\int (f(y)-f(y-1))^2 e^{f(y)}\, d\tau(y) &= \int (f(x+1)-f(x))^2 e^{f(x+1)} \frac{e^{-|y+1|}}{2}\, dy \\ &\leqslant e^{c+1} \int (f(x+1)-f(x))^2 e^{f(x)}\, d\tau(x)\end{aligned}$$

and that $939e^{c+1} \leqslant 2740$ for $c = 1/\sqrt{200}$. This ends the proof. □

5.7 Characterisation of weak transport inequalities

The main result of this section is the following characterization of the transport inequalities $\overline{\mathrm{T}}_\theta$ associated to convex cost functions θ quadratic near 0.

Theorem 5.7.1. *Let $\mu \in \mathcal{P}_1(\mathbb{R})$ and $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function such that $\theta(t) = t^2$ for all $t \leqslant t_o$, for some $t_o > 0$. The following propositions are equivalent :*

i) There exists $a > 0$ such that μ satisfies $\overline{\mathrm{T}}(\theta(a\,\cdot))$.

ii) There exists b such that for all $u > 0$,

$$\sup_x \left(U_\mu(x+u) - U_\mu(x)\right) \leqslant \frac{1}{b}\theta^{-1}(u + t_o^2).$$

Moreover, constants are related as follows : i) implies ii) with $b = a\kappa_1$ and ii) implies i), with $a = b\kappa_2$, where κ_1 and κ_2 are two constants depending only on θ. More precisely,

$$\kappa_1 = \frac{t_o}{8\theta^{-1}(\log(3) + t_o^2)},$$

and

$$\kappa_2 = \frac{1}{2}\min\left(\frac{t_o}{\theta^{-1}(2+t_o^2)}; \frac{\max((c\sqrt{K})/t_o; 1)}{2\sqrt{K}\theta^{-1}(1+t_o^2)}\right).$$

Let Δ_μ denote the modulus of continuity of U_μ defined by

$$\Delta_\mu(h) = \sup\left\{U_\mu(x+u) - U_\mu(x), x \in \mathbb{R}, 0 \leqslant u \leqslant h\right\}, \quad h \geqslant 0.$$

The condition *ii)* asserts that

$$\Delta_\mu(h) \leqslant \frac{1}{b}\theta^{-1}(h + t_o^2).$$

Therefore Δ_μ is bounded around zero and $\Delta_\mu(h)$ goes not necessarily to zero as h goes to zero. This is one of the main differences with the conditions given in [43] to characterize the measure satisfying a usual transport-entropy $\overline{\mathrm{T}}(\theta)$. Actually if the measure μ is discrete and not a Dirac measure, the support of μ is not connected and there exist $a < b$ with a and b in the support of μ such that $\mu(]a,b[) = 0$. In that case, we may easily check that for all $h > 0$,

$$b - a \leqslant \Delta_\mu(h).$$

This show that in discrete setting $\lim_{h\to 0} \Delta_\mu(h) > 0$.

Let us briefly give the main ideas of this proof.

The weak transport-entropy equality $i)$ follows from condition $ii)$ as follows; using equality (5.4.1), we get that

$$\mathcal{T}_{\theta(a\,\cdot)}(\nu|\mu) \leqslant \mathcal{T}_{\theta_1(a\,\cdot)}(\nu|\mu) + \mathcal{T}_{\theta_2(a\,\cdot)}(\nu|\mu) \leqslant 2\mathrm{H}(\nu|\mu),$$

for a good choice of the constant a, by relating the condition $ii)$, either to a weak transport-entropy inequality with the cost function $\theta_2(a\,\cdot)$, either to a weak transport-entropy inequality with the cost function $\theta_1(a\,\cdot)$.

Proof of Theorem 5.7.1. Let $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function such that $\theta(t) = t^2$ on $[0, t_o]$ for some $t_o > 0$. Let us define $\theta_1(t) = t^2$ on $[0, t_o]$ and $\theta_1(t) = 2tt_o - t_o^2$ on $[t_o, +\infty)$ and $\theta_2(t) = [\theta(t) - t^2]_+$. Note that θ_1 and θ_2 are both convex and that θ_2 vanishes on $[0, t_o]$ and that $\max(\theta_1, \theta_2) \leqslant \theta \leqslant \theta_1 + \theta_2$.

First assume that μ satisfies the weak transport-entropy inequality $\overline{\mathrm{T}}(\theta(a\,\cdot))$ for some $a > 0$ (*i.e.* Item (i) of Theorem 5.7.1). Then, since $\theta \geqslant \theta_2$ it clearly satisfies $\overline{\mathrm{T}}(\theta_2(a\,\cdot))$. According to Theorem 5.5.1, the mapping U_μ sending the exponential measure on μ satisfies the condition :

$$\sup_{x\in\mathbb{R}} U_\mu(x+u) - U_\mu(x) \leqslant \frac{1}{b}\theta_2^{-1}(u), \qquad \forall u > 0, \tag{5.7.2}$$

with $b = a\kappa_1$, where $\kappa_1 = t_o/(8\theta_2^{-1}(\log 3))$. Since $\theta_2^{-1}(u) = \theta^{-1}(u + t_o^2)$ this proves Item (ii) of Theorem 5.7.1 with the announced dependency between the constants.

Now assume that μ satisfies Item (ii) of Theorem 5.7.1, or equivalently (5.7.2) for some $b > 0$. Recall that we set, in Theorem 5.6.1, $\kappa := 5480$ and $c := 1/(10\sqrt{2})$. Then, observe that, plugging $u = 1$ into (5.7.2) and using Theorem 5.6.1, one concludes that μ satisfies $\overline{\mathrm{T}}(\alpha)$ with α defined by $\alpha(u) = \bar{\alpha}(u/\sqrt{2D})$, with $D = \kappa\left(\theta^{-1}(1+t_o^2)\right)^2 \frac{1}{b^2}$ and

$$\bar{\alpha}(v) = \begin{cases} v^2 & \text{if } |v| \leqslant c\sqrt{\kappa/2} \\ c\sqrt{\kappa}|v| - \frac{c^2\kappa}{2} & \text{if } |v| > c\sqrt{\kappa/2} \end{cases} = \begin{cases} v^2 & \text{if } |v| \leqslant \sqrt{137/10} \\ 2\sqrt{137}|v| - \frac{137}{5} & \text{if } |v| > c\sqrt{137/10}. \end{cases}$$

It is not difficult to check that $\bar{\alpha}$ compares to θ_1. More precisely, for all $v \in \mathbb{R}$, it holds

$$\bar{\alpha}(v) \geqslant \theta_1\left(\max\left(\frac{c\sqrt{\kappa/2}}{t_o}; 1\right)|v|\right) = \theta_1\left(\max\left(\frac{\sqrt{137/10}}{t_o}; 1\right)|v|\right).$$

Therefore μ satisfies $\overline{\mathrm{T}}(\theta_1(a_1'\,\cdot\,))$, and by monotonicity $\overline{\mathrm{T}}(\theta_1(a_1\,\cdot\,))$ with

$$a_1' := \frac{\max((c\sqrt{\kappa/2})/t_o;1)}{\sqrt{2\kappa}\theta^{-1}(1+t_o^2)}b = \frac{\max\left(\frac{\sqrt{137/10}}{t_o};1\right)}{4\sqrt{685}\theta^{-1}(1+t_o^2)}b \geqslant \frac{1}{105}\frac{\max(1,t_o)}{t_o\theta^{-1}(1+t_o^2)}b =: a_1.$$

On the other hand, according to Theorem 5.5.1, μ also satisfies $\overline{\mathrm{T}}(\theta_2(a_2\,\cdot\,))$, with $a_2 = \frac{t_o}{\theta^{-1}(2+t_o^2)}b$. Letting $a = \min(a_1, a_2)$, one concludes that μ satisfies both $\overline{\mathrm{T}}(\theta_1(a\,\cdot\,))$ and $\overline{\mathrm{T}}(\theta_2(a\,\cdot\,))$. Hence, since $\theta(at) \leqslant \theta_1(at) + \theta_2(at)$ and according to (5.4.1), it holds

$$\begin{aligned}\mathcal{T}_{\theta(a\,\cdot\,)}(\nu|\mu) &\leqslant \mathcal{T}_{\theta_1(a\,\cdot\,)+\theta_2(a\,\cdot\,)}(\nu|\mu) = \mathcal{T}_{\theta_1(a\,\cdot\,)}(\nu|\mu) + \mathcal{T}_{\theta_2(a\,\cdot\,)}(\nu|\mu)\\ &\leqslant 2H(\nu|\mu),\end{aligned}$$

and so μ satisfies $\overline{\mathrm{T}}^+(\frac{1}{2}\theta(a\,\cdot\,))$. By convexity of θ and since $\theta(0) = 0$, it holds $\frac{1}{2}\theta(2at) \geqslant \theta(at)$, and so μ satisfies $\overline{\mathrm{T}}^+(\theta((a/2)\,\cdot\,))$. Finally we observe that

$$\frac{a}{2} = \frac{b}{210}\min\left(\frac{\max(1,t_o)}{t_o\theta^{-1}(1+t_o^2)};\frac{105t_o}{\theta^{-1}(2+t_o^2)}\right) \geqslant \frac{b}{210}\frac{\min(1,t_o)}{\theta^{-1}(2+t_o^2)} =: \kappa_2 b$$

so that, by monotonicity, μ satisfies $\overline{\mathrm{T}}^+(\theta(\kappa_2 b\,\cdot\,))$. The same reasoning yields the conclusion that μ satisfies $\overline{\mathrm{T}}^-(\kappa_2 b\,\cdot\,))$, which completes the proof. □

5.8 Characterisation of convex modified log-Sobolev inequalities

The goal of this section is to characterize the convex modified log-Sobolev inequality. The main theorem is the following :

Theorem 5.8.1. *Let μ be a probability measure on $\mathbb{R}$. The following conditions are equivalent.*

(i) For all convex functions $\varphi : \mathbb{R} \to \mathbb{R}$, there exists $C > 0$ such that the modified log-Sobolev inequality holds

$$\mathrm{Ent}_\mu(e^\varphi) \leqslant C\int_{\mathbb{R}} \varphi'^2 e^\varphi d\mu \qquad \text{(CmLS)}$$

for all convex functions $\varphi : \mathbb{R} \to \mathbb{R}$.

(ii) There exist $a, b > 0$ such that for all $h > 0$,

$$\sup_{x\in\mathbb{R}} U_\mu(x+h) - U_\mu(x) \leqslant \sqrt{a+bh}. \tag{5.8.2}$$

In the result above we did not keep track of the constants, in order not to overload the presentation. Let us simply mention that the relation between the constants C, a and b can be made explicit.

One of the key idea is to use the equivalence between CmLS and $\overline{\mathrm{T}}_2$ from [50] recalled below :

Theorem 5.8.3. *The following assertions are equivalent :*

(i) There exists $C > 0$ such that weak transport inequality $\overline{\mathrm{T}}_2^-(C)$ holds.

(ii) There exists $D > 0$ such that for all convex functions φ, it holds

$$\int \exp\left\{\frac{Q\varphi}{D}\right\} d\mu \leqslant \exp\left\{\frac{1}{D}\int \varphi d\mu\right\}.$$

(iii) There exists $E > 0$ such that the modified log-Sobolev inequality holds

$$\mathrm{Ent}_\mu(e^\varphi) \leqslant E \int_{\mathbb{R}} \varphi'^2 e^\varphi d\mu \qquad \text{(CmLS)}$$

for all convex functions $\varphi : \mathbb{R} \to \mathbb{R}$.

Now combining with Theorem 5.7.1, one has

$$(5.8.2) \Leftrightarrow \overline{\mathrm{T}}_2 \Rightarrow \overline{\mathrm{T}}_2^- \Leftrightarrow \mathrm{CmLS}.$$

Therefore, we only need to prove $\overline{\mathrm{T}}_2^- \Rightarrow (5.8.2)$.

We begin with the following lemma

Lemma 5.8.4. *If the modified log-Sobolev inequality for convex functions (*CmLS*) holds, then there exists some constant c such that*

$$\frac{1}{\mu(x,\infty)}\int_x^\infty (u-x)^2 \mu(dx) \leqslant c$$

for $0 \leqslant x < \sup \mathrm{Supp}(\mu)$.

proof. The modified log-Sobolev inequality implies the Poincaré inequality for convex functions. It follows (from [12]) that there exist $a, b > 0$ such that $U_\mu(x+h) - U_\mu(x) \leqslant a + bh$ for $x \in \mathbb{R}$, $h \geqslant 0$.

Recall that $U_\mu = F_\mu^{-1} \circ F_\tau$, where F_μ is the generalized inverse of F_μ and denote by $V_\mu = F_\tau^{-1} \circ F_\mu$. Even if U_μ is not necessarily invertible (if μ is discrete for instance), the following holds for all x :

$$U_\mu^{-1}((x,\infty)) = (V_\mu(x), \infty),$$

and

$$U_\mu \circ V_\mu(x) \leqslant x.$$

Therefore, letting , it holds

$$\begin{aligned}
\frac{1}{\mu(x,\infty)}\int_x^\infty (u-x)^2 \mu(dx) &= \frac{1}{\tau(V_\mu(x),\infty)}\int_{V_\mu(x)}^\infty \Big(U_\mu(u) - x\Big)^2 \tau(du) \\
&\leqslant e^{V_\mu(x)} \int_{V_\mu(x)}^\infty \Big(U_\mu(u) - U_\mu \circ V_\mu(x)\Big) e^{-u} du \\
&\leqslant \int_{V_\mu(x)}^\infty \Big(a + b(u - V_\mu(x))\Big)^2 e^{-(u - V_\mu(x))} du \\
&= \int_0^\infty (a+bt)^2 e^{-t} dt < \infty,
\end{aligned}$$

□

Proof of the theorem. Assume that the convex modified log-Sobolev inequality holds for μ. Then the weak transport entropy inequality $\overline{\mathrm{T}}_2^-$. Equivalently, according to Theorem 5.8.3, it holds for some $D > 0$

$$\Big(\int_{\mathbb{R}} \exp\left(\tfrac{1}{D}Qf\right)d\mu\Big)^D \exp\Big(-\int_{\mathbb{R}} f d\mu\Big) \leqslant 1, \tag{5.8.5}$$

for all convex Lipschitz function $f : \mathbb{R} \to \mathbb{R}$ bounded from below, where $Qf(t) = \inf_{y\in\mathbb{R}}\{f(y) + (t-y)^2\}$. Fix $x > 0$ and consider the function $f(t) = [t-x]_+^2$. Then $Qf(t) = 0$ if $t \leqslant x$. For $t > x$,

$$Qf(t) = \inf_{y\in\mathbb{R}}\{[y-x]_+^2 + (t-y)^2\} = \frac{(t-x)^2}{2}.$$

Hence the dual formulation (5.8.5) indicates that

$$\begin{aligned}\mu(-\infty, x) + \int_x^\infty \exp\left(\tfrac{1}{2D}(t-x)^2\right)d\mu(t) &= \int_{\mathbb{R}} \exp\left(\tfrac{1}{D}Qf\right)d\mu \\ &\leqslant \exp\Big(\tfrac{1}{D}\int_{\mathbb{R}} f d\mu\Big) = \exp\Big(\tfrac{1}{D}\int_x^\infty (t-x)^2\mu(dt)\Big).\end{aligned} \tag{5.8.6}$$

By Lemma 5.8.4 there exists $c > 0$ such that $\int_x^\infty (t-x)^2\mu(dt) \leqslant c\mu(x,\infty)$. Moreover, $\mu(x,\infty) \to 0$ as $x \to \infty$. Thus, there exists $M > 0$ such that $\mu(M,\infty) \leqslant D/c$ and for all $x \geqslant M$, we have

$$\int_x^\infty (t-x)^2\mu(dt) \leqslant D.$$

Note that $e^u \leqslant 1 + 2u$ for $u \in (0,1)$. As a result, for $x \geqslant M$,

$$\exp\Big(\tfrac{1}{D}\int_x^\infty (t-x)^2\mu(dt)\Big) \leqslant 1 + 2\int_x^\infty (t-x)^2\mu(dt).$$

Combining this with equation (5.8.6), one gets that for $x \geqslant M$

$$\int_x^\infty \exp\left(\tfrac{1}{2D}(t-x)^2\right)\mu(dt) \leqslant \mu(x,\infty) + 2\int_x^\infty (t-x)^2\mu(dt)$$

and hence, by Lemma 5.8.4,

$$\frac{1}{\mu(x,\infty)}\int_x^\infty \exp\left(\tfrac{1}{2D}(t-x)^2\right)\mu(dt) \leqslant 1 + \frac{2}{\mu(x,\infty)}\int_x^\infty (t-x)^2\mu(dt) \leqslant 1 + 2c.$$

Denote $b = 1/2$ and $\beta(u) = \frac{1}{2D}[u-1]_+^2$ for $u > 0$. We have $\beta(bu) \leqslant \frac{1}{2D}u^2$. Therefore

$$\frac{1}{\mu(x,\infty)}\int_x^\infty \exp\Big(\beta\big(b(t-x)\big)\Big)\mu(dt) \leqslant 1 + 2c$$

for all $x \geqslant M$.

On the other hand, for $x \in (0, M)$,

$$\frac{1}{\mu(x,\infty)} \int_x^\infty \exp\Big(\beta(b(t-x))\Big)\mu(dt) \leqslant \frac{c}{D} \int_x^\infty \exp\left(\tfrac{1}{2D}(t-x)^2\right)\mu(dt)$$

according to inequality (5.8.6),

$$\frac{c}{D} \int_x^\infty \exp\left(\tfrac{1}{2D}(t-x)^2\right)\mu(dt) \leqslant \frac{c}{D} \exp\left(\tfrac{1}{D} \int_x^\infty (t-x)^2\mu(dt)\right) \leqslant \frac{c}{D}\exp(c/D).$$

Hence, there exists $K > 0$ such that for all $x \geqslant 0$,

$$\frac{1}{\mu(x,\infty)} \int_{(x,\infty)} \exp\Big(\beta(b(t-x))\Big)\mu(dt) \leqslant K.$$

Thus, following the notations of Theorem 5.5.1, we have

$$K^+ := \sup_{x>m} \frac{1}{\mu(x,\infty)} \int_x^\infty \exp\Big(\beta(b(t-x))\Big)\mu(dt) < \infty,$$

and by a similar argument, we get $K^- < \infty$. According to Theorem 5.5.1, we conclude that (5.8.2) holds. □

Remark 5.8.7. *We have in fact proved that Equation* (5.8.5) *implies* $K^+ < \infty$ *and* $K^- < \infty$*, which is equivalent to the contraction property* (5.8.2) *of theorem 5.8.1. If we consider the cost function* $\beta(x) = [|x| - a]^2, a > 0$ *instead of* x^2 *then, following the same line of reasoning, we see that the implication*

$$(5.8.5) \Rightarrow \max\{K^+, K^-\} < \infty$$

still holds with

$$Qf(t) = \inf_y \{f(y) + [|t-y| - a]_+^2\}.$$

Therefore, the weak transport inequality $\overline{\mathrm{T}}_\beta^-$ *is equivalent to the contraction property. This method can be generalized to more general convex costs, one can refer to [94] for more details.*

5.9 A necessary condition for transport-information inequalities

In this section, we are interested in transport-information inequalities introduced in [54, 53]. We recall that μ satisfies $\mathrm{W_1I}$ if and only if for all probability $\nu \ll \mu$, it holds

$$W_1(\mu,\nu)^2 \leqslant \mathcal{I}(\nu|\mu),$$

where the Fisher information is defined by

$$\mathcal{I}(\nu|\mu) := \frac{1}{4}\int_{\mathbb{R}} \frac{f'^2}{f} d\mu,$$

with $f = d\nu/d\mu$. We say that μ satisfies $\mathrm{W_1H}$ if and only if for all probability $\nu \ll \mu$, it holds

$$W_1(\mu,\nu)^2 \leqslant H(\nu|\mu).$$

The following implications are known

$$\begin{array}{ccccc} \mathrm{LS} & \Rightarrow & \mathrm{T_2I} & \Rightarrow & \mathrm{T_2} \\ & & \Downarrow & & \Downarrow \\ & & \mathrm{W_1I} & \Rightarrow & \mathrm{W_1H} \end{array}$$

A natural question is to ask whether these implications can be reversed. This question is already partially answered. In [22], Cattiaux and Guillin provided an example of a real probability measure, which satisfies Talagrand's inequality $\mathrm{T_2}$ but not the log-Sobolev inequality LS. This probability is defined on $\mathbb{R}$ as follows :

$$\mu(dx) = \frac{1}{Z}\exp\{-|x|^3 - |x|^\beta - 3x^2\sin^2(x)\}dx, \text{ with } 2 < \beta < \frac{5}{2}.$$

Later, Gozlan [43] provided another example which achieved the same task with a tail distribution exactly Gaussian. Guillin, Leonard, Wang and Wu provided an example of a probability measure on $\mathbb{R}$ in [53], such that $\mathrm{W_1I}$ holds but not LS :

$$\mu(dx) = \frac{1}{Z}\exp\{-x^4 - |x|^\beta - 4x^3\sin^2(x)\}dx, \text{ with } 2 < \beta < 3$$

In the same paper, they also gave an example which satisfies $\mathrm{W_1H}$ but not $\mathrm{W_1I}$.

The goal of this section is to construct an example of a real probability measure which satisfies $\mathrm{T_2}$ but not $\mathrm{W_1I}$ and, as a consequence, not $\mathrm{T_2I}$. The idea is to find a necessary condition for the inequality $\mathrm{W_1I}$ and to compare it to the characterisation of $\mathrm{T_2}$ given in [43].

For technical reasons, we will assume that μ is a probability measure absolutely continuous w.r.t Lebesgue measure on $\mathbb{R}$ with median 0. We also denote by $x \mapsto \mu(x)$ the density function of μ with respect to Lebesgue measure, that is to say

$$\mu(dx) = \mu(x)dx.$$

We assume further that for all segment $[a,b], a < b,\ \mu[a,b] > 0$, in order to avoid technical discussions.

5.9.1 Test functions

The idea to obtain a necessary condition is to consider the following family of test functions :

Definition 5.9.1. *Define a family of functions $(f_r)_r, r \in \mathbb{R}_+^*$ by*

$$f_r(x) = \begin{cases} 0 & \text{if } x \leqslant 0, \\ \int_0^x \frac{dt}{\mu(t)} & \text{if } 0 < x < r, \\ \int_0^r \frac{dt}{\mu(t)}, & \text{if } x \geqslant r, \end{cases}$$

and let $(\nu_r)_{r \in \mathbb{R}_+^*}$ denote the family of probability measures defined by

$$\frac{d\nu_r}{d\mu}(x) = g_r(x) := \frac{f_r(x)^2}{Z_r},$$

with

$$Z_r = \int_0^\infty f_r^2(x)\mu(x)dx.$$

Finally, denote by T_r the monotone rearrangement map sending μ to ν_r (which realizes the optimal transport cost for all convex costs).

Note that $\nu_r(0, T_r(0)) = \mu(-\infty, 0) = 1/2$.

This choice of test functions is rather natural. One of the main reason is that the Fisher information is easy to compute : we have

$$\mathcal{I}(\nu_r|\mu) = \frac{1}{4}\int \frac{g_r'^2}{g_r} d\mu = \frac{1}{Z_r}\int f_r'^2 d\mu = \frac{1}{Z_r}\int_0^r \frac{1}{\mu^2(x)} d\mu(x) = \frac{f_r(r)}{Z_r}.$$

It follows that, if we assume that the transport information inequality $\alpha(\mathcal{T}) \leqslant I$ holds for μ, we have immediately

$$\alpha(\mathcal{T}(\nu_r, \mu))\frac{Z_r}{f_r(r)} \leqslant c,$$

according to the item 2. of next lemma, it follows that

$$\alpha(\mathcal{T}(\nu_r, \mu))\int_0^r \frac{dt}{\mu(t)}\mu(r, \infty) < c. \tag{5.9.2}$$

We thus get a condition which is similar to the characterisations of Poincaré inequality and Log-Sobolev inequalities. Our next goal will be to estimate the transportation cost in the previous formula. Before that, let us explore some additional properties of our family of test functions.

Proposition 5.9.3. *With the notations above, the following assertions hold :*

1. *For all $0 < x < y \in \mathbb{R}$, $Z_x < Z_y$.*

2. *For all* $r > 0$, $Z_r/f_r(r) \geqslant f_r(r)\mu(r,\infty)$.
3. *The function* $r \mapsto f_r(r)^2/Z_r, r \in \mathbb{R}_+^*$ *is strictly increasing.*

proof. (1) If $0 < x < y$, then

$$f_x(t) \begin{cases} = f_y(t) & t \leqslant x \\ < f_y(t) & t > x. \end{cases}$$

and the conclusion follows.

(2) We have :

$$\frac{Z_r}{f_r(r)} = \frac{\int_0^\infty f_r(x)^2 d\mu(x)}{f_r(r)} \geqslant \frac{\int_r^\infty f_r(r)^2 d\mu(x)}{f_r(r)} = f_r(r)\mu(r,\infty).$$

(3) Assume that $h > 0$, in order to show

$$\frac{f_r(r)^2}{Z_r} < \frac{f_{r+h}(r+h)^2}{Z_{r+h}},$$

it is enough to show

$$\frac{f_{r+h}(r+h)^2}{f_r(r)^2} > \frac{Z_{r+h}}{Z_r}.$$

Observe that $f_{r+h} = f_r$ on $[0,r]$ and $f_{r+h}(x) \leqslant f_{r+h}(r+h)$ for all $x > r$ so it holds

$$\begin{aligned} Z_{r+h} &= \int_0^r f_{r+h}(x)^2 d\mu + \int_r^\infty f_{r+h}(x)^2 d\mu \\ &< \int_0^r f_r(x)^2 d\mu + f_{r+h}(r+h)^2 \mu(r,\infty) \\ &= Z_r + \mu(r,\infty)(f_{r+h}(r+h)^2 - f_r(r)^2). \end{aligned}$$

Thus,

$$\frac{Z_{r+h}}{Z_r} < 1 + \frac{\mu(r,\infty)}{Z_r}(f_{r+h}(r+h)^2 - f_r(r)^2).$$

On the other hand, according to (2), it holds

$$\frac{\mu(r,\infty)}{Z_r} \leqslant \frac{1}{f_r(r)^2}.$$

Therefore,

$$\begin{aligned} \frac{Z_{r+h}}{Z_r} &< 1 + \frac{\mu(r,\infty)}{Z_r}(f_{r+h}(r+h)^2 - f_r(r)^2) \\ &\leqslant 1 + \frac{1}{f_r(r)^2}(f_{r+h}(r+h)^2 - f_r(r)^2) = \frac{f_{r+h}(r+h)^2}{f_r(r)^2}. \end{aligned}$$

□

In dimension one, according to a well known result in optimal transport, for any convex cost function $\theta : \mathbb{R}^+ \to \mathbb{R}$, it holds

$$\mathcal{T}_\theta(\nu_r, \mu) = \int \theta(|x - T_r(x)|) d\mu,$$

where we recall that T_r is the monotone rearrangement map sending μ on ν_r, which is defined as

$$T_r := F_{\nu_r}^{-1} \circ F_\mu.$$

In order to get a necessary condition for transport information inequalities, we now want to estimate from below the transportation cost. Let us turn to the family of functions T_r and study their properties. The next proposition shows that the function $r \mapsto T_r(x)$ is strictly increasing for all $x \in \mathbb{R}$.

Proposition 5.9.4 (Monotonicity of T_r). *For all $x \in \mathbb{R}$, the function $r \mapsto T_r(x), r \in \mathbb{R}_+^*$ is strictly increasing.*

proof. Let $r, h > 0$, we shall show that for given x, $T_r(x) < T_{r+h}(x)$. To discuss the monotonicity, we consider different cases :

Case (i), $T_{r+h}(x) \leqslant r$.
Since for all $t \in [0, T_{r+h}]$, $f_r(t) = f_{r+h}(t)$, thus

$$\begin{aligned}
\int_0^{T_{r+h}(x)} \frac{f_r^2}{Z_r} d\mu &> \int_0^{T_{r+h}(x)} \frac{f_{r+h}^2}{Z_{r+h}} d\mu \\
&= \nu_{r+h}(0, T_{r+h}(x)) = \mu(-\infty, x) = \nu_r(0, T_r(x)) \\
&= \int_0^{T_r(x)} \frac{f_r^2}{Z_r} d\mu,
\end{aligned}$$

where the first inequality is from the fact that $Z_r < Z_{r+h}$. Hence, $T_r(x) < T_{r+h}(x)$.

Case (ii), $T_r(x) \geqslant r + h$.
Since $f_r(x) = f_r(r)$ on $(r, \infty) \supset (T_r(x), \infty)$, it holds

$$\begin{aligned}
\mu(x, \infty) &= \nu_r(T_r(x), \infty) \\
&= \int_{T_r(x)}^{\infty} \frac{f_r(x)^2}{Z_r} d\mu \\
&= \mu(T_r(x), \infty) \frac{f_r(r)^2}{Z_r} \\
&< \mu(T_r(x), \infty) f_{r+h}(r+h)^2 / Z_{r+h},
\end{aligned}$$

the last inequality is from the item (3) of Proposition 5.9.3. We deduce that there exists $v > T_r(x) \geqslant r + h$, such that

$$\mu(v, \infty) f_{r+h}(r+h)^2 / Z_{r+h} = \mu(x, \infty).$$

Hence,

$$\int_v^{\infty} \frac{f_{r+h}(x)^2}{Z_{r+h}} d\mu = \mu(x,\infty).$$

It follows that $T_{r+h}(x) = v > T_r(x)$.

Case (iii), $T_r(x) \geqslant r$ and $T_{r+h}(x) \geqslant r+h$.
By similar arguments of the previous case, we deduce the existence of $v > T_r(x) \geqslant r$, such that

$$\mu(v,\infty) f_{r+h}(r+h)^2/Z_{r+h} = \mu(x,\infty).$$

On the other hand, since $T_{r+h}(x) \geqslant r+h$, it holds

$$\mu(T_{r+h}(x),\infty) f_{r+h}(r+h)^2/Z_{r+h} = \mu(x,\infty).$$

We conclude from this , that $\mu(v,\infty) = \mu(T_{r+h}(x),\infty)$ and so $T_{r+h}(x) = v > T_r(x)$.

Now it is enough to show the monotonicity in the following last case :
Case (iv), $r < T_r(x), T_{r+h}(x) < r+h$.

We shall prove it by contradiction. Assume that

$$r < T_{r+h}(x) \leqslant T_r(x) < r+h.$$

Let $h' = T_{r+h}(x) - r$. We remark that $r + h' = T_{r+h}(x) < r+h$ and $T_{r+h}(x) \leqslant r+h'$, according to the Case (i), it holds $T_{r+h}(x) > T_{r+h'}(x)$.
On the other hand, since $T_r(x) \geqslant r + h'$, the Case (ii) allows to conclude that $T_{r+h'}(x) > T_r(x)$. As a consequence,

$$T_{r+h}(x) > T_{r+h'}(x) > T_r(x).$$

This is a contradiction since we suppose that $T_{r+h}(x) \leqslant T_r(x)$.

Therefore, by Cases $(i) - (iv)$, one can conclude that for all $r, h > 0$,

$$T_{r+h}(x) > T_r(x).$$

□

For a given $x > 0$, the function $r \in \mathbb{R}^+ \mapsto T_r(x)$ is increasing and is bounded from below by 0. Thus the limit when r goes to 0 exists. We calculate this limit in the next proposition.

Proposition 5.9.5. *Given* $x \geqslant 0$, $\lim_{r\to 0} T_r(x) = F_\mu^{-1}(1 - \frac{\mu(x,\infty)}{2}) > x$.

proof. Let $\varepsilon > 0$, denote $c_\varepsilon = \frac{f_\varepsilon(\varepsilon)^2}{Z_\varepsilon}$. By definition, one can write $\nu_\varepsilon(t) = \mu(t)\frac{f_\varepsilon(t)^2}{Z_\varepsilon}$. Now it holds

$$1 = \nu_\varepsilon(0,\infty) = \int_0^{\varepsilon} \frac{f_\varepsilon(t)^2}{Z_\varepsilon} d\mu + c_\varepsilon \mu(\varepsilon,\infty) \in \Big(c_\varepsilon \mu(\varepsilon,\infty), c_\varepsilon \mu(0,\infty)\Big)$$

Therefore, by the fact that $\mu(0,\infty) = \frac{1}{2}$, it follows that

$$2 < c_\varepsilon < \frac{1}{\mu(\varepsilon,\infty)}$$

which implies that $c_\varepsilon \to 2$, when $\varepsilon \to 0$. Now consider

$$\mu(x,\infty) = \nu(T_\varepsilon(x),\infty) = \int_{T_\varepsilon(x)}^{\infty} \frac{f_\varepsilon(t)^2}{Z_\varepsilon} d\mu, \tag{5.9.6}$$

Noticing that when ε goes to 0,

$$\int_{\varepsilon}^{\infty} \frac{f_\varepsilon(t)^2}{Z_\varepsilon} d\mu = \mu(\varepsilon,\infty)c_\varepsilon \to 1 > \mu(x,\infty) = \int_{T_\varepsilon(x)}^{\infty} \frac{f_\varepsilon(t)^2}{Z_\varepsilon} d\mu.$$

Thus, for ε small enough, we have $T_\varepsilon(x) > \varepsilon$. Together with (5.9.6), for ε small enough, it follows that

$$\mu(x,\infty) = \int_{T_\varepsilon(x)}^{\infty} \frac{f_\varepsilon(t)^2}{Z_\varepsilon} d\mu = c_\varepsilon \mu(T_\varepsilon(x),\infty);$$

Thus

$$\lim_{\varepsilon\to 0} \mu(T_\varepsilon(x),\infty) = \frac{\mu(x,\infty)}{2}.$$

Hence, $\lim_{\varepsilon\to 0} T_\varepsilon(x) = F_\mu^{-1}(1 - \frac{\mu(x,\infty)}{2})$. □

5.9.2 Lower bound of the transportation cost for the test function

Next we will derive a lower bound for the transportation cost in terms of the test function.

Proposition 5.9.5 and the assumption of median 0 immediately yield that

$$T_0(0) := \lim_{\varepsilon\to 0} T_\varepsilon(0) = F_\mu^{-1}(3/4) > 0.$$

Therefore, for $0 < x < T_0(0)$, by monotonicity of $r \mapsto T_r(0)$ and monotonicity of $x \mapsto T_r(x)$, it holds

$$0 < x < T_0(0) < T_r(0) < T_r(x).$$

Hence, for a convex cost θ,

$$\begin{aligned}\mathcal{T}_\theta(\nu_r,\mu) = \int \theta(|x - T_r(x)|)d\mu &> \int_0^{T_r(0)} \theta(|x - T_r(x)|)d\mu \\ &\geqslant \mu(0,T_r(0))\theta(T_r(0) - T_0(0)) > \frac{1}{4}\theta(T_r(0) - T_0(0)).\end{aligned}$$

where the last inequality we used $\mu(0,T_r(0)) > \mu(0,T_0(0)) = \frac{1}{4}$. Plugging it into equation (5.9.2), and now denoting by m the median of μ in the place of 0, one gets

Theorem 5.9.7. *Let μ be symmetric with respect to its median m and absolutely continuous with respect to Lebesgue measure. Assume that μ satisfies the transport information inequality with a convex cost θ and increasing function α, more precisely, for all probability measure $\nu \ll \mu$, it holds :*

$$\alpha\left(\mathcal{T}_\theta(\nu,\mu)\right) \leqslant c\mathcal{I}(\nu|\mu)$$

then

$$B^+ := \sup_{x>m}\left\{\alpha\left(\frac{1}{4}\theta\left(T_x(m) - F_\mu^{-1}\left(\frac{3}{4}\right)\right)\right)\int_m^x \frac{1}{\mu(t)}dt\ \mu(x,\infty)\right\} \leqslant c < \infty.$$

Remark 5.9.8. *The symmetric property is only for a simplification issue, for μ non-symmetric, the following negative counter part part also holds*

$$B^- := \sup_{x<m}\left\{\alpha\left(\frac{1}{4}\theta\left(\tilde{T}_x(m) - F_\mu^{-1}\left(\frac{1}{4}\right)\right)\right)\int_x^m \frac{1}{\mu(t)}dt\ \mu(-\infty,x)\right\} \leqslant c < \infty,$$

with $\tilde{T}_x$ a function analogy of T_x but related to the negative part of μ.

Corollary 5.9.9. *Let μ be symmetric with respect to its median m and absolutely continuous with respect to Lebesgue measure. Assume that there exists a convex cost θ, with $\theta > 0$ on $\mathbb{R}\setminus\{0\}$ and an increasing positive function $\alpha : \mathbb{R}^+ \to \mathbb{R}^{+*}$, such that μ satisfies the transport information inequality :*

$$\forall \nu \ll \mu, \quad \alpha\left(\mathcal{T}_\theta(\nu,\mu)\right) \leqslant c\mathcal{I}(\nu|\mu).$$

Furthermore, if there exists $\varepsilon > 0$ such that (m being the median of μ)

$$M := \int_m^{m+\varepsilon} \frac{1}{\mu(t)}dt < \infty,$$

then μ satisfies the Poincaré inequality with constant C, and it holds

$$C \leqslant \sup\left\{\frac{4c}{T_\varepsilon(m) - F_\mu^{-1}\left(\frac{3}{4}\right)}, 2M\right\}$$

proof. According to the characterisation of Poincaré inequality given in [82] by Muckenhoupt, it is enough to show that

$$A := \sup_{x>m}\left\{\int_m^x \frac{1}{\mu(t)}dt\ \mu(x,\infty)\right\} < \infty.$$

By monotonicity of T_r (Proposition 5.9.4), one has $T_{x+\varepsilon}(m) - F_\mu^{-1}\left(\frac{3}{4}\right) > 0$ for all $x \geqslant 0$. Theorem 5.9.7 indicates that for all $x > m+\varepsilon$,

$$\sup_{x>m+\varepsilon}\left\{\int_m^x \frac{1}{\mu(t)}dt\ \mu(x,\infty)\right\} < \frac{c}{T_\varepsilon(m) - F_\mu^{-1}\left(\frac{3}{4}\right)}.$$

We observe that on the other hand,

$$\sup_{x\in(m,m+\varepsilon)}\left\{\int_m^x \frac{1}{\mu(t)}dt\ \mu(x,\infty)\right\}\leqslant M\mu(m,\infty)=\frac{M}{2}.$$

Thus

$$A\leqslant \sup\left\{\frac{c}{T_\varepsilon(m)-F_\mu^{-1}\left(\frac{3}{4}\right)},M\right\}.$$

The conclusion follows by [82] which state the fact that the optimal constant of Poncaré inequality C satisfies $C\leqslant 4A$.

□

5.9.3 An example of real probability measure which satisfies T_2 but not $\mathrm{W}_1\mathrm{I}$

Consider a real probability measure μ whose density with respect to Lebesgue measure is the following :

$$p(t)=\frac{e^{-t}}{2}\sum_{k=1}^{\infty}\mathbf{1}_{(k,k+1)}(t)e^{(k+1)-(k+1)^2},\forall t\geqslant 0, \tag{5.9.10}$$

and $\mu(\{k\})=e^{-k}-e^{1-(k+1)^2}$.
Now we define $\tilde{\mu}$, a modification of μ, such that $\tilde{\mu}$ is absolutely continuous with respect to Lebesgue measure, $\tilde{\mu}$ is of density :

$$\tilde{p}(t)=\frac{e^{-t}}{2}\sum_{k=1}^{\infty}\mathbf{1}_{(k+\frac{\varepsilon}{2^{k^2}},k+1)}(t)e^{(k+1)-(k+1)^2}+\sum_{k=1}^{\infty}\mathbf{1}_{(k,k+\frac{\varepsilon}{2^{k^2}})}(t)\frac{2^{k^2}}{\varepsilon}(e^{-k}-e^{1-(k+1)^2}),\forall t\geqslant 0.$$

It is easy to see that at neigbhorhood of ∞, $\tilde{\mu}(n,\infty)\sim e^{-n^2}$ and $\int_0^n\frac{1}{\tilde{\mu}(t)}dt\sim(1-1/e)e^{n^2}$. Then we will show that

$$T_x(0)\rightarrow\infty$$

when $x\rightarrow\infty$ and as a consequence, it will follow that the B^+ defined in Theorem 5.9.7 will be ∞ and so $\mathrm{W}_1\mathrm{I}$ will fail.

Assume that in the contrary $T_x(0)$ is bounded, then there exists $A>0$ such that for all $n>A$, $T_n(0)\leqslant A$. It follows that for all $n>A$ and $f_n=f_A$ on $[0,T_n(0)]$. We deduce that for all $n>A$:

$$\frac{1}{2}=\frac{1}{Z_n}\int_0^{T_n(0)}f_n^2(x)d\tilde{\mu}\leqslant\frac{1}{Z_n}\int_0^A f_n^2(x)d\tilde{\mu}=\frac{1}{Z_n}\int_0^A f_A^2(x)d\tilde{\mu}\rightarrow 0$$

Where in the last step we use the fact that $Z_n\rightarrow\infty$ as n goes to ∞. By contradiction, we can conclude that

$$T_x(0)\rightarrow\infty.$$

5.10 Recap.

Here we recap characterisations for functional inequalities in dimension one.

- The weak transport cost $\overline{\mathcal{T}}$ is additive for convex costs.
- The equivalence following holds

$$U(x+h)-U(x)\leqslant\sqrt{a+bh}\Leftrightarrow\overline{\mathrm{T}}_\theta\Leftrightarrow\overline{\mathrm{T}}_2^-\Leftrightarrow\overline{\mathrm{T}}_\beta^-\Leftrightarrow\mathrm{CmLS}\Leftrightarrow\mathrm{T}_\beta,$$

  where β is a convex function vanished in a neighborhood of 0 and U is the push forward of exponential measure with respect to μ.
- We also have the following equivalence

$$U(x+h)-U(x)\leqslant a+bh\Leftrightarrow\mathrm{P}\ \mathit{convex}\Leftrightarrow\overline{\mathrm{T}}_\alpha\Leftrightarrow\overline{\mathrm{T}}_\alpha^+\Leftrightarrow\overline{\mathrm{T}}_\alpha^-,$$

  where α is a quadratic-linear cost.

Bibliographie

[1] E. Abakumov, A. Beaulieu, F. Blanchard, M. Fradelizi, N. Gozlan, B. Host, T. Jeantheau, M. Kobylanski, G. Lecué, M. Martinez, M. Meyer, M. Mourgues, F. Portal, F. Ribaud, C. Roberto, P. Romon, J. Roth, P.-M. Samson, P. Vandekerkhove, and A. Youssfi. The logarithmic Sobolev constant of the lamplighter. *J. Math. Anal. Appl.*, 399(2) :576–585, 2013.

[2] L. Ambrosio, N. Gigli, and G. Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *Invent. Math.*, 195(2) :289–391, 2014.

[3] L. Ambrosio, N. Gigli, and G. Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *Invent. Math.*, 195(2) :289–391, 2014.

[4] L. Ambrosio, N. Gigli, and G. Savaré. Bakry-émery curvature-dimension condition and riemannian ricci curvature bounds. *Ann. Probab.*, 43(1) :339–404, 2015.

[5] C. Ané, S. Blachère, D. Chafai, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Sythèse*. Société Mathématique de France, Paris, 2000.

[6] D. Bakry and M. Émery. Diffusion hypercontractives. *Séminaire de probabilités, Lecture Notes in Math.*, XIX(1123) :177–206, 1983/84.

[7] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Springer, 2014.

[8] Z. M. Balogh, A. Engulatov, L. Hunziker, and O. E. Maasalo. Functional inequalities and Hamilton–Jacobi equations in geodesic spaces. *Potential Anal.*, 36(2) :317–337, 2012.

[9] F. Bauer, P. Horn, Y. Lin, G. Lippner, D. Mangoubi, and S.-T. Yau. Li-yau inequality on graphs. *ArXiv e-prints*, 2013.

[10] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. Ziegler. *Oriented matroids*, volume 46 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, second edition, 1999.

[11] S. Bobkov, I. Gentil, and M. Ledoux. Hypercontractivity of hamilton-jacobi equations. *J.Math.Pures Appl.*, 80(7) :669–696, 2001.

[12] S. Bobkov and F. Götze. Discrete isoperimetric and Poincaré-type inequalities. *Probab. Theory Related Fields*, 114(2) :245–277, 1999.

[13] S. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1) :1–28, 1999.

[14] S. Bobkov and C. Houdré. Weak dimension-free concentration of measure. *Bernouilli*, 6.

[15] S. Bobkov and M. Ledoux. Poincaré's inequalities and talagrand's concentration phenomenon for the exponential distribution. *Probab.Theory Relat. Fields*, (107) :383–400, 1997.

[16] S. Bobkov and M. Ledoux. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields*, 107(3) :383–400, 1997.

[17] S. Bobkov and P. Tetali. Modified logarithmic sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2) :289–335, 2006.

[18] V. Bogachev. *Measure theory.* Springer-Verlag, 2007.

[19] A.-I. Bonciocat and K.-Th. Sturm. Mass transportation and rough curvature bounds for discrete spaces. *Journal of Functional Analysis*, 256(9) :2944–2966, 2009.

[20] 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.

[21] S. Cambanis, G. Simons, and W. Stout. Inequalities for $Ek(X, Y)$ when the marginals are fixed. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 36(4) :285–294, 1976.

[22] P. Cattieux and A. Guillin. On quadratic transportation cost inequalities. *J.Maths Pure Appl.*, 86(4) :341–361, 2006.

[23] F. Chung, Y. Li, and S.-T. Yau. Harnack inequalities for graphs with non-negative ricci curvature. *J. Math. Anal. Appl.*, 415 :25–32, 2014.

[24] Dario Cordero-Erausquin, Robert J. McCann, and Michael Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.*, 146(2) :219–257, 2001.

[25] K. Costello, N. Gozlan, J. Melbourne, W. Perkins, C. Roberto, P-M. Samson, Y. Shu, and P. Tetali. Discrete ricci curvature and functional inequalities, on graphs. In preparation, 2016.

[26] G. Dall'Aglio. Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Ann. Scuoloa Norm. Sup. Pisa (3)*, 10 :35–74, 1956.

[27] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3) :695–750, 1996.

[28] 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.

[29] R. L. Dobrushin. Definition of a system of random variables by means of conditional distributions. *(Russian) Teor. Verojatnost. i Primenen.*, 15 :469–497, 1970.

[30] R.M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambrige, 2002.

[31] R. Eldan, J. Lee, and J. Lehec. personal communication.

[32] M. Erbar and J. Maas. Ricci curvature of finite markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.*, 206 :997–1038, 2012.

[33] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[34] A. Faggionato, D. Gabrielli, M. Mariani, and Y. Shu. Large deviations for the invariant distribution of markov chains with exponen- tially small jump rates. *In preparation.*

[35] M. Fathi and Y. Shu. Curvature and transport inequalities for markov chains in discrete spaces. *ArXiv e-prints*, 2015.

[36] F.Barthe and C.Roberto. Soblev inequalities for probability measures on the real line. september 2003.

[37] N. Feldheim, A. Marsiglietti, P. Nayar, and J. Wang. A note on the convex infimum convolution inequality. *ArXiv e-prints*, May 2015.

[38] M. Fréchet. Sur les tableaux dont les marges et des bornes sont données. *Rev. Inst. Internat. Statist.*, 28 :10–32, 1960.

[39] M. Freidlin and A. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, 2012.

[40] N. Gigli and M. Ledoux. From log Sobolev to Talagrand : a quick proof. *Discrete Contin. Dyn. Syst.*, 33(5) :1927–1935, 2013.

[41] E. Giné, G.R. Grimmett, and L. Saloff-Coste. *Lectures on Probability Theory and Statistics*. Springer, 1997.

[42] N. Gozlan. A characterization of dimension free concentration in terms of transportation inequalities. *Ann. Probab.*, 37(6) :2480–2498, 2009.

[43] N. Gozlan. Transport-entropy inequalities on the line. *Electron. J. Probab.*, 17 :no. 49, 18, 2012.

[44] N. Gozlan and C. Léonard. Transport inequalities. a survey. *Markov Processes Relat. Fields*, (16) :635–736, 2010.

[45] N. Gozlan, C. Roberto, and P.-M. Samson. A new characterization of Talagrand's transport-entropy inequalities and applications. *Ann. Probab.*, 39(3) :857–880, 2011.

[46] N. Gozlan, C. Roberto, and P.-M. Samson. Characterization of Talagrand's transport-entropy inequalities in metric spaces. *Ann. Probab.*, 41(5) :3112–3139, 2013.

[47] N. Gozlan, C. Roberto, and P.-M. Samson. Hamilton Jacobi equations on metric spaces and transport entropy inequalities. *Rev. Mat. Iberoam.*, 30(1) :133–163, 2014.

[48] N. Gozlan, C. Roberto, and P.-M. Samson. Hamilton Jacobi equations on metric spaces and transport entropy inequalities. 2014.

[49] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali. Characterization of a class of weak transport-entropy inequalities on the line. 2015.

[50] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. 2014.

[51] N. Gozlan, C. Roberto, and P.M. Samson. From concentration to logarithmic sobolev and poincaré inequalities. *J. Funct. Anal.*, 260(5) :1491–1522, 2011.

[52] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4) :1061–1083, 1975.

[53] A. Guillin, C. Léonard, F.Y. Wang, and L.M. Wu. Transportation-information inequalities for markov processes : relations with other functional inequalities. Febuary 2009.

[54] A. Guillin, C. Léonard, L. Wu, and N. Yao. Transportation-information inequalities for markov processes. *PTRF*, 144 :669–695, 2009.

[55] A. Guionnet and B. Zegarlinski. Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Math.*, pages 1–134. Springer, Berlin, 2003.

[56] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Grundlehren Text Editions. Springer-Verlag, Berlin, 2001. Abridged version of ıt Convex analysis and minimization algorithms. I [Springer, Berlin, 1993 ; MR1261420 (95m :90001)] and ıt II [ibid. ; MR1295240 (95m :90002)].

[57] F. Hirsch, C. Profeta, B. Roynette, and M. Yor. *Peacocks and associated martingales, with explicit constructions*, volume 3 of *Bocconi & Springer Series*. Springer, Milan ; Bocconi University Press, Milan, 2011.

[58] W. Hoeffding. Maßstabinvariante korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5 :181–233, 1940.

[59] P. Horn, Y. Lin, S. Liu, and S-T. Yau. Poincaré inequality and guassian heat kernel estimate for nonnegative curvature graphs. *ArXiv e-prints*, 2014.

[60] J. Jost and S. Liu. Ollivier's ricci curvature, local clustering and curvature dimension inequalities on graphs. *Discrete and Computational Geometry*, 51(2) :300–322, 2014.

[61] A. Joulin. Poisson-type deviation inequalities for curved continuous time markov chains. *Bernoulli*, 13(3) :782–798, 2007.

[62] A. Joulin and Y. Ollivier. Curvature, concentration, and error estimates for markov chain monte carlo. *Ann. Probab.*, 38 :2418–2442, 2010.

[63] B. Klartag, G. Kozma, P. Ralli, and P. Tetali. Discrete curvature and abelian groups. *ArXiv e-prints*, 2015.

[64] M. Ledoux. The concentration of measure phenomenon. *Mathematical Surveys and Monographs*, 2001.

[65] T.-Y. Lee and H.-T. Yau. Logarithmic Sobolev inequality for some models of random walks. *Ann. Probab.*, 26(4) :1855–1873, 1998.

[66] C. Léonard. On the convexity of the entropy along entropic interpolations. *ArXiv e-prints*, 2013.

[67] Y. Lin and S.-T. Yau. Ricci curvature and eigenvalue estimate on locally. *Math. Res. Lett.*, 17(2) :343–356, 2010.

[68] J. Lott and C. Villani. Hamilton-Jacobi semigroup on length spaces and applications. *J. Math. Pures Appl. (9)*, 88(3) :219–229, 2007.

[69] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.*, 169(3) :903–991, 2009.

[70] J. Maas. Gradient flows of the entropy for finite markov chains. *J. Funct. Anal.*, 261, 2011.

[71] A.-W. Marshall, I. Olkin, and B.-C. Arnold. *Inequalities : theory of majorization and its applications.* Springer Series in Statistics. Springer, New York, second edition, 2011.

[72] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics (Saint-Flour, 1997)*, volume 1717 of *Lecture Notes in Math.*, pages 93–191. Springer, Berlin, 1999.

[73] F. Martinelli. Relaxation times of Markov chains in statistical mechanics and combinatorial structures. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 175–262. Springer, Berlin, 2004.

[74] K. Marton. A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory*, 32(3) :445–446, 1986.

[75] K. Marton. Bounding $\overline{d}$-distance by informational divergence : a method to prove measure concentration. *Ann. Probab.*, 24(2) :857–866, 1996.

[76] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3) :556–571, 1996.

[77] K. Marton. An inequality for relative entropy and logarithmic Sobolev inequalities in Euclidean spaces. *J. Funct. Anal.*, 264(1) :34–61, 2013.

[78] B. Maurey. Some deviation inequalities. *Geom. Funct. Anal.*, 1(2) :188–197, 1991.

[79] L. Miclo. An example of application of discrete hardy's inequality. *Markov Process. Related Fields*, 5(3) :319–330, 1999.

[80] A. Mielke. Geodesic convexity of the relative entropy in reversible markov chains. *Calc. Var. Part. Diff. Equ.*, 48(1) :1–31, 2013.

[81] E. Milman. Isoperimetric and concentration inequalities : equivalence under curvature lower bound. *Duke Math J.*, 154(2) :207–239, 2010.

[82] B. Muckenhoupt. Hardy's inequality with weights. *Studia Math.*, 44 :31–38, 1972.

[83] F. Munch. Remarks on curvature dimension conditions on graphs. *ArXiv e-prints*, 2015.

[84] R. I. Oliveira. On the convergence to equilibrium of kac's random walk on matrices. *Ann. Appl. Probab.*, 19(3) :1200–1231, 2009.

[85] Y. Ollivier. Ricci curvature of markov chains on metric spaces. *J. Funct. Anal.*, 256(3) :810–864, 2009.

[86] Y. Ollivier. A survey of ricci curvature for metric spaces and markov chains, in probabilistic approach to geometry. *Adv. Stud. Pure Math.*, 57 :343–381, 2010.

[87] 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.

[88] R. Rado. An inequality. *J. London Math. Soc.*, 27 :1–6, 1952.

[89] C. Roberto. Inégalité de hardy et de sobolev logarithmique. *Phd Thesis*, 2001.

[90] M.D. Sammer. Aspects of mass transportation in discrete concentration inequalities. *PhD thesis*, 2005.

[91] P.-M. Samson. Concentration inequalities for convex functions on product spaces. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 33–52. Birkhäuser, Basel, 2003.

[92] M. Schmuckenschlager. Curvature of nonlocal markov generators. *Convex geometric analysis*, 34 :189–197, 1999.

[93] Y. Shu. Hamilton-jacobi equations on graphs and applications. Hal Preprint, 2014.

[94] Y. Shu and M. Strzelecki. Characterization of a class of convex modified log-sobolev inequalities on the line. 2016.

[95] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36 :423–439, 1965.

[96] K.-T. Sturm. On the geometry of metric measure spaces. i, ii. *Acta Math.*, 196(1) :65–177, 2006.

[97] K.-T. Sturm and M. Von Renesse. Transport inequalities, gradient estimates, entropy and ricci curvature. *Communications on Pure and Applied Mathematics*, 58.

[98] M. Talagrand. Transportations cost for gaussian and other product measures. *Geom. Funct. Anal*, 6(3) :587–600, 1996.

[99] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[100] N-Y. Wang. Concentration inequalities for Gibbs sampling under d_{l_2}-metric. *Electron. Commun. Probab.*, 19 :no. 63, 11, 2014.