



HAL
open science

Quelques applications du transport optimal en analyse et en probabilités

Max Fathi

► **To cite this version:**

Max Fathi. Quelques applications du transport optimal en analyse et en probabilités. Probabilités [math.PR]. Université Paul Sabatier (Toulouse 3), 2019. tel-02105311

HAL Id: tel-02105311

<https://hal.science/tel-02105311>

Submitted on 20 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PAUL SABATIER TOULOUSE 3



Manuscrit
présenté pour l'obtention de
l'Habilitation à Diriger des Recherches

Max Fathi



Quelques applications du transport optimal en analyse et en probabilités

Soutenue le 04/04/2019,

après avis des rapporteurs

Pr. Eric A. Carlen	Rutgers University
Pr. Dario Cordero-Erausquin	Université Pierre et Marie Curie
Pr. Alessio Figalli	ETH Zürich

devant le jury formé de

Pr. Eric A. Carlen	Rutgers University
Pr. Dario Cordero-Erausquin	Université Pierre et Marie Curie
Pr. Laure Coutin	Université de Toulouse
Pr. Alessio Figalli	ETH Zürich
Pr. Arnaud Guillin	Université Clermont Auvergne
Pr. Michel Ledoux	Université de Toulouse

-What do you want to be when you grow up?
-An honest, brave, compassionate human being.
-No... I mean, how do you want to sell your labor?
Corey Mohler, *Existential Comics*

*This was in a lecture room, and I invited him to consider the proposition:
'There is no hippopotamus in this room at present'.
When he refused to believe this, I looked under all the desks without finding one;
but he remained unconvinced.*
Bertrand Russell on Ludwig Wittgenstein

Vos exergues sont absurdes.
Catherine K., *Principia Impressoricae*

Résumé

Ce mémoire présente la majeure partie de mes travaux effectués depuis la fin de ma thèse. Mon travail porte principalement sur des questions mathématiques à l'interface entre analyse et probabilités. Les résultats présentés ici portent principalement sur trois thèmes : la méthode de Stein et ses applications à l'estimation de distances entre mesures de probabilité, la stabilité des inégalités fonctionnelles, et les notions de courbure de Ricci pour les espaces discrets et leurs applications. Chacun de ces thèmes est l'objet d'un des chapitres de ce mémoire. Dans mes travaux sur chacun de ces thèmes, on retrouve des idées et méthodes issues de deux sujets à l'interface entre analyse et probabilités : le transport optimal de mesures et les inégalités fonctionnelles.

Une question qui revient fréquemment dans mes travaux est celle d'estimer quantitativement la proximité entre différentes mesures de probabilités. Ce type de problème apparaît sous diverses formes dans de nombreux domaines des mathématiques et des sciences, par exemple pour évaluer la qualité d'un algorithme numérique ou d'un estimateur statistique, la vitesse de convergence d'un processus vers un état d'équilibre, ou encore pour étudier la rigidité d'un problème variationnel. L'accent sera donc sur les estimées quantitatives que l'on obtient, et leur dépendance en les différents paramètres des problèmes considérés.

Summary

This document presents a large part of my work since the end of my PhD thesis. I have mostly worked on mathematical problems at the boundary between analysis and probability. The results described here mostly deal with the following three topics: Stein's method and its uses in estimating distances between probability measures, stability of functional inequalities, and notions of Ricci curvature bounds for discrete spaces and their applications. Each of these topics will be the focus of a chapter of this manuscript. In all of them we shall see ideas and techniques from two fields at the interface of analysis and probability: optimal transportation of measures and functional inequalities.

A question that frequently appear in my work is to quantitatively estimate how close various probability measures are. This type of problem arises in various forms in many areas of mathematics and natural sciences, for example in order to evaluate the efficiency of a numerical scheme or a statistical estimator, the rate of convergence of some process towards a stationary state, or to study rigidity in variational problems. There will therefore be an emphasis on the quantitative bounds we can obtain, and the way they depend on the various parameters of the problems considered.

Contents

1	Introduction	12
1.1	Publication list	12
1.2	Content of the manuscript	14
1.3	Notations	15
2	Background	16
2.1	Optimal transport	16
2.2	Functional inequalities	18
2.3	Curvature	21
3	Stein’s method: an analytic viewpoint	24
3.1	Stein kernels	25
3.2	Rate of convergence in the central limit theorem	28
3.3	Free probability	29
3.4	Perspectives	31
3.4.1	Dependent random variables	31
3.4.2	Stein kernels in non-Gaussian settings	31
3.4.3	CLT for random eigenfunctions in positive curvature	32
4	Refinements of functional inequalities	34
4.1	Gaussian functional inequalities	35
4.2	Spectral inequalities	37
4.3	Perspectives	39
4.3.1	Stability for variational problems	39
4.3.2	Estimates on optimal transport maps	40
5	Discrete Ricci curvature	41
5.1	Entropic curvature	41
5.2	Discrete transport-information inequalities	44
5.3	Perspectives	45
5.3.1	The Peres-Tetali conjecture	45
5.3.2	Geometry and censoring for Markov chains on discrete spaces	46
5.3.3	Long-time behavior for mean-field equations on discrete spaces	46
5.3.4	Rigidity of curvature bounds in the discrete setting	47

Remerciements

Je tiens en premier lieu à remercier Michel Ledoux, pour son bon accueil, ses conseils avisés et toutes les discussions que nous avons eues depuis mon arrivée à Toulouse. Qu'il ait accepté de parrainer cette habilitation est un honneur et un plaisir pour moi.

Je voudrais exprimer ma reconnaissance à Eric Carlen, Dario Cordero-Erausquin et Alessio Figalli pour avoir bien voulu donner leur avis sur ce texte. Leurs travaux sont un modèle et une source d'inspiration pour moi. Merci aussi à Laure Coutin et Arnaud Guillin d'avoir accepté de faire partie du jury.

Ce document résume des travaux faits depuis la fin de ma thèse, sur lesquels j'ai eu la chance et le plaisir de collaborer avec Guillaume Cébron, Maria Colombo, Thomas Courtade, Matthias Erbar, Emanuel Indrei, Vaios Laschos, Michel Ledoux, Jan Maas, Tobias Mai, Brent Nelson, Ashwin Pananjady, André Schlichting, Marielle Simon, Yan Shu et Gabriel Stoltz. Au-delà de ces collaborations, ce travail est le fruit de nombreuses discussions, à Toulouse et en conférence, d'échanges d'emails, de séminaires de collègues... Je ne saurais tout énumérer, mais je voudrais remercier tout particulièrement Cédric Villani pour avoir guidé mes premiers pas en recherche, Craig Evans et Fraydoun Rezakhanlou pour m'avoir encadré lors de mon séjour postdoctoral à Berkeley, ainsi que Scott Armstrong, Franck Barthe, Jérôme Bertrand, Thierry Bodineau, Guillaume Carlier, Patrick Cattiaux, Djalil Chafaï, Jean Dolbeault, Yuxin Ge, Ivan Gentil, Nathael Gozlan, Tyler Helmuth, Nicolas Juillet, Joseph Lehec, Tony Lelièvre, Bastien Mallein, Georg Menz, Laurent Miclo, Pierre Monmarché, Stefano Olla, Julien Reygner, Silvain Rideau, Dmitri Shlyakhtenko et Prasad Tetali pour les nombreuses discussions.

Je remercie les collègues de l'IMT, un laboratoire où non seulement on travaille bien, mais aussi où il fait bon vivre. Je voudrais remercier en particulier Réda, François, Guillaume, Cécilia, Xavier, Nicolas, Radu, Clément, Laurent, Philippe, Jérôme, Lorick et Manon pour les pizzas, restos, bières et parties de coinche. Merci aussi à Françoise Michel, Delphine Dalla-Riva, Alexandra Moran et Sabrina Belmelih pour leur aide toujours efficace.

Je remercie enfin mes amis et ma famille, pour leur soutien constant.

Chapter 1

Introduction

This text describes some of my post-PhD research work, since 2014. I have mostly worked on the five following topics:

1. Convergence of interacting particle systems to some scaling limit;
2. Rigorous analysis of error bounds for stochastic numerical schemes;
3. Stein's method and error bounds for approximation of probability measures;
4. Functional inequalities in probability theory, and their stability;
5. Notions of Ricci curvature for discrete spaces and their applications.

The distinction made here between these topics is somewhat arbitrary, and many of my works mix several of them. The first two topics were the main focus of my PhD thesis. Although I have continued working on both of them, I will not discuss here my post-PhD work on them (items 11, 12 and 13 in the list below), and focus on the last three, each of which will be the main focus of one chapter of this document.

A common feature in most of my work is the use of tools and ideas from analysis (calculus of variations, PDEs, regularity estimates) to derive *quantitative* bounds in problems from the field of probability, in a broad sense, such as error estimates in random approximation problems, rates of convergence to equilibrium for stochastic dynamics, rigorous error bounds from a scaling limits for discrete models in statistical physics, non-asymptotic confidence intervals for statistical estimators...

1.1 Publication list

Articles described in this manuscript

1. T.A. Courtade and M. Fathi, Stability of the Bakry-Émery theorem on \mathbb{R}^n . *Arxiv preprint* 2018.
2. M. Fathi, A sharp symmetrized form of Talagrand's transport-entropy inequality for the Gaussian measure. *to appear in Electronic Communications in Probability*, 2018.
3. M. Fathi, Stein kernels and moment maps. *to appear in Annals of Probability*, 2018.

4. T.A. Courtade, M. Fathi and A. Pananjady, Existence of Stein Kernels under a Spectral Gap, and Discrepancy Bounds *to appear in Annales de l'IHP: Probabilités et Statistiques*, 2018
5. T.A. Courtade, M. Fathi and A. Pananjady, Wasserstein stability of the entropy power inequality for log-concave densities *IEEE Transactions on Information Theory* 64, 8, 5691-5703, 2018. (a short version was also published in the proceedings of the 2017 *International Symposium on Information Theory*).
6. M. Erbar and M. Fathi, Poincaré, modified logarithmic Sobolev and isoperimetric inequalities for Markov chains with non-negative Ricci curvature. *Journal of Functional Analysis* 274 (2018), no. 11, 3056-3089
7. M. Fathi and B. Nelson, Free Stein kernels and an improvement of the free logarithmic Sobolev inequality, *Advances in Mathematics* 317 (2017) pp. 193-223.
8. M. Fathi and Y. Shu, Curvature and transport inequalities for Markov chains in discrete spaces, *Bernoulli*, 24(1), 2018, 672-698
9. M. Fathi, E. Indrei and M. Ledoux. Quantitative logarithmic Sobolev inequalities and stability estimates. *Discrete and Continuous Dynamical Systems* 36 (2016), no. 12, 6835-6853.
10. M. Fathi and J. Maas, Entropic Ricci curvature bounds for discrete interacting systems, *Annals of Applied Probability* 26 (2016), no. 3, 1774-1806.

Other articles

11. M. Erbar, M. Fathi, V. Laschos and A. Schlichting, Gradient flow structure for McKean-Vlasov equations on discrete spaces, *Discrete and Continuous Dynamical Systems* 36 (2016), no. 12, 6799-6833.
12. M. Fathi and M. Simon, The gradient flow approach to hydrodynamic limits for the simple exclusion process. Particle Systems and Partial Differential Equations III, ed. A. J. Soares and P. Goncalves, Springer Proceedings in Mathematics & Statistics.
13. M. Fathi and G. Stoltz, Improving dynamical properties of stabilized discretizations of overdamped Langevin dynamics. *Numerische Mathematike* 136 (2017), no. 2, 545-602.
14. M. Fathi. A gradient flow approach to large deviations for diffusion processes. *Journal de Mathématiques Pures et Appliquées* (9) 106 (2016) , no. 5, 957-993.
15. M.H. Duong and M. Fathi. The two-scale approach for non-reversible dynamics. *Markov Processes and Related Fields* Volume 22, Issue 1, 1-36, 2016.
16. M. Fathi. Modified logarithmic Sobolev inequalities for canonical ensembles. *ESAIM: Probab. Stat.* Vol. 19 (2015) 544–559
17. M. Fathi and G. Menz. Hydrodynamic limit for conservative spin systems with super-quadratic, partially inhomogeneous single-site potential. *unpublished manuscript*, 2014
18. M. Fathi, A.-A. Homman and G. Stoltz. Error analysis of the transport properties of Metropolized schemes. Proceeding of the 2013 CEMRACS summer school, *ESAIM: Proceedings and Surveys*, vol. 48, p.341-363, 2015.

19. M. Fathi and N. Frikha. Transport-Entropy inequalities and deviation estimates for stochastic approximation schemes. *Electronic Journal of Probability*, Volume 18 (2013), no. 67, 1-36.
20. M. Fathi. A two-scale approach to the hydrodynamic limit. Part II: local Gibbs behavior. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 10(2): 625-651, 2013.

In this list, items 14 to 20 correspond to my PhD thesis.

1.2 Content of the manuscript

We now briefly describe the content of the rest of this manuscript.

Chapter 2 is a (minimal) introduction to several topics that play an important role in my work: optimal transport, information-theoretic functional inequalities, and Ricci curvature lower bounds. It mostly contains background material, the basic definitions and some important theorems on these topics that are either used in my work, or motivate some of the questions I study.

Chapter 3 is about Stein kernels and their applications to quantitative central limit theorems. It describes results from the publications 3, 4 and 6 in the above list. The main topic is a particular way of implementing a strategy for estimating distances between probability measures, known as Stein's method. It consists in characterizing probability measures via integration by parts formulas, and using these to compare the measures. For example, the standard Gaussian measure on \mathbb{R}^d can be characterized as the only probability measure such that

$$\int x \cdot f(x) d\gamma = \int \operatorname{div}(f)(x) d\gamma$$

for all nice smooth vector-valued test functions f . It then turns out that if another probability measure satisfies this integration by parts inequality up to a small error, for some large enough class of test functions, then it is close to being Gaussian, and this statement can be made explicitly quantitative. My main contribution has been two constructions of Stein kernels, a class of objects that can be used to define approximate integration by parts formulas for probability measures. As an application, I derived various estimates on the rate of convergence in the classical central limit theorem, with a particular focus on the dependence in the dimension. I have also worked on the counterpart of these Stein kernels in free probability, with applications to free functional inequalities and rates of convergence in the free central limit theorem.

Chapter 4 is about refinements and stability of functional inequalities, describing results from the (pre)publications 1, 2, 4, 5 and 9. A typical functional inequality compares two integral quantities. Classical examples include Sobolev inequalities and isoperimetric inequalities. The main example of interest in my work is the logarithmic Sobolev inequality, which, given a reference probability measure μ , states (if it holds) that

$$\operatorname{Ent}_\mu(f) \leq C_{LSI}(\mu) \int \frac{|\nabla f|^2}{2f} d\mu$$

for all positive smooth function f . While in general the sharp constant $C_{LSI}(\mu)$ is hard to compute, for the standard Gaussian measure we know that it is equal to 1 in all dimensions. Moreover, equality holds if and only if f is of the form $C \exp(\lambda \cdot x)$ for some $\lambda \in \mathbb{R}^d$. We can then try to see if further

improvements are possible, for example by extracting an extra remainder term measuring how far f is from being an exponential, or by improving the constant over some restricted class of functions that exclude exponentials. Similar questions can be raised for many other inequalities, such as isoperimetric inequalities, spectral inequalities, etc... Beyond the Gaussian LSI, I have worked on transport-entropy inequalities, the Shannon-Stam entropy power inequality and spectral inequalities.

Chapter 5 is about Ricci curvature bounds in discrete spaces, describing results from articles 6, 8 and 10. Ricci curvature bounds play an important role in differential geometry, and over the past few years there have been several attempts at defining an analogous notion in a discrete setting, motivated by the development of tools for proving functional inequalities on graphs. I have worked on several of these notions, although mostly on the so-called entropic Ricci curvature, and proved several discrete counterparts to classical results in Riemannian geometry, such as a discrete Buser theorem relating spectral gap and Cheeger isoperimetric constant in nonnegative curvature, as well as a relation between diameter and (modified) logarithmic Sobolev constant for such graphs. I have also worked on methods for computing curvature lower bounds for concrete examples.

1.3 Notations

- $\gamma_{d,p}$ stands for the d -dimensional Gaussian probability measure with barycenter at $p \in \mathbb{R}^d$. If the subscript is omitted, it stands for the standard centered Gaussian measure.
- $Ent_\mu(f)$ stands for the relative entropy of the nonnegative function f with respect to the measure μ , given by $\int f \log f d\mu - (\int f d\mu) \log (\int f d\mu)$. Similarly, if ν is another probability measure, $Ent_\mu(\nu)$ is the entropy of the density of ν with respect to μ (and takes value $+\infty$ if ν is not absolutely continuous w.r.t. μ).
- $I_\mu(f)$ stands for the relative Fisher information of the nonnegative function f with respect to the measure μ , given by $\int \frac{|\nabla f|^2}{2f} d\mu$. Similarly, $I_\mu(\nu)$ is the relative Fisher information of the density of ν with respect to μ (and takes value $+\infty$ if ν is not absolutely continuous w.r.t. μ).
- id stands for the identity map.
- Id stands for the identity matrix (with dimension specified by the context).
- $\mu_\# F$ is the pushforward of the measure μ by the function F .
- $\Pi(\mu, \nu)$ is the set of all couplings between the two probability measures μ and ν .
- C is a constant that may change from line to line, or even within a line.

Chapter 2

Background

2.1 Optimal transport

The aim of this section is to provide a brief introduction to optimal transport, with a focus on the aspects most relevant to my own work. For a more extensive presentation of the field, we refer to the monographs [165, 166].

The optimal transport problem was first proposed by Monge in [136]. Its relaxed form, proposed by Kantorovitch, takes the following form: given two probability measures (which represent the location of the mass to be transported, and the target allocation) and a cost function c (representing the cost of moving some mass from a point x to some point y), we look for a coupling π of μ and ν that minimizes the quantity

$$\int c(x, y) d\pi.$$

Equivalently, we are looking for the distribution of a pair of random variables (X, Y) such that X has law μ and Y law ν , that minimizes among all such pairs the quantity $\mathbb{E}[c(X, Y)]$.

Over the past 30 years, there has been a huge amount of progress, both on the theory of optimal transport (structure, existence, uniqueness, behavior under additional constraints...) and its applications to various areas of mathematics (PDEs, probability, statistics, geometry, operator algebras...) and other natural sciences (statistical physics, engineering, economics...). The purpose of this section is to review a few highlights of the theory, with a particular focus on the case where the cost is the squared distance on some metric space, which is the one most relevant to my work.

The first structural result we would like to highlight is the Brenier theorem [27] (later improved by McCann [129]), which states that on the Euclidean space, and when considering the quadratic cost $c(x, y) = |x - y|^2$, if one of the two measures, say μ , has a density (or more generally, does not give positive mass to some lower-dimensional subset), the optimal transport plan has a very specific structure: it is supported on the graph of the gradient of some convex function, i.e. it takes the form

$$\pi(dx, dy) = \mu_{\#}(id, \nabla\varphi)$$

with $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ convex. In particular, this result implies the existence of a map sending μ onto ν with this specific structure. In some of the results and techniques we shall later discuss, the optimality of the map will in some sense play no direct role, we will only exploit the fact that there exists a map with such a specific geometry.

As a consequence, if we assume that both measures have a density and write the change-of-variable formula for the optimal transport map $\nabla\varphi$ sending $\mu = \rho_0(x)dx$ onto $\nu = \rho_1(x)dx$, we see that it solves (in a weak sense) the following PDE:

$$\rho_0(x) = \rho_1(\nabla\varphi(x)) \det \text{Hess } \varphi(x) \quad (2.1)$$

This type of PDE, involving the determinant of the Hessian of the unknown function, is known as a Monge-Ampère PDE. McCann [130] showed that the optimal transport map solves this PDE almost everywhere, if the Hessian is understood in the sense of Alexandrov.

If we consider a cost function of the form $c(x, y) = d(x, y)^p$, it turns out that the optimal transport cost

$$W_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu, \nu)} \int d(x, y)^p d\pi \right)^{1/p}$$

is a distance on the (sub)space of probability measures with finite p -moment \mathcal{P}_p . These distances are known as Wasserstein or Monge-Kantorovitch distances, and control the weak- $*$ convergence of probability measures (plus convergence of the p -moment). A particular feature of this family of distances is that they trivially provide an isometric embedding of the base space into the space of probability measures. We will later see how the geometry of the space of probability measures endowed with a Wasserstein distance remembers some features of the geometry of the base space.

We now move on to a dynamic viewpoint on optimal transport, based on work of Benamou and Brenier [14], and further refined by Otto [99, 147]. The basic idea is that instead of minimizing over all possible way of allocating the mass of the source to its target, we can minimize over all ways of continuously moving the mass from the source to the target.

$$W_2(\mu_0, \mu_1)^2 = \inf \int_0^1 \int |v_t(x)|^2 d\mu_t dt$$

where the infimum is taken over all pairs (μ_t, v_t) of curves of probability measures (with endpoints μ_0 and μ_1) and of vector fields satisfying the relation

$$\dot{\mu}_t + \text{div}(v_t \mu_t) = 0. \quad (2.2)$$

If we analyze the curve (μ_t, v_t) that attains the minimum, we can show it is actually induced by the optimal coupling in the original formulation of W_2 , via the relations $\mu_t = ((1-t)id + t\nabla\varphi)\#\mu_0$ and $v_t(x) = \nabla\varphi(((1-t)id + t\nabla\varphi)^{-1}(x))$, i.e. the flow follows the geodesic between x and $\nabla\varphi(x)$. This point of view extends to the Riemannian setting.

We can view the Benamou-Brenier formula as allowing to endow the space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ with a formal (infinite-dimensional) Riemannian structure. This point of view has been developed by Otto [147], following earlier work by Jordan, Kinderlehrer and Otto [99]. The relation (2.2) allows to identify the tangent space at a measure μ with vector fields, and the quantity $\int |v(x)|^2 d\mu$ is then an L^2 norm on this tangent space. The Benamou-Brenier formula can then be viewed as stating that the Wasserstein distance between measures is obtained by minimizing the action over all possible curves linking the two measures, which matches with the usual definition of Riemannian distances. Curves of probability measures achieving the minimum in the Benamou-Brenier formula are then geodesics in the space of probability measures.

Finally, I would like to mention Kantorovitch's dual formulation of the optimal transport problem. Under some assumptions on the cost function c , we have

$$\inf_{\pi \in \Pi(\mu, \nu)} \int c(x, y) d\pi(x, y) = \sup_{\phi(x) + \psi(y) \leq c(x, y)} \int \phi(x) d\mu(x) + \int \psi(y) d\nu(y).$$

In the case of the L^1 Wasserstein distance, the above dual formula is known as the Kantorovitch-Rubinstein formula, and can be written as

$$W_1(\mu, \nu) = \sup_{f \text{ 1-lip}} \int f d\mu - \int f d\nu \quad (2.3)$$

where the property of being Lipschitz is understood relative to the distance used to define the cost. It will play an important role in several of the results we shall discuss in this manuscript.

2.2 Functional inequalities

Functional inequalities are inequalities comparing integral quantities over classes of functions or measures. Emblematic examples include Sobolev inequalities in PDE theory and geometric functional analysis inequalities, such as the Brunn-Minkowski inequality or isoperimetric inequalities. They can be used to *quantify* various phenomenon, by using them to compare certain distances or other quantities of interest.

I will start by defining the three main type of functional inequalities I am interested in:

Definition 1. A probability measure μ on \mathbb{R}^d is said to satisfy:

- a Poincaré inequality with constant ρ (denoted by $PI(\rho)$) if for any smooth test function

$$\text{Var}_\mu(f) \leq \frac{1}{\rho} \int |\nabla f|^2 d\mu;$$

- a logarithmic Sobolev inequality with constant ρ (denoted by $LSI(\rho)$) if for any smooth nonnegative test function

$$\text{Ent}_\mu(f) \leq \frac{1}{2\rho} \int \frac{|\nabla f|^2}{f} d\mu;$$

- a transport-entropy, or Talagrand inequality, of order p and constant ρ (denoted by $\mathcal{T}_p(\rho)$) if for any probability measure ν

$$W_p(\nu, \mu)^2 \leq \frac{2}{\rho} \text{Ent}_\mu(\nu).$$

For the Talagrand inequality, we shall mostly be interested in the cases $p = 1$ and $p = 2$. A basic example of probability measure satisfying these inequality is the standard Gaussian measure, for which the sharp constant is $\rho = 1$ in any dimension. All three of these inequalities compare two kinds of distances to the reference measure.

We have already seen Wasserstein distances in the previous section. Before commenting on the functional inequalities we just defined, I would like to briefly explain why we are interested in the other quantities playing a role in those inequalities.

The relative entropy $\text{Ent}_\mu(\nu) = \int f \log f d\mu$ of a probability measure $\nu = f\mu$ with respect to a reference probability measure μ plays an important role as a measure of information in several areas, most notably kinetic theory and information theory. It was first introduced by Boltzmann in his work on the Boltzmann equation in kinetic theory, as a monotone quantity along the flow, in accordance with the second principle of thermodynamics, and hence a good way of quantifying the trend to equilibrium. Note that the way we define entropy is as the negative of the physical entropy, so it will end up being decreasing rather than increasing. The Pinsker inequality asserts it dominates the squared total variation distance between μ and ν , so that it indeed controls convergence in distribution. It then appeared in

information theory, where Shannon [157] used it as a measure of the rate at which a stochastic source of data produces information. Informally, a source with low (mathematical) entropy makes it harder for the receiver to identify data from the signal he receives. Later on, it appeared in the theory of large deviations, as a sharp way of estimating the exponential decay rate of the probability of observing an abnormal behavior for a large collection of i.i.d. random variables sampled from the reference measure. We refer to the nice survey by Chafaï [39] for more about the history of entropy and its applications to many fields of mathematics and natural sciences.

Historically, Fisher information was introduced in statistics as a way of quantifying the information an observable contains about an unknown parameter on which its distribution depends. It plays an important role in quantifying the risk of using certain estimators to infer the value of the target parameter. The two main applications that motivate our interest in it come however from slightly different problems: it arises as the rate of dissipation of entropy for certain stochastic dynamics, and as the large deviations rate function for the occupation measure of Markov processes.

The LSI was first introduced by Gross [94] in 1975, in relation with the hypercontractivity of certain Markovian semigroups. It can be viewed as a type of Sobolev inequality, where the embedding of the weighted Sobolev space $H^1(\mu)$ is not in an L^p space but in an Orlicz-type space $L^2 \log L$ (i.e. we gain a logarithmic factor in integrability, rather than improve the exponent).

Transport-entropy inequalities were introduced by Talagrand [164] and Marton [128] to study concentration of measure. We also refer to the survey [91]. These functional inequalities and their variants have found applications to many concrete problems, among which we can mention:

- Long-time behavior of stochastic processes and PDEs [8]
- Statistics: confidence intervals, concentration of measure [25]
- Statistical physics: interacting particle or spin systems [171], hydrodynamic limits [95, 104], computational statistical physics [118]
- Computer science: analyzing algorithms, optimization problems, information theory [150]

A striking feature of LSI and Poincaré inequalities, as well as \mathcal{T}_2 inequalities, is that they *tensorize*: a product measure whose factors satisfy one of these inequalities also does. In particular, a product measure $\mu^{\otimes n}$ satisfies these inequalities with the same constant as μ . This observation allows us to use functional inequalities to derive *dimension-free* quantitative estimates. These are particularly useful when analyzing the behavior of high-dimensional systems, such as interacting particle systems in statistical physics (where the dimension is typically of the order of the Avogadro number, i.e. around 10^{24}), certain numerical schemes such as spatial discretizations of PDE (where the effective dimension is of order ϵ^{-d} , where ϵ is the size of the mesh, which typically has to be taken quite small) or various models in statistics. See also [138] about the role of large numbers in mathematics.

The Poincaré and logarithmic Sobolev inequalities are naturally connected with long-time behavior of Markov process: they respectively control the rate of convergence to equilibrium in L^2 distance and relative entropy of certain diffusion processes. Transport-entropy inequalities lie between the two, for example we have

$$LSI(\rho) \implies \mathcal{T}_2(\rho) \implies PI(\rho).$$

We can show that \mathcal{T}_2 and LSI imply PI by linearizing. The fact that the LSI implies \mathcal{T}_2 is a celebrated theorem of Otto and Villani [146], extended in full generality by Gozlan [92]. Transport-entropy inequalities are also a functional way of encoding concentration inequalities. For example, a measure satisfies a

\mathcal{T}_1 inequality if and only if for all 1-lipschitz functions f we have

$$\mu\left(f \geq \int f d\mu + r\right) \leq C \exp(-cr^2) \tag{2.4}$$

for all $r \geq 0$ and some positive constants c and C [17]. Such inequalities are known as Gaussian concentration inequalities. \mathcal{T}_2 inequalities are equivalent to a dimension-free form of such inequalities [92]. Concentration of measure has been an active field of research since V. Milman’s proof of Dvoretzky’s theorem on sections of convex bodies [135], and has found applications in geometry, statistical physics and statistics. We refer to Ledoux’s monograph [110] for an introduction to the field.

To clarify the connection between functional inequalities, I would like to briefly described the now-classical Herbst argument, which shows how the LSI implies the Gaussian concentration estimate (2.4). Consider a 1-Lipschitz function f with $\int f d\mu = 0$. If we introduce the log-Laplace transform on \mathbb{R}_+

$$H(\lambda) = \lambda^{-1} \log \int e^{\lambda f} d\mu$$

we have

$$H'(\lambda) = \frac{1}{\lambda} Ent_{\mu} \left(\frac{e^{\lambda f}}{\int e^{\lambda f} d\mu} \right)$$

so if we can apply the LSI, using the fact that f is Lipschitz we get

$$H'(\lambda) \leq 1/(2\rho)$$

and hence

$$\log \int e^{\lambda f} d\mu \leq \lambda^2/(2\rho).$$

We then apply the usual Markov inequality

$$\mu\left(f \geq \int f d\mu + r\right) \leq e^{-\lambda r} \exp(\lambda^2/(2\rho))$$

and optimize in λ to get the Gaussian concentration bound (2.4).

We will see in the next subsection the Bakry-Emery theorem, one of the main results on sufficient conditions for a LSI to hold, based on curvature bounds for an underlying weighted space. Other methods for proving functional inequalities include Lyapunov-functional approaches [38], coupling techniques [148], Muckenhoupt-type criteria in dimension one [137] and multi-scale expansions [95, 117].

To conclude this section, I would like to mention one of the main conjectures in the field.

Conjecture 1 (KLS conjecture). *There exists a universal constant that bounds the Poincaré constant of isotropic log-concave measures. In particular, this constant does not depend on the dimension.*

The currently known best upper bound is of order $d^{1/2}$, due to works of Eldan, Lee and Vempala [114]. The original motivation behind this conjecture was the study of the mixing time of random algorithms used to approximate the volume of convex sets in high dimension. We refer to the recent survey [115] for an overview of this conjecture, its relation with other open problems in geometric functional analysis, and its applications in computer science. We shall later discuss the Bakry-Emery theorem, which gives a dimension-free bound under a stronger assumption.

Let us now give a brief explanation of the role of convexity in this conjecture. If we only consider uniform densities on sets, log-concavity is equivalent to convexity of the support. Recall that the Poincaré

constant controls the mixing time of the Brownian motion on the set. By deforming a convex set to form a bottleneck, we create an obstruction to convergence, and hence slow down the convergence to equilibrium, so the Poincaré constant should get worse. The thinner the bottleneck, the worse the constant gets. Assuming convexity rules out this behavior. Similarly, if we add a potential, creating a double well slows down the diffusion process, since it must overcome an energy barrier to explore the whole space. Convexity in some sense rules out the presence of bottlenecks, in a fairly strong sense, and the conjecture predicts that this is enough to guarantee a dimension-free rate of convergence to equilibrium.

2.3 Curvature

The main purpose of this section is to give a brief introduction to the Lott-Sturm-Villani synthetic notion of lower bounds on Ricci curvature for metric spaces. It will mostly play a role as inspiration for the work presented in Chapter 5 of this thesis, but curvature bounds also play an implicit role in many of the results presented in chapters 3 and 4, in particular via the Bakry-Emery theorem on the connection between positive curvature and functional inequalities.

The Ricci curvature tensor is a tensor associated to a Riemannian manifold. We will not give a presentation of its standard definition, since it will not directly play a role here, as we will mostly be interested in lower bounds on curvature, and we will soon see an equivalent synthetic definition. The picture to keep in mind is that there are model spaces having constant Ricci curvature, which are spheres when the curvature is positive, flat Euclidean spaces for null curvature and hyperbolic spaces for negative curvature [170]. Having curvature bounded from below by some constant κ will mean that certain quantities of geometric interest, such as diameter, spectral gap and sharp constants in certain functional inequalities will necessarily be better than those of the model space with constant curvature and same dimension. Interestingly, Ricci curvature tensor is a local notion, but will still allow us to deduce quantitative bounds on global quantities.

The connection between the Ricci curvature tensor and probability is made via the Laplace-Beltrami operator Δ , which is the generator of the Brownian motion on the manifold (up to a scaling factor $1/2$, depending on conventions). The Bochner-Lichnerowicz formula

$$\frac{1}{2}\Delta(|\nabla f|^2) = \nabla f \cdot \nabla(\Delta f) + |\nabla^2 f|^2 + \text{Ric}(\nabla f, \nabla f)$$

then connects the Ricci curvature tensor, the Laplacian and the Riemannian metric. In particular, if curvature is nonnegative

$$\frac{1}{2}\Delta(|\nabla f|^2) - \nabla f \cdot \nabla(\Delta f) \geq 0.$$

If we replace the Laplacian by the generator of a Markov process, the left-hand side above corresponds to the Bakry-Emery Γ_2 tensor, and its nonnegativity is known as the curvature-dimension condition $\text{CD}(0, \infty)$.

Bakry and Emery [6], in their study of long-time behavior and functional inequalities for diffusion processes on manifolds, proposed an extension of the Ricci curvature tensor to weighted manifold. The motivation is to take into account the effect of an extra gradient vector field on the heat flow, i.e. diffusion PDEs of the form

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

where Δ is the Laplace-Beltrami operator and ∇V is a gradient vector field. This PDE generates a semigroup P_t , acting on functions, and its adjoint P_t^* , acting on probability measures. When the vector field is zero, the heat flow semigroup corresponds to the flow generated by Brownian motion.

Lott, Sturm and Villani [124, 162, 161] defined a synthetic notion of Ricci curvature via optimal transport, building up on previous works [146, 54, 168].

Theorem 1. *Let M be a smooth, connected complete Riemannian manifold, and μ a positive measure with density e^{-V} with respect to the volume measure. The following are equivalent:*

1. *The curvature dimension condition $\text{Ric} + \text{Hess } V \geq K$ holds with $K \in \mathbb{R}$ (which we shall denote by $CD(K, \infty)$);*
2. *For any path (ν_t) achieving the minimum in the Benamou-Brenier formula among all curves linking any two fixed probability measures ν_0 and ν_1 , we have*

$$\text{Ent}_\mu(\nu_t) \leq (1-t)\text{Ent}_\mu(\nu_0) + t\text{Ent}_\mu(\nu_1) - \frac{Kt(1-t)}{2}W_2(\nu_0, \nu_1)^2.$$

The first property is the Bakry-Emery curvature-dimension condition for weighted manifolds. The second property expresses the semi-convexity of the entropy along those paths. In particular, nonnegative curvature for an unweighted manifold is equivalent to plain convexity of the entropy (with respect to the volume measure). The full theory also takes into account the dimension, but we will not discuss that aspect here.

Since the two properties are equivalent, we can adopt the second one as a *definition* of lower bounds on Ricci curvature, known as the entropic curvature-dimension condition $CD(K, \infty)$. However, unlike the first one, which a priori requires a genuine Riemannian structure to be well-defined (although this can be relaxed a bit, see [2]), the second one has a meaning as soon as the underlying space is geodesic.

An explanation of this equivalence can be made via the so-called Otto calculus [146, 147] that we have briefly mentioned earlier. Formally, we can interpret the Wasserstein distance W_2 as a Riemannian distance on the space of probability measures, and view the Benamou-Brenier formula as a reformulation of the distance as the minimal energy of a path linking two probability measures. Hence these curves can be viewed as geodesics, and the convexity property (ii) corresponds to geodesic (semi-)convexity of the entropy. A formal computation of the Hessian of the entropy in this Riemannian structure shows that it is given by the Bakry-Emery curvature tensor, and hence convexity properties are indeed expected to be equivalent to curvature bounds. This formal Riemannian calculus can be made rigorous, see for example [1] and the references therein.

The main reason why I am interested in curvature is that it provides a convenient tool to study and prove functional inequalities. The basic result in that direction is the Bakry-Emery theorem [6], which we formulate here in the more general setting of metric spaces [166]:

Theorem 2. *Assume that the weighted metric space (X, d, μ) has Ricci curvature is bounded from below by $\kappa > 0$. Then the measure μ satisfies $LSI(\kappa^{-1})$.*

There is actually a stronger version of this theorem, by Bakry and Ledoux [7], that derives an isoperimetric inequality. If the curvature is merely nonnegative, this result does not hold, but there are some results relating strong functional inequalities to weaker concentration inequalities [29, 120, 109, 133], with more recent extensions to slightly negative curvature [134, 93]. Other families of functional inequalities where curvature bounds play a role include Sobolev inequalities, capacity inequalities in potential theory, heat kernel estimates, Harnack gradient estimates...

A crucial ingredient in the theory are the gradient bounds, which quantify the interplay between derivatives and semigroups.

Theorem 3. *Assume that curvature is bounded from below by $\kappa \in \mathbb{R}$, and let P_t be the semigroup with generator $\Delta - \nabla V \cdot \nabla$. Then we have the subcommutation relation*

$$|\nabla P_t f| \leq e^{-\kappa t} P_t |\nabla f|.$$

These gradient bounds, which are also related to the Li-Yau parabolic Harnack inequalities [120, 9], allow to quantitatively control norms and regularity bounds along the flow. They can also be used to prove functional inequalities. For example, for the Poincaré inequality when $\kappa > 0$, we have

$$\begin{aligned} \text{Var}_\mu(f) &= 2 \int_0^\infty \int |\nabla P_t f|^2 d\mu dt \\ &\leq 2 \int_0^\infty \int e^{-2\kappa t} P_t |\nabla f|^2 d\mu dt = \frac{1}{\kappa} \int |\nabla f|^2 d\mu. \end{aligned}$$

There are similar proofs of logarithmic Sobolev inequalities and many other interesting estimates via this gradient bound and its variants [8].

Chapter 3

Stein's method: an analytic viewpoint

Stein's method is a set of techniques for bounding distances between probability measures, implemented via integration by parts formula. It was originally proposed by Stein [159, 160], and has developed in a large field of research, with many variants and applications. We refer to [153, 42] for recent surveys of the field.

Let us start by presenting the methods in the Gaussian case. As remarked by Stein, the standard Gaussian measure γ on \mathbb{R}^d is the only probability measure satisfying the integration by parts formula

$$\int x \cdot f d\gamma = \int \text{Tr}(\nabla f) d\gamma \quad \forall f : \mathbb{R}^d \rightarrow \mathbb{R}^d. \quad (3.1)$$

It is often enough to consider the weaker formula

$$\int x \cdot \nabla g d\gamma = \int \Delta g d\gamma \quad \forall g : \mathbb{R}^d \rightarrow \mathbb{R} \quad (3.2)$$

which expresses the fact that the standard Gaussian measure is the invariant measure of the Ornstein-Uhlenbeck process, a diffusion process with generator

$$\mathcal{L}g = \Delta g - x \cdot \nabla g.$$

At a heuristic level, Stein's idea was that a measure that approximately satisfies the above integration by parts formula for a large enough class of test functions should be close in some sense to γ . The next step to apply this heuristic to obtain concrete estimates is to find a way to associate error estimates in the Stein equation (3.1) to distance bounds. The easiest case is the Wasserstein distance W_1 , because of the Kantorovitch-Rubinstein duality formula (2.3).

If we consider a given 1-lipschitz function g , we can solve the Poisson equation

$$\mathcal{L}h_g = g - \int g d\gamma$$

and get

$$W_1(\mu, \gamma) \leq \sup_g \int \Delta h_g - x \cdot \nabla h_g d\mu.$$

Due to the elliptic nature of \mathcal{L} , we can show that h_g inherits some regularity properties from g . In particular $\|\text{Hess } h_g\|_{HS} \leq 1$ pointwise [43]. So if we have an error bound on the integration by parts formula over the class of test functions satisfying this regularity property, we deduce a bound on $W_1(\mu, \gamma)$. This strategy adapts well to distances of the form

$$d(\mu, \nu) = \sup_{g \in \mathcal{A}} \int g d\mu - \int g d\nu$$

with \mathcal{A} some class of test functions, such as the total variation distance or the Kolmogorov distance. It is also possible to estimate some distances that do not admit a dual formula of this form, such as W_2 , by using more involved extensions of this scheme.

The concrete challenge is then to find bounds on $\int \Delta h_g - x \cdot \nabla h_g d\mu$. While concrete proofs are problem-specific, there are broad classes of tools and techniques that have been devised to obtain such bounds, such as zero-bias transforms and exchangeable pairs methods. My work has been focused on one particular way of implementing Stein's method, using an object known as a Stein kernel.

3.1 Stein kernels

Definition 2. A function $\tau : \mathbb{R}^d \rightarrow \mathcal{M}_d(\mathbb{R})$ is said to be a Stein kernel for a given centered probability measure ν on \mathbb{R}^d (relative to the standard Gaussian measure) if for all smooth test functions $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ we have

$$\int x \cdot f(x) d\nu = \int \langle \tau, \nabla f \rangle_{HS} d\nu. \quad (3.3)$$

Stein kernels were used somewhat implicitly in some of Stein's works. They started getting used more explicitly in works of Chatterjee [40] and Nourdin and Peccati [139, 141, 140] for random variables that can be realized as the image of a Gaussian variable by some smooth function, either via direct formulas or via the tools and formalism of Malliavin calculus.

We start by remarking that Stein kernels may not exist. Indeed, if the measure ν is purely atomic, there cannot be a matrix-valued function satisfying (3.3) for all smooth compactly-supported functions. We shall give several sufficient conditions for existence, but we do not know of a synthetic characterization of measures admitting a Stein kernel. Moreover, they are not unique in dimension $d \geq 2$.

In dimension one, for measures with nice density, which we shall denote by p , Stein kernels are unique, and we have the explicit formula

$$\tau(x) = \frac{1}{p(x)} \int_x^\infty yp(y) dy.$$

This formula and its variants for other target measures have been extensively used, see [119] and the references therein. My own contributions have mostly focused on the multidimensional setting, where we do not have an explicit formula anymore, so I will not develop features specific to the one-dimensional situation here.

An immediate reformulation of Stein's characterization of the normal law is that the identity matrix is a Stein kernel for ν if and only if ν is exactly the standard centered Gaussian distribution. Hence we expect that if the measure ν admits a Stein kernel that is close in some sense to the identity matrix, then the measure itself will be close to Gaussian. We will later see various inequalities that quantify this heuristic. The most natural way of quantifying proximity of a Stein kernel to the identity matrix is to use an L^2 norm, which leads to the following definition:

Definition 3. *The Stein discrepancy $S(\nu)$ of a probability measure ν is given by*

$$S(\nu) := \inf_{\tau} \left(\int |\tau - Id|_{HS}^2 d\nu \right)^{1/2}$$

where the infimum is taken over all Stein kernels for μ . More generally, we can define an L^p Stein discrepancy for $p \geq 1$ as

$$S_p(\nu) := \inf_{\tau} \left(\int |\tau - Id|_{HS}^p d\nu \right)^{1/p}.$$

Ledoux, Nourdin and Peccati [112], proved several functional inequalities involving the Stein discrepancy. Among them, we should mention:

$$\begin{aligned} Ent_{\gamma}(\nu) &\leq \frac{S(\nu)^2}{2} \log \left(1 + \frac{I_{\gamma}(\nu)}{S(\nu)^2} \right); \\ W_2(\nu, \gamma) &\leq S(\nu). \end{aligned}$$

The first inequality, called the HSI inequality, is an improvement of the Gaussian LSI, allowing to deduce entropy bounds from discrepancy bounds for measures with bounded Fisher information. The second inequality states that convergence in discrepancy is stronger than convergence to the Gaussian in W_2 distance, and hence justifies using such discrepancies as a way of estimating how close two probability measures are, since the topologies controlled by transport distances and entropy are well understood. It is also possible to control L^p transport distances by L^p discrepancies [112, 79]. The case $p = 1$ is an almost immediate consequence of the regularity bound on solutions to the Poisson equation we discussed earlier. We also mention that earlier work of Nourdin, Peccati and Swan [142] connected Stein discrepancy and Fisher information.

With Courtade and Pananjady, we established existence of Stein kernels under a spectral gap assumption:

Theorem 4. *Assume ν is a centered probability measure satisfying $PI(\rho)$. Then there exists a vector valued function g such that its Jacobian matrix ∇g is a Stein kernel for ν , and moreover the discrepancy satisfies*

$$S_2(\nu)^2 \leq (\rho^{-1} - 2) \int |x|^2 d\nu + d$$

In particular, if ν is isotropic

$$S_2(\nu) \leq \sqrt{d(\rho^{-1} - 1)}.$$

We actually proved a more general version, using converse weighted Poincaré inequalities, which encompasses some measures with heavy tails, such as generalized Cauchy distributions with density $\nu(dx) = C(1 + |x|^2)^{-\beta}$ with β large enough.

The proof is based on the observation that the Stein identity characterizing Stein kernels can be recast as the Euler-Lagrange equation for minimizing the convex functional $f \rightarrow \int ((\frac{1}{2}|\nabla f|^2 - x \cdot f) d\nu$. Weighted Poincaré type inequalities are enough to guarantee existence of a minimizer, and also provide an estimate on their norm, which leads to the above theorem.

The second construction I have found [79] is based on a variant of the optimal transport problem, known as the moment measure/map problem. It is based on a result of Cordero-Erausquin and Klartag [51], stating that given a measure $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ that is centered and not supported on a hyperplane, there exists a convex function φ such that

$$e^{-\varphi} dx_{\#}(\nabla \varphi) = \nu. \tag{3.4}$$

If we assume ν has a density h with respect to the Lebesgue measure, this condition boils down to the Monge-Ampère equation

$$e^{-\varphi} = h(\nabla\varphi) \det(\text{Hess } \varphi); \quad (3.5)$$

The map $\nabla\varphi$ is also the Brenier map sending $e^{-\varphi}$ onto ν , so this problem can be viewed as a variant of the optimal transport problem, where instead of specifying both measures, we only specify the target measure, and require a relation between the source measure and the map. We also refer to [154] for a construction of these moment maps that highlights the connection with optimal transport. This problem was originally motivated by applications in complex geometry, where the maps associated to convex polytopes play a role in the construction of Kähler-Einstein metrics on toric varieties [15, 169]. Because of this connection with geometry, the associated Monge-Ampère equation $e^{-\varphi} = \rho(\nabla\varphi) \det \text{Hess } \varphi$ is sometimes called the toric Kähler-Einstein equation.

Let us now formally show that we can construct a Stein kernel for ν from this moment map. By writing the change of variable formula, integrating by parts and recalling that the Legendre transform φ^* of φ satisfies $\nabla\varphi^* = (\nabla\varphi)^{-1}$, we have

$$\begin{aligned} \int x \cdot f(x) d\nu &= \int \nabla\varphi(y) \cdot f(\nabla\varphi(y)) e^{-\varphi(y)} dy \\ &= \int \text{Tr}(\text{Hess } \varphi(y) \nabla f(\nabla\varphi(y))) e^{-\varphi(y)} dy = \int \text{Tr}(\text{Hess } \varphi(\nabla\varphi^*(x)) \nabla f(x)) d\nu. \end{aligned}$$

Hence, we see that up to regularity issues $\text{Hess } \varphi \circ \nabla\varphi^*$ is a Stein kernel. The integration by parts formula above can be established rigorously for measures with a nice density and a convex support.

A nice feature of this particular construction is that it yields a Stein kernel whose values are positive, symmetric matrices. It is not clear at all that the first construction above has this property. In particular, this allows to introduce the process

$$dX_t = -X_t dt + \sqrt{2 \text{Hess } \varphi(\nabla\varphi^*(X_t))} dB_t$$

and see that the Stein identity for ν implies it is an invariant measure for this process. This can be compared to the classical drift-diffusion equation

$$dX_t = -\nabla V(X_t) dt + \sqrt{2} dB_t$$

where V is the potential of ν . Both SDEs are modification of the Ornstein-Uhlenbeck process that leave ν invariant. For the second one, we modify the drift without changing the diffusion matrix. This alternative process introduced above in some sense follows the opposite logic: we modify the diffusion coefficient without changing the drift to target some given probability measure. Both are equally natural choices, at least at a conceptual level, yet it is the second one that is mostly used, both as a theoretical tool to study properties of ν , and for numerical simulations.

Another nice feature of this construction is that it implies that the Stein discrepancy can be controlled by a suitable regularity estimate on the solution to (3.5). If we look at the discrepancy associated to this kernel, we have

$$S_p(\nu)^p \leq \int \|\text{Hess } \varphi(\nabla\varphi^*) - \text{Id}\|_{HS}^p d\nu = \int \|\text{Hess } \varphi - \text{Id}\|_{HS}^p e^{-\varphi} dx.$$

So this construction leads to a bound on the discrepancy given by a regularity estimate on an associated Monge-Ampère type equation. While in general, such global regularity estimates are hard to obtain, in the log-concave setting, Klartag obtained regularity bounds [107] of the form $\int (\partial_{ii}\varphi)^p e^{-\varphi} dx \leq C(p)$, leading to the following estimate:

Theorem 5. Assume that ν is a log-concave, isotropic probability measure on \mathbb{R}^d and satisfies some regularity assumptions. For any $p \geq 2$, we have a universal bound on the L^p Stein discrepancy

$$S_p(\nu) \leq Cp^3d.$$

A last interesting consequence of this construction is a weighted version of the Poincaré inequality involving this Stein kernel, generalizing to arbitrary dimension a result of Saumard [155].

3.2 Rate of convergence in the central limit theorem

The main application of Stein kernels is that they allow to get clean bounds on rates of convergence in the central limit theorem, with little effort. In all this section, ν will be a centered, isotropic probability measure on \mathbb{R}^d , and ν_n will be the law of $n^{-1/2} \sum_{i=1}^n X_i$ where the X_i are i.i.d. random variables distributed according to ν . The central limit theorem states that ν_n converges to the standard Gaussian distribution γ , and we wish to quantify the rate of convergence. A hint as to why Stein kernels work so well for this is given by the following result, obtained in collaboration with Courtade and Pananjady:

Theorem 6. For any $n \geq \ell \geq 1$, we have

$$S_2(\nu_n) \leq \sqrt{\frac{\ell}{n}} S_2(\nu_\ell).$$

In particular, the discrepancy is monotone along the CLT, and converges to zero at rate $n^{-1/2}$.

This result generalizes a classical known bound when $l = 1$, which was already sufficient to get a rate of convergence in the CLT. The rate is sharp. Monotonicity results along the CLT have a long history, going back to Shannon’s conjecture on monotonicity of entropy [157], which was solved fifty years later by Artstein, Ball, Barthe and Naor [4]. Unlike the above result on Stein discrepancy, monotonicity results for entropy (and related results in Fisher information) do not include a contraction rate. In particular, it is *much* easier to derive sharp rates of convergence for the discrepancy than for entropy.

The proof is based on the following representation of a Stein kernel for a sum of i.i.d. random variables:

$$\tau_n(x) = \mathbb{E}[\tau(X_1) | S_n = x]$$

where τ is a Stein kernel for a single random variable. To show this, we simply have to remark that due to independence, a Stein kernel for the whole vector (X_1, \dots, X_n) is given by the block-diagonal matrix whose blocks are Stein kernels for each random variable, and then project by the map $(x_1, \dots, x_n) \rightarrow n^{-1/2} \sum x_i$. The estimate then follows from a general correlation bound for partial sums along the CLT, due to Dembo, Kagan and Shepp [60].

When $p > 2$, we can also obtain decay rates of order $n^{-1/2}$, but it seems unlikely that the associated discrepancies would be monotone in full generality, since the picture looks very similar to the Rosenthal inequality for sums of i.i.d. random variables.

Combining the various results we described, we get the following rates of convergence in the CLT:

Theorem 7. If ν satisfies $PI(\rho)$, then $W_2(\nu_n, \gamma) \leq \frac{\sqrt{d(\rho^{-1}-1)}}{\sqrt{n}}$;

If ν is log-concave, then for all $p \geq 2$ $W_p(\nu_n, \gamma) \leq \frac{Cp^4d}{\sqrt{n}}$, where C is a universal constant, independent of ν .

If ν is uniformly log-concave, i.e. satisfies $CD(K, \infty)$ with $K > 0$, then $W_p(\nu_n, \gamma) \leq \frac{Cp^4\sqrt{d}}{\sqrt{Kn}}$.

Once again, the rate $n^{-1/2}$ is optimal in general, although it can be improved under extra moment conditions. The best possible dependence on the dimension is \sqrt{d} , as it is the correct bound for random variables with independent coordinates. Bonis [23] recently proved the sharp rate under much weaker, sharp moment conditions on the random variables, albeit with a worse dependence on the dimension. In dimension one, sharp rates of convergence in transport distance have been derived by Rio [151] and Bobkov [19]. The sharp entropic CLT in dimension higher than one is still an open problem [3, 10, 69].

We would like to conclude this section with some ongoing work. If we look at the Edgeworth asymptotic expansion for various distances in the CLT [20, 19], we see that the rate of convergence should be better for random variables whose moments of order $k \geq 2$ match with those of the standard Gaussian. It is possible to use Stein's method to see this, by introducing the integration by parts formula associated to Hermite polynomials. Indeed, the Gaussian measure (in dimension one, for simplicity) satisfies

$$\int H_k(x)f(x)d\gamma = \int H_{k-1}(x)f'(x)d\gamma$$

where H_k is the k -th Hermite polynomial $H_k(x) = e^{x^2/2} \frac{d^k}{dx^k}(e^{-x^2/2})$. For each moment of order $k \geq 2$ that match with the Gaussian, we can introduce a Stein kernel representing associated to H_k , and it turns out that we can build a linear combination of these kernels whose variance decays like n^{-k} . By following the same line of arguments as described above for standard Stein kernels when $k = 2$, we can get for example the following rate of convergence:

Theorem 8 (F. 2018, in preparation). *If ν is centered, isotropic, satisfies $PI(\rho)$ and if moreover its mixed third moments are equal to zero, then we have*

$$W_2(\nu_n, \gamma) \leq \frac{\sqrt{d\rho^{-1}(\rho^{-1} - 1)}}{2n} (1 + \log n - \log \rho)$$

Note that we miss the expected sharp rate n^{-1} for such random variables by a logarithmic factor. On the other hand, the dependence on the dimension is once again sharp. It is possible to get a rate of n^{-1} , at the cost of weakening the distance used. For variables satisfying higher moment constraints, it is also possible to get a sharp rate under a spectral gap, but with a distance that is further modified. In the case $k = 2$ and in dimension one, Goldstein and Reinert also proved a similar rate of convergence for a weaker distance using another variant of Stein's method [90].

3.3 Free probability

Free probability is an area of research pioneered by Voiculescu that aims at understanding the structure of Von Neumann algebras related to free groups. The starting point is that there is an analogy between traces of products of operators $\text{Tr}(A_1^{k_1} \cdots A_n^{k_n})$ and mixed moments of random variables $\mathbb{E}[X_1^{k_1} \cdots X_n^{k_n}]$. In particular, if we only consider a single operator, we can represent the collection of traces $\text{Tr}(A^k)$ as the moments of a probability measure on \mathbb{R} . However, we do not wish to assume that the operators commute, so that unlike moments of random variables the above traces depend on the order in which the operators are written.

The framework is to consider an algebra of operators endowed with a linear functional $\varphi : \mathcal{A} \rightarrow \mathbb{R}$ such that $\varphi(1) = 1$, that plays the role of a (normalized) trace.

The counterpart to independence in free probability is the notion of free independence. Two non-commuting random variables are freely independent if when considering the algebras \mathcal{A}_i ($i = 1, 2$) they generate, for any sequence x_1, \dots, x_n with $x_j \in \mathcal{A}_{\ell_j}$, $\ell_{j+1} \neq \ell_j$ and $\varphi(x_j) = 0$ for all j we have

$$\varphi(x_1 \cdots x_n) = 0.$$

That is, alternated products of centered elements are also centered. The definition extends similarly to a larger number of variables. Note that since the variables do not commute, we cannot just exchange the order in the trace. An important discovery by Voiculescu is that large random matrices with independent entries are asymptotically freely independent, as the size of the matrices goes to infinity.

To keep notations simple, we shall mostly describe our results in the one-dimensional case, that is consider an algebra generated by a single operator, and associate to it a probability measure whose moments correspond to the trace of powers of the element generating the algebra.

In the basic analogy between classical probability and free probability, the measure that plays the role of the standard Gaussian distribution is the semicircular law, whose density on \mathbb{R} is given by

$$d\sigma(x) = \frac{\sqrt{4-x^2}}{2\pi} \mathbb{1}_{|x| \leq 2} dx.$$

This measure plays two important roles:

- It describes the asymptotic limit of the spectrum of large random matrices with independent entries.
- It is the universal limit in the free analogue of the central limit theorem. In particular, there is a free analogue of Shannon's problem on monotonicity of entropy, solved by Shlyakhtenko [158].

The counterpart to entropy is Voiculescu's free entropy, defined as

$$\mathcal{E}(\mu) = - \int \int \log |x-y| d\mu(x) d\mu(y).$$

Part of the analogy between the semicircular law and the Gaussian law is that the semicircular law minimizes the free entropy, under a constraint of fixed second moment. In dimension higher than one, there are two ways of defining free entropy (one as a large deviations rate function, the other as a monotone quantity for convolution with semicircular laws of increasing variance), and it is an open problem to show that they match. They were originally introduced with the aim of constructing new invariants for von Neumann algebras.

In joint work with Nelson [82], we introduced free Stein kernels, which are counterparts of classical Stein kernels adapted to free probability. In dimension one, and relative to a semicircular law, they are defined as a function $\tau : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that

$$\int x f(x) d\mu = \int \int \tau(x, y) \frac{f(x) - f(y)}{x - y} d\mu(x) d\mu(y)$$

for all nice test functions f . The right-hand side involves the free difference quotient $\frac{f(x)-f(y)}{x-y}$, which plays the role of a derivative in free probability. As in the classical case, we can take $\tau = 1$ iff μ is a semicircular law with unit variance. It turns out that these kernels were implicitly used by Guionnet and Shlyakhtenko in their construction of free monotone transport maps for small deformations of semicircular laws [96].

Using these free Stein kernels, we can introduce a Stein discrepancy, and prove the free counterparts of the functional inequalities of [112]. In particular, we obtain a free HSI inequality, that improves on the free logarithmic Sobolev inequality of Biane and Speicher [16]. We also deduced as a straightforward application of these inequalities new rates of convergence in the free CLT. In dimension one, the improved free LSI takes the following form:

Theorem 9. *For any centered probability measure μ on \mathbb{R} , given a free Stein kernel τ , we have*

$$\begin{aligned} & \frac{1}{2} \int |x|^2 d\mu - \int \int \log |x - y| d\mu(x) d\mu(y) - \frac{3}{4} \\ & \leq \frac{1}{2} \left(\int |\tau(x, y) - 1|^2 d\mu(x) d\mu(y) \right) \log \left(1 + \frac{\int (H\mu(x) - x)^2 d\mu(x)}{\int |\tau(x, y) - 1|^2 d\mu(x) d\mu(y)} \right). \end{aligned}$$

The same result holds in higher dimension, if we use the so-called non-microstates free entropy.

After we communicated this result, Mai realized that in free probability existence of Stein kernels is a much easier problem than in classical probability, and that Stein kernels always exist, with an explicit formula. Indeed, in dimension one, it is easy to check that if the measure is centered, $\tau(x, y) = (x - y)^2$ is always a free Stein kernel relative to the semicircular law. A similar formula holds in higher dimension, and for other reference measures. In particular, the prefactor in the rate of convergence in the free CLT in Wasserstein distance is controlled by the fourth moment. The same holds in classical probability, but the proof is surprisingly much more difficult. With Cébron, we realized that the arguments of [57] also extend to the free setting with no assumption beyond centeredness, and yet the Stein kernel obtained this way never coincides with the one Mai discovered (it is necessarily a free difference quotient). So in particular free Stein kernels are never unique.

3.4 Perspectives

3.4.1 Dependent random variables

While the classical CLT is the most standard situation where a sum of random variables converges to a Gaussian, there is a vast literature on CLTs for sums of non-independent random variables. Important examples include the CLT for additive functionals of Markov chains [105], combinatorial CLTs [22] and fluctuation bounds for interacting systems in statistical physics [41]. Some of these can be proved using variants of Stein's method. It is natural to ask whether we can also use Stein kernels in some of these situations.

In Section 3.2, we have discussed rates of convergence for the classical CLT, with i.i.d. random variables. The fact that they are identically distributed is not essential, for example we could still get similar results with a uniform bound on the Poincaré constant of each random variable. However, we make a strong use of the assumption of independence. The main reason is that if we look at a Stein kernel for the joint laws, we get a matrix that is block-diagonal. If the random variables are correlated, this is not the case anymore. However, if the off-diagonal blocks were small enough, their contribution to the discrepancy for the law of the sum would also be small, and we could still hope to get a rate of convergence. At this point, the main question would be to figure out how to translate the classical conditions that ensure a CLT holds (for example, stationary strongly mixing sequences under moment conditions) into a bound on the off-diagonal elements of this joint Stein kernel. Further questions of interest include the rate of convergence in the martingale CLT and error bounds for the convergence of equilibrium fluctuations of interacting particle systems to their Gaussian limit.

3.4.2 Stein kernels in non-Gaussian settings

Until now, I have mostly worked on Stein's method when the target distribution is Gaussian. However, the abstract setup can be easily adapted to other target distributions. What we need to adapt the abstract framework is to have an integration by parts formula that characterizes it.

A first natural example of non-Gaussian limit theorem where the methods presented above should find some application are stable CLTs. For certain random variables with heavy tails and infinite second moment, the classical CLT does not apply, but nevertheless at a different scale we observe convergence in distribution for the normalized average $n^{-\alpha/4} \sum X_i$, for some exponent $\alpha \in (0, 2)$. The one parameter family of possible limits are known as stable laws. These are infinitely divisible distributions, related to alpha-stable Lévy processes. In particular, (up to normalization) they are the invariant measures for Lévy-Ornstein-Uhlenbeck processes, whose generator are given by

$$L_\alpha = -(-\Delta)^{\alpha/2} + x \cdot \nabla.$$

Hence they satisfy an integration by parts formula involving the fractional Laplacian $(-\Delta)^{\alpha/2}$. It is natural to expect that the technique of [56] could be used to derive rates of convergence for stable CLTs, by working in some appropriate fractional Sobolev space.

In the discrete setting, once can use random walks or other Markov chains to define a discrete integration by parts formula that characterizes a given measure. For example, the Poisson measure on \mathbb{N} is the invariant measure of the simple birth and death process, and hence satisfies the discrete integration by parts formula

$$\mathbb{E}[X f(X)] = \mathbb{E}[f(X + 1) - f(X)].$$

Once again, this is enough to set up the abstract framework of [112]. Natural questions are whether there are natural conditions to ensure the validity of discrete functional inequalities mimicking those of [112] for discrete measures, and bounds on distances between discrete measures via Stein's method. A first example would be to recover the bounds of [46] on the quantitative law of small numbers via arguments from functional analysis, in the spirit of [56]. It seems likely that there is a connection with intertwining of semigroups [49].

Finally, if we consider a general probability measure μ with density e^{-V} on \mathbb{R}^d , Barbour's generator approach to Stein's method suggests to use the integration by parts formula

$$\int \nabla V(x) \cdot f(x) d\mu = \int \operatorname{div} f(x) d\mu$$

which suggests to define Stein kernels relative to μ as matrix-valued functions satisfying

$$\int \nabla V \cdot f d\nu = \int \langle \tau, \nabla f \rangle_{HS} d\nu$$

for a large enough class of smooth vector-valued test functions f . Under some strong assumptions on V it is possible to prove functional inequalities involving the discrepancy [112]. The arguments of [57] can be extended to establish existence of such kernels under a spectral gap and integrability assumptions on ∇V . A first example of application would be estimating the rate of convergence to equilibrium of the diffusion

$$dX_t = -\nabla V(X_t) dt + \sqrt{2} dB_t$$

in Stein discrepancy. The quantity should be well-adapted, since it is the generator of the diffusion that has been used to provide the integration by parts formula as a starting point.

3.4.3 CLT for random eigenfunctions in positive curvature

I would like to conclude this chapter with a conjecture motivated by some of the results discussed above.

Conjecture 2. *Let (M_n, g_n, μ_n) be a sequence of weighted Riemannian manifolds with a uniform positive lower bound on their Ricci curvature, and assume that there is a fixed value $-\lambda > 0$ that is an eigenvalue of the Laplacian on any of these spaces, and that the associated sequence of eigenspaces E_n have dimension d_n growing to infinity. Is it true that as n goes to infinity a typical random eigenfunction (w.r.t. the uniform measure on the unit sphere of E_n) has distribution close to the (standard) Gaussian measure?*

The main motivation for this conjecture is that it would give a single geometric, unified framework that would include as corollaries a version of the fourth moment theorem of Nualart and Peccati [143], and Klartag’s central limit theorem for convex bodies [106]. This first result can be derived using Stein’s method and the representation of polynomials as eigenfunctions for the Gaussian space [111]. The connection with the CLT for random projections of convex bodies is that if we consider the moment map φ associated with a uniform measure on a centered, isotropic convex body, this CLT is equivalent to saying that the pushforward of $e^{-\varphi}$ by $\partial_{e_i}\varphi$ is approximately Gaussian for most directions e_i . However, if we introduce the weighted Riemannian manifold $(\mathbb{R}^d, (\text{Hess } \varphi)^{-1}, e^{-\varphi})$, it turns out that the functions $\partial_{e_i}\varphi$ are all eigenfunctions for the Laplacian on this manifold, and moreover the convexity of the set implies that curvature is bounded from below by $1/2$ [108]. Hence a positive answer to the above conjecture would indeed recover Klartag’s result. To our knowledge, there is no proof of this CLT that exploits the structure we just described. Moreover, we do not know in general the sharp rate of convergence for this CLT, and this problem is related to the KLS conjecture.

In [131], a criterion for asymptotic normality of eigenfunctions is given, and applied for spheres of growing dimensions (which are indeed spaces of constant positive curvature), but we still lack an argument to apply it to general manifolds of positive curvature.

Chapter 4

Refinements of functional inequalities

In the past few years, there has been a lot of interest in the problem of establishing *deficit estimates* for sharp functional inequalities. The problem can be stated as follows: given a functional inequality for which *all* cases of equality are known, is there a quantitative estimate that shows how an object for which equality almost holds is close to the set of equalizers? For example, for the classical isoperimetric inequality, it is known that for a given perimeter, the sets that maximize the volume are balls, and it turns out that sets with *almost* maximal volume are nearly round.

I have been interested in two ways of refining functional inequalities:

- Deficit estimates: given a functional inequality for which *all* cases of equality are known, is there a quantitative estimate that shows how an object for which equality almost holds is close to the set of equalizers? An example to keep in mind is the isoperimetric inequality, which we mentioned above.
- Improved functional inequalities for subclasses of functions. For example, it is sometimes possible to further improve the sharp constant in a functional inequality by considering some class of arguments that satisfy extra regularity or symmetry assumptions.

Several important breakthroughs have been obtained in the past decade, including deficit estimates for isoperimetric inequalities, Sobolev inequalities, concentration inequalities and the Brunn-Minkowski inequality [11, 65, 67, 88, 86, 87]. We refer to [85] for an overview. Techniques that have been used include optimal transport, calculus of variations, PDE methods and stochastic calculus.

Let us list a few applications of refined functional inequalities:

- Convergence to equilibrium, see for example [34] where a stability estimate is used to study the Keller-Segel model at the critical mass.
- Improved mixing times for specific initial data.
- Information theory [58].
- Quantum mechanics [35].

4.1 Gaussian functional inequalities

This section describes results from [80, 57, 78] on refinements of Gaussian functional inequalities.

The first inequality I have studied is the Gaussian logarithmic Sobolev inequality, originally proved by Gross in 1975 [94], and which states that

$$Ent_\gamma(\mu) \leq \frac{1}{2}I_\gamma(\mu).$$

The constant $1/2$ is sharp, as can be shown by considering non-centered standard Gaussians, for which the above inequality is an equality. Carlen [33] showed that these are the *only* equality cases, and in particular that among centered measures equality holds only for γ .

There have been several results on estimating from below the deficit in the Gaussian LSI

$$\delta_{LSI}(\mu) := \frac{1}{2}I_\gamma(\mu) - Ent_\gamma(\mu),$$

among which we can mention [18, 63, 84, 113]. Applications include faster rates of convergence for diffusion processes and refined inequalities in quantum mechanics. More recently, Eldan has studied *upper* bounds on the deficit [66], motivated by applications to large deviations for mean-field models in statistical physics.

My main contribution to this problem, obtained in joint work with Indrei and Ledoux [80], is a sharp dimension-free deficit estimate for measures satisfying a Poincaré inequality:

Theorem 10. *Assume that μ is centered and satisfies a Poincaré inequality with constant $\rho > 0$. Then the deficit in the Gaussian LSI satisfies*

$$\delta_{LSI}(\mu) \geq c(\rho)I_\gamma(\mu)$$

where $c(\rho) := \frac{(1-\rho)^2 - 1 + \rho - \rho \log \rho}{2(1-\rho)^2}$ (and $c(1) = 1/4$).

This result can also be recast as an improvement of the constant in the LSI for a subclass of measures. Lower bounds on the deficit in other metrics (entropy, transport, total variation) immediately follow by comparison. The constant we obtain is sharp. When $\rho = 1$, this result detects the sharp threshold for cutoff of independent Ornstein-Uhlenbeck processes starting from the origin, so in some sense for applications to convergence to equilibrium there cannot be a better entropy bound valid for all measures in $PI(1)$. An earlier version was obtained in [98], with a control on the Wasserstein distance and assuming extra two-sided bounds on the Hessian of the potential.

The proof we came up with is based on a semigroup argument, by extracting an extra positive term in the Bakry-Emery proof of the LSI [6]. Let us briefly describe how it goes. We consider the Ornstein-Uhlenbeck dynamic, whose distributions at a given time are given by the identity in law

$$X_t \equiv e^{-t}X + \sqrt{1 - e^{-2t}}U$$

where X is the initial data and U is a standard Gaussian random variable, independent from X . Alternatively, this describes in closed form the solution to the stochastic differential equation

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

If we write for ν_t the law of X_t , the relative entropy $Ent_\gamma(\nu_t)$ is decreasing along the flow, and its dissipation is given by the relative Fisher information $I_\gamma(\nu_t)$. To prove the LSI, Bakry and Emery

computed the derivative of the Fisher information (which, up to the sign, is the second derivative of the entropy), and obtained the formula

$$\frac{d}{dt}I_\gamma(\nu_t) = -2I_\gamma(\nu_t) - 2 \int |\text{Hess log } d\nu_t/d\gamma|^2 d\nu_t.$$

To prove the LSI, we can neglect the second term on the left-hand side, which is nonpositive, and obtain the decay rate $I_\gamma(\nu_t) \leq e^{-2t}I_\gamma(\nu_0)$. Integrating in time yields the LSI, and the constant turns out to be sharp in general. However, the second term that was neglected contains some information about the way entropy decays along the flow. In particular, if ν_t satisfies a Poincaré inequality with a constant ρ_t , then

$$\int |\text{Hess log } d\nu_t/d\gamma|^2 d\nu_t \geq \rho_t I_\gamma(\nu_t)$$

and we could improve the rate of decay of the Fisher information. It turns out that, using gradient estimates in positive curvature from Theorem 3, we can prove that the flow propagates in some sense Poincaré inequalities: if the initial data ν satisfies it with a constant ρ , then at later times ν_t also does, with a constant ρ_t that is bounded in an explicit way, depending on time and ρ . Using explicit estimates on how the constants involved depend on time leads to Theorem 10.

Kolesnikov and Kosov recently gave an alternative proof of Theorem 10, based on optimal transport, refining Cordero-Erausquin's proof of the LSI [52]. Other deficit estimates include [18, 113, 97].

We have also investigated lower bounds for the deficit in the Gaussian Talagrand transport-entropy inequality

$$\delta_{Tal}(\mu) := 2Ent_\gamma(\mu) - W_2(\mu, \gamma)^2.$$

As for the LSI, the constant 2 is sharp in general and equality holds if and only if μ is a standard Gaussian measure with arbitrary mean. δ_{LSI} and δ_{Tal} can be related via the HWI inequality [146]

$$Ent_\gamma(\mu) \leq W_2(\mu, \gamma)\sqrt{I_\gamma(\mu)} - \frac{1}{2}W_2(\mu, \gamma)^2,$$

yielding

$$\delta_{LSI}(\nu) \geq \frac{\delta_{Tal}(\nu)^2}{16Ent_\gamma(\nu)}.$$

Our main result is the following deficit estimate

Theorem 11. *For any centered probability measure ν , the deficit in the Gaussian Talagrand inequality satisfies*

$$\delta_{Tal}(\nu) \geq c \min \left(\frac{W_{1,1}(\nu, \gamma)^2}{d}; \frac{W_{1,1}(\nu, \gamma)}{\sqrt{d}} \right)$$

where $W_{1,1}$ stands for the L^1 Wasserstein distance associated to the L^1 distance on \mathbb{R}^d .

This result was improved by Cordero-Erausquin [53], who replaced the quantity $W_{1,1}/\sqrt{d}$ by the stronger W_1 distance with an L^2 cost. These estimates are obtained by refining Cordero-Erausquin's transport-based proof of the inequality, in the spirit of the work [88] on isoperimetric inequalities. However, the estimate is not entirely satisfactory: while unlike Theorem 10 we have no strong regularity assumptions on the data, if we test the estimate on a high-dimensional product measure $\mu^{\otimes d}$, we get a lower bound of order \sqrt{d} , while the deficit tensorizes and is of order d . The lower bound on δ_{LSI} obtained as a corollary behaves even worse, as it is of order 1 instead of d .

In another direction, I have shown the following strengthening of the Gaussian Talagrand inequality, as a dual formulation of the functional Santaló inequality in convex geometry [5].

Theorem 12. *Let μ be a centered probability measure on \mathbb{R}^d , and ν another probability measure. We have*

$$W_2(\mu, \nu)^2 \leq 2(Ent_\gamma(\mu) + Ent_\gamma(\nu)).$$

This inequality improves on Talagrand's inequality, since we recover it when taking $\mu = \gamma$. Note that the Talagrand inequality implied this inequality up to a factor two by just applying the triangle inequality, so the point here is the sharp improved constant. It allows to slightly refine some Gaussian concentration estimates, and can also be extended to symmetric probability measures satisfying a curvature condition, by pushing forward the Gaussian inequality by the optimal transport map, which is 1-lipschitz according to Caffarelli's contraction theorem [30].

Finally, with Courtade and Pananjady [57], we looked into stability for the Shannon-Stam entropy power inequality. This inequality states that

$$\exp\left(-\frac{2}{d}Ent_{dx}(\mu * \nu)\right) \geq \exp\left(-\frac{2}{d}Ent_{dx}(\mu)\right) + \exp\left(-\frac{2}{d}Ent_{dx}(\nu)\right)$$

for any pair of probability measures μ and ν , and where $\mu * \nu$ stands for their convolution. The quantity $\exp\left(-\frac{2}{d}Ent_{dx}(\mu)\right)$ is usually denoted by $N(\mu)$, and is called the entropy power of μ . It was originally proposed by Shannon in his pioneering work on communication channels, and rigorously proved by Stam. See also [116] for a very simple stochastic proof.

Carlen and Soffer [32] showed that for isotropic measures with bounded Fisher information, the deficit in the EPI controls an increasing function of the entropy. This function is non-explicit (it is built via a compactness argument) and depends on the dimension and the information bound. Later on, Toscani derived a (somewhat complicated) formula for the deficit, and raised the question of whether it could be bounded from below by some explicit distance to the set of Gaussian measures, at least for log-concave measures. Our main deficit estimate is the following answer to Toscani's question:

Theorem 13. *Let μ and ν be probability measures satisfying the curvature condition $CD(\rho, \infty)$ with $\rho > 0$. Then*

$$N(\mu * \nu) \geq (N(\mu) + N(\nu))\Delta_{EPI}(\mu, \nu)$$

with

$$\Delta_{EPI}(\mu, \nu) := \exp\left(\frac{\rho \min(\theta, 1 - \theta)}{4d} \left((1 - \theta)W_2(\mu, \Gamma)^2 + \theta W_2(\nu, \Gamma)^2 + d_F(\mu, \nu)^2\right)\right)$$

where $W_2(\mu, \Gamma)$ is the distance of μ to the set of centered Gaussian measures, and $d_F(\mu, \nu)^2 = \inf_{s \in (0, 1)} \|\sqrt{s}\Sigma_\mu^{1/2} - \sqrt{1-s}\Sigma_\nu^{1/2}\|_{HS}^2$ where Σ_μ is the covariance matrix of μ , and θ is chosen such that $\theta/(1-\theta) = N(\mu)/N(\nu)$.

Following the general ideas of [88], the proof consists in extracting from an optimal transport proof of the EPI, due to Rioul [152], an extra reminder term, which can be compared to the Wasserstein distance when the transport map from the Gaussian onto the argument satisfies some uniform C^2 regularity bound. Caffarelli's contraction theorem [30] provides uniform pointwise bounds on those maps when the measures are uniformly log-concave, which results in the above bound. Weaker estimates can also be obtained under weaker assumptions, as detailed in [57].

4.2 Spectral inequalities

There are several areas of mathematics and applications where we consider some family of positive symmetric operators, and seek a uniform bound on their spectrum. A classical example is the Faber-Krahn inequality, which states that the spectrum of $-\Delta$ on sets satisfying a volume constraint, and

enforcing Dirichlet boundary condition, is bounded from above by the spectral gap on a ball with suitable volume. Stability estimates for the spectrum of the Laplacian plays a role in shape optimization, as well as long-time behavior of stochastic processes [163, 149].

As mentioned previously, the Poincaré inequality has a spectral interpretation: the best constant for the measure $\mu = e^{-V} dx$ is also the smallest eigenvalue of the operator $L = -\Delta + \nabla V \cdot \nabla$, viewed as a symmetric operator on $L(\mu)/\{\text{constants}\}$.

A first result arises as a direct consequence of the estimate on the Stein discrepancy from Theorem 4. Assume that a given measure μ is isotropic. The normalization of the second moments of each coordinate implies that the best constant in the Poincaré inequality for μ must satisfy $\rho \leq 1$. Hence among all isotropic measures, the one with the best Poincaré constant is the standard Gaussian measure. Borovkov and Utev [24] showed that this property actually characterizes the Gaussian law among all centered, isotropic measures. With Courtade and Pananjady [56], we showed that this remark can be improved into the following result:

Theorem 14. *Let μ be a centered, isotropic probability measure. Then its Poincaré constant satisfies*

$$C_P(\mu) \geq 1 + \frac{W_2(\mu, \gamma)^2}{d}.$$

Both the exponent 2 and the dependence on the dimension here are sharp, as can be seen when testing this estimate on high-dimensional product measures.

In another direction, as a consequence of either the Brascamp-Lieb inequality or the Bakry-Emery theorem, any probability measure of the form $e^{-V} dx$ on a convex set with $\text{Hess } V \geq \text{Id}$ is in PI(1), i.e. its Poincaré constant is better than that of the normal law. It was shown [47] that any such measure for which the Poincaré constant is one must have a Gaussian factor. De Philippis and Figalli [59] later gave a quantitative dimension-dependent version of this statement, and recently with Courtade [55] we improved the bounds and eliminated the dependence on the dimension in that last result, obtaining the following estimate:

Theorem 15. *Let $\mu = e^{-V} dx$ be a probability measure with $\text{Hess } V \geq I_n$, and assume that there exists k functions $u_i \in H^1(\mu)$, $k \leq n$, such that for any $i \in \{1, \dots, k\}$ we have*

$$\int u_i d\mu = 0; \quad \int u_i^2 d\mu = 1; \quad \int \nabla u_i \cdot \nabla u_j d\mu = 0, \quad \forall j \neq i$$

and

$$\int |\nabla u_i|^2 d\mu \leq 1 + \epsilon$$

for some $\epsilon \geq 0$. Then there exists a vector $p \in \mathbb{R}^k$ and a function $W : \mathbb{R}^{n-k} \rightarrow \mathbb{R}$ satisfying $\text{Hess } W \geq I_{n-k}$ such that

$$W_1(\mu, \gamma_{k,p} \otimes e^{-W}) \leq 26k^{3/2}\sqrt{\epsilon}.$$

In particular, if $C_P(\mu)$ is the sharp Poincaré constant of μ , we have

$$W_1(\mu, \gamma_{k,p} \otimes e^{-W}) \leq 26\sqrt{C_P(\mu)^{-1} - 1}.$$

The proof uses three main ingredients. First, we show that the fact that the u_i are approximate optimizers in the Poincaré inequality gives rise to an approximate integration by parts formula

$$\left| \int u_i f d\mu - \int \nabla u_i \cdot \nabla f d\mu \right| \leq C\sqrt{\epsilon} \int |\nabla f|^2 d\mu.$$

The second step, which was already proved in [59], is to show that the u_i are close to affine functions. This allows to compare the above approximate integration by parts identity to the Stein identity, and hence compare μ to a Gaussian measure via Stein's method. The main reason why we get a better bound than [59] is that we mostly remain at a linear level, while the proof of [59] uses nonlinear estimates at the level of the Monge-Ampère equation for the transport map between the standard Gaussian measure and μ . It is also possible to derive a similar stability result for logarithmic Sobolev inequalities and Gaussian concentration bounds for uniformly log-concave measures with the same broad technique, albeit under strong regularity assumptions on the near-extremizers and a worse dependence on ϵ .

This result leaves open several interesting questions. The first is about optimality: some computations of [59] suggest that the sharp rate should actually be of order ϵ instead of $\sqrt{\epsilon}$. Another natural question is the extension to a geometric setting. For example, is there an analogue of this theorem in the setting of metric spaces satisfying a curvature-dimension condition $CD(K,N)$. Cavaletti and Mondino have shown that such a space with almost maximal spectral gap is close to a spherical suspension, which corresponds to the case $k = 1$ in the above result. The proof uses Klartag's needle decomposition, a non-smooth variant of Obata's maximal diameter theorem due to Ketterer, and a compactness argument. In particular, the estimates they obtain are not constructive. We do not know of a variant of Theorem 15 in a geometric setting when $k \geq 2$, or of constructive estimates even when $k = 1$. One of the difficulties in obtaining a constructive stability estimate is that we have to work in the Gromov-Hausdorff topology, and hence look at embedded representations of the space we consider and the target space in some unknown ambient metric space.

4.3 Perspectives

4.3.1 Stability for variational problems

At an abstract level, the argument for deriving an integration by parts formula for near-minimizers in the Poincaré inequality is a particular instance of a much more general phenomenon. If we seek to minimize a functional of the form $\int H(f, \nabla f) d\mu$, the Euler-Lagrange equation associated to a minimizer u is

$$\int h \partial_1 H(u, \nabla u) + \nabla h \cdot \partial_2 H(u, \nabla u) d\mu = 0 \quad \forall h$$

which is indeed an integration by parts formula for μ . Hence if we know that minimizing the above quantity with respect to two different measures leads to approximately the same minimizers, we expect that Stein's method could be used to prove the measures are close, at least if the minimizers are nice enough.

There are many interesting problems in the calculus of variations that take this form. For example, it would be interesting to see if there is a variant of Stein's method allowing to give easy proofs of stability of spectral inequalities for sets, or for best constants in Sobolev-type inequalities [127]. For instance, the Faber-Krahn inequality on eigenvalues of the Laplacian with Dirichlet boundary condition states that the minimal value is given when the domain is a ball. See [26] for a recent survey. It is easy to see that the variational formulation for the smallest eigenfunction gives rise to an integration by parts formula for the set under consideration, with coefficients given by the associated eigenfunction, so maybe Stein's method could be used to compare this formula to the one characterizing the uniform measure on the ball (with coefficients given by a Bessel function) and extract quantitative estimates. As a motivation, we note that there are some open questions remaining on stability for certain boundary conditions.

4.3.2 Estimates on optimal transport maps

This section is partly based on discussions with Maria Colombo.

In [56] (and many other works on functional inequalities and concentration for uniformly log-concave measures), we made use of Caffarelli's contraction theorem [30], which states that the optimal Brenier map sending the standard Gaussian measure onto a probability measure on \mathbb{R}^d satisfying $\text{CD}(\rho, \infty)$ for some $\rho > 0$ is $\rho^{-1/2}$ -Lipschitz. This statement actually implies the Bakry-Emery bound on the LSI constant. Motivated by the KLS conjecture, it is natural to ask what regularity estimates can be proved when the target measure is only log-concave, without a uniform bound. The example of the exponential measure shows that the map cannot be expected to always be Lipschitz. However, in dimension one, we can obtain the following estimate [56]:

Proposition 1. *Let T be the optimal transport map sending the standard Gaussian measure onto a centered log-concave probability measure μ with unit variance. There is a numerical constant c independent of μ such that $T'(x) \leq c\sqrt{1 + |x|^2}$.*

In view of this statement, it is natural to conjecture the following:

Conjecture 3. *The optimal transport map T sending the standard d -dimensional Gaussian measure onto a centered, isotropic log-concave probability measure μ satisfies*

$$\|\nabla T(x)\|_{op} \leq c\sqrt{d + |x|^2}.$$

The guess of the constant d is because the norm $|X|$ of an isotropic log-concave random variable is highly concentrated around its expectation, which is of order \sqrt{d} , so taking something smaller would not change the typical size of the upper bound in this conjecture. A positive answer would imply weighted Poincaré and logarithmic Sobolev inequalities for the Gaussian measure. As an additional motivation, we mention that E. Milman showed that if we can get a universal upper bound on $\int \|\nabla T(x)\|_{op} d\gamma$ for all isotropic log-concave measures, then the KLS conjecture would hold true [133].

With Maria Colombo, we proved the following universal estimate on such transport maps:

Theorem 16 (Colombo & F., in preparation). *The optimal transport map sending the standard Gaussian measure onto a centered isotropic log-concave probability measure satisfies*

$$|T(x)| \leq cd^{1/4}(d + |x|^2)$$

with c a universal constant.

We actually prove an upper bound with a prefactor that depends on the Poincaré constant of the target distribution, the above statement is then obtained by combining with the best known result estimating these constants. It can be seen as a (hopefully non-sharp) integrated form of the regularity bounds we would like to prove.

Chapter 5

Discrete Ricci curvature

The question of how to define a notion of Ricci curvature lower bound for discrete spaces, such as the discrete hypercube was raised in the 1990s by Gromov and Stroock. The motivation was that many consequences of Ricci curvature bounds were known to hold for the hypercube. Several definitions have been proposed in the past fifteen years. In this section, I will discuss some of them, and their consequences.

There are two motivations to study curvature in a discrete setting:

- Using the insight from classical differential geometry to guess and prove discrete geometric statements.
- Develop discrete analogues of the tools used to prove functional inequalities in a continuous setting, and use them to study the mixing time and other properties of Markov chains on discrete spaces.

Several possible definitions of curvature bounds have been proposed, each based on adapting to the discrete setting one particular property that characterizes these bounds in the continuous setting. We shall first discuss entropic Ricci curvature, and Ollivier's coarse Ricci curvature.

5.1 Entropic curvature

The synthetic notion of Ricci curvature lower bound of Lott, Sturm and Villani (LSV) does not rely on a genuine Riemannian structure, but requires the underlying space to have a geodesic structure. This is not the case for discrete spaces, so some modification is required.

In [125, 132, 75, 48], Erbar, Maas, Mielke and Chow *et al.* proposed a natural extension of the LSV definition of curvature. It views the generator of a given reversible Markov chain as the counterpart to the Laplacian for a manifold (which is the generator of Brownian motion). Their next step is to derive a counterpart to the Jordan-Kinderlehrer-Otto point of view on the heat equation as the gradient flow of the entropy with respect to the L^2 Wasserstein distance. Loosely speaking, the main element can be described as follows:

Given a reversible irreducible Markov chain with kernel K on a finite space \mathcal{X} , with invariant probability measure π , there exists a Riemannian distance on the space of probability measures (identified with a simplex) such that the evolution of the law of the Markov chain is the gradient flow of the relative entropy Ent_π with respect to \mathcal{W} .

The distance \mathcal{W} is defined by mimicking the Benamou-Brenier formula, and will play the role of W_2 . Let us recall this formula:

$$W_2(\mu_0, \mu_1)^2 = \inf_{(\mu_t, v_t)} \int_0^1 \int |v_t(x)|^2 d\mu_t dt,$$

where (μ_t, v_t) solves the continuity equation $\dot{\mu}_t + \operatorname{div}(v_t \mu_t) = 0$. To mimic this formula, we must first define the analog of the continuity equation $\dot{\mu} + \operatorname{div}(v\mu) = 0$. The discrete analog of a vector field v would be a function defined on the edges of the graph, i.e. $\Phi : \mathcal{X} \times \mathcal{E} \rightarrow \mathbb{R}$, that represents the flux in each direction. But since it is defined on edges and not on vertices, to define the analog of $\operatorname{div}(v\mu)$ we must associate to the measure μ a measure $\hat{\mu}$ on edges. The divergence is then naturally given by

$$\operatorname{div}(\Phi \hat{\mu})(x) = \sum_{y \in \mathcal{X}} \Phi(x, y) \hat{\mu}(x, y).$$

We can then associate to a path of probability measures μ_t at least one time-dependent discrete vector field Φ_t such that

$$\dot{\mu}_t + \operatorname{div}(\Phi_t \hat{\mu}_t) = 0$$

and then define the discrete transport distance as

$$\mathcal{W}(\mu_0, \mu_1)^2 = \inf_{(\mu_t, \Phi_t)} \int_0^1 \sum_{x, y} \Phi_t(x, y)^2 \hat{\mu}_t(x, y) dt.$$

All that is left is to make explicit the choice of measure $\hat{\mu}$. It turns out that the right choice of $\hat{\mu}$ associated to a measure $\mu = \rho\pi$ is to define it as $\hat{\mu}(x, y) = \Lambda(\rho(x), \rho(y))K(x, y)\pi(x) := \hat{\rho}(x, y)K(x, y)\pi(x)$ with $\Lambda(a, b) := \frac{a-b}{\log a - \log b}$ the logarithmic mean. This choice ensures that the discrete gradient applied to the logarithm of the density satisfies the chain rule

$$\nabla \log \rho(x, y) = \hat{\rho}(x, y) \nabla \log \rho(x, y)$$

which is the discrete counterpart to the usual chain rule for the logarithm. This relation is crucial to ensure that the gradient flow of the relative entropy (which is the expectation of $\log \rho$ with respect to μ) is indeed the discrete heat equation. This gradient flow structure has been adapted to other types of dynamics, such as jump processes [70], spatially homogeneous Boltzmann dynamics [71], discrete mean-field dynamics [73] and quantum Markovian dynamics [36, 37]. It is also a suitable framework for the study of spatial discretization schemes for PDEs [126, 77].

At this point, the definition of Ricci curvature suggested by this approach is clear: we define a lower bound on the (entropic) Ricci curvature as a lower bound on the second derivative of Ent_π along all \mathcal{W} -geodesics.

Several applications of curvature to functional inequalities have immediate discrete counterparts with this definition. The most important one is that a positive lower bound implies a modified logarithmic Sobolev inequality (mLSI):

$$\operatorname{Ent}_\pi(\rho) \leq \frac{1}{2\kappa} \sum_{x \neq y} (\rho(x) - \rho(y)) (\log \rho(x) - \log \rho(y)) K(x, y) \pi(x). \quad (5.1)$$

This inequality corresponds to a bound on the entropy by the entropy production functional for the Markovian dynamic with rates K . In particular, positive curvature implies an exponential rate of convergence to equilibrium in relative entropy for the underlying Markov chain.

Due to the lack of chain rule for discrete derivatives, there are several inequalities that would formally be a discrete counterpart to the classical LSI, and they are not equivalent. What is usually called a LSI for discrete spaces is the inequality

$$Ent_{\pi}(\rho) \leq C \sum_{x \neq y} (\sqrt{\rho(x)} - \sqrt{\rho(y)})^2 K(x, y) \pi(x)$$

which is strictly stronger than the mLSI, and does not necessarily hold with a constant of order κ^{-1} for Markov chains with curvature bounded from below by κ . For example, for the random walk on the complete graph with N sites, the curvature is bounded by a constant independent of N , but the above LSI holds with constant $(\log N)^{-1}$ (while the mLSI constant is indeed of order 1).

As in the continuous setting, the discrete mLSI implies a discrete Poincaré inequality

$$Var_{\pi}(f) \leq \frac{1}{2c} \sum_{x, y} (f(x) - f(y))^2 K(x, y) \pi(x),$$

as well as Gaussian concentration bounds and a discrete version of Talagrand's transport-entropy inequality.

With Erbar [72], we have studied functional inequalities for Markov chains when the curvature is nonnegative, but not strictly positive. The main goal was to get discrete analogs of the results of E. Milman on the equivalence between isoperimetry and concentration under curvature bounds. While we have not been able to prove a full discrete counterpart to that theory, we obtained a few weaker results. The two main results are the following:

Theorem 17. *Consider a Markov chain with nonnegative entropic Ricci curvature.*

- *The sharp constant C_P in the discrete Poincaré inequality and the Cheeger constant*

$$h = \sup_{A \subset \mathcal{X}} \frac{\pi^+(\partial A)}{\pi(A)(1 - \pi(A))}$$

where the perimeter measure of A is given by $\pi^+(\partial A) := \sum_{x \in A, y \in A^c} K(x, y) \pi(x)$, satisfy

$$h \geq \sqrt{9K_* C_P}$$

where $K_* = \min\{K(x, y) : K(x, y) > 0\}$ is the minimal transition rate.

- *The mLSI constant satisfies $\lambda_{mLSI} \geq cD^{-2}$, where D is an upper bound on the diameter (with respect to the distance \mathcal{W}) and c is a universal constant. The same holds for the Poincaré constant.*

The first part is a discrete analog of a theorem of Buser [29], and a matching upper bound up to a universal constant holds in full generality, without any curvature assumption. The Cheeger isoperimetric inequality is equivalent to an L^1 version of the Poincaré inequality. The proof follows a strategy of [109]. The second part combines ideas from [93] and [12], where we first obtain a non-tight mLSI, with an extra constant, using a discrete HWI inequality implied by the curvature bound, and then tighten it by carefully estimating the contribution to the entropy of regions where the function considered takes large values. Both proofs use discrete gradient estimates that arise as a discrete counterpart to Theorem 3.

My second contribution to the study of entropic curvature is about examples of Markov chains with curvature bounds. With Maas, we have worked on how to compute lower bounds for the entropic Ricci

curvature for concrete Markov chains. While the Hessian of the entropy is computable, the expression is somewhat complicated, and directly computing its infimum over a high-dimensional simplex is not feasible in practice. The tensorization property of Ricci curvature allows for computing bounds for high-dimensional non-interacting systems. However, as soon as there is an interaction, this does not apply anymore. By mimicking a strategy proposed by Caputo, Dai Pra and Posta [31] for proving mLSI for discrete Markov chains, we have developed a method for estimating the Hessian of the entropy in presence of non-trivial interactions. One of the main examples for which we were able to derive a bound is the zero-range process on a complete graph:

Theorem 18. *Consider the following system of K interacting particles on the complete graph with L vertices. For each site x , we denote by η_x the number of particles present at a given time. At rate $c_x(\eta_x)$, a single particle chooses another site uniformly at random, and jumps there. Then, assuming that for any x, k we have*

$$0 < c \leq c_x(k+1) - c_x(k) \leq c + \delta$$

with $\delta < c$, the entropic Ricci curvature of the system is bounded from below by $(2c - 5\delta)/4$. In particular, we have positive curvature if $\delta < 2c/5$.

Note that the bound is dimension-free: it depends neither on the size of the graph, nor on the number of particles.

Our method has been pushed further in [74] to cover more examples, such as the Ising model at some high enough temperature. The techniques have also been used in [102, 103] to establish Beckner inequalities, and to study convergence to equilibrium for spatial discretizations of certain PDEs.

5.2 Discrete transport-information inequalities

In joint work with Shu [83], I have also studied functional inequalities under other discrete notions of curvature bounds, namely discrete versions of the Bakry-Emery curvature condition and Ollivier's coarse Ricci curvature condition. To avoid introducing too many definitions, I will mostly focus on our results for the second one, and only briefly describe afterward what we did for the first one.

Ollivier's coarse Ricci curvature for Markov chains on general metric spaces [144, 145] is defined as a contraction rate for the dynamic. More precisely, curvature is bounded from below by a constant $\kappa \in \mathbb{R}$ if for any two probability measures μ and ν and any time $t \geq 0$ we have

$$W_1(P_t^* \mu, P_t^* \nu) \leq e^{-\kappa t}$$

where W_1 is the usual L^1 Wasserstein distance and P_t^* is the semigroup, acting on measures, generated by the Markov dynamic. A similar definition is possible for discrete-time processes. The motivation for this definition is that in the Riemannian setting curvature bounds are equivalent to contraction of the semigroup generated by Brownian motion in Wasserstein distance W_2 . The choice of W_1 instead of W_2 for the coarse definition is motivated by applicability to discrete Markov dynamics. Contraction rates for Markov chains have been studied since Dobrushin's work on the Ising model [62], where it was used as a tool to give bounds on the critical temperature. It has since been revisited for various applications, for example [100]. We remark that this definition is in some sense less intrinsic than the entropic curvature condition, since the choice of the distance is not specified by the Markov process, although in many concrete situations there is a particular natural distance, such as the graph distance. It is sometimes possible to deform the natural distance to make curvature positive, even though at the beginning it was not [64].

In practice, estimating this curvature is done by constructing a well-chosen coupling between arbitrary initial data, and showing that the trajectories indeed get closer at a strictly positive rate. The literature on constructing such couplings is extremely vast, going back to Bubley and Dyer’s work on Markov chain mixing time [28]. However, there is no completely general method, they are usually constructed in an ad hoc way.

When coarse curvature is bounded from below by a strictly positive constant, we can derive a lower bound on the spectral gap, as well as certain exponential concentration estimates [144], with applications for example to error bounds for Markov Chain Monte Carlo algorithms [101]. In general, it seems this curvature condition does not recover the strongest functional inequalities that can be derived using the entropic curvature condition we described in the previous section.

With Yan Shu [83], we proved that for random walks on graphs, if the coarse Ricci curvature is bounded from below by a constant $\kappa > 0$, then the invariant measure π satisfies the transport-information inequality

$$W_1(\rho\pi, \pi)^2 \leq \frac{1}{\kappa^2} \sum_{x,y} (\rho(x) - \rho(y))(\log \rho(x) - \log \rho(y))K(x, y)\pi(x).$$

This inequality is stronger than a transport-entropy inequality, but weaker than the mLSI. It controls the Gaussian concentration for fluctuations of additive functionals of the Markov chain. The proof mimics the Bakry-Emery argument, by adapting it to the discrete setting, where the lack of chain rule causes some issues.

Note that a similar functional inequality where W_1 would be replaced by the stronger W_2 distance would *never* be true in the discrete setting, except in the trivial situation of a measure whose support is a single point.

As a corollary, this result implies a Talagrand inequality \mathcal{T}_1 holds, with a constant proportional to κ^{-1} , and with the transport distance constructed from the standard graph distance. This result was first proved by Eldan, Lee and Lehec [68], using a more information-theoretical approach.

Actually, most of the work in [83] deals with proving transport-information and transport-entropy inequalities under discrete variants of the Bakry-Emery curvature condition [156, 121, 13, 123]. It turns out that under a variant of that condition, the transport-entropy inequality can be improved to control a transport-like distance that is stronger than W_1 (but still weaker than W_2).

5.3 Perspectives

5.3.1 The Peres-Tetali conjecture

One of the main advantage of the entropic curvature condition compared to the coarse curvature condition is that it implies a modified logarithmic Sobolev inequality, and hence exponential convergence to equilibrium in relative entropy. While in full generality it is not true that a coarse Ricci curvature bound implies the same bound on the mLSI constant, Peres and Tetali asked the following question in the restricted setting of random walks:

Conjecture 4. *Consider a simple random walk on some finite graph, and assume that the coarse Ricci curvature is bounded from below by some positive constant κ . Is this enough to guarantee that a modified logarithmic Sobolev inequality holds, with constant of order κ^{-1} ?*

This is actually the question that motivated [68] and [83], but current results only concern weaker transport inequalities, rather than the mLSI.

This conjecture highlights the difference of understanding between the entropic curvature described in Section 5.1 and Ollivier’s coarse Ricci curvature. For the first one, as we have already seen the analogue of this conjecture holds (and the proof is a straightforward adaptation of its continuous counterpart). So as an abstract tool, entropic curvature seems stronger than coarse Ricci curvature. But we have many examples where we know a lower bound on the coarse curvature holds, but do not know how to estimate the entropic curvature. For example, for the Ising model, the temperature threshold at which we can prove positive curvature is much higher for the entropic curvature than for the coarse curvature (where the bound matches with the Dobrushin uniqueness condition [62]).

One of the motivations for raising this question is that a positive answer would imply new bounds on the mixing times of certain algorithms for counting proper colorings of graphs. Moreover, a positive answer to this conjecture would imply that there is no family of expander graph with a uniform positive bound on the coarse Ricci curvature, answering a question of E. Milman and Naor [145].

5.3.2 Geometry and censoring for Markov chains on discrete spaces

To motivate the discussion in this section, let us start by stating a conjecture of Ding and Mossel [61]:

Conjecture 5. *A subset A of the hypercube $\{0, 1\}^d$ is said to be monotone increasing if for any $x \in A$ and any $y \in \{0, 1\}^d$ such that $x_i \leq y_i \forall i = 1, \dots, d$ we have $y \in A$.*

Let A be a monotone increasing subset of the d -dimensional hypercube, with $\pi(A) = c$, where π stands for the uniform probability measure on the hypercube, and consider a censored random walk restricted to A , that is a simple random walk, where any jump attempt that leaves A is suppressed. Is it true that the mixing time of this random walk is bounded by a quantity of the form $\varphi(c)d \log d$, with φ independent of d ?

This conjecture is a particular instance of a much more general question: given a Markov chain, is it possible to modify it by forbidding it to visit certain regions without significantly worsening the rate of convergence to equilibrium?

If we consider diffusions on manifolds, Ricci curvature bounds give a nice sufficient condition for this to happen: if we restrict a diffusion with positive curvature to a geodesically complete subset, then the curvature bound still controls the rate of convergence to equilibrium. Similar results are possible when curvature is only nonnegative, or even slightly negative if the set is not too small, by combining the results of [134] and [21].

At a heuristic level, it would seem like Conjecture 5 should fit in this framework: the discrete hypercube has positive curvature, and given two points inside a monotone set, there is at least one path of minimal length (w.r.t. the graph distance) that remains inside that set. However, if we use entropic curvature, the set of probability measures supported on some monotone set is in general not geodesically complete, and hence a proof would require new ideas. Recently, Erbar, Maas and Wirth [76] have started investigating which subsets of some discrete space support a geodesically complete space of probability measures, with respect to the discrete transport metric \mathcal{W} .

5.3.3 Long-time behavior for mean-field equations on discrete spaces

This section is based on discussions with Matthias Erbar and Andre Schlichting.

One of the main successes of the optimal transport viewpoint on curvature bounds is the study of long-time behavior for McKean-Vlasov equations. Carrillo, McCann and Villani noticed that these dynamics can be written as a gradient flow of a certain entropy functional in the Wasserstein space, and convexity properties can be easily characterized via convexity properties of the underlying potential, in the spirit of

the Bakry-Emery theorem. If we consider a spatial discretization of these PDEs, we are led to consider discrete mean-field equations on graphs, and recently there have been some attempts at understanding their long-time behavior [50, 167].

With Erbar, Laschos and Schlichting, we have uncovered a gradient flow structure for certain families of nonlinear mean-field dynamics on graphs [73]. Just as in the case of linear Markov dynamics, if the driving entropy functional is uniformly convex along geodesics for the associated distance, we automatically deduce a rate of convergence to equilibrium. We are currently working on adapting the techniques of [81, 74] to compute curvature bounds in this setting.

5.3.4 Rigidity of curvature bounds in the discrete setting

One of the active mathematical field that grew out of the study of Ricci curvature bounds is the study of geometric rigidity theorems under curvature bound assumptions, pioneered by Cheeger and his collaborators. A typical statement of this type asserts that if a certain geometric quantity (diameter, volume, spectral gap...) is as large as possible given a certain bound on the (Ricci) curvature, then the manifold under consideration must be of a certain type. We have already seen an example of such a phenomenon in Section 4.2, where we discussed stability of the Bakry-Emery bound on the spectral gap of a weighted Euclidean space.

Since we have notions of Ricci curvature bounds for discrete spaces, it is natural to wonder if there are similar rigidity theorems in the graph setting. At this point, I do not know of any rigidity/stability results for entropic Ricci curvature in discrete spaces.

Let us now present one particular problem. The splitting theorem of Cheeger and Gromoll [45, 44] asserts that a manifold with nonnegative curvature that contains a line is isometric to a product manifold $\mathbb{R} \times M'$. It was extended to the non-smooth setting in [89]. A first guess for a discrete analog would be the following question: consider a Markov chain with nonnegative (entropic) curvature, and assume that there exists a function $f : X \rightarrow \mathbb{Z}$ such that the process $f(X_t)$ is a simple random walk on \mathbb{Z} . Is there a relabeling of X such that the Markov chain is a random walk on a product space $\mathbb{Z} \times X'$, with coordinates that evolve independently conditionally on their initial values, and such that the first coordinate performs a simple random walk, and the second coordinate is a Markov chain with nonnegative (entropic) curvature.

Matthias Erbar pointed out to me that the answer to this question is negative: the simple random walk on the two-dimensional triangle lattice has nonnegative curvature, and its projection on a line performs a simple random walk, yet it is not equivalent up to relabeling to a random walk on a product space with independent coordinates. So the naive guess does not hold. But maybe it is possible to classify the graphs with nonnegative curvature that contain a line?

In a different direction, [122] has recently considered a discrete counterpart to the maximal diameter theorem in the setting of graphs with positive discrete Bakry-Emery curvature and the Lichnerowicz eigenvalue estimate.

Bibliography

- [1] Luigi Ambrosio, Nicola Gigli, and Giuseppe 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.
- [2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Bakry-émery curvature-dimension condition and Riemannian Ricci curvature bounds. *Ann. Probab.*, 43(1):339–404, 2015.
- [3] Shiri Artstein, Keith M. Ball, Franck Barthe, and Assaf Naor. On the rate of convergence in the entropic central limit theorem. *Probab. Theory Related Fields*, 129(3):381–390, 2004.
- [4] Shiri Artstein, Keith M. Ball, Franck Barthe, and Assaf Naor. Solution of Shannon’s problem on the monotonicity of entropy. *J. Amer. Math. Soc.*, 17(4):975–982, 2004.
- [5] S. Artstein-Avidan, B. Klartag, and V. Milman. The Santaló point of a function, and a functional form of the Santaló inequality. *Mathematika*, 51(1-2):33–48 (2005), 2004.
- [6] D. Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.
- [7] D. Bakry and M. Ledoux. Lévy-Gromov’s isoperimetric inequality for an infinite-dimensional diffusion generator. *Invent. Math.*, 123(2):259–281, 1996.
- [8] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.
- [9] Dominique Bakry, Ivan Gentil, and Michel Ledoux. On harnack inequalities and optimal transportation. *Annali della Scuola Normale Superiore di Pisa*, 14:705–727, 2015.
- [10] Keith Ball and Van Hoang Nguyen. Entropy jumps for isotropic log-concave random vectors and spectral gap. *Studia Math.*, 213(1):81–96, 2012.
- [11] Marco Barchiesi, Alessio Brancolini, and Vesa Julin. Sharp dimension free quantitative estimates for the Gaussian isoperimetric inequality. *Ann. Probab.*, 45(2):668–697, 2017.
- [12] Franck Barthe and Alexander V. Kolesnikov. Mass transport and variants of the logarithmic Sobolev inequality. *J. Geom. Anal.*, 18(4):921–979, 2008.
- [13] Frank Bauer, Paul Horn, Yong Lin, Gabor Lippner, Dan Mangoubi, and Shing-Tung Yau. Li-Yau inequality on graphs. *J. Differential Geom.*, 99(3):359–405, 2015.

- [14] Jean-David Benamou and Yann Brenier. A numerical method for the optimal time-continuous mass transport problem and related problems. In *Monge Ampère equation: applications to geometry and optimization (Deerfield Beach, FL, 1997)*, volume 226 of *Contemp. Math.*, pages 1–11. Amer. Math. Soc., Providence, RI, 1999.
- [15] Robert J. Berman and Bo Berndtsson. Real Monge-Ampère equations and Kähler-Ricci solitons on toric log Fano varieties. *Ann. Fac. Sci. Toulouse Math. (6)*, 22(4):649–711, 2013.
- [16] Philippe Biane and Roland Speicher. Free diffusions, free entropy and free Fisher information. *Ann. Inst. H. Poincaré Probab. Statist.*, 37(5):581–606, 2001.
- [17] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.
- [18] S. G. Bobkov, N. Gozlan, C. Roberto, and P.-M. Samson. Bounds on the deficit in the logarithmic Sobolev inequality. *J. Funct. Anal.*, 267(11):4110–4138, 2014.
- [19] Sergey G. Bobkov. Berry-Esseen bounds and Edgeworth expansions in the central limit theorem for transport distances. *Probab. Theory Related Fields*, 170(1-2):229–262, 2018.
- [20] Sergey G. Bobkov, Gennadiy P. Chistyakov, and Friedrich Götze. Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. *Ann. Probab.*, 41(4):2479–2512, 2013.
- [21] Sergey G. Bobkov, Piotr Nayar, and Prasad Tetali. Concentration properties of restricted measures with applications to non-Lipschitz functions. In *Geometric aspects of functional analysis*, volume 2169 of *Lecture Notes in Math.*, pages 25–53. Springer, Cham, 2017.
- [22] E. Bolthausen. An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete*, 66(3):379–386, 1984.
- [23] Thomas Bonis. Rates in the central limit theorem and diffusion approximation via Stein’s method. *Arxiv preprint*, 2018.
- [24] A. A. Borovkov and S. A. Utev. An inequality and a characterization of the normal distribution connected with it. *Teor. Veroyatnost. i Primenen.*, 28(2):209–218, 1983.
- [25] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.
- [26] Lorenzo Brasco and Guido De Philippis. Spectral inequalities in quantitative form. In *Shape optimization and spectral theory*, pages 201–281. De Gruyter Open, Warsaw, 2017.
- [27] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.
- [28] R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in markov chains. in *38th Annual Symposium on Foundations of Computer Science, Los Alamitos, CA, IEEE*, page pp. 223231., 1997.
- [29] Peter Buser. A note on the isoperimetric constant. *Ann. Sci. École Norm. Sup. (4)*, 15(2):213–230, 1982.

- [30] Luis A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.*, 214(3):547–563, 2000.
- [31] Pietro Caputo, Paolo Dai Pra, and Gustavo Posta. Convex entropy decay via the Bochner-Bakry-Emery approach. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):734–753, 2009.
- [32] E. A. Carlen and A. Soffer. Entropy production by block variable summation and central limit theorems. *Comm. Math. Phys.*, 140(2):339–371, 1991.
- [33] Eric A. Carlen. Superadditivity of Fisher’s information and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 101(1):194–211, 1991.
- [34] Eric A. Carlen and Alessio Figalli. Stability for a GNS inequality and the log-HLS inequality, with application to the critical mass Keller-Segel equation. *Duke Math. J.*, 162(3):579–625, 2013.
- [35] Eric A. Carlen, Rupert L. Frank, and Elliott H. Lieb. Stability estimates for the lowest eigenvalue of a Schrödinger operator. *Geom. Funct. Anal.*, 24(1):63–84, 2014.
- [36] Eric A. Carlen and Jan Maas. An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker-Planck equation is gradient flow for the entropy. *Comm. Math. Phys.*, 331(3):887–926, 2014.
- [37] Eric A. Carlen and Jan Maas. Gradient flow and entropy inequalities for quantum Markov semi-groups with detailed balance. *J. Funct. Anal.*, 273(5):1810–1869, 2017.
- [38] Patrick Cattiaux and Arnaud Guillin. Functional inequalities via Lyapunov conditions. In *Optimal transportation*, volume 413 of *London Math. Soc. Lecture Note Ser.*, pages 274–287. Cambridge Univ. Press, Cambridge, 2014.
- [39] Djalil Chafaï. From Boltzmann to random matrices and beyond. *Ann. Fac. Sci. Toulouse Math. (6)*, 24(4):641–689, 2015.
- [40] Sourav Chatterjee. Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields*, 143(1-2):1–40, 2009.
- [41] Sourav Chatterjee. Spin glasses and Stein’s method. *Probab. Theory Related Fields*, 148(3-4):567–600, 2010.
- [42] Sourav Chatterjee. A short survey of Stein’s method. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV*, pages 1–24. Kyung Moon Sa, Seoul, 2014.
- [43] Sourav Chatterjee and Elizabeth Meckes. Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 4:257–283, 2008.
- [44] Jeff Cheeger and Tobias H. Colding. Lower bounds on Ricci curvature and the almost rigidity of warped products. *Ann. of Math. (2)*, 144(1):189–237, 1996.
- [45] Jeff Cheeger and Detlef Gromoll. The splitting theorem for manifolds of nonnegative Ricci curvature. *J. Differential Geometry*, 6:119–128, 1971/72.
- [46] L. H. Y. Chen. Poisson approximation for dependent trials. *Ann. Probab.*, 3(3):534–545, 1975.
- [47] Xu Cheng and Detang Zhou. Eigenvalues of the drifted Laplacian on complete metric measure spaces. *Commun. Contemp. Math.*, 19(1):1650001, 17, 2017.

- [48] Shui-Nee Chow, Wen Huang, Yao Li, and Haomin Zhou. Fokker-Planck equations for a free energy functional or Markov process on a graph. *Arch. Ration. Mech. Anal.*, 203(3):969–1008, 2012.
- [49] Bertrand Cloez and Claire Delplancke. Intertwinings and stein’s magic factors for birth-death processes. *to appear in Ann. IHP: Probab. Stat.*, 2018.
- [50] Bertrand Cloez and Marie-Noémie Thai. Quantitative results for the fleming-viot particle system in discrete space. *Stoch. Proc. Appl.*, 2013.
- [51] D. Cordero-Erausquin and B. Klartag. Moment measures. *J. Funct. Anal.*, 268(12):3834–3866, 2015.
- [52] Dario Cordero-Erausquin. Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.*, 161(3):257–269, 2002.
- [53] Dario Cordero-Erausquin. Transport inequalities for log-concave measures, quantitative forms, and applications. *Canad. J. Math.*, 69(3):481–501, 2017.
- [54] 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.
- [55] T.A. Courtade and M. Fathi. Stability of the Bakry-Emery theorem on \mathbb{R}^n . *Arxiv preprint*, 2018.
- [56] T.A. Courtade, A. Pananjady, and M. Fathi. Existence of stein kernels under a spectral gap, and discrepancy bounds. *to appear in Ann. IHP Probab. Stat.*, 2018.
- [57] T.A. Courtade, A. Pananjady, and M. Fathi. Wasserstein stability of the entropy power inequality for log-concave densities. *IEEE Transactions on Information Theory*, 64(8):5691–5703, 2018.
- [58] Thomas A. Courtade. A strong entropy power inequality. *IEEE Trans. Inform. Theory*, 64(4, part 1):2173–2192, 2018.
- [59] Guido De Philippis and Alessio Figalli. Rigidity and stability of Caffarelli’s log-concave perturbation theorem. *Nonlinear Anal.*, 154:59–70, 2017.
- [60] Amir Dembo, Abram Kagan, and Lawrence A. Shepp. Remarks on the maximum correlation coefficient. *Bernoulli*, 7(2):343–350, 2001.
- [61] Jian Ding and Elchanan Mossel. Mixing under monotone censoring. *Electron. Commun. Probab.*, 19:no. 46, 6, 2014.
- [62] R. L. Dobrusin. Definition of a system of random variables by means of conditional distributions. *Teor. Veroyatnost. i Primenen.*, 15:469–497, 1970.
- [63] Jean Dolbeault and Giuseppe Toscani. Stability results for logarithmic Sobolev and Gagliardo-Nirenberg inequalities. *Int. Math. Res. Not. IMRN*, (2):473–498, 2016.
- [64] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields*, 166(3-4):851–886, 2016.
- [65] Ronen Eldan. A two-sided estimate for the Gaussian noise stability deficit. *Invent. Math.*, 201(2):561–624, 2015.

- [66] Ronen Eldan. Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *to appear in Geom. Funct. Anal.*, 2018.
- [67] Ronen Eldan and Bo‘az Klartag. Dimensionality and the stability of the Brunn-Minkowski inequality. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 13(4):975–1007, 2014.
- [68] Ronen Eldan, James R. Lee, and Joseph Lehec. Transport-entropy inequalities and curvature in discrete-space Markov chains. In *A journey through discrete mathematics*, pages 391–406. Springer, Cham, 2017.
- [69] Ronen Eldan, Dan Mikulincer, and Alex Zhai. The CLT in high dimensions: quantitative bounds via martingale embedding. *Arxiv preprint*, 2018.
- [70] Matthias Erbar. Gradient flows of the entropy for jump processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 50(3):920–945, 2014.
- [71] Matthias Erbar. A gradient flow approach to the Boltzmann equation. *Arxiv preprint*, 2016.
- [72] Matthias Erbar and Max Fathi. Poincaré, modified logarithmic Sobolev and isoperimetric inequalities for Markov chains with non-negative Ricci curvature. *J. Funct. Anal.*, 274(11):3056–3089, 2018.
- [73] Matthias Erbar, Max Fathi, Vaios Laschos, and André Schlichting. Gradient flow structure for McKean-Vlasov equations on discrete spaces. *Discrete Contin. Dyn. Syst.*, 36(12):6799–6833, 2016.
- [74] Matthias Erbar, Chris Henderson, Georg Menz, and Prasad Tetali. Ricci bounds for weakly interacting markov chains. *Electron. J. Probabl.*, 22(40):1–23, 2017.
- [75] Matthias Erbar and Jan Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.*, 206(3):997–1038, 2012.
- [76] Matthias Erbar, Jan Maas, and Melchior Wirth. On the geometry of geodesics in discrete optimal transport. *Arxiv preprint*, 2018.
- [77] Matthias Erbar, Martin Rumpf, Bernhard Schmitzer, and Stefan Simon. Computation of optimal transport on discrete metric measure spaces. *Arxiv preprint*, 2017.
- [78] Max Fathi. A sharp symmetrized form of talagrand’s transport-entropy inequality for the gaussian measure. *to appear in Elect. Comm. Probab.*, 2018.
- [79] Max Fathi. Stein kernels and moment maps. *to appear in Ann. Probab.*, 2018.
- [80] Max Fathi, Emanuel Indrei, and Michel Ledoux. Quantitative logarithmic Sobolev inequalities and stability estimates. *Discrete Contin. Dyn. Syst.*, 36(12):6835–6853, 2016.
- [81] Max Fathi and Jan Maas. Entropic Ricci curvature bounds for discrete interacting systems. *Ann. Appl. Probab.*, 26(3):1774–1806, 2016.
- [82] Max Fathi and Brent Nelson. Free Stein kernels and an improvement of the free logarithmic Sobolev inequality. *Adv. Math.*, 317:193–223, 2017.
- [83] Max Fathi and Yan Shu. Curvature and transport inequalities for Markov chains in discrete spaces. *Bernoulli*, 24(1):672–698, 2018.

- [84] F. Feo, E. Indrei, M. R. Posteraro, and C. Roberto. Some remarks on the stability of the log-Sobolev inequality for the Gaussian measure. *Potential Anal.*, 47(1):37–52, 2017.
- [85] Alessio Figalli. Stability in geometric and functional inequalities. *Proceedings of the 6th European Congress of Mathematics*, 2012.
- [86] Alessio Figalli and David Jerison. Quantitative stability for the Brunn-Minkowski inequality. *Adv. Math.*, 314:1–47, 2017.
- [87] Alessio Figalli, Francesco Maggi, and Connor Mooney. The sharp quantitative Euclidean concentration inequality. *Camb. J. Math.*, 6(1):59–87, 2018.
- [88] Alessio Figalli, Francesco Maggi, and Aldo Pratelli. A mass transportation approach to quantitative isoperimetric inequalities. *Invent. Math.*, 182(1):167–211, 2010.
- [89] Nicola Gigli. An overview of the proof of the splitting theorem in spaces with non-negative Ricci curvature. *Anal. Geom. Metr. Spaces*, 2:169–213, 2014.
- [90] Larry Goldstein and Gesine Reinert. Stein’s method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.*, 7(4):935–952, 1997.
- [91] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Process. Related Fields*, 16(4):635–736, 2010.
- [92] Nathael Gozlan. A characterization of dimension free concentration in terms of transportation inequalities. *Ann. Probab.*, 37(6):2480–2498, 2009.
- [93] Nathael Gozlan, Cyril Roberto, and Paul-Marie Samson. From concentration to logarithmic Sobolev and Poincaré inequalities. *J. Funct. Anal.*, 260(5):1491–1522, 2011.
- [94] Leonard Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.
- [95] Natalie Grunewald, Felix Otto, Cédric Villani, and Maria G. Westdickenberg. A two-scale approach to logarithmic Sobolev inequalities and the hydrodynamic limit. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(2):302–351, 2009.
- [96] Alice Guionnet and Dmitri Shlyakhtenko. Free monotone transport. *Invent. Math.*, 197(3):613–661, 2014.
- [97] Emanuel Indrei and Daesung Kim. Deficit estimates for the logarithmic sobolev inequality. *Arxiv preprint*, 2018.
- [98] Emanuel Indrei and Diego Marcon. A quantitative log-Sobolev inequality for a two parameter family of functions. *Int. Math. Res. Not. IMRN*, (20):5563–5580, 2014.
- [99] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.
- [100] Aldéric Joulin. A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature. *Bernoulli*, 15(2):532–549, 2009.
- [101] Aldéric Joulin and Yann Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.*, 38(6):2418–2442, 2010.

- [102] Ansgar Jüngel and Stefan Schuchnigg. A discrete Bakry-Emery method and its application to the porous-medium equation. *Discrete Contin. Dyn. Syst.*, 37(11):5541–5560, 2017.
- [103] Ansgar Jüngel and Wen Yue. Discrete Beckner inequalities via the Bochner-Bakry-Emery approach for Markov chains. *Ann. Appl. Probab.*, 27(4):2238–2269, 2017.
- [104] Claude Kipnis and Claudio Landim. *Scaling limits of interacting particle systems*, volume 320 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.
- [105] Claude Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1):1–19, 1986.
- [106] Bo’az Klartag. A central limit theorem for convex sets. *Invent. Math.*, 168(1):91–131, 2007.
- [107] Bo’az Klartag. Logarithmically-concave moment measures I. In *Geometric aspects of functional analysis*, volume 2116 of *Lecture Notes in Math.*, pages 231–260. Springer, Cham, 2014.
- [108] Alexander V. Kolesnikov. Hessian metrics, $CD(K, N)$ -spaces, and optimal transportation of log-concave measures. *Discrete Contin. Dyn. Syst.*, 34(4):1511–1532, 2014.
- [109] Michel Ledoux. A simple analytic proof of an inequality by P. Buser. *Proc. Amer. Math. Soc.*, 121(3):951–959, 1994.
- [110] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
- [111] Michel Ledoux. Chaos of a Markov operator and the fourth moment condition. *Ann. Probab.*, 40(6):2439–2459, 2012.
- [112] Michel Ledoux, Ivan Nourdin, and Giovanni Peccati. Stein’s method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.*, 25(1):256–306, 2015.
- [113] Michel Ledoux, Ivan Nourdin, and Giovanni Peccati. A Stein deficit for the logarithmic Sobolev inequality. *Sci. China Math.*, 60(7):1163–1180, 2017.
- [114] Yin Tat Lee and Santosh S. Vempala. Eldan’s stochastic localization and the KLS hyperplane conjecture: an improved lower bound for expansion. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017*, pages 998–1007. IEEE Computer Soc., Los Alamitos, CA, 2017.
- [115] Yin Tat Lee and Santosh S. Vempala. The kannan-lovász-simonovits conjecture. *Arxiv preprint*, 2017.
- [116] Joseph Lehec. Representation formula for the entropy and functional inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(3):885–899, 2013.
- [117] Tony Lelièvre. A general two-scale criteria for logarithmic Sobolev inequalities. *J. Funct. Anal.*, 256(7):2211–2221, 2009.
- [118] Tony Lelièvre and Gabriel Stoltz. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numer.*, 25:681–880, 2016.
- [119] Christophe Ley, Gesine Reinert, and Yvik Swan. Stein’s method for comparison of univariate distributions. *Probab. Surv.*, 14:1–52, 2017.

- [120] Peter Li and Shing-Tung Yau. On the parabolic kernel of the Schrödinger operator. *Acta Math.*, 156(3-4):153–201, 1986.
- [121] Yong Lin and Shing-Tung Yau. Ricci curvature and eigenvalue estimate on locally finite graphs. *Math. Res. Lett.*, 17(2):343–356, 2010.
- [122] Shiping Liu, Florentin Münch, and Norbert Peyerimhoff. Rigidity properties of the hypercube via bakry-emery curvature. *arxiv preprint*, 2017.
- [123] Shiping Liu, Florentin Münch, and Norbert Peyerimhoff. Bakry-émery curvature and diameter bounds on graphs. *Calc. Var. Partial Differential Equations*, 57(2):Art. 67, 9, 2018.
- [124] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.
- [125] Jan Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.
- [126] Jan Maas and Daniel Matthes. Long-time behavior of a finite volume discretization for a fourth order diffusion equation. *Nonlinearity*, 29(7):1992–2023, 2016.
- [127] Francesco Maggi and Cédric Villani. Balls have the worst best Sobolev inequalities. *J. Geom. Anal.*, 15(1):83–121, 2005.
- [128] Katalin Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3):556–571, 1996.
- [129] Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.
- [130] Robert J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.
- [131] Elizabeth Meckes. On the approximate normality of eigenfunctions of the Laplacian. *Trans. Amer. Math. Soc.*, 361(10):5377–5399, 2009.
- [132] Alexander Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Partial Differential Equations*, 48(1-2):1–31, 2013.
- [133] Emanuel Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.*, 177(1):1–43, 2009.
- [134] Emanuel Milman. Isoperimetric and concentration inequalities: equivalence under curvature lower bound. *Duke Math. J.*, 154(2):207–239, 2010.
- [135] V. D. Milman. New proof of the theorem of Dvoretzky on sections of convex bodies. *Funct. Anal. Appl.*, 5:28–37, 1971.
- [136] G. Monge. *Mémoire sur la théorie des déblais et des remblais*. 1784.
- [137] Benjamin Muckenhoupt. Hardy’s inequality with weights. *Studia Math.*, 44:31–38, 1972. Collection of articles honoring the completion by Antoni Zygmund of 50 years of scientific activity, I.
- [138] Randall Munroe. Number line. <https://xkcd.com/899/>.

- [139] Ivan Nourdin and Giovanni Peccati. Stein’s method on Wiener chaos. *Probab. Theory Related Fields*, 145(1-2):75–118, 2009.
- [140] Ivan Nourdin and Giovanni Peccati. *Normal approximations with Malliavin calculus*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2012. From Stein’s method to universality.
- [141] Ivan Nourdin, Giovanni Peccati, and Anthony Réveillac. Multivariate normal approximation using Stein’s method and Malliavin calculus. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(1):45–58, 2010.
- [142] Ivan Nourdin, Giovanni Peccati, and Yvik Swan. Entropy and the fourth moment phenomenon. *J. Funct. Anal.*, 266(5):3170–3207, 2014.
- [143] David Nualart and Giovanni Peccati. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.*, 33(1):177–193, 2005.
- [144] Yann Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.
- [145] Yann Ollivier. A survey of Ricci curvature for metric spaces and Markov chains. In *Probabilistic approach to geometry*, volume 57 of *Adv. Stud. Pure Math.*, pages 343–381. Math. Soc. Japan, Tokyo, 2010.
- [146] 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.
- [147] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.
- [148] Yury Polyanskiy and Yihong Wu. Wasserstein continuity of entropy and outer bounds for interference channels. *IEEE Trans. Inform. Theory*, 62(7):3992–4002, 2016.
- [149] Tobias Povel. Confinement of Brownian motion among Poissonian obstacles in \mathbf{R}^d , $d \geq 3$. *Probab. Theory Related Fields*, 114(2):177–205, 1999.
- [150] M. Raginsky and I. Sason. Concentration of measure inequalities in information theory, communications, and coding. *Foundations and Trends in Communications and Information Theory*, 10(1-2):1–246, 2013.
- [151] Emmanuel Rio. Upper bounds for minimal distances in the central limit theorem. *Ann. IHP: Probab. Stat.*, 45(3):802–817, 2009.
- [152] Olivier Rioul. Yet another proof of the entropy power inequality. *IEEE Trans. Inform. Theory*, 63(6):3595–3599, 2017.
- [153] Nathan Ross. Fundamentals of Stein’s method. *Probab. Surv.*, 8:210–293, 2011.
- [154] Filippo Santambrogio. Dealing with moment measures via entropy and optimal transport. *J. Funct. Anal.*, 271(2):418–436, 2016.
- [155] Adrien Saumard. Weighted poincaré inequalities, concentration inequalities and tail bounds related to the behavior of the stein kernel in dimension one. *Arxiv preprint*, 2018.

- [156] Michael Schmuckenschläger. Curvature of nonlocal Markov generators. In *Convex geometric analysis (Berkeley, CA, 1996)*, volume 34 of *Math. Sci. Res. Inst. Publ.*, pages 189–197. Cambridge Univ. Press, Cambridge, 1999.
- [157] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, Ill., 1949.
- [158] Dimitri Shlyakhtenko. Shannon’s monotonicity problem for free and classical entropy. *Proc. Natl. Acad. Sci. USA*, 104(39):15254–15258, 2007. With an appendix by Hanne Schultz.
- [159] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602. Univ. California Press, Berkeley, Calif., 1972.
- [160] Charles Stein. *Approximate computation of expectations*, volume 7 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1986.
- [161] Karl-Theodor Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.
- [162] Karl-Theodor Sturm. On the geometry of metric measure spaces. II. *Acta Math.*, 196(1):133–177, 2006.
- [163] Alain-Sol Sznitman. Fluctuations of principal eigenvalues and random scales. *Comm. Math. Phys.*, 189(2):337–363, 1997.
- [164] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.
- [165] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [166] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
- [167] Denis Villemonais. Lower bound for the coarse Ricci curvature of continuous-time pure jump processes. *Arxiv preprint*, 2017.
- [168] Max-K. von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.
- [169] Xu-Jia Wang and Xiaohua Zhu. Kähler-Ricci solitons on toric manifolds with positive first Chern class. *Adv. Math.*, 188(1):87–103, 2004.
- [170] Zach Weinersmith. <http://www.smbc-comics.com/comic/angles>.
- [171] Boguslaw Zegarliński. The strong decay to equilibrium for the stochastic dynamics of unbounded spin systems on a lattice. *Comm. Math. Phys.*, 175(2):401–432, 1996.