

## SOME RECENT DEVELOPMENTS IN FUNCTIONAL INEQUALITIES \*

EMMANUEL BOISSARD<sup>1</sup>, NATHAEL GOZLAN<sup>2</sup>, JOSEPH LEHEC<sup>3</sup>, CHRISTIAN LÉONARD<sup>4</sup>,  
GEORG MENZ<sup>5</sup> AND ANDRÉ SCHLICHTING<sup>6</sup>

**Abstract.** We present new directions of investigations of several classical functional inequalities.

### 1. INTRODUCTION

This paper presents different recent directions in the study of some classical functional inequalities: the Poincaré inequality, the logarithmic Sobolev inequality and Talagrand’s quadratic transport-entropy inequality. The results presented in this text were the subjects of a series of talks given by the authors in the session “Functional inequalities” during the Journées MAS 2012 in Clermont-Ferrand.

Let us recall the definitions of the above mentioned inequalities and fix some notation. In what follows,  $\mu$  is a probability measure on some smooth Riemannian manifold  $(\mathcal{X}, g)$  equipped with its geodesic distance  $d$ . In all the paper,  $P(Y)$  will always denote the space of Borel probability measures on a measurable space  $Y$ .

The Poincaré inequality compares the variance of a function  $f$  to the  $L_2$  norm of its derivative:  $\mu$  verifies Poincaré inequality with constant  $C \geq 0$ , in short **PI**( $C$ ), if

$$C \operatorname{Var}_\mu(f) \leq \int |\nabla f|^2 d\mu, \quad (1.1)$$

for all  $f$  smooth enough.

---

\* The second and the fourth authors were partially supported by the ANR project GeMeCoD. ANR 2011 BS01 007 01

<sup>1</sup> Emmanuel Boissard  
Weierstrass Institute, Mohrenstrasse 39, 10117 Berlin, Germany

<sup>2</sup> Nathael Gozlan  
Université Paris-Est, LAMA (UMR 8050), UPEMLV, UPEC, CNRS, F-77454, Marne-la-Vallée, France

<sup>3</sup> Joseph Lehec  
CEREMADE (UMR CNRS 7534) Université Paris-Dauphine, Place de Lattre de Tassigny, 75016 Paris

<sup>4</sup> Christian Léonard  
Modal-X. Université Paris Ouest. Bât. G, 200 av. de la République. 92001 Nanterre, France

<sup>5</sup> Georg Menz  
Stanford University, Department of Mathematics, 450 Serra Mall, Stanford, California 94305

<sup>6</sup> André Schlichting  
Institute for Applied Mathematics, University of Bonn, Endenicher Allee 60, D-53115 Bonn

In the same spirit, the logarithmic Sobolev inequality gives a similar comparison involving the entropy functional instead of the variance:  $\mu$  verifies the logarithmic Sobolev inequality  $\mathbf{LSI}(C)$ ,  $C \geq 0$  if

$$\text{Ent}_\mu(f^2) := \int f^2 \log \left( \frac{f^2}{\int f^2 d\mu} \right) d\mu \leq C \int |\nabla f|^2 d\mu, \tag{1.2}$$

for all smooth enough function  $f$ .

Finally, Talagrand’s inequality compares the quadratic transport cost between  $\mu$  and another probability measure  $\nu$  to the relative entropy of  $\nu$  with respect to  $\mu$ :  $\mu$  verifies Talagrand’s inequality  $\mathbf{T}_2(C)$ ,  $C \geq 0$ , if

$$W_2^2(\nu, \mu) \leq CH(\nu|\mu), \quad \forall \nu \in \mathbf{P}(\mathcal{X}). \tag{1.3}$$

In the equation above,  $W_2$  denotes the quadratic Wasserstein distance. Recall that for  $p \geq 1$ ,  $W_p$  is defined, for all  $\mu_0, \mu_1 \in \mathbf{P}(\mathcal{X})$ , by

$$W_p^p(\mu_0, \mu_1) := \inf \left\{ \int_{\mathcal{X}^2} d^p(x, y) \pi(dx dy); \pi \in \mathbf{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1 \right\} \tag{1.4}$$

where  $\pi_0(dx) := \pi(dx \times \mathcal{X})$  and  $\pi_1(dy) := \pi(\mathcal{X} \times dy)$  are the marginal measures on  $\mathcal{X}$  of  $\pi \in \mathbf{P}(\mathcal{X}^2)$ . The relative entropy  $H(\nu|m)$  of a probability measure  $\nu$  with respect to a general positive measure  $m$  is defined by

$$H(\nu|m) = \int \log \frac{d\nu}{dm} d\nu, \tag{1.5}$$

if  $\nu$  is absolutely continuous with respect to  $m$  and  $+\infty$  otherwise. When  $m$  is a probability measure,  $H(\nu|m) \geq 0$ .

We refer to [BE85], [ABC<sup>+</sup>00], [Led01], [Roy07], [GL10], [Vil03, Vil09], [BGL13] for general introductions to these functional inequalities and their applications to the study of Markov semigroups, concentration of measure, statistical mechanics, geometric inequalities... Let us just mention the following well known results. First, the basic hierarchy between these three inequalities, which was discovered by Otto and Villani in [OV00], is the following

$$\mathbf{LSI}(C) \Rightarrow \mathbf{T}_2(C) \Rightarrow \mathbf{PI}(2/C).$$

Second, a well known sufficient condition for LSI is the Bakry-Emery condition  $\text{CD}(\kappa, \infty)$  [BE85]: if  $\mu = e^{-V(x)} dx$ , where  $dx$  is the Riemannian measure and  $V$  a function of class  $\mathcal{C}^2$  verifies

$$\text{Ric} + \text{Hess } V \geq \kappa, \tag{1.6}$$

then  $\mu$  verifies  $\mathbf{LSI}(2/\kappa)$ . Finally, let us mention that the works by Lott-Villani [LV09] and Sturm [Stu06a, Stu06b] introduced a new definition of the condition  $\text{CD}(\kappa, \infty)$  that makes sense for general geodesic metric spaces. This new definition is formulated in terms of convexity properties of the relative entropy functional along the geodesics for the Wasserstein distance  $W_2$  (see Section 2 for more informations). When the metric space is a Riemannian manifold, the  $\text{CD}(\kappa, \infty)$  condition of Lott-Sturm-Villani coincides with the Bakry-Emery condition (1.6).

The paper is organized as follows.

Section 2 contains the contribution by Christian Léonard based on his papers [Léo<sub>a</sub>, Léo<sub>b</sub>, Léo<sub>c</sub>]. The general question considered there is how to extend the  $\text{CD}(\kappa, \infty)$  condition to discrete spaces. Indeed, for such spaces, the Lott-Sturm-Villani  $\text{CD}(\kappa, \infty)$  condition can not be satisfied. Section 2 presents some new ideas related to discrete Markov processes that could yield to an interesting discrete version of  $\text{CD}(\kappa, \infty)$ .

Section 3 contains the contribution of Emmanuel Boissard based on his paper [BLG12] (joint with Thomas Le Gouic) and deals with the question of finding a good approximation of a continuous distribution by a discrete one. The quality of the approximation is measured with respect to the Wasserstein distances  $W_p$ . Transport inequalities play a role to give some quantified version of Varadarajan theorem, namely to show that with high probability the empirical distribution of an i.i.d sequence is close in  $W_2$  distance to the reference probability.

Section 4 contains Joseph Lehec's contribution based on [Leh12]. A representation formula of the relative entropy with respect to the Wiener measure is used to provide unified proofs of several classical functional inequalities for Gaussian measures. In Section 4, the technique is illustrated by showing a short proof of the  $T_2$  inequality for the Wiener process.

Section 5 contains the contribution by Georg Menz and André Schlichting based on [MS12]. The question is to give a precise asymptotic for the Log-Sobolev or Poincaré constants of a Gibbs measure  $\mu_T(dx) = \frac{1}{Z_T} e^{-H(x)/T} dx$  when  $T \rightarrow 0$  (low temperature). Under different regularity assumptions, the Eyring-Kramer [GLE41] formula gives the first order behavior of the Poincaré constant. In section 5, an alternative proof of this formula is sketched using a variance decomposition formula and a transport argument.

#### *Acknowledgements*

Many thanks to the organizers of the *Journées MAS 2012* at Clermont-Ferrand, with a special mention for Arnaud Guillin.

## 2. TWO TRANSFORMS OF MARKOV PROCESSES RELATED TO CURVATURE

### **Introduction**

We sketch some ideas related to the problem of defining a type of Ricci curvature for metric measure spaces. We especially have in mind the case of a discrete space. Our guideline is standard: we rely on convexity properties of some relative entropy along interpolations. Our thesis is that not only displacement interpolations (related to optimal transport), but also *entropic* interpolations (related to the Schrödinger problem) could be useful for the purpose of understanding curvature. To construct these entropic interpolations, we introduce two transforms: (i) the  $(f, g)$ -transform of a Markov measure and (ii) a slowing down procedure which allows to recover displacement interpolations as limits of entropic ones.

At Section 2.1, basic facts about the Lott-Sturm-Villani theory are recalled. Then, we give some informal arguments at Section 2.2 in favor of a specific construction of some displacement interpolation on an abstract space which admits some reversible Markov process. Two transforms of Markov measures appear naturally. The first one which we call  $(f, g)$ -transform, is developed at Section 2.3 and the second one, which is a slowing down procedure, is commented on at Section 2.4.

Most of the present material comes from the papers [Léo12, Léoa, Léob]. Very little about literature is proposed; more is given in the previously cited papers.

### **2.1. Lower bounded curvature and its connections with some functional inequalities**

We recall the basis of the Lott-Sturm-Villani theory of lower bounded curvature on length spaces. The reference textbook on the subject is Villani's textbook [Vil09].

#### *Basic facts about the Lott-Sturm-Villani theory*

Let  $\mathcal{X}$  be a Riemannian manifold with Riemannian distance  $d$ . The Wasserstein pseudo-distance  $W_2$  of order 2 between two probability measures  $\mu_0$  and  $\mu_1$  on  $\mathcal{P}(\mathcal{X})$  is defined by (1.4). In restriction to  $\mathcal{P}_2(\mathcal{X}) := \{p \in \mathcal{P}(\mathcal{X}); \int_{\mathcal{X}} d^2(x_o, x) p(dx) < \infty\}$ , for some  $x_o \in \mathcal{X}$ ,  $W_2$  is a distance. Suppose for simplicity that  $\mu_0, \mu_1 \in \mathcal{P}_2(\mathcal{X})$  are such that the infimum in the expression of  $W_2^2(\mu_0, \mu_1)$  is uniquely achieved (for instance when  $\mu_0$  or  $\mu_1$  is absolutely continuous). Denote  $\hat{\pi} \in \mathcal{P}(\mathcal{X}^2)$  this optimal coupling and suppose in addition that  $\hat{\pi}$  doesn't

give any mass to the cut-locus. Then, the  $W_2$ -displacement interpolation  $[\mu_0, \mu_1] := (\mu_t)_{t \in [0,1]} \in P_2(\mathcal{X})^{[0,1]}$  between  $\mu_0$  and  $\mu_1$  is unique and is given by

$$\mu_t(dz) = \int_{\mathcal{X}^2} \delta_{\gamma_t^{xy}} \widehat{\pi}(dxdy), \quad 0 \leq t \leq 1 \tag{2.1}$$

where for any  $x; y \in \mathcal{X}$ ,  $\gamma^{xy}$  is the constant speed geodesic path between  $x$  and  $y$ . It follows from  $d(\gamma_s^{xy}, \gamma_t^{xy}) = |t - s|d(\gamma_s^{xy}, \gamma_t^{xy})$ ,  $0 \leq s, t \leq 1$ , that

$$W_2(\mu_s, \mu_t) = |t - s|W_2(\mu_0, \mu_1), \quad 0 \leq s, t \leq 1,$$

meaning that  $[\mu_0, \mu_1]$  is a constant speed geodesic path on  $(P_2(\mathcal{X}), W_2)$ . This interpolation was introduced by McCann in [McC97].

Recall that the relative entropy functional  $H$  is defined by (1.5).

**Theorem 2.1** (Sturm-von Renesse, [Sv05]). *The compact Riemannian manifold  $\mathcal{X}$  has a Ricci curvature bounded from below by  $\kappa \in \mathbb{R}$  if and only if for any  $W_2$ -displacement interpolation  $[\mu_0, \mu_1]$ ,  $\mu_0, \mu_1 \in P_2(\mathcal{X})$ , we have*

$$H(\mu_t|\text{vol}) \leq (1 - t)H(\mu_0|\text{vol}) + tH(\mu_1|\text{vol}) - \kappa W_2^2(\mu_0, \mu_1)t(1 - t)/2, \quad \forall 0 \leq t \leq 1,$$

where  $\text{vol}$  is the normalized volume measure on  $\mathcal{X}$ .

As  $t \mapsto \mu_t$  is not twice differentiable, this is a nonsmooth statement for the informal convexity inequality  $\frac{d^2}{dt^2}H(\mu_t|\text{vol}) \geq \kappa W_2^2(\mu_0, \mu_1)$ . Some interesting consequences of this powerful result are the following *convergence to equilibrium* and *concentration of the equilibrium measure* properties.

- (1) Exponential  $W_2$ -contraction of the heat flow: Let  $\nu_t = \nu_0 e^{t\Delta}$  denote the heat flow starting at  $\nu_0 \in P_2(\mathcal{X})$ ;  $\Delta$  is the Laplace-Beltrami operator. Then, for all  $\alpha, \beta \in P_2(\mathcal{X})$ ,

$$W_2(\alpha e^{t\Delta}, \beta e^{t\Delta}) \leq e^{-\kappa t}W_2(\alpha, \beta), \quad t \geq 0.$$

It follows immediately that when  $\kappa > 0$ , the heat flow performs an exponentially fast convergence to equilibrium:  $W_2(\mu e^{t\Delta}, \text{vol}) \leq e^{-\kappa t}W_2(\mu, \text{vol})$ ,  $t \geq 0$ .

- (2) Logarithmic Sobolev inequality: When  $\kappa > 0$ , then **LSI**( $2/\kappa$ ) is satisfied. This can be restated as follows: for all  $\mu \in P(\mathcal{X})$ ,

$$H(\mu|\text{vol}) \leq \kappa^{-1}I(\mu|\text{vol}), \tag{2.2}$$

where  $I(\mu|\text{vol}) := \frac{d}{dt}|_{t=0}H(\mu e^{t\Delta}|\text{vol}) = \frac{1}{2} \int_{\mathcal{X}} |\nabla \log \frac{d\mu}{d\text{vol}}|^2 d\mu$  is the Fisher information of  $\mu$  with respect to  $\text{vol}$  (entropy production). Again, we see that the heat flow performs the following exponentially fast convergence to equilibrium:

$$H(\mu e^{t\Delta}|\text{vol}) \leq e^{-\kappa t}H(\mu|\text{vol}), \quad t \geq 0. \tag{2.3}$$

- (3) Talagrand’s transport inequality: When  $\kappa > 0$ , then **T<sub>2</sub>**( $2/\kappa$ ) is verified. As was recalled in the introduction, **LSI** implies **T<sub>2</sub>**, but the entropy convexity allows for a direct proof. Furthermore, for each  $n \geq 1$ , **T<sub>2</sub>** also holds for the volume measure  $\text{vol}^{\otimes n}$  on the product space  $\mathcal{X}^n$ : For all  $\mu \in P(\mathcal{X}^n)$ ,

$$\frac{\kappa}{2} W_2^2(\mu, \text{vol}^{\otimes n}) \leq H(\mu|\text{vol}^{\otimes n}),$$

with the same constant  $\kappa$ . Marton’s argument leads us to a *dimension-free* Gaussian concentration inequality for the equilibrium measure  $\text{vol}^{\otimes n}$  on  $\mathcal{X}^n$ : For any  $n \geq 1$ ,  $\text{vol}^{\otimes n}$  concentrates as the Gaussian measure with variance  $\kappa^{-1}$ . This means that for any measurable subset  $A \subset \mathcal{X}^n$  such that  $\text{vol}^{\otimes n}(A) \geq 1/2$ ,

$$\text{vol}^{\otimes n}(A^r) \geq 1 - \exp(-\kappa(r - r_o)^2/2), \quad r \geq r_o := \sqrt{\log(2)}$$

where  $A^r := \{x \in \mathcal{X}^n, d_2(x, y) \leq r, \text{ for some } y \in \mathcal{X}^n\}$  is the  $r$ -blowup of  $A$  with respect to the distance  $d_2$  on  $\mathcal{X}^n$  defined by  $d_2^2(x, y) := \sum_{1 \leq i \leq n} d^2(x_i, y_i)$ . Note that it is shown in [Goz09] that dimension-free Gaussian concentration is equivalent to  $\mathbf{T}_2$ . For more details about concentration of measure and transport inequalities, including  $\mathbf{LSI} \implies \mathbf{T}_2$ , Marton's argument and Gozlan's dimension-free concentration result, see [GL10] for instance.

The Lott-Sturm-Villani theory [LV09, Stu06a, Stu06b] extends these results from the Riemannian manifold setting to any *length space*  $\mathcal{X}$ , i.e. a metric space which admits constant speed geodesic paths for arbitrary endpoints. Typically, one can think of  $\mathcal{X}$  as the Gromov-Hausdorff limit of Riemannian manifolds. Unfortunately, this rules out discrete spaces since they are not geodesic. This theory also extends the above results to weighted length spaces where the reference measure  $\text{vol}$  is replaced by  $m = e^{-V} \text{vol}$  for some nonnegative function  $V : \mathcal{X} \rightarrow [0, \infty)$ .

## 2.2. Constructing displacement interpolations

We wish to find a unifying approach, embracing the settings of Riemannian manifolds and weighted valued graphs that leads to curvature results as above.

### Notation

Our basic ingredients are a Polish state space  $\mathcal{X}$  with its Borel  $\sigma$ -field, the path space  $\Omega \subset \mathcal{X}^{[0,1]}$ , the state  $\mathbb{P}(\Omega)$  of all probability measures on  $\Omega$ . We denote by  $X = (X_t)_{0 \leq t \leq 1}$  the canonical process that is defined by  $X_t(\omega) = \omega_t$  for any  $t \in [0, 1]$  and any  $\omega = (\omega_s)_{0 \leq s \leq 1} \in \Omega$ . The law of the random position at time  $0 \leq t \leq 1$  under  $P \in \mathbb{P}(\Omega)$  is  $P_t = (X_t)_\# P \in \mathbb{P}(\mathcal{X})$ .

### An expected result

We also choose a reference measure  $m \in \mathbb{M}_+(\mathcal{X})$ , the space of nonnegative measures on  $\mathcal{X}$ , which is going to play the same role as  $\text{vol}$  or  $e^{-V} \text{vol}$  in previous section. Our aim is to design some interpolation  $[\mu_0, \mu_1] = (\mu_t)_{0 \leq t \leq 1} \in \mathbb{P}(\mathcal{X})^{[0,1]}$  and consider the relative entropy as a function of time along the interpolation:

$$h_{(\mu_0, \mu_1)}(t) := H(\mu_t | m), \quad t \in [0, 1].$$

In analogy with Theorem 2.1, the expected type of result is as follows.

**Informal statement 2.2.** *If there exists  $\kappa \in \mathbb{R}$  such that for any  $(\mu_0, \mu_1)$ ,*

$$h_{(\mu_0, \mu_1)}(t) \leq (1-t)h_{(\mu_0, \mu_1)}(0) + th_{(\mu_0, \mu_1)}(1) - \Phi_L(t, \mu_0, \mu_1; \kappa), \quad \forall t \in [0, 1]$$

*where  $\Phi_L$  is some well-chosen function, then the “curvature” of the generator  $L$  is bounded from below by  $\kappa$ .*

When saying that a Markov generator  $L$  has a lower bounded curvature, we identify the lower bound of the Ricci curvature of a Riemannian manifold with that of its Laplace-Beltrami operator.

A specific question is: “What about a discrete state space  $\mathcal{X}$  with a graph structure?”

### Two transforms

How to interpolate properly between two probability measures  $\mu_0$  and  $\mu_1 \in \mathbb{P}(\mathcal{X})$ ? To validate an interpolation, it is necessary that it shares some analogues of the specific properties of the usual  $W_2$ -displacement interpolation. The main idea is as follows.

### A thought experiment

Suppose you observe at time  $t = 0$  a very large collection of particles that are distributed with a profile close to the probability measure  $\mu_0 \in \mathbb{P}(\mathcal{X})$  on the state space  $\mathcal{X}$ .

- (1) As in the lazy gas experiment proposed by Schrödinger in 1931 [Sch31] and also described in Villani’s textbook [Vil09, p. 445] in a different context, ask them to rearrange into a new profile close to some  $\mu_1 \in P(\mathcal{X})$  at some later time  $t = 1$ . Since the particles are able to create mutual optimality (Gibbs conditioning principle), they will find an optimal transference plan between  $\mu_0$  and  $\mu_1$ . For details about the Schrödinger problem, see [Léod].
- (2) Now, we add something new with respect to Schrödinger’s original problem: we suppose that the typical speed of these particles is close to zero (the particles are lazy), then each particle will decide to travel at the lowest possible cost, meaning that it chooses an almost *geodesic* path. Indeed, since it is very slow, it is typically expected that its final position is close to its initial one. But it is required by the optimal transference plan that it reaches a distant final position.

Let us go back to (2.1). Informally, the Schrödinger problem described at (1) is responsible for the transference plan  $\hat{\pi}$  while the slowing procedure at (2) enforces the appearance of the geodesic paths in  $\delta_{\gamma xy}$ . Therefore, it is worth defining the displacement interpolation to be the limit of the flow of  $t$ -marginal profiles on  $\mathcal{X}$  of this lazy particle system, as the number of particles tends to infinity and the slowing down tends to zero-speed.

Let us give a short analytic description of its construction. Let  $R \in M_+(\Omega)$  be a reversible Markov process with generator  $L$  and reversing measure  $m \in M_+(\mathcal{X})$ . For instance, one can think of the reversible Brownian motion on  $\mathcal{X} = \mathbb{R}^n$ , i.e. the Wiener measure with Lebesgue measure as its initial marginal or a random walk on a graph  $\mathcal{X}$  satisfying the detailed balance conditions. Introduce the following two elementary transforms of  $R$ .

- (1)  $(f, g)$ -transform:

$$P = f_0(X_0)g_1(X_1) R. \tag{2.4}$$

This transform is attached to the Schrödinger problem in the following sense. If the nonnegative measurable functions  $f_0, g_1 : \mathcal{X} \rightarrow [0, \infty)$  are such that  $P_0 = \mu_0$  and  $P_1 = \mu_1$ , then  $P \in P(\Omega)$  corresponds to the most likely random behavior under the constraints of Schrödinger’s thought experiment. It is the unique solution of the so-called Schrödinger problem, see [Léod, Thm. 2.12]:

$$H(P|R) \rightarrow \min; \quad P \in P(\Omega) : P_0 = \mu_0, P_1 = \mu_1. \tag{2.5}$$

The  $(f, g)$ -transform  $P$  given at (2.4) is the response to the need of a collective optimal rearrangement.

- (2) Slowing down :

$$L^k := k^{-1}L, \quad k \rightarrow \infty.$$

To see that this corresponds to slowing down, write the Markov semi-group  $e^{tL^k} = e^{(t/k)L}$ ,  $t \geq 0$ . We introduce the sequence  $(R^k)_{k \geq 1}$  in  $M_+(\Omega)$  of Markov measures with initial measure  $R_0^k = m$  and generator  $L^k$  for each  $k \geq 1$ .

Mixing this two transforms leads us to

$$P^k := f_0^k(X_0)g_1^k(X_1) R^k$$

where  $f_0^k$  and  $g_1^k$  are chosen such that  $\lim_k P_0^k = \mu_0$  and  $\lim_k P_1^k = \mu_1$ .

**Definition 2.3.** *The displacement interpolation  $[\mu_0, \mu_1] = (\mu_t)_{t \in [0,1]}$  between  $\mu_0$  and  $\mu_1$  is given by*

$$\mu_t := \lim_{k \rightarrow \infty} P_t^k, \quad t \in [0, 1]$$

*if this limit exists.*

It is shown in the author’s paper [Léo12] that when the particles perform Brownian motions on  $\mathcal{X} = \mathbb{R}^n$ , the resulting interpolation is the usual  $W_2$ -displacement interpolation. In [Léo0a], this approach leads to displacement interpolations on a discrete metric graph  $(\mathcal{X}, d)$  which are constant speed geodesic paths on the Wasserstein space  $(P(\mathcal{X}), W_1)$  of order 1 with  $W_1(\mu_0, \mu_1) := \inf \{ \int_{\mathcal{X}^2} d(x, y) \pi(dx dy); \pi \in P(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1 \}$ .

### 2.3. $(f, g)$ -transform

We provide some details about the first transform related to the Schrödinger problem. Let us introduce a time-symmetric version of Doob’s  $h$ -transform.

**Definition 2.4.** *In view of (2.4), we call  $P := f_0(X_0)g_1(X_1)R \in \mathcal{P}(\Omega)$  an  $(f, g)$ -transform of the Markov path measure  $R \in \mathcal{M}_+(\Omega)$ .*

Let  $f_0, g_1$  as above be given. We introduce for each  $t \in [0, 1]$ , the functions  $f_t, g_t : \mathcal{X} \rightarrow [0, \infty)$  defined by

$$\begin{cases} f_t(z) & := E_R(f_0(X_0) \mid X_t = z) \\ g_t(z) & := E_R(g_1(X_1) \mid X_t = z) \end{cases}, \quad \text{for } P_t\text{-a.e. } z \in \mathcal{X}. \tag{2.6}$$

The Markov property of the reference path measure  $R$  implies the following result.

**Theorem 2.5.** *The path measure  $P = f_0(X_0)g_1(X_1)R$  is Markov and for each  $0 \leq t \leq 1$ , its time marginal  $P_t \in \mathcal{P}(\mathcal{X})$  is given by  $P_t = f_t g_t m$ .*

*Proof.* See [Léod, Thm. 3.4]. □

**Definition 2.6.** *We call  $[\mu_0, \mu_1]^R = (\mu_t)_{t \in [0,1]}$  with  $\mu_t = f_t g_t m$  the  $R$ -entropic interpolation between  $\mu_0$  and  $\mu_1$ .*

As  $P$  is Markov, it admits forward and backward generators  $\overrightarrow{A}$  and  $\overleftarrow{A}$ . To express them, we need to introduce the carré du champ<sup>1</sup> of  $R$ . It is defined for any functions  $u, v$  on  $\mathcal{X}$  such that  $u, v$  and  $uv$  are in  $\text{dom } L$ , by

$$\tilde{\Gamma}(u, v) := L(uv) - uLv - vLu.$$

In general, the forward and backward generators  $(\partial_t + \overrightarrow{A}_t)_{0 \leq t \leq 1}$  and  $(-\partial_t + \overleftarrow{A}_t)_{0 \leq t \leq 1}$  of  $P$  depend explicitly on  $t$ . The following informal statement is known for long in specific situations. Rigorous statement and proof are given in [Léoc] for instance.

**Informal statement 2.7.** *Under some hypotheses on  $R$ , the forward and backward generators of the  $(f, g)$ -transform  $P$  are given for any function  $u : [0, 1] \times \mathcal{X} \rightarrow \mathbb{R}$  belonging to some class of regular functions, by*

$$\begin{cases} \overrightarrow{A}_t u(x) & = Lu(x) + \frac{\tilde{\Gamma}(g_t, u)(x)}{g_t(x)}, \quad (t, x) \in [0, 1] \times \mathcal{X} \\ \overleftarrow{A}_t u(x) & = Lu(x) + \frac{\tilde{\Gamma}(f_t, u)(x)}{f_t(x)}, \quad (t, x) \in (0, 1] \times \mathcal{X} \end{cases}$$

where  $f_t, g_t$  are defined at (2.6). Because of Theorem 2.5, for any  $t$  no division by zero occurs  $P_t$ -a.e.

It is worthwhile describing these dynamics in terms of  $\begin{cases} \varphi := \log f, \\ \psi := \log g \end{cases}$ . It gives us

$$\begin{cases} \overleftarrow{A}_t & = A_{\varphi_t} := L + e^{-\varphi_t} \tilde{\Gamma}(e^{\varphi_t}, \cdot) \\ \overrightarrow{A}_t & = A_{\psi_t} := L + e^{-\psi_t} \tilde{\Gamma}(e^{\psi_t}, \cdot) \end{cases} \tag{2.7}$$

where  $\varphi$  and  $\psi$  are solutions of the Hamilton-Jacobi equations

$$\begin{cases} (-\partial_t + B)\varphi(t, x) = 0, & 0 < t < 1, & (\partial_t + B)\psi(t, x) = 0, & 0 \leq t < 1, \\ \varphi_0 = \log f_0, & t = 0, & \psi_1 = \log g_1, & t = 1, \end{cases} \tag{2.8}$$

where the non-linear operator  $B$  is defined by  $Bu := e^{-u}Le^u$  for any function  $u$  such that  $e^u \in \text{dom } L$ .

---

<sup>1</sup>Its standard definition among the “functional inequalities” community is  $\Gamma = \tilde{\Gamma}/2$ .

*The Brownian case*

In the special case where  $R = R^o$  is the reversible Brownian motion on  $\mathcal{X} = \mathbb{R}^n$ , we have  $Lu = L^o u := \Delta u/2$ ,  $Bu = \Delta u/2 + |\nabla u|^2/2$ ,  $\tilde{\Gamma}(u, v) = \nabla u \cdot \nabla v$  for any  $u, v \in \mathcal{C}^2(\mathcal{X})$ . The expressions (2.7) are

$$\begin{cases} \overrightarrow{A}_t &= \Delta/2 + \nabla\psi_t \cdot \nabla \\ \overleftarrow{A}_t &= \Delta/2 + \nabla\varphi_t \cdot \nabla \end{cases}$$

where  $\psi$  solves (2.8):

$$\begin{cases} (\partial_t + \Delta/2)\psi_t(x) + |\nabla\psi_t(x)|^2/2 = 0, & (t, x) \in [0, 1] \times \mathcal{X} \\ \psi_1 = \log g_1, & t = 1, \end{cases}$$

and  $\varphi$  solves

$$\begin{cases} (-\partial_t + \Delta/2)\varphi_t(x) + |\nabla\varphi_t(x)|^2/2 = 0, & (t, x) \in (0, 1] \times \mathcal{X} \\ \varphi_0 = \log f_0, & t = 0. \end{cases}$$

Let us go back to the general case and define the operators

$$\begin{cases} C &:= B - L \quad (\text{nonlinear part of the HJ operator } B), \\ \Theta(u) &:= e^{-u}\tilde{\Gamma}(e^u, u) - Cu, \\ \Theta_2(u) &:= L\Theta u + e^{-u}\tilde{\Gamma}(e^u, \Theta u) + Bu e^{-u}\tilde{\Gamma}(e^u, u) - e^{-u}\tilde{\Gamma}(e^u Bu, u). \end{cases}$$

**Theorem 2.8** (Second derivative of the entropy along the entropic interpolations). *Fix the  $R$ -entropic interpolation  $[\mu_0, \mu_1]^R = (\mu_t)_{t \in [0,1]}$  and consider the function*

$$h_{(\mu_0, \mu_1)}(t) := H(\mu_t|m), \quad t \in [0, 1].$$

Then,  $h_{(\mu_0, \mu_1)}$  is twice differentiable on  $(0, 1)$  and

$$\begin{aligned} h'_{(\mu_0, \mu_1)}(t) &= \langle -\Theta(\varphi_t) + \Theta(\psi_t), \mu_t \rangle, \\ h''_{(\mu_0, \mu_1)}(t) &= \langle \Theta_2(\varphi_t) + \Theta_2(\psi_t), \mu_t \rangle. \end{aligned}$$

*Proof.* See [Léob]. □

In the Brownian case with some potential  $V$ , i.e.  $L = \frac{1}{2}(\Delta - \nabla V \cdot \nabla)$ , we have

$$\begin{aligned} \Theta(u) &= \Gamma(u)/2 := |\nabla u|^2/2 \\ \Theta_2(u) &= L\Gamma(u) - 2\Gamma(Lu, u) =: \Gamma_2(u)/2 \\ &= (\|\nabla^2 u\|_{\text{HS}}^2 + [\nabla^2 V + \text{Ric}](\nabla u))/2 \end{aligned}$$

where  $\text{Ric}_x(v)$  is the Ricci curvature at  $x \in \mathcal{X}$  in direction  $v \in T_x\mathcal{X}$  and  $\Gamma, \Gamma_2$  refer to the standard definitions of the carré du champ and its iteration associated with the generator  $2L$  that were introduced by Bakry and Emery in [BE85]. We note that  $\Theta$  and  $\Theta_2$  are half the Bakry-Emery operators  $\Gamma$  and  $\Gamma_2$ . Last equality is a consequence of Bochner formula. We see immediately that the curvature-dimension bound  $\text{CD}(\kappa, \infty) \stackrel{\text{def}}{\iff} \Gamma_2 \geq \kappa\Gamma$  writes as

$$\Theta_2 \geq \kappa\Theta.$$

An immediate consequence of Theorem 2.8 is the following

**Corollary 2.9.** *Let  $R^o$  be the reversible Brownian motion on the Riemannian manifold  $\mathcal{X}$ . Then,  $\mathcal{X}$  has nonnegative Ricci curvature if and only if for any  $R^o$ -entropic interpolation  $(\mu_t)_{t \in [0,1]} = [\mu_0, \mu_1]^{R^o}$ , the function  $t \in [0, 1] \mapsto H(\mu_t|m)$  is convex.*



Let us define the Fisher information by  $I(\rho m|m) := \int_{\mathcal{X}} \Theta(\log \rho) \rho \, dm$ ,  $\rho m \in P(\mathcal{X})$

**Corollary 2.10** (Logarithmic Sobolev inequality, general case). *Let  $\mathcal{X}$  be any state space admitting an  $m$ -reversible Markov measure with generator  $L$ . Denote  $\nu_t := \nu_0 e^{tL}$ ,  $t \geq 0$  the corresponding heat flow. Assume that  $\Theta_2 \geq \kappa \Theta$ . Then,*

- (1)  $\kappa \in \mathbb{R}$ :  $I(\nu_t|m) \leq I(\nu_0|m)e^{-\kappa t}$ ,  $t \geq 0$ .
- (2)  $\kappa > 0$ :  $H(\nu_t|m) \leq H(\nu_0|m)e^{-\kappa t}$ ,  $t \geq 0$
- (3)  $\kappa > 0$ :  $H(\alpha|m) \leq \kappa^{-1}I(\alpha|m)$ ,  $\forall \alpha \in P(\mathcal{X})$ .

Last inequality is the already encountered logarithmic Sobolev inequality (2.2).

*Sketch of proof.* Instead of the time interval  $[0, 1]$ , we consider  $[0, T]$  and let  $T \rightarrow \infty$ . We see that  $\nu_t = P_t$  with  $P = \rho_0(X_0)R$ . This corresponds to  $\begin{cases} f_0 = \rho_0 \\ g_T \equiv 1 \end{cases}$ , that is  $\begin{cases} \varphi_t = \log \rho_t, \\ \psi_t \equiv 0. \end{cases}$  for all  $t \geq 0$ . Denoting  $h(t) = H(\nu_t|m)$ , it follows with Theorems 2.5 and 2.8 that  $I(t) := -h'(t) = \langle \Theta(\varphi_t), \nu_t \rangle = I(\nu_t|m)$ . Our assumption  $\Theta_2 \geq \kappa \Theta$  implies  $I'(t) \leq -\kappa I(t)$ . Hence, we obtain  $I(t) \leq I(0)e^{-\kappa t}$  which is inequality (1). On the other hand, we have  $H(\nu_0|m) = h(0) - h(\infty) = \int_0^\infty I(t) \, dt \leq I(0) \int_0^\infty e^{-\kappa t} \, dt = I(0)/\kappa$  which is inequality (3) with  $\nu_0 = \alpha$ . Now, plugging  $\alpha = \nu_t$  into (3) gives us:  $h(t) \leq -h'(t)/\kappa$ . This implies  $h(t) \leq h(0)e^{-\kappa t}$  which is inequality (2).  $\square$

### 2.4. Slowing an $(f, g)$ -transform down

Let  $R$  be Markov with generator  $L$ . The slowed down process is represented by the sequence  $(R^k)_{k \geq 1}$  in  $M_+(\Omega)$  of Markov measures associated with the generators  $L^k := L/k$ ,  $k \geq 1$ . Remark that slowing the process down doesn't modify its reversible measure  $m$ ; one converges more slowly towards the same equilibrium. Suppose also that the sequence  $(R^k)_{k \geq 1}$  in  $M_+(\Omega)$  obeys the large deviation principle in  $\Omega$  with speed  $\alpha_k$  and rate function  $C$ , meaning approximately that for a "large class" of measurable subsets  $A$  of  $\Omega$ , we have

$$R^k(A) \underset{k \rightarrow \infty}{\asymp} \exp \left[ -\alpha_k \inf_{\omega \in A} C(\omega) \right]. \tag{2.9}$$

Recall that the  $(f, g)$ -transform of  $R$  given at (2.4) is the unique minimizer, if it exists, of the Schrödinger problem (2.5). Therefore, the convergence of a sequence of  $(f, g)$ -transforms may result from the variational convergence of the corresponding Schrödinger problems. It happens that (2.5) must be normalized as follows

$$H(P|R^k)/\alpha_k \rightarrow \min; \quad P \in P(\Omega) : P_0 = \mu_0, P_1 = \mu_1 \tag{S_{dyn}^k}$$

to  $\Gamma$ -converge as  $k$  tends to infinity to the dynamical Monge-Kantorovich optimal transport problem

$$\int_{\Omega} C \, dP \rightarrow \min; \quad P \in P(\Omega) : P_0 = \mu_0, P_1 = \mu_1. \tag{MK_{dyn}}$$

It is easily seen that any minimizer of  $(\text{MK}_{\text{dyn}})$  concentrates on minimizing geodesic paths. Similarly, the static analogue of  $(\text{S}_{\text{dyn}}^k)$  which is

$$H(\pi|R_{01}^k)/\alpha_k \rightarrow \min; \quad \pi \in P(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1 \tag{S^k}$$

where  $R_{01}^k \in P(\mathcal{X}^2)$  is the joint law of the endpoint  $(X_0, X_1)$  under  $R^k$ ,  $\Gamma$ -converges to the Monge-Kantorovich optimal transport problem

$$\int_{\mathcal{X}^2} c \, d\pi \rightarrow \min; \quad \pi \in P(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1 \tag{MK}$$

with  $c$  defined by  $c(x, y) := \inf \{C(\omega); \omega \in \Omega : \omega_0 = x, \omega_1 = y\}$ .

**Informal statement 2.11** (See [Léo12, Léoa]). *Suppose that the slowed down Markov measure  $R^k \in M_+(\Omega)$  associated with the generator  $L^k := L/k$  satisfies the large deviation principle (2.9) with speed  $\alpha_k$  and rate function  $C$  in  $\Omega$ . Then,*

$$\Gamma\text{-}\lim_{k \rightarrow \infty} (S_{\text{dyn}}^k) = (\text{MK}_{\text{dyn}}) \quad \text{and} \quad \Gamma\text{-}\lim_{k \rightarrow \infty} (S^k) = (\text{MK}).$$

*In particular:*

- (1) *In the reversible Brownian motion case in  $\mathcal{X} = \mathbb{R}^n$ , we have:  
 $\alpha_k = k$ ,  $C(\omega) = \int_{[0,1]} |\dot{\omega}_t|^2/2 dt$ ,  $c(x, y) = |y - x|^2/2$ .*
- (2) *In the case of a random walk on a graph  $\mathcal{X}$ , we have:  
 $\alpha_k = \log k$ ,  $C(\omega) = \sum_{0 \leq t \leq 1} \mathbf{1}_{\{\omega_t \neq \omega_{t-}\}}$  and  $c = d$  the standard graph distance.*

An important consequence of this convergence result is that the limit  $\mu_t := \lim_{k \rightarrow \infty} P_t^k$  in the Definition 2.3 of the displacement interpolation is effective under mild hypotheses, see [Léo12, Léoa].

*The role of slowing down*

We have just seen that slowing down is necessary for constructing displacement interpolations as limits of entropic interpolations. But it is not mandatory in every problem about curvature. Indeed, denoting  $\Theta^k$  and  $\Theta_2^k$  the operators  $\Theta$  and  $\Theta_2$  associated with the slowed down generator  $L^k = k^{-1}L$ , we immediately remark that

$$\begin{cases} \Theta^k &= k^{-1}\Theta, \\ \Theta_2^k &= k^{-2}\Theta, \end{cases}$$

so that the  $\text{CD}(\kappa, \infty)$  bound at level  $k$ :  $\Theta_2^k \geq \kappa\Theta^k$ , boils down to  $\Theta_2 \geq \kappa k\Theta$ . We see that the role of  $k$  is irrelevant since it only changes the time scale: think of the rate  $e^{-\kappa kt}$  of convergence to equilibrium in (2.3).

Both entropic and displacement interpolations admit conserved quantities, but they are much simpler for the displacement interpolations than for the entropic interpolations. In particular, the displacement interpolations have a *constant speed* so that computing with them is easier. In relation with these conserved quantities, remark that unlike the displacement interpolation  $[\mu_0, \mu_0]$  with  $\mu_1 = \mu_0$ , the entropic interpolation  $[\mu_0, \mu_0]^R$  is not constantly equal to  $\mu_0$ . It seems that there is no way of deriving transport-entropy inequalities with optimal constants such as (1.3) from entropic interpolations. The inherent fluctuation of these interpolations weakens the constants which are optimal at the “no-motion limit”:  $k \rightarrow \infty$ , i.e. when working with displacement, rather than entropic, interpolations.

A more sophisticated slowing down than  $L/k$  is performed in [Léoa] which allows for recovering any distance  $d$  on a graph. It is shown that  $L/k$  leads to the standard graph distance, while one needs to slow down at rate  $k^{-d(x,y)}$  along the edge  $(x, y)$  in the general case. Therefore, slowing down allows for recovering a given distance on a graph.

**Conclusion**

There is still work to be done.

In particular, it is necessary to make precise the function  $\Phi_L$  that appears at Statement 2.2 in the case where  $\mathcal{X}$  is a discrete space. This will require to transfer the computation rules of the dynamics of the entropic interpolations  $[\mu_0, \mu_1]^R$  which are given at (2.7) and (2.8), to the displacement interpolations by letting  $k$  tend to infinity.

Another challenge is to interpret when  $\mathcal{X}$  is a discrete space, the  $\Theta_2$ -inequality:  $\Theta_2 \geq \kappa\Theta$ , in terms of some discrete curvature to be defined. Part of the difficulty comes from the complexity of the expression of  $\Theta_2$  which is more involved than its analogue  $\Gamma_2/2$ . In particular, it would be interesting to isolate a pertinent analogue of the Ricci curvature by means of a Bochner-like formula. The comparison with Ollivier’s coarse curvature [Oll09] should also be investigated.

### 3. ON THE MEAN SPEED OF CONVERGENCE OF EMPIRICAL MEASURES IN WASSERSTEIN DISTANCE

#### 3.1. Introduction.

We recall the so-called *empirical law of large numbers* : let  $(E, d, \mu)$  denote a measured Polish space, and let  $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$  denote the empirical measure associated with the i.i.d. sample  $(X_i)_{1 \leq i \leq n}$  of law  $\mu$ . With probability 1,  $L_n \rightarrow \mu$  as  $n \rightarrow +\infty$  (convergence is understood in the sense of the weak topology of measures). This theorem is known as Glivenko-Cantelli theorem and is due in this form to Varadarajan [Var58].

Our purpose here is to give bounds on the speed of convergence in  $W_p$  distance (defined by (1.4)) for the Glivenko-Cantelli theorem, i.e. bounds for the a.s. convergence  $W_p(L_n, \mu) \rightarrow 0$ , following [BLG12].

Such results are desirable notably in view of numerical and statistical applications. There are many works in statistics devoted to convergence rates in some metric associated with the weak convergence of measures, see e.g. [Dud68], [vdVW96]. For Wasserstein metrics, such bounds can be deduced from [AKT84], [Tal92], [HK94], [DY95], [?], [BB11]. Additionally, the approximation of a given probability measure by a measure with finite support in Wasserstein distance is a topic that appears in various guises in the literature, see for example [GL00]. The first motivation for this work was to extend the results obtained by F. Bolley, A. Guillin and C. Villani [BGV07] in the case of variables with support in  $\mathbb{R}^d$ .

The problem of convergence of  $W_p(L_n, \mu)$  to 0 can be divided in two separate questions :

- the first one is to estimate the *mean rate of convergence*, that is the convergence rate of  $\mathbb{E}[W_p(L_n, \mu)]$ ,
- while the second one is to study the concentration properties of  $W_p(L_n, \mu)$  around its mean, that is to find bounds on the quantities

$$\mathbb{P}(W_p(L_n, \mu) - \mathbb{E}[W_p(L_n, \mu)] \geq t).$$

#### 3.2. Convergence in mean in Polish spaces and applications.

We recall that for  $S \subset E$ , the covering number of order  $\delta$  for  $S$ , denoted by  $N(S, \delta)$ , is defined as the minimal  $n \in \mathbb{N}$  such that there exist  $x_1, \dots, x_n$  in  $S$  with

$$S \subset \cup_{i=1}^n B(x_i, \delta).$$

Our main statement is summed up in the following result.

**Theorem 3.1.** *Choose  $t > 0$ . Let  $\mu$  be a probability measure on  $E$  with support included in  $S \subset E$  with finite diameter  $\delta_S$  such that  $N(S, t) < +\infty$ . We have the bound :*

$$\mathbb{E}[W_p(L_n, \mu)] \leq c \left( t + n^{-1/2p} \int_t^{\delta_S/4} N(S, \delta)^{1/2p} d\delta \right).$$

with  $c \leq 22$ .

This result can then be applied by fine-tuning the choice of the scale  $t$  when additional information on the covering numbers is available.

We now quote one important application from [BLG12]. We tackle the case where  $E$  is a separable Banach space with norm  $\|\cdot\|$ , and  $\mu$  a centered Gaussian random variable with values in  $E$ . The couple  $(E, \mu)$  is called a (separable) Gaussian Banach space.

Let  $X$  be an  $E$ -valued r.v. with law  $\mu$ , and define the weak variance of  $\mu$  as  $\sigma = \sup_{f \in E^*, |f| \leq 1} \mathbb{E}[f^2(X)]^{1/2}$ . The small ball function of a Gaussian Banach space  $(E, \mu)$  is defined as

$$\psi(t) = -\log \mu(B(0, t)).$$

We assume that

- (1) there exists  $\kappa > 1$  such that  $\psi(t) \leq \kappa\psi(2t)$ , for  $0 < t \leq t_0$ ,
- (2) for all  $\varepsilon > 0$ ,  $n^{-\varepsilon} = o(\psi^{-1}(\log n))$ .

**Theorem 3.2.** *Let  $(E, \mu)$  be a Gaussian Banach space with weak variance  $\sigma$  and small ball function  $\psi$ . Assume that Assumptions (1) and (2) hold. Then there is a  $C = C(\mu)$  (depending explicitly on the quantities above) such that*

$$\mathbb{E}[W_2(L_n, \mu)] \leq C\psi^{-1}(\log n). \tag{3.1}$$

In order to underline the interest of the result above, we introduce some definitions from the field of optimal quantization, see [GL00]. For  $n \geq 1$  and  $1 \leq r < +\infty$ , define the optimal quantization error at rate  $n$  as

$$\delta_{n,r}(\mu) = \inf_{\nu \in \Theta_n} W_r(\mu, \nu)$$

where the infimum runs over the set  $\Theta_n$  of probability measures with finite support of cardinal bounded by  $n$ . Roughly speaking, Theorem 4.1 in [DFMS03] and Theorem 2 in [GLP03] imply that there exist  $c, c' > 0$  such that

$$c\psi^{-1}(\log n) \leq \delta_{n,r} \leq c'\psi^{-1}(\log n).$$

We can restate Theorem 3.2 by saying that (under some assumptions on the small ball function and when the distortion index is  $r = 2$ ) the empirical measure is a rate-optimal quantizer in average - and in fact with high probability, see below. Using empirical measures as candidates for quantization is also envisioned in the recent work [DSS11]. As an illustration, we consider the classical case where  $E = (L_2([0, 1]), \|\cdot\|_2)$  and  $\mu$  is the Wiener measure. In this case, we quote [DFMS03] to get  $\psi(t) \sim_{t \rightarrow 0} \frac{1}{8t^2}$ . Thus,

$$\frac{1}{\sqrt{8 \log n}} \leq \delta_{n,r} \leq \frac{1}{\sqrt{\log n}}$$

Actually, a sharper result is  $\delta_{n,r} \sim \sqrt{2}/\pi\sqrt{\log n}$ , c.f. [LP04]. In our case, we get the bound  $\mathbb{E}[W_2(L_n, \mu)] = O\left(\frac{1}{\sqrt{\log n}}\right)$ .

### 3.3. Concentration around the mean *via* transportation inequalities.

So-called *transportation inequalities* (as defined e.g. in [GL10]) provide a powerful tool to tackle concentration of  $W_p(L_n, \mu)$  around its mean. The next result states that a  $\mathbf{T}_2$  inequality on  $\mu$  implies a Gaussian concentration inequality for  $W_2(L_n, \mu)$ . We reproduce a particular case of more general results of N. Gozlan and C. Léonard ([GL07, GL10]).

**Theorem 3.3.** ([GL07], Theorem 12). *Let a probability measure  $\mu$  on  $E$  satisfy the inequality  $\mathbf{T}_2(C)$  inequality. The following holds :*

$$\mathbb{P}(W_2(L_n, \mu) \geq \mathbb{E}[W_2(L_n, \mu)] + t) \leq \exp\left(-n\frac{t^2}{C}\right). \tag{3.2}$$

**Proposition 3.4.** *Let  $(E, \mu)$  be a Gaussian Banach space and let  $\sigma^2$  denote the weak variance of  $\mu$ . Then  $\mu$  satisfies the inequality  $\mathbf{T}_2(2\sigma^2)$ . Hence,*

$$\mathbb{P}(W_2(L_n, \mu) \geq \mathbb{E}[W_2(L_n, \mu)] + t) \leq \exp\left(-n\frac{t^2}{2\sigma^2}\right).$$

## 4. REPRESENTATION FORMULA FOR THE ENTROPY AND FUNCTIONAL INEQUALITIES

### 4.1. Introduction

The purpose of these notes is to describe a representation formula for the relative entropy in the Wiener space, in the spirit of Borell's formula for the Laplace transform. As an application, we give unified and simple proofs of a number of Gaussian inequalities. These lecture notes are mainly based on the article [Leh12].

## 4.2. The representation formula

Let  $\mathbb{W}$  be the Wiener space of continuous paths taking values in  $\mathbb{R}^d$ , let  $\gamma$  be the Wiener measure and  $X$  be the coordinate process. Let  $\mathbb{H} = \mathbb{H}_0^1(\mathbb{R}_+; \mathbb{R}^d)$  be the associated Cameron-Martin space: a path  $u$  belongs to  $\mathbb{H}$  if its energy

$$\|u\|^2 = \int_0^{+\infty} |\dot{u}_t|^2 dt$$

is finite. Given a probability space  $(\Omega, \mathbb{P})$  equipped with a filtration  $(\mathcal{F}_t)_{t \geq 0}$  we call *drift* any adapted process  $U$  which belongs to  $\mathbb{H}$  almost surely. The following proposition is a straightforward consequence of Girsanov's formula.

**Proposition 4.1.** *Let  $B$  be a Brownian motion defined on some filtered probability space and let  $U$  be a drift. Letting  $\mu$  be the law of  $B + U$ , we have*

$$H(\mu|\gamma) \leq \frac{1}{2} \mathbb{E} \|U\|^2.$$

A natural question then arises: given a probability measure  $\mu$  on  $W$ , can equality be achieved in the previous inequality? The following result of Föllmer [Fol85] answers positively, up to a change of probability space.

**Proposition 4.2.** *Let  $\mu$  be a measure on  $\mathbb{W}$  absolutely continuous with respect to  $\gamma$ . There is a drift  $U$  such that under  $\mu$  the process  $Y = X - U$  is a Brownian motion and*

$$H(\mu|\gamma) = \frac{1}{2} \int_{\mathbb{W}} \|U\|^2 d\mu. \quad (4.1)$$

Thus  $Y$  is Brownian motion on  $(\mathbb{W}, \mu)$  and the drift  $U$  is such that  $Y + U$  has law  $\mu$  and  $H(\mu|\gamma) = \frac{1}{2} \mathbb{E} \|U\|^2$ .

Given two probability measures  $\mu, \nu$  on the Wiener space  $\mathbb{W}$ , we consider the quadratic transport cost with respect to the pseudo distance  $d(w, w') = \|w - w'\|$  on  $\mathbb{W}$ :

$$W_2^2(\mu, \nu) = \inf \left\{ \int_{\mathbb{W} \times \mathbb{W}} \|w - w'\|^2 \pi(dw, dw') \right\}$$

where the infimum is taken on all couplings  $\pi$  of  $\mu$  and  $\nu$ .

According to Proposition 4.2, if  $\mu$  is absolutely continuous with respect to  $\gamma$  there exists a Brownian motion  $B$  and a drift  $U$  such that  $B + U$  has law  $\mu$  and

$$H(\mu|\gamma) = \frac{1}{2} \mathbb{E} \|U\|^2.$$

Then  $(B, B + U)$  is a coupling of  $(\gamma, \mu)$  and by definition of the transportation distance  $W_2$

$$W_2^2(\mu, \gamma) \leq \mathbb{E} \|U\|^2 = 2H(\mu|\gamma).$$

We thus have proved the Wiener space version of Talagrand's transportation inequality [Tal96]. Also, if the density  $F$  of  $\mu$  belongs to the domain of the Malliavin operator  $D$  then the Föllmer drift can be explicitly written in terms of  $DF$  and it is easy to derive from (4.1) the Wiener space version of the classical log-Sobolev inequality [Gro75]

$$H(\mu|\gamma) \leq \frac{1}{2} \int_{\mathbb{W}} \|D \log(F)\|^2 d\mu.$$

Under some technical assumptions on the measure  $\mu$ , Föllmer's result can be strengthened as follows.

**Theorem 4.3.** *Given a Brownian motion  $B$*

$$H(\mu|\gamma) = \min_U \left( \frac{1}{2} \mathbb{E} \|U\|^2 \right)$$

where the minimum is on all drifts  $U$  such that  $B + U$  has law  $\mu$ .

Together with the following Legendre duality

$$\log\left(\int_{\mathbb{W}} e^f d\gamma\right) = \sup_{\mu} \left[ \int_{\mathbb{W}} f d\mu - H(\mu|\gamma) \right].$$

we get the Boué and Dupuis formula [BD98]:

**Theorem 4.4.** *Given a Brownian motion  $B$  we have for every function  $f: \mathbb{W} \rightarrow \mathbb{R}$  measurable and bounded from below*

$$\log\left(\int_{\mathbb{W}} e^f d\gamma\right) = \sup_U \left[ \mathbb{E}\left(f(B + U) - \frac{1}{2}\|U\|^2\right) \right],$$

where the supremum is taken over all drifts  $U$ .

A weaker version of this formula was put forward by Borell [Bor00] in the context of functional inequalities. Following his approach it is easy to derive from Theorem 4.4 the following Wiener space version of the Brunn-Minkowski inequality.

**Theorem 4.5.** *For every subsets  $K, L$  of  $\mathbb{W}$  and  $s \in [0, 1]$*

$$\gamma(K)^{1-s}\gamma(L)^s \leq \gamma((1-s)K + sL)e^{-s(1-s)d(K,L)^2/2}, \tag{4.2}$$

where  $d(K, L) = \inf(\|w - \tilde{w}\|, w \in K, \tilde{w} \in L)$ .

As an application, let us derive from this inequality the Brunn-Minkowski inequality for the principal frequency (due to Brascamp and Lieb [BL76]): for every compact (say)  $A, B \subset \mathbb{R}^d$

$$\lambda(A)^{-1/2} + \lambda(B)^{-1/2} \leq \lambda(A + B)^{-1/2}, \tag{4.3}$$

where  $\lambda(A)$  is the first eigenvalue of the operator  $-\Delta$  on  $A$  with Dirichlet boundary condition. Let  $x \in A, y \in B$  and  $s \in [0, 1]$ , inequality (4.2) yields easily

$$\mathbb{P}_x(T_A > t)^{1-s}\mathbb{P}_y(T_B > t)^s \leq \mathbb{P}_{(1-s)x+sy}(T_{(1-s)A+sB} > t) \tag{4.4}$$

where  $\mathbb{P}_x(T_A > t)$  is the probability that a Brownian motion starting from  $x$  has stayed in  $A$  up to time  $t$ . On the other hand, as  $t$  tends to  $+\infty$

$$\mathbb{P}_x(T_A > t) = e^{-\frac{\lambda(A)t}{2} + o(t)}$$

Therefore letting  $t \rightarrow +\infty$  in (4.4) and optimizing in  $s$  yields (4.3).

## 5. A TWO-SCALE PROOF OF THE EYRING-KRAMERS FORMULA.

### 5.1. Introduction

We consider a diffusion on a potential landscape which is given by a smooth Hamiltonian in the regime of small noise. We sketch a new approach to derive the Eyring-Kramers formula for the spectral gap of the associated generator of the diffusion. The new approach is based on a refinement of the two-scale approach introduced by Grunewald, Otto, Villani, and Westdickenberg [GOVW09] and of the mean-difference estimate introduced by Chafaï and Malrieu [CM10]. The Eyring-Kramers formula follows as a simple corollary from two main ingredients: The first one shows that the Gibbs measures restricted to a domain of attraction has a “good” Poincaré constant mimicking the fast convergence of the diffusion to metastable states. The second ingredient is the estimation of the mean-difference by a new weighted transportation distance. It contains the main contribution of the spectral gap, resulting from exponential long waiting times of jumps between metastable states of the diffusion. This new approach also allows to derive sharp estimates on the log-Sobolev constant. For details we refer the reader to the preprint [MS12] of both authors.

## 5.2. Results and sketch of proof

In what follows, we sketch the strategy to apply a combination of the two-scale approach (cf. [GOVW09]) and a transportation technique (cf. [CM10]) to give an alternative proof of the Eyring-Kramers formula. The first rigorous proof in full generality was given by Bovier, Gayraud, and Klein [BGK05] using potential-theoretic ideas. Slightly later, a different proof was given by Helffer, Klein, and Nier [HKN04] via the Witten complex approach. The Eyring-Kramers formula asymptotically determines the Poincaré constant of an arbitrary Gibbs measure  $\mu(dx) = \exp(-H(x)/T) dx$  in the low temperature limit i.e. sending  $T \rightarrow 0$ . The Poincaré constant  $\varrho$  of the Gibbs measure  $\mu$  is the largest constant  $C > 0$  such that  $\mu$  verifies the Poincaré inequality  $\mathbf{PI}(C)$  defined in (1.1). Let us only consider an one-dimensional Hamiltonian  $H$ , even if the results holds in any dimension. Additionally for this short note, we will not discuss some standard growth conditions and nondegeneracy conditions on the Hamiltonian  $H$ . For precise statements, we refer the reader to the preprint [MS12] of both authors. With these simplifications, the Eyring-Kramers formula becomes:

**Proposition 5.1** (Eyring-Kramers formula [GLE41]). *Assume that the Hamiltonian  $H$  has two local minima at  $m_0$  and  $m_1$  such that  $H(m_0) < H(m_1)$ . Then the Poincaré constant  $\varrho$  of the Gibbs measure  $\mu$  is given by*

$$\varrho = \frac{\sqrt{|H''(z)| H''(m_1)}}{2\pi T} \exp\left(-\frac{H(z) - H(m_1)}{T}\right) \left(1 + O(\sqrt{T} |\ln T|)\right),$$

where  $z$  is the saddle between both minima.

The last theorem states that at low temperature  $T \ll 1$  the SG constant  $\varrho$  is essentially determined by the saddle height  $(H(z) - H(m_1))$ . The Eyring-Kramers formula also holds in the case of finitely many local minima. However, to understand the idea of the new approach, the best is to consider only two local minima.

Let us sketch the main idea of the new approach: Let  $\mu_0$  and  $\mu_1$  denote the restriction of the Gibbs measure  $\mu$  to the domain of attraction of the local minima  $m_0$  and  $m_1$  respectively i.e.

$$\mu_0(dx) = \frac{1}{Z_0} 1_{\{x < z\}} \exp\left(-\frac{H(x)}{T}\right) dx, \quad \text{with } Z_0 = \int_{-\infty}^z \exp\left(-\frac{H(x)}{T}\right) dx$$

and

$$\mu_1(dx) = \frac{1}{Z_1} 1_{\{x > z\}} \exp\left(-\frac{H(x)}{T}\right) dx \quad \text{with } Z_1 = \int_{-\infty}^z \exp\left(-\frac{H(x)}{T}\right) dx.$$

As in the two scale-approach [GOVW09], the starting point for our proof is the decomposition of  $\text{Var}_\mu(f)$  into local variances with respect to  $\mu_0$  and  $\mu_1$  and into the variance of a Bernoulli variable i.e.

$$\text{Var}_\mu(f) = Z_0 \text{Var}_{\mu_0}(f) + Z_1 \text{Var}_{\mu_1}(f) + Z_0 Z_1 \left( \int f d\mu_0 - \int f d\mu_1 \right)^2. \quad (5.1)$$

The first two terms on the right hand side of (5.1) are estimated by an application of the Poincaré constants for the restricted measures  $\mu_0$  and  $\mu_1$ . It turns out that these constants scale like  $T^{-1}$ . Heuristically, this fact seems to be plausible because there are no metastabilities on the sets  $\{x < z\}$  and  $\{x > z\}$ . However, the rigorous proof of this fact in higher dimensions is technically challenging because of the lack of convexity of  $H$ . The main contribution to the Poincaré constant comes from the third term on the right hand side of (5.1). Motivated by the transportation technique of Chafaï and Malrieu [CM10], the third term is represented by using a transport  $\Phi_t$  between  $\mu_0$  and  $\mu_1$  as

$$\int f d\mu_0 - \int f d\mu_1 = - \int_0^1 \int \frac{d}{dt} f(\Phi_t(x)) \mu_0(dx) dt = - \int_0^1 \int \nabla f(x) \cdot \dot{\Phi}_t(\Phi^{-1}(x)) \mu_t(dx) dt,$$

where  $\mu_t$  denotes the push forward  $(\Phi_t)_\# \mu_0$ . Using the last identity and Cauchy-Schwarz inequality we get the estimate

$$\begin{aligned} \left( \int f d\mu_0 - \int f d\mu_1 \right)^2 &= \left( \int \nabla f(x) \cdot \int_0^1 \dot{\Phi}_t(\Phi^{-1}(x)) \frac{\mu_t(x)}{\mu(x)} dt \mu(x) dx \right)^2 \\ &\leq \int \left( \int_0^1 |\dot{\Phi}_t(\Phi^{-1}(x))| \frac{\mu_t(x)}{\mu(x)} dt \right)^2 \mu(dx) \int |\nabla f|^2 d\mu. \end{aligned}$$

Recalling the definition of the Poincaré constant  $\varrho$ , it is only left to estimate the weighted transportation cost on the right hand side of the last inequality. For low temperatures (i.e.  $T \ll 1$ ) it suffices to consider truncated Gaussian measures around the local minima  $m_0$  and  $m_1$  instead of the restricted measures  $\mu_0$  and  $\mu_1$ . This simplifies the estimate of the weighted transportation cost yielding the Eyring-Kramers formula after an optimization procedure.

A nice feature of this proof is that it replicates the behavior of the corresponding stochastic process: The fast convergence to local minima is expressed by the good local Poincaré estimate of the restricted measures  $\mu_0$  and  $\mu_1$ , whereas the main contribution to the overall Poincaré constant comes from a Markov chain jumping from one local minima to another.

Most important, this approach is also applicable to the logarithmic Sobolev inequality **LSI** i.e. one can derive estimates on the **LSI** constant in the low temperature regime. Surprisingly, it turns out that these estimates on the **LSI** constant do not coincide with the Eyring-Cramer formula. Additionally, we expect that the estimates on **LSI** constant obtained by this approach are sharp (for details see [MS12]).

## REFERENCES

- [ABC<sup>+</sup>00] Cécile Ané, Sébastien Blachère, Djalil Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.
- [AKT84] M. Ajtai, J. Komlós, and G. Tusnády. On optimal matchings. *Combinatorica*, 4(4):259–264, 1984.
- [BE85] D. Bakry and M. Emery. Diffusions hypercontractives. In *Séminaire de Probabilités*, number 1123 in Lecture Notes in Mathematics. Springer-Verlag, 1985.
- [BGL13] D. Bakry, I. Gentil and M. Ledoux. Analysis and geometry of Markov diffusion operators. In preparation. 2013.
- [BB11] F. Barthe and C. Bordenave. Combinatorial optimization over two random point sets. *ArXiv preprint arXiv:1103.2734v1*, 2011.
- [BLG12] E. Boissard and T. Le Gouic. On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. *Ann. Inst. H. Poincaré Probab. Stat.* (to appear), 2013.
- [BGV07] François Bolley, Arnaud Guillin, and Cédric Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields*, 137(3-4):541–593, 2007.
- [Bor00] C. Borell. Diffusion equations and geometric inequalities, *Potential Anal.* 12 (1) (2000) 49–71.
- [BD98] M. Boué and P. Dupuis. A variational representation for certain functionals of Brownian motion, *Ann. Probab.* 26 (4) (1998) 1641–1659.
- [BGK05] A. Bovier, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues. *J. Eur. Math. Soc. (JEMS)*, 7(1):69–99, 2005.
- [BL76] H.J. Brascamp and E.H. Lieb, On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation, *J. Functional Analysis* 22 (4) (1976) 366–389.
- [CM10] D. Chafaï and F. Malrieu. On fine properties of mixtures with respect to concentration of measure and Sobolev type inequalities. *Ann. Inst. H. Poincaré Probab. Stat.*, 46(1):72–96, 2010.
- [DFMS03] S. Dereich, F. Fehrer, A. Matoussi, and M. Scheutzow. On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces. *J. Theoret. Probab.*, 16(1):249–265, 2003.
- [DSS11] S. Dereich, M. Scheutzow and R. Schottstedt. Constructive quantization: approximation by empirical measures. *ArXiv preprint arXiv:1108.5346*, 2011.
- [DSS11] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: approximation by empirical measures. *ArXiv preprint arXiv:1108.5346*, 2011.
- [Dud68] R. M. Dudley. The speed of mean Glivenko-Cantelli convergence. *Ann. Math. Statist.*, 40:40–50, 1968.



- [DY95] V. Dobrić and J. E. Yukich. Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.*, 8(1):97–118, 1995.
- [Fol85] H. Föllmer, An entropy approach to the time reversal of diffusion processes, in *Stochastic differential systems (Marseille-Luminy, 1984)*, Lecture Notes in Control and Inform. Sci. 69, Springer (1985) 156–163.
- [GLE41] S. Glasstone, K. J. Laidler, and H. Eyring. *The Theory of Rate Processes*. McGraw-Hill, 1941.
- [GL07] Nathael Gozlan and Christian Léonard. A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields*, 139(1-2):235–283, 2007.
- [GL10] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Processes and Related Fields*, 16:635–736, 2010.
- [Goz09] N. Gozlan. A characterization of dimension free concentration in terms of transportation inequalities. *Ann. Probab.*, 37(6):2480–2498, 2009.
- [GL00] Siegfried Graf and Harald Luschgy. *Foundations of quantization for probability distributions*, volume 1730 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2000.
- [GLP03] S. Graf, H. Luschgy and G. Pagès. Functional quantization and small ball probabilities for Gaussian processes. *J. Theoret. Probab.*, 16 (4): 1047–1062, 2003.
- [Gro75] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.* 97 (4) (1975) 1061–1083.
- [GOVW09] N. Grunewald, F. Otto, C. Villani, and M. Westdickenberg. A two-scale approach to logarithmic Sobolev inequalities and the hydrodynamic limit. *Ann. Inst. H. Poincaré Probab. Statist.*, 45(2):302–351, 2009.
- [HKN04] B. Helffer, M. Klein, and F. Nier. Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach. *Mat. Contemp.*, 26:41–85, 2004.
- [HK94] Joseph Horowitz and Rajeeva L. Karandikar. Mean rates of convergence of empirical measures in the Wasserstein metric. *J. Comput. Appl. Math.*, 55(3):261–273, 1994.
- [Led01] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
- [Leh12] J. Lehec. Representation formula for the entropy and functional inequalities, *Ann. Inst. Henri Poincaré Probab. Stat.* (to appear).
- [Léoa] C. Léonard. Lazy random walks and optimal transport on graphs. Preprint.
- [Léob] C. Léonard. On the convexity of the entropy along entropic interpolations. Preprint.
- [Léoc] C. Léonard. Stochastic derivatives and  $h$ -transforms of Markov processes. Preprint.
- [Léod] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. To appear in *Discrete & Continuous Dynamical Systems - Series A*. Special issue on “Optimal Transport and Applications”.
- [Léo12] C. Léonard. From the Schrödinger problem to the Monge-Kantorovich problem. *J. Funct. Anal.*, 262(1879-1920), 2012.
- [LP04] Harald Luschgy and Gilles Pagès. Sharp asymptotics of the functional quantization problem for Gaussian processes. *Ann. Probab.*, 32(2):1574–1599, 2004.
- [LV09] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.*, 169(3):903–991, 2009.
- [MS12] G. Menz and A. Schlichting. Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape. *preprint arXiv:1202.1510v2*, 2012.
- [McC97] R.J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128:153–179, 1997.
- [Oll09] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.
- [OV00] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.
- [Roy07] Gilles Royer. *An initiation to logarithmic Sobolev inequalities*, volume 14 of *SMF/AMS Texts and Monographs*. American Mathematical Society, Providence, RI, 2007. Translated from the 1999 French original by Donald Babbitt.
- [Sch31] E. Schrödinger. Über die Umkehrung der Naturgesetze. *Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math.*, 144:144–153, 1931.
- [Stu06a] K-T. Sturm. On the geometry of metric measure spaces, I. *Acta Math*, 196:65–131, 2006.
- [Stu06b] K-T. Sturm. On the geometry of metric measure spaces, II. *Acta Math*, 196:133–177, 2006.
- [Sv05] K-T. Sturm and M-K. von Renesse. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.
- [Tal92] Michel Talagrand. Matching random samples in many dimensions. *Ann. Appl. Probab.*, 2(4):846–856, 1992.
- [Tal96] M. Talagrand. Transportation cost for Gaussian and other product measures, *Geom. Funct. Anal.* 6 (3) (1996) 587–600.
- [Var58] V. S. Varadarajan. On the convergence of sample probability distributions. *Sankhyā*, 19:23–26, 1958.
- [vdVW96] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.
- [Vil03] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [Vil09] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.