Simple bounds for convergence of empirical and occupation measures in 1- Wasserstein distance

Emmanuel Boissard

Institut de mathématiques de Toulouse
Université Paul Sabatier
118 route de Narbonne, 31062 Toulouse
emmanuel.boissard@math.univ-toulouse.fr

Abstract

We study the problem of non-asymptotic deviations between a reference measure μ and its empirical version L_n, in the 1-Wasserstein metric, under the standing assumption that μ satisfies a transport-entropy inequality. We extend some results of F. Bolley, A. Guillin and C. Villani [8] with simple proofs. Our methods are based on concentration inequalities and extend to the general setting of measures on a Polish space. Deviation bounds for the occupation measure of a contracting Markov chain in W_1 distance are also given.

Throughout the text, several examples are worked out, including the cases of Gaussian measures on separable Banach spaces, and laws of diffusion processe .

1 Introduction

1.1 Generalities

In the whole paper, (E,d) will denote a Polish space with metric d, equipped with its Borel σ-field and $\mathscr{P}(E)$ will denote the set of probability measures over E. Consider $\mu \in \mathscr{P}(E)$ and a sequence of i.i.d. variables X_i, $1 \leq i \leq n$, with common law μ. Let

$$L_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \tag{1}$$

denote the empirical measure associated with the i.i.d. sample $(X_i)_{1 \leq i \leq n}$, then with probability 1, $L_n \to \mu$ as $n \to +\infty$ (here the arrow denotes narrow convergence, or convergence against all bounded continuous functions over E). This theorem is known as the empirical law of large number or Glivenko-Cantelli theorem and is due in this form to Varadarajan [33]. Quantifying the speed of convergence for an appropriate notion of distance between probability measures is an old problem, with notable importance in statistics. For many examples, we refer to the book of Van der Vaart and Wellner [32] and the Saint-Flour course of P.Massart [27].

Our aim here is to study non-asymptotic deviations in 1-Wasserstein distance. This is a problem of interest in the fields of statistics and numerical probability. More specifically, we provide bounds for the quantity $\mathbb{P}(W_1(L_n,\mu) \geq t)$ for $t > 0$, i.e. we quantify the speed of convergence of the variable $W_1(L_n,\mu)$ to 0 in probability.

This paper seeks to complement the work of F.Bolley, A.Guillin and C.Villani in [8] where such estimates are obtained for measures supported in $\mathbb{R}^d$. We sum up (part of) their result here. Suppose that μ is a probability measure on $\mathbb{R}^d$ for $1 \leq p \leq 2$ that satisfies a $\mathbf{T}_p(C)$ transportation-entropy inequality, that is

$$W_p(\nu,\mu) \leq \sqrt{CH(\nu|\mu)} \text{ for all } \nu \in \mathscr{P}_p(\mathbb{R}^d)$$

(see below for definitions). They obtain a non-asymptotic Gaussian deviation estimate for the $p-$Wasserstein distance between the empirical and true measures :

$$\mathbb{P}(W_p(L_n,\mu) \geq t) \leq C(t)\exp(-Knt^2).$$

This is an effective result : the constants K and $C(t)$ may be explicitely computed from the value of some square-exponential moment of μ and the constant C appearing in the transportation inequality.

The strategy used in [8] relies on a non-asymptotic version of (the upper bound in) Sanov's theorem. Roughly speaking, Sanov's theorem states that the proper rate function for the deviations of empirical measures is the entropy functional, or in other words that for 'good' subsets $A \in \mathscr{P}(E)$,

$$\mathbb{P}(L_n \in A) \asymp e^{-nH(A|\mu)}$$

where $H(A|\mu) = \inf_{\nu \in A} H(\nu|\mu)$ (see [11] for a full statement of the theorem).

In a companion work [6], we derive sharper bounds for this problem, using a construction originally due to R.M. Dudley [14]. The interested reader may refer to [6] for a summary of existing results.

Here, our purpose is to show that in the case $p = 1$, the results of [8] can be recovered with simple arguments of measure concentration, and to give various extensions of interest.

- We would like to consider spaces more general than $\mathbb{R}^d$.

- We would like to encompass a wide class of measures in a synthetic treatment. In order to do so we will consider more general transportation inequalities, see below.

- Another interesting feature is to extend the result to dependent sequences such as the occupation measure of a Markov chain. This is a particularly desirable feature in applications : one may wish to approximate a distribution that is unknown, or from which it is practically impossible to sample uniformly, but that is known to be the invariant measure of a simulable Markov chain.

In the remainder of this section, we introduce the tools necessary in our framework : transportation distances and transportation-entropy inequalities. In Section 2, we give our main results, as well as explicit estimates in several relevant cases. Section 3 is devoted to the proof of the main result. Section 4 is devoted to the proof of Theorem 2.6. In Section 5 we show how our strategy of proof can extend to the dependent case.

1.2 A short introduction to transportation inequalities

1.2.1 Transportation costs and Wasserstein distances

We recall here basic definitions and propositions ; for proofs and a thorough account of this rich theory, the reader may refer to [34]. Define $\mathscr{P}_p$, $1 \leq p < +\infty$, as the set of probability measures with a finite p-th moment. The p-Wasserstein metric $W_p(\mu, \nu)$ between $\mu, \nu \in \mathscr{P}_p$ is defined by

$$W_p^p(\mu, \nu) = \inf \int d^p(x, y) \pi(dx, dy)$$

where the infimum is on probability measures $\pi \in \mathscr{P}(E \times E)$ with marginals μ and ν. The topology induced by this metric is slightly stronger than the weak topology : namely, convergence of a sequence $(\mu_n)_{n \in \mathbb{N}}$ to a measure $\mu \in \mathscr{P}_p$ in the p-Wasserstein metric is equivalent to the weak convergence of the sequence plus a uniform bound on the p-th moments of the measures μ_n, $n \in \mathbb{N}$.

We also recall the well-known Kantorovich-Rubinstein dual characterization of W_1 : let $\mathscr{F}$ denote the set of 1-Lipschitz functions $f : E \to \mathbb{R}$ that vanish at some fixed point x_0. We have :

$$W_1(\mu, \nu) = \sup_{f \in \mathscr{F}} \int f \, d\mu - \int f \, d\nu. \tag{2}$$

1.2.2 Transportation-entropy inequalities

For a very complete overview of the subject, the reader is invited to consult the review [17]. More facts and criteria are gathered in Appendix A. For $\mu, \nu \in \mathscr{P}(E)$, define the relative entropy $H(\nu|\mu)$ as

$$H(\nu|\mu) = \int_E \log \frac{d\nu}{d\mu} \nu(dx)$$

if ν is absolutely continuous relatively to μ, and $H(\nu|\mu) = +\infty$ otherwise. Let $\alpha : [0, +\infty) \to \mathbb{R}$ denote a convex, increasing, left-continous function such that $\alpha(0) = 0$.

Definition 1.1. *We say that $\mu \in \mathscr{P}(E)$ satisfies a $\alpha(\mathscr{T}_d)$ inequality if for all $\nu \in \mathscr{P}(E)$,*

$$\alpha(W_1(\mu, \nu)) \leq H(\nu|\mu). \tag{3}$$

We say that $\mu \in \mathscr{P}(E)$ satisfies a $\mathbf{T}_p(C)$ inequality for some $C > 0$ if it satisfies a $\alpha(\mathscr{T}_d)$ inequality with

$$\alpha(t) = \frac{1}{C} t^{2/p}.$$

2 Results and applications

2.1 General bounds in the independent case

Let us first introduce some notation : if $K \subset E$ is compact and $x_0 \in K$, we define the set $\mathscr{F}_K$ of 1-Lipschitz functions over K vanishing at x_0, which is is also compact w.r.t. the uniform distance (as a consequence of the Ascoli-Arzela theorem). We will also need the following definitions.

Definition 2.1. *Let (A, d) be a totally bounded metric space. For every $\delta > 0$, define the covering number $\mathscr{N}(A, \delta)$ of order δ for A as the minimal number of balls of radius δ needed to cover A.*

Definition 2.2. *Let $\alpha : [0, +\infty) \to \mathbb{R}$ be convex, increasing, left-continuous and vanishing at 0. The monotone conjugate of α is*

$$\alpha^{\circledast}(s) = \sup_{t \geq 0} st - \alpha(t).$$

We state our first result in a fairly general fashion.

Theorem 2.3. *Suppose that $\mu \in \mathscr{P}(E)$ satisfies a $\alpha(\mathscr{T}_d)$ inequality. Let $a > 0$ be such that $E_{a,1} = \int e^{ad(x_0, x)} \mu(dx) \leq 2$. Choose a compact $K = K(a, t, u) \subset E$ such that*

$$\mu(K^c) \leq \left[\frac{32}{at} \log \frac{32}{at} - \frac{32}{at} + 1 \right]^{-1}.$$

Denote

$$\mathscr{C}_t = \mathcal{N}(\mathscr{F}_K, t/8). \tag{4}$$

We have

$$\mathbb{P}(W_1(L_n, \mu) \geq t) \leq \exp -n\alpha \left[t/2 - \Gamma(\mathscr{C}_t, n) \right] \tag{5}$$

where $\Gamma(\mathscr{C}_t, n) = \inf_{\lambda > 0} 1/\lambda [\log \mathscr{C}_t + n\alpha^\circledast(\lambda/n)]$, and with the convention that $\alpha(x) = 0$ for $x < 0$.

Remark. It may be convenient to express the condition stated on $E_{a,1}$ through a transportation inequality : on the one hand, they are equivalent as shown in Appendix A, and on the other hand, transportation inequalities can be more adimensional, in the sense that they behave better with respect to tensorization than integrability conditions.

We give an explicit connexion here. Assume that μ satisfies a $\alpha(\mathcal{T}_d)$ inequality. By Proposition A.3, the condition $E_{a,1} \leq 2$ is fulfilled as soon as

$$a \int d(x, x_0) d\mu + \alpha^\circledast(a) \leq \log 2.$$

Remark. With a mild change in the proof, one may replace in (5) the term $t/2$ by ct for any $c < 1$, with the trade-off of choosing a larger compact set, and thus a larger value of $\mathscr{C}_t$. For the sake of readability we do not make further mention of this.

The result in its general form is abtruse, but it yields interesting results as soon as one knows more about α. Let us give a few examples.

Corollary 2.4. *If μ satisfies $\mathbf{T}_1(C)$, we have*

$$\mathbb{P}(W_1(L_n, \mu) \geq t) \leq \mathscr{C}_t \exp -\frac{1}{8C} n t^2.$$

Corollary 2.5. *Suppose that μ verifies the modified transport inequality*

$$W_1(\nu, \mu) \leq C \left(H(\nu|\mu) + \sqrt{H(\nu|\mu)} \right)$$

(as observed in paragraph A.2, this is equivalent to the finiteness of an exponential moment for μ). Then, for $t \leq C/2$,

$$\mathbb{P}(W_1(L_n, \mu) \geq t) \leq A(n, t) \exp -\frac{(\sqrt{2} - 1)^2}{2C^2} n t^2$$

where

$$A(n, t) = \exp \left[4(\sqrt{2} - 1)^2 n (\sqrt{1 + \frac{n}{\log \mathscr{C}_t}} - 1)^{-2} \right]$$

(observe that $A(n, t) \to \mathscr{C}_t$ when $n \to +\infty$).

Remark. Corollary 2.5 states that for sub-exponential measures we have a square-exponential decay in t of the right-most term for t close enough to 0 only, whereas this holds everywhere for sub-Gaussian measures. This is in keeping with other known concentration results for the double-exponential measure, which have a quadratic-then-linear exponential dependence in the enlargement parameter t, see for example [4]. For larger t, the dependence should be exponential here as well.

Proof of Corollary 2.4. In this case, we have $\alpha(t) = \frac{1}{C}t^2$, and so

$$\Gamma(\mathscr{C}_t, n) = \sqrt{\frac{C \log \mathscr{C}_t}{n}},$$

so that we get

$$\mathbb{P}(W_1(L_n, \mu) \geq t) \leq \exp -\frac{n}{C}(\frac{t}{2} - \sqrt{\frac{C \log \mathscr{C}_t}{n}})^2$$

and conclude with the elementary inequality $(a-b)^2 \geq \frac{1}{2}a^2 - b^2$.

$\square$

Proof of Corollary 2.5. Here, $\alpha(x) = \frac{1}{4}(\sqrt{1 + \frac{4x}{C}} - 1)^2$, and one can get the bound

$$\Gamma(\mathscr{C}_t, n) \leq \frac{C}{\sqrt{1 + \frac{n}{\log \mathscr{C}_t}} - 1}.$$

By concavity of the square root function, for $u \leq 1$, we have $\sqrt{1+u} - 1 \geq (\sqrt{2} - 1)u$. Thus, for $t \leq \frac{C}{2}$, we have

$$\begin{aligned}
\alpha(\frac{t}{2} - \Gamma(\mathscr{C}_t, n)) &\geq \frac{(\sqrt{2}-1)^2}{4}(\frac{2}{C}t - \frac{4}{\sqrt{1 + \frac{n}{\log \mathscr{C}_t}} - 1})^2 \\
&\geq \frac{(\sqrt{2}-1)^2}{2C^2}t^2 - 4(\sqrt{2}-1)^2(\sqrt{1 + \frac{n}{\log \mathscr{C}_t}} - 1)^{-2}
\end{aligned}$$

(in the last line we have used again the inequality $(a-b)^2 \geq \frac{a^2}{2} - b^2$). This in turn gives the announced result.

$\square$

Our technique of proof, though related to the one in [8], is based on different arguments : we make use of the tensorization properties of transportation inequalities as well as functional equivalents of transportation inequalities of Bobkov-Götze type (see (19)), instead of a Sanov-type bound. The notion that is key here is the phenomenon of concentration of measure (see e.g. [23]) : its relevance in statistics was put forth very explicitly in [27]. We may sum up our approach as follows : first, we rely on existing tensorization results to obtain concentration of $W_1(L_n, \mu)$ around its

mean $\mathbb{E}[W_1(L_n, \mu)]$, and in a second time we estimate the decay of the mean as $n \to +\infty$. Despite technical difficulties, the arguments are mostly elementary.

The next theorem is a variation on Corollary 2.4. Its proof is based on different arguments, and it The next proposition is a bound on the size of the tail of a Gaussian measure. A proof may be found e.g. in [22], Chapter 4. is postponed to Section 4. We will use this theorem to obtain bounds for Gaussian measures in Theorem 2.6.

Theorem 2.6. *Let $\mu \in \mathscr{P}(E)$ satisfy a $\mathbf{T}_1(C)$ inequality. Then :*

$$\mathbb{P}(W_1(\mu, L_n) \geq t) \leq K_t e^{-nt^2/8C}$$

where

$$K_t = \exp\left[\frac{1}{C} \inf_{\nu} Card\ (Supp\ \nu)(Diam\ Supp\ \nu)^2\right]$$

and ν runs over all probability measures with finite support s The next proposition is a bound on the size of the tail of a Gaussian measure. A proof may be found e.g. in [22], Chapter 4. uch that $W_1(\mu, \nu) \leq t/4$.

Remark. As earlier, we could improve the factor $1/8C$ in the statement above to any constant $c < 1/C$, with the trade-off of a larger constant K_t.

2.2 Comments

We give some comments on the pertinence of the results above. First of all, we argue that the asymptotic order of magnitude of our estimates is the correct one. The term "asymptotic" here means that we consider the regime $n \to +\infty$, and the relevant tool in this setting is Sanov's large deviation principle for empirical measures. A technical point needs to be stressed : there are several variations of Sanov's theorem, and the most common ones (see e.g. [11]) deal with the weak topology on probability measures. What we require is a version of the principle that holds for the stronger topology induced by the 1-Wasserstein metric, which leads to slightly more stringent assumptions on the measure than in Theorem 2.3. With this in mind, we quote the following result from R. Wang, X. Wang and L. Wu [35] :

Proposition 2.7. *Suppose that $\mu \in \mathscr{P}(E)$ satisfies $\int e^{ad(x,x_0)}\mu(dx) < +\infty$ for all $a > 0$ and some $x_0 \in E$, and a $\alpha(\mathscr{T}_d)$ inequality. Then :*

- *for all $A \subset \mathscr{P}(E)$ closed for the W_1 topology,*

$$\limsup_{n \to +\infty} \frac{1}{n} \log \mu(A) \leq -\inf_{\nu \in A} H(\nu|\mu)$$

- *for all $B \subset \mathscr{P}(E)$ open w.r.t. the W_1 topology,*

$$\liminf_{n \to +\infty} \frac{1}{n} \log \mu(B) \geq -\inf_{\nu \in B} H(\nu|\mu).$$

Consider the closed set $A = \{v \in \mathscr{P}(E), W_1(\mu, v) \geq t\}$, then we have according to the above

$$\limsup_{n \to +\infty} \frac{1}{n} \log \mathbb{P}(W_1(L_n, \mu) \geq t) \leq -\alpha(t).$$

With Theorem 2.3 (and the remark following it), we obtain the bound

$$\limsup_{n \to +\infty} \frac{1}{n} \log \mathbb{P}(W_1(L_n, \mu) \geq t) \leq -\alpha(ct)$$

for all $c < 1$, and since α is left-continuous, we indeed obtain the same asymptotic bound as from Sanov's theorem.

Let us come back to the non-asymptotic regime. When we assume for example a $\mathbf{T}_1$ inequality, we get a bound in the form $\mathbb{P}(W_1(L_n, \mu) \geq t) \leq C(t)e^{-Cnt^2}$ involving the large constant $C(t)$. By the Kantorovich-Rubinstein dual formulation of W_1

$$W_1(\mu, v) = \sup_{f \in \mathscr{F}} \int f \, d\mu - \int f \, dv,$$

this amounts to simultaneous deviation inequalities for all 1-Lipschitz observables. We recall briefly the well-known fact that it is fairly easy to obtain a deviation inequality for one Lipschitz observable without a constant depending on the deviation scale t. Indeed, consider a 1-Lipschitz function f and a sequence X_i of i.i.d. variables with law μ. By Chebyshev's bound, for $\theta > 0$,

$$\mathbb{P}(\frac{1}{n}\sum f(X_i) - \int f\mu \geq \varepsilon) \quad \leq \quad \exp -n[\theta\varepsilon - \log(\int e^{\theta f(x)}\mu(dx)e^{-\theta \int f\mu})]$$

According to Bobkov-Götze's dual characterization of $\mathbf{T}_1$, the term inside the log is bounded above by $e^{C\theta^2}$, for some positive C, whence $\mathbb{P}(\frac{1}{n}\sum f(X_i) - \int f\mu \geq \varepsilon) \leq \exp -n[\theta\varepsilon - C\theta^2]$. Finally, take $\theta = \frac{1}{2C}\varepsilon$ to get

$$\mathbb{P}(\frac{1}{n}\sum f(X_i) - \int f\mu \geq \varepsilon) \leq e^{-Cnt^2/2}.$$

Thus, we may see the multiplicative constant that we obtain as a trade-off for the obtention of uniform deviation estimates on all Lipschitz observables.

2.3 Examples of application

For practical purposes, it is important to give the order of magnitude of the multiplicative constant $\mathscr{C}_t$ depending on t. We address this question on several important examples in this paragraph.

2.3.1 The $\mathbb{R}^d$ case

Example 2.8. *Denote $\theta(x) = 32x \log \left[2 \left(32x \log 32x - 32x + 1 \right) \right]$. In the case $E = \mathbb{R}^d$, the numerical constant $\mathscr{C}_t$ appearing in Theorem 2.3 satisfies :*

$$\mathscr{C}_t \leq 2 \left(1 + \theta(\frac{1}{at}) \right) 2^{C_d \theta(\frac{1}{at})^d} \tag{6}$$

where C_d only depends on d. In particular, for all $t \leq \frac{1}{2a}$, there exist numerical constants C_1 and C_2 such that

$$\mathscr{C}_t \leq C_1(1 + \frac{1}{at} \log \frac{1}{at}) e^{C_d C_2^d (\frac{1}{at} \log \frac{1}{at})^d}.$$

Remark. The constants C_d, C_1, C_2 may be explicitly determined from the proof. We do not do so and only state that C_d grows exponentially with d.

Proof. For a measure $\mu \in \mathscr{P}(\mathbb{R}^d)$, a convenient natural choice for a compact set of large measure is a Euclidean ball. Denote $B_R = \{x \in \mathbb{R}^d, |x| \leq R\}$. We will denote by C_d a constant depending only on the dimension d, that may change from line to line. Suppose that μ satisfies the assumptions in Theorem 2.3. By Chebyshev's bound, $\mu(B_R^c) \leq 2e^{-aR}$, so we may choose $K = B_{R_t}$ with

$$R_t \geq \frac{1}{a} \log \left[2 \left(\frac{32}{at} \log \frac{32}{at} - \frac{32}{at} + 1 \right) \right].$$

Next, the covering numbers for B_R are bounded by :

$$\mathscr{N}(B_R, \delta) \leq C_d \left(\frac{R}{\delta} \right)^d.$$

Using the bound (21) of Proposition B.2, we have

$$\mathscr{C}_t \leq \left(2 + 2\lfloor \frac{32R_t}{t} \rfloor \right) 2^{C_d \left(\frac{32R_t}{t} \right)^d}.$$

This concludes the proof for the first part of the proposition. The second claim derives from the fact that for $x > 2$, there exists a numerical constant k such that $\theta(x) \leq kx \log x$.

$\square$

Example 2.8 improves slightly upon the result for the W_1 metric in [8]. One may wonder whether this order of magnitude is close to optimality. It is in fact not sharp, and we point out where better results may be found.

In the case $d = 1$, $W_1(L_n, \mu)$ is equal to the L^1 norm of the difference $F_n - F$, where F_n and F denote respectively the cumulative distribution functions (c.d.f.) of L_n and μ (see e.g. [10]), and it is bounded above by the Kolmogorov-Smirnov divergence $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$. As a consequence

of the celebrated Dvorestky-Kiefer-Wolfowitz theorem (see [26], [32]), we have the following : if $\mu \in \mathscr{P}(\mathbb{R})$ has a continuous c.d.f., then

$$\mathbb{P}(W_1(L_n, \mu) > t) \leq 2e^{-2nt^2}.$$

The behaviour of the Wasserstein distance between empirical and true distribution in one dimension has been very thoroughly studied by del Barrio, Giné, Matran, see [10].

In dimensions greater than 1, the result is also not sharp. Integrating (6), one recovers a bound of the type $\mathbb{E}(W_1(L_n, \mu)) \leq Cn^{-1/(d+2)}(\log n)^c$. Looking into the proof of our main result, one sees that any improvement of this bound will automatically give a sharper result than (6). For the uniform measure over the unit cube, results have been known for a while. The pioneering work in this framework is the celebrated article of Ajtai, Komlos and Tusnády [1]. M.Talagrand [31] showed that when μ is the uniform distribution on the unit cube (in which case it clearly satisfies a $\mathbf{T}_1$ inequality) and $d \geq 3$, there exists $c_d \leq C_d$ such that

$$c_d n^{-1/d} \leq \mathbb{E}W_1(L_n, \mu) \leq C_d n^{-1/d}.$$

Sharp results for general measures are more recent. In general, under some polynomial moment condition, one may get an estimate of the form $\mathbb{E}W_1(L_n, \mu) \leq cn^{-1/d}$: see the article of Dobrić and Yukich [13] for the compactly supported case, and the recent preprint by F. Barthe and C. Bordenave [3] for the unbounded case. More precisely, these articles contain sharp *asymptotic* estimates of the form

$$\mathbb{E}W_1(L_n^1, L_n^2) \sim C(\mu)n^{-1/d}$$

where L_n^1 and L_n^2 are two independent copies of the empirical measure, that hold as soon as $d > 2$, and it is possible to deduce from the proofs non-asymptotic bounds with explicit constants of the form $\mathbb{E}W_1(L_n, \mu) \leq C'(\mu)n^{-1/d}$. Plugging such bounds into our proof would yield a bound on $\mathscr{C}_t$ of the form

$$\mathscr{C}_t \leq \exp[c/t^{d-2}]$$

for some $c > 0$ depending on the data. We do not develop this point in full here. We point out that sharper results for spaces of finite-dimensional type may also be found in the preprint [6].

2.3.2   A first bound for Standard Brownian motion

We wish now to illustrate our results on an infinite-dimensional case. A first natural candidate is the law of the standard Brownian motion, with the sup-norm as reference metric. The natural idea that we put in place in this paragraph is to choose as large compact sets the α-Hölder balls, which are compact for the sup-norm. However the remainder of this paragraph serves mainly an illustrative purpose : we will obtain sharper results, valid for general Gaussian measures on (separable) Banach spaces, in paragraph 2.3.4.

We consider the canonical Wiener space $(\mathscr{C}([0,1], \mathbb{R}), \gamma, \|.\|_\infty)$, where γ denotes the Wiener measure, under which the coordinate process $B_t : \omega \to \omega(t)$ is a standard Brownian motion.

Example 2.9. *Denote by γ the Wiener measure on $\left(\mathscr{C}([0,1],\mathbb{R}),\gamma,\|.\|_\infty\right)$, and for $\alpha < 1/2$, define*

$$C_\alpha = 2^{1+\alpha}\frac{2^{(1-2\alpha)/4}}{1-2^{4/(1-2\alpha)}}\|Z\|_{4/(1-2\alpha)}$$

where $\|Z\|_p$ denotes the L_p norm of a $\mathcal{N}(0,1)$ variable Z. There exists $k > 0$ such that for every $t \le 144/\sqrt{2\log 2}$, γ satisfies

$$\mathbb{P}(W_1(L_n,\gamma) \ge t) \le \mathscr{C}_t e^{-nt^2/64}$$

with

$$\mathscr{C}_t \le \exp\exp(kC_\alpha\frac{\sqrt{\log 1/t}}{t})^{1/\alpha}.$$

Proof. For $0 < \alpha \le 1$, define the α-Hölder semi-norm as

$$|x|_\alpha = \sup_{t,s\in[0,1]} \frac{|x(t)-x(s)|}{|t-s|^\alpha}.$$

Let $0 < \alpha \le 1$ and denote by C_α the Banach space of α-Hölder continuous functions vanishing at 0, endowed with the norm $\|.\|_\alpha$. It is a classical fact that the Wiener measure is concentrated on C_α for all $\alpha \in]0,1/2[$. By Ascoli-Arzela's theorem, C_α is compactly embedded in $\mathscr{C}([0,1],\mathbb{R})$, or in other words the α-Hölder balls $B_{\alpha,R} = \{x \in \mathscr{C}([0,1],\mathbb{R}), \|x\|_\alpha \le R\}$ are totally bounded for the uniform norm. This makes $B(\alpha,R)$ good candidates for compact spaces of large measure. We need to evaluate how big $B(\alpha,R)$ is w.r.t. γ.

To this end we use the fact that the Wiener measure is also a Gaussian measure on C_α (see [2]). Therefore Lemma D.1 applies : denote

$$m_\alpha = \mathbb{E}\sup_t \|B_t\|_\alpha, \quad s_\alpha^2 = \mathbb{E}(\sup_t \|B_t\|_\alpha)^2,$$

we have

$$\gamma(B(\alpha,R)^c) \le 2e^{-(R-m_\alpha)^2/2s_\alpha^2}$$

for $R \ge m_\alpha$. Choosing

$$R_t \ge m_\alpha + \left[2s_\alpha^2\log 2(\frac{32}{at}\log\frac{32}{at}-\frac{32}{at}+1)\right]^{1/2} \tag{7}$$

guarantees that

$$\gamma(B(\alpha,R_t)^c) \le \left(\frac{32}{at}\log\frac{32}{at}-\frac{32}{at}+1\right)^{-1}.$$

On the other hand, according to Corollary C.2, m_α and s_α are bounded by C_α. And Lemma D.3 shows that choosing $a = \sqrt{2\log 2}/3$ ensures $\mathbb{E}e^{a\sup_t |B_t|} \le 2$.

Elementary computations show that for $t \le 144/\sqrt{2\log 2}$, we can pick

$$R_t = 3C_\alpha \sqrt{\log(96/(\sqrt{2\log 2}t))}$$

to comply with the requirement in (7).

Bounds for the covering numbers in α-Hölder balls are computed in [7] :

$$\mathcal{N}(B(\alpha,R),\delta) \le 10\frac{R}{\delta} \exp\left[\log(3)5^{\frac{1}{\alpha}}\left(\frac{R}{\delta}\right)^{\frac{1}{\alpha}}\right]. \tag{8}$$

We recover the (unpretty !) bound

$$\mathscr{C}_t \le 2(1+96\frac{C_\alpha}{t}\sqrt{\log 96/(\sqrt{2\log 2}t)})\exp\left[240\log 2\frac{C_\alpha}{t}\sqrt{\log 96/(\sqrt{2\log 2}t)}\right.$$

$$\left. \times \exp\log 3\left(120\frac{C_\alpha}{t}\sqrt{\log 96/(\sqrt{2\log 2}t)}\right)^{1/\alpha}\right].$$

The final claim in the Proposition is obtained by elementary majorizations.

$\square$

2.3.3 Paths of S.D.E.s

H.Djellout, A.Guillin and L.Wu established a $\mathbf{T}_1$ inequality for paths of S.D.E.s that allows us to work as in the case of Brownian motion. We quote their result from [19].

Consider the S.D.E. on $\mathbb{R}^d$

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad X_0 = x_0 \in \mathbb{R}^d \tag{9}$$

with $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathscr{M}_{d\times m}$ and (B_t) is a standard m-dimensional Brownian motion. We assume that b and σ are locally Lipschitz and that for all $x,y \in \mathbb{R}^d$,

$$\sup_x |\sqrt{\operatorname{tr}\sigma(x)^t\sigma(x)}| \le A, \quad \langle y-x, b(y)-b(x)\rangle \le B(1+|y-x|^2)$$

For each starting point x it has a unique non-explosive solution denoted $(X_t(x))_{t\ge 0}$ and we denote its law on $\mathscr{C}([0,1],\mathbb{R}^d)$ by $\mathbb{P}_x$.

Theorem 2.10 ([19])**.** *Assume the conditions above. There exists C depending on A and B only such that for every $x \in \mathbb{R}^d$, $\mathbb{P}_x$ satisfies a $\mathbf{T}_1(C)$ inequality on the space $\mathscr{C}([0,1],\mathbb{R}^d)$ endowed with the sup-norm.*

We will now state our result. A word of caution : in order to balance readability, the following computations are neither optimized nor made fully explicit. However it should be a simple, though dull, task for the reader to track the dependence of the numerical constants on the parameters.

From now on we make the simplifying assumption that the drift coefficient is globally bounded by B (this assumption is certainly not minimal).

Example 2.11. *Let μ denote the law of the solution of the S.D.E. (9) on the Banach space $C([0,1], \mathbb{R}^d)$ endowed with the sup-norm. Let C be such that μ satisfies $\mathbf{T}_1(C)$. For all $0 < \alpha < 1/2$ there exist C_α and c depending only on A, B, α and d, and such that for $t \leq c$,*

$$\mathbb{P}(W_1(L_n, \mu) \geq t) \leq \mathscr{C}_t e^{-nt^2/8C}$$

and

$$\mathscr{C}_t \leq \exp \exp \left[C_\alpha \left(\log \frac{1}{t} \right)^{-1+1/2\alpha} \left(\frac{1}{t} \right)^{-1+3/2\alpha} \right].$$

Proof. The proof goes along the same lines as the Brownian motion case, so we only outline the important steps. First, there exists a depending explicitly on A, B, d such that $\mathbb{E}_{\mathbb{P}_x} e^{a\|X.\|_\infty} \leq 2$: this can be seen by checking that the proof of Djellout-Guillin-Wu actually gives the value of a Gaussian moment for μ as a function of A, B, d, and using standard bounds.

Corollary C.3 applies for $\alpha < 1/2$ and p such that $1/p = 1/2 - \alpha$: there exists $C' < +\infty$ depending explicitly on A, B, α, d, such that $\mathbb{E}\|X.\|_\alpha^p \leq C'$. Consequently,

$$\mu(B(\alpha, R)^c) \leq C'/R^p.$$

So choosing

$$R = \left(C'(\frac{32}{at} \log \frac{32}{at} - \frac{32}{at} + 1) \right)^{1/p}$$

guarantees that

$$\mu(B(\alpha, R_t)^c) \leq \left(\frac{32}{at} \log \frac{32}{at} - \frac{32}{at} + 1 \right)^{-1}.$$

For $t \leq c$ small enough, $R \leq C'' \left(\frac{1}{t} \log \frac{1}{t} \right)^{1/p}$ with c, C'' depending on A, B, α, d. The conclusion is reached again by using estimate (8) on the covering numbers of Hölder balls.

$\square$

2.3.4   Gaussian r.v.s in Banach spaces

In this paragraph we apply Theorem 2.6 to the case where E is a separable Banach space with norm $\|.\|$, and μ is a centered Gaussian random variable with values in E, meaning that the image of μ by

every continuous linear functional $f \in E^*$ is a centered Gaussian variable in $\mathbb{R}$. The couple (E, μ) is said to be a Gaussian Banach space.

Let X be a E-valued r.v. with law μ, and define the weak variance of μ as

$$\sigma = \sup_{f \in E^*, |f| \leq 1} \left(\mathbb{E} f^2(X) \right)^{1/2}.$$

The small ball function of a Gaussian Banach space (E, μ) is the function

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

We can associate to the couple (E, μ) their Cameron-Martin Hilbert space $H \subset E$, see e.g. [22] for a reference. It is known that the small ball function has deep links with the covering numbers of the unit ball of H, see e.g. Kuelbs-Li [21] and Li-Linde [24], as well as with the approximation of μ by measures with finite support in Wasserstein distance (the quantization or optimal quantization problem), see Fehringer's Ph.D. thesis [15], Dereich-Fehringer-Matoussi-Scheutzow [12], Graf-Luschgy-Pagès [18]. It should thus come as no surprise that we can give a bound on the constant K_t depending solely on ψ and σ. This is the content of the next example.

Example 2.12. *Let (E, μ) be a Gaussian Banach space. Denote by ψ its small ball function and by σ its weak variance. Assume that t is such that $\psi(t/16) \geq \log 2$ and $t/\sigma \leq 8\sqrt{2 \log 2}$. Then*

$$\mathbb{P}(W_1(L_n, \mu) \geq t) \leq K_t e^{-nt^2/16\sigma^2}$$

with

$$K_t = \exp \exp \left[c(\psi(t/32) + \log(\sigma/t)) \right]$$

for some universal constant c.

A bound for c may be tracked in the proof.

Proof. Step 1. Building an approximating measure of finite support.

Denote by K the unit ball of the Cameron-Martin space associated to E and μ, and by B the unit ball of E. According to the Gaussian isoperimetric inequality (see [22]), for all $\lambda > 0$ and $\varepsilon > 0$,

$$\mu(\lambda K + \varepsilon B) \geq \Phi \left(\lambda + \Phi^{-1}(\mu(\varepsilon B)) \right)$$

where $\Phi(t) = \int_{-\infty}^{t} e^{-u^2/2} du / \sqrt{2\pi}$ is the Gaussian c.d.f.. Note

$$\mu' = \frac{1}{\mu(\lambda K + \varepsilon B)} \mathbf{1}_{\lambda K + \varepsilon B} \mu$$

the restriction of μ to the enlarged ball. As proved in [6], Appendix 1, the Gaussian measure μ satisfies a $\mathbf{T}_2(2\sigma^2)$ inequality, hence a $\mathbf{T}_1$ inequality with the same constant. We have

$$W_1(\mu, \mu') \leq \sqrt{2\sigma^2 H(\mu'|\mu)} = \sqrt{-2\sigma^2 \log \mu(\lambda K + \varepsilon B)}$$
$$\leq \sqrt{-2\sigma^2 \log \Phi(\lambda + \Phi^{-1}(\mu(\varepsilon B)))}.$$

On the other hand, denote $k = \mathcal{N}(\lambda K, \varepsilon)$ the covering number of λK (w.r.t. the norm of E). Let $x_1, \ldots, x_k \in K$ be such that union of the balls $B(x_i, \varepsilon)$ contains λK. From the triangle inequality we get the inclusion

$$\lambda K + \varepsilon B \subset \bigcup_{i=1}^{k} B(x_i, 2\varepsilon).$$

Choose a measurable map $T : \lambda K + \varepsilon B \to \{x_1, \ldots, x_k\}$ such that for all x, $|x - T(x)| \leq 2\varepsilon$. The push-forward measure $\mu^k = T_\# \mu'$ has support in the finite set $\{x_1, \ldots, x_k\}$, and clearly

$$W_1(\mu', \mu^k) \leq 2\varepsilon.$$

Choose $\varepsilon = t/16$, and

$$\lambda = \Phi^{-1}(e^{-t^2/(128\sigma^2)}) - \Phi^{-1}(\mu(\varepsilon B)) \tag{10}$$
$$= \Upsilon^{-1}(e^{-\psi(t/16)}) + \Phi^{-1}(e^{-t^2/(128\sigma^2)}) \tag{11}$$

where $\Upsilon(t) = \int_t^{+\infty} e^{-u^2/2} du / \sqrt{2\pi}$ is the tail of the Gaussian distribution (we have used the fact that $\Phi^{-1} + \Upsilon^{-1} = 0$, which comes from symmetry of the Gaussian distribution).

Altogether, this ensures that $W_1(\mu, \mu^k) \leq t/4$.

Step 2. Bounding λ.

We can use the elementary bound $\Upsilon(t) \leq e^{-t^2/2}$, $t \geq 0$ to get

$$\Upsilon^{-1}(u) \leq \sqrt{-2 \log u}, \quad 0 < u \leq 1/2$$

which yields $\Upsilon^{-1}(e^{-\psi(t/16)}) \leq \sqrt{\psi(t/16)}$ as soon as $\psi(t/16) \geq \log 2$. Likewise,

$$\Phi^{-1}(e^{-t^2/128\sigma^2}) = \Upsilon^{-1}(1 - e^{-t^2/128\sigma^2})$$
$$\leq \sqrt{2 \log \frac{1}{1 - e^{-t^2/128\sigma^2}}}$$

as soon as $t^2/128\sigma^2 \leq \log 2$. Moreover, for $u \leq \log 2$, we have $1/(1 - e^{-u}) \leq 2 \log 2 / u$. Putting everything together, we get

$$\lambda \leq \sqrt{\psi(t/16)} + c \sqrt{\log \sigma / t} \tag{12}$$

for some universal constant $c > 0$. Observe that the first term in (12) will usually be much larger than the second one.

Step 3.

From Theorem 2.6 we know that

$$\mathbb{P}(W_2(\mu, L_n) \geq t) \leq K_t e^{-nt^2/16\sigma^2}$$

with

$$K_t = \exp\left[\frac{1}{2\sigma^2}\frac{k}{2}(\text{Diam }\{x_1, \ldots, x_k\})^2\right].$$

The diameter is bounded by $\text{Diam } K = 2\sigma\lambda \leq c\sigma(\sqrt{\psi(t/16)} + c\sqrt{\log\sigma/t})$.

We wish now to control $k = \mathcal{N}(\lambda K, t/16)$ in terms of the small ball function ψ. The two quantities are known to be connected : for example, Lemma 1 in [21] gives the bound

$$\mathcal{N}(\lambda K, \varepsilon) \leq e^{\lambda^2/2 + \psi(\varepsilon/2)}.$$

Thus

$$k \leq \exp\left[\psi(t/16) + \psi(t/32) + c\log\sigma/t\right].$$

With some elementary majorizations, this ends the proof.

$\square$

We can now sharpen the results of Proposition 2.9. Let γ denote the Wiener measure on $\mathscr{C}([0,1], \mathbb{R}^d)$ endowed with the sup-norm, and denote by σ^2 its weak variance. Let λ_1 be the first nonzero eigenvalue of the Laplacian operator on the ball of $\mathbb{R}^d$ with homogeneous Dirichlet boundary conditions : it is well-known that the small ball function for the Brownian motion on $\mathbb{R}^d$ is equivalent to λ_1/t^2 when $t \to 0$.

As a consequence, there exists $C = C(d)$ such that for small enough $t > 0$ we have

$$W_1(L_n, \gamma) \leq \exp\left[\exp(C\lambda_1/t^2)\right]\exp\left[-nt^2/16\sigma^2\right]. \tag{13}$$

2.4 Bounds in the dependent case : occupation measures of contractive Markov chains

The results above can be extended to the convergence of the occupation measure for a Markov process. As an example, we establish the following result.

Theorem 2.13. *Let $P(x, dy)$ be a Markov kernel on $\mathbb{R}^d$ such that*

1. *the measures $P(x, .)$ satisfy a $\mathbf{T}_1(C)$ inequality*

2. *$W_1(P(x, .), P(y, .)) \leq r|x - y|$ for some $r < 1$.*

Let π denote its invariant measure. Let $(X_i)_{i\geq 0}$ denote the Markov chain associated with P under $X_0 = 0$. Let $m_1 = \int |x| d\mu$.

Set $a = \frac{2}{C}\left(\sqrt{4m_1^2 + C\log 2} - 2m_1\right)$. There exists $C_d > 0$ depending only on d such that for $t \leq 2/a$,

$$\mathbb{P}(W_1(L_n, \pi) \geq t) \leq K(n, t)\exp -n\frac{(1-r)^2}{8C}t^2$$

where

$$K(n, t) = \exp\left[\frac{m_1}{\sqrt{nC}} + C_d(\frac{1}{at}\log\frac{1}{at})^{\frac{d}{2}}\right]^2.$$

Remark. The result is close to the one obtained in the independent case, and, as stressed in the introduction, it holds interest from the perspective of numerical simulation, in cases where one cannot sample uniformly from a given probability distribution π but may build a Markov chain that admits π as its invariant measure.

Remark. We comment on the assumptions on the transition kernel. The first one ensures that the $\mathbf{T}_1$ inequality is propagated to the laws of X_n, $n \geq 1$. As for the second one, it has appeared several times in the Markov chain literature (see e.g. [19], [28], [20]) as a particular variant of the Dobrushin uniqueness condition for Gibbs measures. It has a nice geometric interpretation as a positive lower bound on the Ricci curvature of the Markov chain, put forward for example in [28]. Heuristically, this condition implies that the Markov chains started from two different points and suitably coupled tend to get closer.

Remark. In the preprint [6], we put forward a different approach under the assumption that the Markov transition kernel satisfies a discrete-time Poincaré inequality. This requirement is weaker than the contractivity condition that we ask for here, as shown in [28], Corollary 30. On the other hand, it only allows to obtain a control on the average distance $\mathbb{E}(W_1(L_n, \mu))$, and it requires more stringent regularity conditions on the initial law (it should have a density with respect to the invariant measure of the Markov chain).

3   Proof of Theorem 2.3

The starting point is the following result, obtained by Gozlan and Leonard ([16], see Chapter 6) by studying the tensorization properties of transportation inequalities.

Lemma 3.1. *Suppose that* $\mu \in \mathscr{P}(E)$ *verifies a* $\alpha(\mathscr{T}_d)$ *inequality. Define on* E^n *the metric*

$$d^{\oplus n}((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \sum_{i=1}^{n} d(x_i, y_i).$$

Then $\mu^{\otimes n} \in \mathscr{P}(E^n)$ *verifies a* $\alpha'(\mathscr{T}_{d^{\oplus n}})$ *inequality, where* $\alpha'(t) = \frac{1}{n}\alpha(nt)$. *Hence, for all Lipschitz functionals* $Z : E^n \to \mathbb{R}$ *(w.r.t. the distance* $d^{\oplus n}$), *we have the concentration inequality*

$$\mu^{\otimes n}(Z \geq \int Z d\mu^{\otimes n} + t) \leq \exp -n\alpha(\frac{t}{n\|Z\|_{Lip}}) \quad \text{for all } t \geq 0.$$

Let X_i be an i.i.d. sample of μ. Recalling that

$$W_1(L_n, \mu) = \sup_{f\,1-\mathrm{Lip}} \frac{1}{n}\sum_{i=1}^n f(X_i) - \int f\,d\mu$$

and that

$$(x_1, \ldots, x_n) \mapsto \sup_{f\,1-\mathrm{Lip}} \frac{1}{n}\sum_{i=1}^n f(x_i) - \int f\,d\mu$$

is $\frac{1}{n}$-Lipschitz w.r.t. the distance $d^{\oplus n}$ on E^n (as a supremum of $\frac{1}{n}$-Lipschitz functions), the following ensues :

$$\mathbb{P}(W_1(L_n, \mu) \geq \mathbb{E}[W_1(L_n, \mu)] + t) \leq \exp -n\alpha(t). \tag{14}$$

Therefore, we are led to seek a control on $\mathbb{E}[W_1(L_n, \mu)]$. This is what we do in the next lemma.

Lemma 3.2. *Let $a > 0$ be such that $E_{a,1} = \int e^{ad(x,x_0)}\mu(dx) \leq 2$.*

Let $\delta > 0$ and $K \in E$ be a compact subset containing x_0. Let $\mathcal{N}_\delta$ denote the covering number of order δ for the set $\mathcal{F}_K$ of 1-Lipschitz functions on K vanishing at x_0 (endowed with the uniform distance).

Also define $\sigma : [0, +\infty) \to [1, +\infty)$ as the inverse function of $x \mapsto x \ln x - x + 1$ on $[1, +\infty)$.

The following holds :

$$\mathbb{E}[W_1(L_n, \mu)] \leq 2\delta + 8\frac{1}{a}\frac{1}{\sigma(\frac{1}{\mu(K^c)})} + \Gamma(\mathcal{N}_\delta, n)$$

where

$$\Gamma(\mathcal{N}_\delta, n) = \inf_{\lambda > 0} \frac{1}{\lambda}[\log \mathcal{N}_\delta + n\alpha^*(\frac{\lambda}{n})].$$

Proof. We denote by $\mathcal{F}$ the set of 1-Lipschitz functions f over E such that $f(x_0) = 0$. Let us denote

$$\Psi(f) = \int f\,d\mu - \int f\,dL_n,$$

we have for $f, g \in \mathcal{F}$:

$$
\begin{aligned}
|\Psi(f) - \Psi(g)| \quad \leq \quad & \int |f - g|1_K\,d\mu + \int |f - g|1_K\,dL_n \\
& + \int (|f| + |g|)1_{K^c}\,d\mu + \int (|f| + |g|)1_{K^c}\,dL_n \\
\leq \quad & 2\|f - g\|_{L^\infty(K)} + 2\int d(x, x_0)1_{K^c}\,d\mu + 2\int d(x, x_0)1_{K^c}\,dL_n
\end{aligned}
$$

2313

When $f : E \to \mathbb{R}$ is a measurable function, denote by $f|_K$ its restriction to K. Notice that for every $g \in \mathscr{F}_K$, there exists $f \in \mathscr{F}$ such that $f|_K = g$. Indeed, one may set

$$f(x) = \begin{cases} g(x) \text{ if } x \in K \\ \inf_{y \in K} f(y) + d(x, y) \text{ otherwise} \end{cases}$$

and check that f is 1-Lipschitz over E.

By definition of $\mathscr{N}_\delta$, there exist functions $g_1, \ldots, g_{\mathscr{N}_\delta} \in \mathscr{F}_K$ such that the balls of center g_i and radius δ (for the uniform distance) cover $\mathscr{F}_K$. We can extend these functions to functions $f_i \in \mathscr{F}$ as noted above.

Consider $f \in \mathscr{F}$ and choose f_i such that $|f - f_i| \leq \delta$ on K :

$$
\begin{aligned}
\Psi(f) &\leq |\Psi(f) - \Psi(f_i)| + \Psi(f_i) \\
&\leq \Psi(f_i) + 2\delta + 2 \int d(x, x_0) 1_{K^c} d\mu + 2 \int d(x, x_0) 1_{K^c} dL_n \\
&\leq \max_{j=1, \ldots, \mathscr{N}_\delta} \Psi(f_j) + 2\delta + 2 \int d(x, x_0) 1_{K^c} d\mu + 2 \int d(x, x_0) 1_{K^c} dL_n
\end{aligned}
$$

The right-hand side in the last line does not depend on f, so it is also greater than $W_1(L_n, \mu) = \sup_{\mathscr{F}} \Psi(f)$.

We pass to expectations, and bound the terms on the right. We use Orlicz-Hölder's inequality with the pair of conjugate Young functions

$$
\begin{aligned}
\tau(x) &= \begin{cases} 0 \text{ if } x \leq 1 \\ x \log x - x + 1 \text{ otherwise} \end{cases} \\
\tau^*(x) &= e^x - 1
\end{aligned}
$$

(for definitions and a proof of Orlicz-Hölder's inequality, the reader may refer to [29], Chapter 10). We get

$$\int d(x, x_0) 1_{K^c} d\mu \leq 2 \|1_{K^c}\|_\tau \|d(x, x_0)\|_{\tau^*}$$

where

$$\|1_{K^c}\|_\tau = \inf\{\theta > 0, \quad \int \tau\left(\frac{1_{K^c}}{\theta}\right) d\mu \leq 1\}$$

and

$$\|d(x, x_0)\|_{\tau^*} = \inf\{\theta > 0, \quad \int \left[e^{\frac{d(x, x_0)}{\theta}} - 1\right] d\mu \leq 1\}.$$

It is easily seen that $\|1_{K^c}\|_\tau = 1/\sigma(1/\mu(K^c))$. And we assumed that a is such that $E_{a,1} = \int \exp a d(x, x_0) d\mu \leq 2$, so $\|d(x, x_0)\|_{\tau^*} \leq 1/a$. Altogether, this yields

$$\int d(x, x_0) 1_{K^c} d\mu \leq 2\frac{1}{a}\frac{1}{\sigma(\frac{1}{\mu(K^c)})}.$$

Also, if $X_1, \ldots, X_n$ are i.i.d. variables of law μ,

$$\mathbb{E}\left[\int d(x, x_0) 1_{K^c} dL_n\right] = \mathbb{E}[d(X_1, x_0) 1_{K^c}(X_1)] \leq \frac{2}{a}\frac{1}{\sigma(1/\mu(K^c))}$$

as seen above. Putting this together yields the inequality

$$\mathbb{E}[W_1(L_n, \mu)] \leq 2\delta + \frac{8}{a}\frac{1}{\sigma(1/\mu(K^c))} + \mathbb{E}[\max_{j=1,\ldots,\mathcal{N}_\delta} \Psi(f_j)].$$

The remaining term can be bounded by a form of maximal inequality. First fix some i and $\lambda > 0$:
we have

$$
\begin{aligned}
\mathbb{E}[\exp \lambda \Psi(f_i)] &= \mathbb{E}[\exp \frac{\lambda}{n} \sum_{j=1}^{n} (f(X_j) - \int f d\mu)] \\
&= (\mathbb{E}[\exp \frac{\lambda}{n}(f(X_1) - \int f d\mu)])^n \\
&\leq e^{n\alpha^\circledast(\lambda/n)}.
\end{aligned}
$$

In the last line, we have used estimate (19). Using Jensen's inequality, we may then write

$$
\begin{aligned}
\mathbb{E}[\max_{j=1,\ldots,\mathcal{N}_\delta} \Psi(f_j)] &\leq \frac{1}{\lambda} \log \mathbb{E}[\max_{j=1,\ldots,\mathcal{N}_\delta} \exp \lambda \Psi(f_j)] \\
&\leq \frac{1}{\lambda} \log \sum_{j=1}^{\mathcal{N}_\delta} \mathbb{E}[\exp \lambda \Psi(f_j)] \\
&\leq \frac{1}{\lambda}[\log \mathcal{N}_\delta + n\alpha^*(\frac{\lambda}{n})]
\end{aligned}
$$

So minimizing in λ we have

$$\mathbb{E}[\max_{j=1,\ldots,\mathcal{N}_\delta} \Psi(f_j)] \leq \Gamma(\mathcal{N}_\delta, n).$$

Bringing it all together finishes the proof of the lemma.

$\square$

We can now finish the proof of Theorem 2.3.

Proof. Come back to the deviation bound (14). Choose $\delta = t/8$, and choose K such that

$$\mu(K^c) \leq \left[\frac{32}{at} \log \frac{32}{at} - \frac{32}{at} + 1 \right]^{-1}.$$

We thus have $2\delta + 8[a\sigma(1/\mu(K^c))]^{-1} \leq t/2$, which implies

$$\mathbb{E}(W_1(L_n, \mu)) \leq t/2 + \Gamma(\mathscr{C}_t, n) \tag{15}$$

and so

$$\mathbb{P}(W_1(L_n, \mu) \geq t) \leq \exp -n\alpha(\frac{t}{2} - \Gamma(\mathscr{N}_\delta, n)),$$

with the convention $\alpha(y) = 0$ if $y < 0$.

$\square$

4   Proof of Theorem 2.6

In this section, we provide a different approach to our result in the independent case. As earlier we first aim to get a bound on the speed of convergence on the average W_1 distance between empirical and true measure. The lemma below provides another way to obtain such an estimate.

Lemma 4.1. *Let $\mu^k \in \mathscr{P}(E)$ be a finitely supported measure such that $|Supp \, \mu^k| \leq k$. Let $D(\mu^k) = Diam \, Supp \, \mu^k$ be the diameter of $Supp \, \mu^k$. The following holds :*

$$\mathbb{E}W_1(\mu, L_n) \leq 2W_1(\mu, \mu^k) + D(\mu^k)\sqrt{k/n}.$$

Proof. Let π_{opt} be an optimal coupling of μ and μ^k (it exists : see e.g. Theorem 4.1 in [34]), and let (X_i, Y_i), $1 \leq i \leq n$, be i.i.d. variables on $E \times E$ with common law π_{opt}.

Let $L_n = 1/n \sum_{i=1}^n \delta_{X_i}$ and $L_n^k = 1/n \sum_{i=1}^n \delta_{Y_i}$. By the triangle inequality, we have

$$W_1(L_n, \mu) \leq W_1(L_n, L_n^k) + W_1(\mu, \mu^k) + W_1(\mu^k, L_n^k).$$

With our choice of coupling for L_n and L_n^k it is easily seen that

$$\mathbb{E}W_1(L_n, L_n^k) \leq W_1(\mu, \mu^k)$$

Let us take care of the last term. We use Lemma 4.2 below to obtain that

$$\mathbb{E}W_1(L_n^k, \mu^k) \leq D(\mu^k)\mathbb{E}\left(1 - \sum_{i=1}^{k} \mu^k(x_i) \wedge L_n^k(x_i)\right)$$

$$= D(\mu^k)\sum_{i=1}^{k} \mathbb{E}(\mu^k(x_i) - \mu^k(x_i) \wedge L_n^k(x_i))$$

$$\leq D(\mu^k)\sum_{i=1}^{k} \mathbb{E}|\mu^k(x_i) - L_n^k(x_i)|$$

$$\leq \frac{D(\mu^k)}{n}\sum_{i=1}^{k} \sqrt{\mathbb{E}|n\mu^k(x_i) - nL_n^k(x_i)|^2}.$$

Observe that the variables $nL_n^k(x_i)$ follow binomial laws with parameter $\mu^k(x_i)$ and n. We get :

$$\mathbb{E}W_1(\mu^k, L_n^k) \leq \frac{D(\mu^k)}{n}\sum_{i=1}^{k} \sqrt{n\mu^k(x_i)(1 - \mu^k(x_i))} \leq D(\mu^k)\sqrt{k/n}$$

(the last inequality being a consequence of the Cauchy-Schwarz inequality).

$\square$

Lemma 4.2. *Let μ, ν be probability measures with support in a finite metric space $\{x_1, \ldots, x_k\}$ of diameter bounded by D. Then*

$$W_1(\mu, \nu) \leq D\left(1 - \sum_{i=1}^{k} (\mu(x_i) \wedge \nu(x_i))\right).$$

Proof. We build a coupling of μ and ν that leaves as much mass in place as possible, in the following fashion : set $f_i = \mu(x_i) \wedge \nu(x_i)$ and $\lambda = \sum_{i_1}^{k} f_i$. Define the measures

$$f = \sum_{i=1}^{k} f_i \delta_{x_i},$$

$$\mu_1 = \frac{1}{1-\lambda}(\mu - f),$$

$$\nu_1 = \frac{1}{1-\lambda}(\nu - f).$$

Finally, build independent random variables $X_1 \sim \mu_1$, $Y_1 \sim \nu_1$, $Z \sim q$ and B with Bernoulli law of parameter λ. Define

$$X = (1-B)X_1 + BZ, \quad Y = (1-B)Y_1 + BZ.$$

It is an easy check that $X \sim \mu$, $Y \sim \nu$.

Thus we have the bound

$$W_1(\mu, v) \leq \mathbb{E}|X - Y| = (1 - \lambda)\mathbb{E}|X_1 - Y_1| \leq D(1 - \lambda)$$

and this concludes the proof.

$\square$

Proof of Theorem 2.6. As stated earlier, we have the concentration bound

$$\mathbb{P}(W_p(L_n, \mu) \geq t + \mathbb{E}W_p(L_n, \mu)) \leq e^{-nt^2/C}.$$

The proof is concluded by arguments similar to the ones used before, calling upon Lemma 4.1 to bound the mean.

$\square$

5 Proofs in the dependent case

Before proving Theorem 2.13, we establish a more general result in the spirit of Lemma 3.2.

As earlier, the first ingredient we need to apply our strategy of proof is a tensorization property for the transport-entropy inequalities in the case of non-independent sequences. To this end, we restate results from [19], where only $\mathbf{T}_1$ inequalities were investigated, in our framework.

For $x = (x_1, \ldots, x_n) \in E^n$, and $1 \leq i \leq n$, denote $x^i = (x_1, \ldots, x_i)$. Endow E^n with the distance $d_1(x, y) = \sum_{i_1}^n d(x_i, y_i)$. Let $v \in \mathscr{P}(E^n)$, the notation $v^i(dx_1, \ldots, dx_i)$ stands for the marginal measure on E^i, and $v^i(.|x^{i-1})$ stands for the regular conditional law of x_i knowing x^{i-1}, or in other words the conditional disintegration of v^i with respect to v^{i-1} at x^{I-1}(its existence is assumed throughout).

The next theorem is a slight extension of Theorem 2.11 in [19]. Its proof can be adapted without difficulty, and we omit it here.

Theorem 5.1. *Let $v \in \mathscr{P}(E^n)$ be a probability measure such that*

1. *For all $i \geq 1$ and all $x^{i-1} \in E^{i-1}$ ($E^0 = \{x_0\}$), $v^i(.|x^{i-1})$ satisfies a $\alpha(\mathscr{T}_d)$ inequality, and*

2. *There exists $S > 0$ such that for every 1-Lipschitz function*

$$f : (x_{k+1}, \ldots, x_n) \to f(x_{k+1}, \ldots, x_n),$$

for all $x^{k-1} \in E^{k-1}$ and $x_k, y_k \in E$, we have

$$\left| \mathbb{E}_v \left[f(X_{k+1}, \ldots, X_n)|X^k = (x^{k-1}, x_k) \right] - \mathbb{E}_v \left[f(X_{k+1}, \ldots, X_n)|X^k = (x^{k-1}, y_k) \right] \right| \leq Sd(x_k, y_k) \tag{16}$$

2318

Then v verifies the transportation inequality $\tilde{\alpha}(\mathscr{T}_d) \leq H$ with

$$\tilde{\alpha}(t) = n\alpha\left(\frac{1}{n(1+S)}t\right).$$

In the case of a homogeneous Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition kernel $P(x,dy)$, the next proposition gives sufficient conditions on the transition probabilities for the laws of the variables X_n and the path-level law of $(X_1,\ldots,X_n)$ to satisfy some transportation inequalities. Once again the statement and its proof are adaptations of the corresponding Proposition 2.10 of [19].

Proposition 5.2. *Let $P(x,dy)$ be a Markov kernel such that*

1. *the transition measures $P(x,.)$ satisfies $\alpha(\mathscr{T}_d) \leq H$ for all $x \in E$, and*

2. *$W_1(P(x,.),P(y,.)) \leq rd(x,y)$ for all $x,y \in E$ and some $r < 1$.*

Then there exists a unique invariant probability measure π for the Markov chain associated to P, and the measures $P^n(x,.)$ and π satisfy $\alpha'(\mathscr{T}_d) \leq H$, where $\alpha'(t) = \frac{1}{1-r}\alpha((1-r)t)$.

Moreover, under these hypotheses, the conditions of Theorem 5.1 are verified with $S = \frac{r}{1-r}$ so that the law P_n of the n-uple $(X_1,\ldots,X_n)$ under $X_0 = x_0 \in E$ verifies $\tilde{\alpha}(\mathscr{T}_d) \leq H$ where $\tilde{\alpha}(t) = n\alpha(\frac{1-r}{n}t)$.

Proof. The first claim is obtained exactly as in the proof of Proposition 2.10 in [19], observing that the contraction condition 2 is equivalent to

$$W_1(v_1 P, v_2 P) \leq rW_1(v_1, v_2) \quad \text{for all } v_1, v_2 \in \mathscr{P}_1(E)$$

and also to

$$\|Pf\|_{\text{Lip}} \leq r\|f\|_{\text{Lip}} \quad \text{for all } f.$$

This entails that whenever f is 1-Lipschitz, $P^n f$ is r^n-Lipschitz. Now, by condition 1, we have

$$\begin{aligned}
P^n(e^{sf}) &\leq P^{n-1}\left(\exp\left(sPf + \alpha^{\circledast}(s)\right)\right) \\
&\leq P^{n-2}\left(\left(sP^2 f + \alpha^{\circledast}(s) + \alpha^{\circledast}(rs)\right)\right) \\
&\leq \ldots \\
&\leq \exp\left(\left(sP^n f + \alpha^{\circledast}(s) + \ldots + \alpha^{\circledast}(r^n s)\right)\right).
\end{aligned}$$

As $\alpha^{\circledast}$ is convex and vanishes at 0, we have $\alpha^{\circledast}(rt) \leq r\alpha^{\circledast}(t)$ for all $t \geq 0$. Thus,

$$P^n(e^{sf} \leq \exp\left(sP^n f + \sum_{k=0}^{+\infty} r^k \alpha^{\circledast}(s)\right) = \exp\left(sP^n f + \frac{1}{1-r}\alpha^{\circledast}(s)\right)$$

It remains only to check that $\frac{1}{1-r}\alpha^{\circledast}$ is the monotone conjugate of α' and to invoke Proposition A.2.

Moving on to the final claim, since the process is homogeneous, to ensure that (16) is satisfied, we need only show that for all $k \geq 1$, for all $f : E^k \to \mathbb{R}$ 1-Lipschitz w.r.t. d_1, the function

$$x \mapsto \mathbb{E}\left[f(X_1, \ldots, X_k)|X_0 = x\right]$$

is $\frac{r}{1-r}$-Lipschitz. We show the following : if $g : E^2 \to \mathbb{R}$ is such that for all $x_1, x_2 \in E$ the functions $g(., x_2)$, resp. $g(x_1, .)$, are 1-Lipschitz, resp. λ-Lipschitz, then the function

$$x_1 \mapsto \int g(x_1, x_2) P(x_1, dx_2)$$

is $(1 + \lambda r)$-Lipschitz. Indeed,

$$|\int g(x_1, x_2) P(x_1, dx_2) - \int g(y_1, x_2) P(y_1, dx_2)|$$
$$\leq \int |g(x_1, x_2) - g(y_1, x_2)| P(x_1, dx_2) + |\int g(y_1, x_2)(P(x_1, dx_2) - P(y_1, dx_2))|$$
$$\leq (1 + \lambda r) d(x_1, y_1).$$

It follows easily by induction that the function

$$f_k : x_1 \mapsto \int f(x_1, \ldots, x_k) P(x_{k-1}, dx_k) \ldots P(x_1, dx_2)$$

has Lipschitz norm bounded by $1 + r + \ldots r^k \leq \frac{1}{1-r}$. Hence the function $x \mapsto \int f_k(x_1) P(x, dx_1)$ has Lipschitz norm bounded by $\frac{r}{1-r}$. But this function is precisely

$$x \mapsto \mathbb{E}\left[f(X_1, \ldots, X_k)|X_0 = x\right]$$

and the proof is complete. $\qquad\square$

We are in position to prove the analogue of Lemma 3.2 in the Markov case.

Lemma 5.3. *Consider the Markov chain associated to a transition kernel P as in Proposition 5.2. Let P_n denote the law of the Markov path $(X_1, \ldots, X_n)$ associated to P under $X_0 = x_0$. Introduce the averaged occupation measure $\pi_n = \mathbb{E}_{P_n}(L_n)$ and the invariant measure π. Let $m_1 = \int d(x, x_0) \pi(dx)$.*

Suppose that there exists $a > 0$ such that for all $i \geq 1$ $E_{a,i} = \int e^{ad(x,x_0)} P^i(dx) \leq 2$.

Let $\delta > 0$ and $K \Subset E$ be a compact subset containing x_0. Let $\mathcal{N}_\delta$ denote the metric entropy of order δ for the set $\mathcal{F}_K$ of 1-Lipschitz functions on K vanishing at x_0 (endowed with the uniform distance). Also define $\sigma : [0, +\infty) \to [1, +\infty)$ as the inverse function of $x \mapsto x \ln x - x + 1$ on $[1, +\infty)$.

The following holds :

$$\mathbb{E}_{P_n}[W_1(L_n, \pi_n)] \leq 2\delta + \frac{8}{a} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sigma(\frac{1}{P^i(K^c)})} + \Gamma(\mathcal{N}_\delta, n)$$
$$W_1(\pi_n, \pi) \leq \frac{m_1}{n(1-r)}.$$

where $\Gamma(\mathcal{N}_\delta, n) = \inf_{\lambda>0} \frac{1}{\lambda} \left[\log \mathcal{N}_\delta + n\alpha^\circledast(\frac{\lambda}{n(1-r)}) \right]$

Proof. Convergence to the equilibrium measure is dealt with using the contraction hypothesis. Indeed, by convexity of the map $\mu \mapsto W_1(\mu, \pi)$, we first have

$$W_1(\pi_n, \pi) \leq \frac{1}{n} \sum_{i=1}^{n} W_1(P^i(x_0, .), \pi).$$

Now, using that the contraction property (2) in Proposition 5.2 is equivalent to the inequality $W_1(\mu_1 P, \mu_2 P) \leq r W_1(\mu_1, \mu_2)$ for all $\mu_1, \mu_2 \in \mathscr{P}_1(E)$, and using the fact that π is P-invariant,

$$W_1(\pi_n, \pi) \leq \frac{1}{n} \sum_{i=1}^{n} r^i W_1(\delta_{x_0}, \pi) \leq \frac{W_1(\delta_{x_0}, \pi)}{n(1-r)} = \frac{m_1}{n(1-r)}.$$

In order to take care of the second term, we will use the same strategy (and notations) as in the independent case. Introduce once again a compact subset $K \subset E$ and a covering of $\mathscr{F}_K$ by functions $f_1, \ldots, f_{\mathcal{N}_\delta}$ suitably extendend to E. With the same arguments as before, we get

$$\mathbb{E}_{P_n} W_1(L_n, \pi_n) \leq \mathbb{E}_{P_n} \left(\max_{j=1,\ldots,\mathcal{N}_\delta} \Psi(f_j) \right) + 2\delta + 2 \int d(x, x_0) 1_{K^c} d\pi_n$$
$$+ 2\mathbb{E}_{P_n} \left(\int d(x, x_0) 1_{K^c} dL_n \right)$$

Then,

$$\int d(x_0, y) \pi_n(dy) = \frac{1}{n} \sum_{i=1}^{n} \int d(x_0, y) 1_{K^c} P^i(x_0, dy).$$

As before we can use Orlicz-Hölder's inequality to recover the bound

$$\int d(x_0, y) d\pi_n \leq \frac{2}{a} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sigma(\frac{1}{P^i(K^c)})}.$$

And likewise,

$$\mathbb{E}\left(\int d(x, x_0) 1_{K^c} dL_n \right) = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} d(x_0, X_i) 1_{K^c} \right]$$
$$= \frac{1}{n} \sum_{i=1}^{n} \int d(x_0, y) 1_{K^c} P^i(x_0, dy)$$

and we have the same bound as above.

As for the last term remaining : it will be possible to use the maximal inequality techniques just as in the proof of Theorem 2.3, provided that we can find bounds on the terms $\mathbb{E}\left[\exp \lambda \Psi(f_j)\right]$, where this time

$$\Psi(f) = \int f \, dL_n - \int f \, d\pi_n.$$

Denote

$$F_j(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} f_j(x_i).$$

This is a $\frac{1}{n}$-Lipschitz function on E^n. Since P_n satisfies a $\tilde{\alpha}(\mathcal{T}_d) \le H$ inequality, we have

$$\int \exp \lambda F_j \, dP_n \le \exp\left[\lambda \int F_j \, dP_n + n\alpha^{\circledast}(\frac{\lambda}{n(1-r)})\right].$$

But this is exactly the bound

$$\mathbb{E}\left[\exp \lambda \Psi(f_j)\right] \le e^{n\alpha^{\circledast}(\frac{\lambda}{n(1-r)})}.$$

We may then proceed as in the independent case and obtain

$$\mathbb{E}[\max_{j=1,\ldots,\mathcal{N}_\delta} \Psi(f_j)] \le \inf_{\lambda > 0} \frac{1}{\lambda}\left[\log \mathcal{N}_\delta + n\alpha^{\circledast}(\frac{\lambda}{n(1-r)})\right].$$

$\square$

For any Lipschitz function $f : E^n \to \mathbb{R}$ (w.r.t. d_1), we have the concentration inequality

$$P_n(x \in E^n, \quad f(x) \ge \int f \, dP_n + t) \le \exp -n\alpha\left(\frac{(1-r)t}{n\|f\|_{\text{Lip}}}\right).$$

Remembering that $E^n \ni x \mapsto W_1(L_n^x, \pi_n)$ is $\frac{1}{n}$-Lipschitz, we get the bound

$$\mathbb{P}(W_1(L_n, \pi_n) \ge \mathbb{E}_{P_n}[W_1(L_n, \pi_n)] + t) \le \exp -n\alpha((1-r)t). \tag{17}$$

Thanks to the triangular inequality $W_1(L_n, \pi_n) \ge W_1(L_n, \pi) - W_1(\pi_n, \pi)$, it holds that

$$\mathbb{P}(W_1(L_n, \pi) \ge W_1(\pi_n, \pi) + \mathbb{E}_{P_n}[W_1(L_n, \pi_n)] + t) \le \exp -n\alpha((1-r)t). \tag{18}$$

This in turn leads to an estimate on the deviations, under the condition that we may exhibit a compact set with large measure for all the measures P^i. We now move on to the proof of Theorem 2.13.

Proof of Theorem 2.13. Fix $\delta = t/8$. Set $m_1^i = \int |x| P^i(dx)$. We have

$$m_1^i \leq m_1 + W_1(P^i, \pi) \leq m_1 + r^i W_1(\delta_0, \pi)$$
$$\leq 2m_1.$$

Thus

$$\int e^{a|x|} P^i(dx) \leq e^{am_1^i + Ca^2/4} \leq e^{2m_1 a + Ca^2/4}.$$

With a as in the theorem, the above ensures that $\int e^{a|x|} P^i(dx) \leq 2$.

Let B_R denote the ball of center 0 and radius R : we have $P^i(B_R^c) \leq 2e^{-aR}$. Let

$$R = \frac{1}{a} \log 2\sigma^{-1}(\frac{32}{at}).$$

so that $2\delta + \frac{8}{a}\frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sigma(\frac{1}{P^i(K^c)})} \leq t/2$.

As $\alpha(t) = \frac{1}{C} t^2$ we can compute

$$\Gamma(\mathcal{N}_\delta, n) = \frac{1}{1-r}\sqrt{\frac{C}{n}}\sqrt{\log \mathcal{N}_\delta}.$$

We have chosen $K = B_R$ and $\delta = t/8$. Working as in the proof of Proposition 2.8, when $t \leq 2/a$, we can bound $\log \mathcal{N}_\delta$ by

$$\log \mathcal{N}_\delta \leq C_d (\frac{1}{at} \log \frac{1}{at})^d$$

where C_d is a numerical constant depending on the dimension d. Plugging the above estimates in (18) and using the inequality $(u-v)^2 \geq u^2/2 - v^2$ gives the desired result.

$\square$

A Some facts on transportation inequalities

A.1 Links with transportation inequalities

A crucial feature of transportation inequalities is that they imply the concentration of measure phenomenon, a fact first discovered by Marton ([25]). The following proposition is obtained by a straightforward adaptation of her famous argument :

Proposition A.1. *If μ verifies a $\alpha(\mathcal{T}_d)$ inequality, then for all measurable sets $A \subset E$ with $\mu(A) \geq \frac{1}{2}$ and $r \geq r_0 = \alpha^{-1}(\log 2)$,*

$$\mu(A^r) \geq 1 - e^{-\alpha(r-r_0)}$$

where $A^r = \{x \in E, d(x, A) \le r\}$.

Moreover, let X be a r.v. with law μ. For all 1-Lipschitz functions $f : E \to \mathbb{R}$ and all $r \ge r_0$, we have

$$\mathbb{P}(f(X) \ge m_f + r) \le e^{-\alpha(r - r_0)}$$

where m_f denotes a median of f.

Bobkov and Götze ([5]) were the first to obtain an equivalent dual formulation of transportation inequalities. We present it here in a more general form obtained by Gozlan and Leonard (see [17]), in the case when the transportation cost function is the distance.

Proposition A.2 ([17]). *Assume that d is a metric defining the topology of E, and that μ admits some finite exponential moment.*

Then μ satisfies the $\alpha(\mathcal{T}_d)$ inequality

$$\alpha(\mathcal{T}_d(\mu, \nu)) \le H(\nu | \mu)$$

for all $\nu \in \mathscr{P}(E)$ with finite first moment if and only if for all $f : E \to \mathbb{R}$ 1-Lipschitz and all $\lambda > 0$,

$$\int e^{\lambda(f(x) - \int f d\mu)} \mu(dx) \le e^{\alpha^{\circledast}(\lambda)}. \tag{19}$$

Recall that $\alpha^{\circledast}$ is the monotone conjugate of α, see Definition 2.2.

In the case $\mathbf{T}_1(C)$, Condition (19) becomes : for all 1-Lipschitz $f : E \to \mathbb{R}$ and $\lambda > 0$,

$$\int e^{\lambda(f - \int f d\mu)} \mu(x) \le e^{C\lambda^2/4}. \tag{20}$$

Observe that finiteness of a single-exponential moment is a necessary condition for the existence of a $\alpha(\mathcal{T}_d)$ inequality with α a convex increasing positive function. To see this, consider the statement of Proposition A.1, and notice that α must be super-linear.

The next proposition is a straightforward application of the Bobkiv-Götze criterion.

Proposition A.3. *Assume that μ satisfies a $\alpha(\mathcal{T}_d)$ inequality. Denote*

$$m_1 = \int d(x_0, x) d\mu.$$

As soon as $\lambda m_1 + \alpha^{\circledast}(\lambda) \le \log z$, we have

$$E_{\lambda, 1} = \int e^{\lambda d(x, x_0)} d\mu \le z.$$

A.2 Integral criteria

An interesting feature of transportation inequalities is that some of them are characterized by simple moment conditions, making it tractable to obtain their existence. In [19], Djellout, Guillin and Wu showed that μ satisfies a $\mathbf{T}_1$ inequality if and only if $\int \exp[a_0 d^2(x_0, y)]\mu(dy) < +\infty$ for some a_0 and some x_0. They also connect the value of a_0 and of the Gaussian moment with the value of the constant C appearing in the transportation inequality. More generally, Gozlan and Leonard provide in [16] a nice criterion to ensure that a $\alpha(\mathbf{T}_d)$ inequality holds. We only quote here one side of what is actually an equivalence :

Theorem A.4. *Let $\mu \in \mathscr{P}(E)$. Suppose there exists $a > 0$ with $\int e^{ad(x_0,x)}\mu(dx) \leq 2$ for some $x_0 \in E$, and a convex, increasing function γ on $[0, +\infty)$ vanishing at 0 and $x_1 \in E$ such that*

$$\int \exp \gamma(d(x_1, x))\mu(dx) \leq B < +\infty.$$

Then μ satisfies the $\alpha(\mathscr{T}_d)$ inequality

$$\alpha(W_1(\mu, \nu)) \leq H(\nu|\mu)$$

for all $\nu \in \mathscr{P}(E)$ with finite first moment, with

$$\alpha(x) = \max\left((\sqrt{ax+1} - 1)^2, \, 2\gamma(\frac{x}{2} - 2\log B)\right).$$

One particular instance of the result above was first obtained by Bolley and Villani, with sharper constants, in the case when μ only has a finite exponential moment ([9], Corollary 2.6). Their technique involves the study of weighted Pinsker inequalities, and encompasses more generally costs of the form d^p, $p \geq 1$ (we give only the case $p = 1$ here).

Theorem A.5. *Let $a > 0$ be such that $E_{a,1} = \int e^{ad(x_0,x)}\mu(dx) < +\infty$. Then for $\nu \in \mathscr{P}_1(E)$, we have*

$$W_1(\mu, \nu) \leq C\left(H(\nu|\mu) + \sqrt{H(\nu|\mu)}\right)$$

where $C = \frac{2}{a}\left(\frac{3}{2} + \log E_{a,1}\right) < +\infty$.

And in the case when μ admits a finite Gaussian moment, the following holds ([9], Corollary 2.4) :

Theorem A.6. *Let $a > 0$ be such that $E_{a,2} = \int e^{ad^2(x_0,x)}\mu(dx) < +\infty$. Then μ satisfies a $\mathbf{T}_1(C)$ inequality where $C = \frac{2}{a}\left(1 + \log E_{a,2}\right) < +\infty$.*

B   Covering numbers of the set of 1-Lipschitz functions

In this section, we provide bounds for the covering numbers of the set of 1-Lipschitz functions over a precompact space.

Note that these results are likely not new. However, we have been unable to find an original work, so we provide proofs for completeness.

Let (K, d) be a precompact metric space, and let $\mathcal{N}(K, \delta)$ denote the minimal number of balls of radius δ necessary to cover K. Let $x_0 \in K$ be a fixed point, and let $\mathcal{F}$ denote the set of 1-Lipschitz functions over K vanishing at x_0. This is also a precompact space when endowed with the metric of uniform convergence. We denote by $\mathcal{N}(\mathcal{F}, \delta)$ the minimal number of balls of radius δ necessary to cover $\mathcal{F}$. Finally, we set $R = \max_{x \in K} d(x, x_0)$.

Our first estimate is a fairly crude one.

Proposition B.1. *We have*

$$\mathcal{N}(\mathcal{F}, \varepsilon) \leq \left(2 + 2 \lfloor \frac{3R}{\varepsilon} \rfloor \right)^{\mathcal{N}(K, \frac{\varepsilon}{3})}.$$

Proof. For simplicity, write $n = \mathcal{N}(K, \varepsilon)$. Let $x_1, \ldots, x_n$ be the centers of a set of balls covering K. For any $f \in \mathcal{F}$ and $1 \leq i \leq n$, we have

$$|f(x_i)| = |f(x_i) - f(x_0)| \leq R.$$

For any n-uple of integers $k = (k_1, \ldots, k_n)$ such that $-\lfloor \frac{R}{\varepsilon} \rfloor - 1 \leq k_i \leq \lfloor \frac{R}{\varepsilon} \rfloor$, $1 \leq i \leq n$, choose a function $f_k \in \mathcal{F}$ such that $k_i \varepsilon \leq f_k(x_i) \leq (k_i + 1)\varepsilon$ if there exists one.

Consider $f \in \mathcal{F}$. Let $l_i = \lfloor \frac{f(x_i)}{\varepsilon} \rfloor$ and $l = (l_1, \ldots, l_n)$. Then the function f_l defined above exists and $|f(x_i) - f_l(x_i)| \leq \varepsilon$ for $1 \leq i \leq n$. But then for any $x \in K$ there exists i, $1 \leq i \leq n$, such that $x \in B(x_i, \varepsilon)$, and thus

$$|f(x) - f_l(x)| \leq |f(x) - f(x_i)| + |f(x_i) - f_l(x_i)| + |f_l(x_i) - f_l(x)| \leq 3\varepsilon.$$

This implies that $\mathcal{F}$ is covered by the balls of center f_k and radius 3ε. As there are at most $(2 + 2\lfloor \frac{R}{\varepsilon} \rfloor)^n$ choices for k, this ends the proof.

$\square$

However, this bound is quite weak : as one can see by considering the case of a segment, for most choices of a n-uple, there will not exist a function in $\mathcal{F}$ satisfying the requirements in the proof. With the extra assumption that K is connected, we can get a more refined result.

Proposition B.2. *If K is connected, then*

$$\mathcal{N}(\mathcal{F}, \varepsilon) \leq \left(2 + 2 \lfloor \frac{4R}{\varepsilon} \rfloor \right) 2^{\mathcal{N}(K, \frac{\varepsilon}{16})}. \tag{21}$$

Remark. The simple idea in this proposition is first to bring the problem to a discrete metric space (graph), and then to bound the number of Lipschitz functions on this graph by the number of Lipschitz functions on a spanning tree of the graph.

Proof. In the following, we will denote $n = \mathcal{N}(K, \varepsilon)$ for simplicity. Let x_i, $1 \leq i \leq n$ be the centers of a set of n balls $B_1, \ldots B_n$ covering K. Consider the graph G built on the n vertices $a_1, \ldots, a_n$, where vertices a_i and a_j are connected if and only if $i \neq j$ and the balls B_i and B_j have a non-empty intersection.

Lemma B.3. *The graph G is connected. Moreover, there exists a subgraph G' with the same set of vertices and whose edges are edges of G, which is a tree.*

Proof. Suppose that G were not connected . Upon exchanging the labels of the balls, there would exist k, $1 \leq k < n$, such that for $i \leq k < j$ the balls B_i and B_j have empty intersection. But then K would be equal to the disjoint reunion of the sets $\bigcup_{i=1}^{k} B_i$ and $\bigcup_{j=k+1}^{n} B_j$, which are both closed and non-empty, contradicting the connectedness of K.

The second part of the claim is obtained by an easy induction on the size of the graph. $\qquad\square$

Introduce the set $\mathscr{A}$ of functions $g : \{a_1, \ldots, a_n\} \to \mathbb{R}$ such that $g(a_1) = 0$ and $|g(a_i) - g(a_j)| = 4\varepsilon$ whenever a_i and a_j are connected in G'. Using the fact that G' is a tree, it is easy to see that $\mathscr{A}$ contains at most 2^n elements.

Define a partition of K by setting $C_1 = B_1$, $C_2 = B_2 \backslash C_1$, …, $C_n = B_n \backslash C_{n-1}$ (remark that none of the C_i is empty since the B_i are supposed to constitute a minimal covering). Also fix for each i, $1 \leq i \leq n$, a point $y_i \in C_i$ (choosing $y_1 = x_1$). Notice that C_i is included in the ball of center y_i and radius 2ε, and that $d(y_i, y_j) \leq 4\varepsilon$ whenever a_i and a_j are connected in G (and therefore in G').

To every 1-Lipschitz function $f : K \to \mathbb{R}$ we associate $T(f) : \{a_1, \ldots, a_n\} \to \mathbb{R}$ defined by $T(f)(a_i) = f(y_i)$. For any $x \in K$, and $f_1, f_2 \in \mathscr{F}$, we have the following :

$$
\begin{aligned}
|f_1(x) - f_2(x)| &\leq |f_1(x) - f_1(y_i)| + |f_1(y_i) - f_2(y_i)| + |f_2(y_i) - f_2(x)| \\
&\leq 4\varepsilon + \|T(f_1) - T(f_2)\|_{\ell_\infty(G')}
\end{aligned}
$$

where i is such that $x \in C_i$. We now make the following claim :

Lemma B.4. *For every 1-Lipschitz function $f : K \to \mathbb{R}$ such that $f(y_1) = 0$, there exists $g \in \mathscr{A}$ such that $\|T(f) - g\|_{\ell_\infty(G')} \leq 4\varepsilon$.*

Assume for the moment that this holds. As there are at most 2^n functions in $\mathscr{A}$, it is possible to choose at most 2^n 1-Lipschitz functions $f_1, \ldots, f_{2^n}$ vanishing at x_1 such that for any 1-Lipschitz function f vanishing at x_1 there exists f_i such that $|T(f) - T(f_i)| \leq 8\varepsilon$. Using the inequality above, this implies that the balls of center f_i and radius 12ε for the uniform distance cover the set of 1-Lipschitz functions vanishing at x_1.

Finally, consider $f \in \mathscr{F}$. We may write

$$
f = f - f(x_1) + f(x_1)
$$

and observe that on the one hand, $f - f(x_1)$ is a 1-Lipschitz function vanishing at x_1, and that on the other hand, $|f(x_1)| \leq R$. Thus the set $\mathscr{F}$ is covered by the balls of center $f_i + 4k\varepsilon$ and radius 16ε, where $-\lfloor \frac{R}{4\varepsilon} \rfloor - 1 \leq k \leq \lfloor \frac{R}{4\varepsilon} \rfloor$. There are at most $(2 + 2\lfloor \frac{R}{4\varepsilon} \rfloor)2^n$ such balls, which proves the desired result.

$\qquad\square$

We now prove Lemma B.4.

Proof. Let us use induction again. If $K = B_1$ then $T(f) = 0$ and the property is straightforward. Now if $K = C_1 \cup \ldots \cup C_n$, we may assume without loss of generality that a_n is a leaf in G', that is a vertex with exactly one neighbor, and that it is connected to a_{n-1}. By hypothesis there exists $\tilde{g} : \{a_1, \ldots, a_{n-1}\} \to \mathbb{R}$ such that $|\tilde{g}(a_i) - \tilde{g}(a_j)| = 4\varepsilon$ whenever a_i and a_j are connected in G', and $|\tilde{g}(a_i) - f(a_i)| \leq 4\varepsilon$ for $1 \leq i < n$. Set $g = \tilde{g}$ on $\{a_1, \ldots, a_{n-1}\}$, and

- $g(a_n) = g(a_{n-1}) + 4\varepsilon$ if $f(y_n) - g(a_{n-1}) < 0$,

- $g(a_n) = g(a_{n-1}) - 4\varepsilon$ otherwise.

Since

$$|f(y_n) - g(a_{n-1})| \leq |f(y_n) - f(y_{n-1})| + |f(y_{n-1}) - g(a_{n-1})| \leq 8\varepsilon$$

it is easily checked that $|f(y_n) - g(a_n)| \leq 4\varepsilon$. The function g belongs to $\mathscr{A}$ and our claim is proved.

$\square$

C Hölder moments of stochastic processes

We quote the following result from Revuz and Yor's book [30] (actually the value of the constant is not given in their statement but is easily tracked in the proof).

Theorem C.1. *Let X_t, $t \in [0, 1]$ be a Banach-valued process such that there exist $\gamma, \varepsilon, c > 0$ with*

$$\mathbb{E}\left[|X_t - X_s|^\gamma\right] \leq c|t - s|^{1+\varepsilon},$$

then there exists a modification $\tilde{X}$ of X such that

$$\mathbb{E}\left[\left(\sup_{s \neq t} \frac{|\tilde{X}_t - \tilde{X}_s|}{|t - s|^\alpha}\right)^\gamma\right]^{1/\gamma} \leq 2^{1+\alpha}(2c)^{1/\gamma}\frac{1}{1 - 2^{\alpha - \varepsilon/\gamma}}$$

for all $0 \leq \alpha < \varepsilon/\gamma$.

Corollary C.2. *Let $(B_t)_{0 \leq t \leq 1}$ denote the standard Brownian motion on $[0, 1]$. Let $m_\alpha = \mathbb{E}\sup_t \|B_t\|_\alpha$ and $s_\alpha^2 = \mathbb{E}(\sup_t \|B_t\|_\alpha)^2$, then m_α and s_α are bounded by*

$$C_\alpha = 2^{1+\alpha}\frac{2^{(1-2\alpha)/4}}{1 - 2^{(2\alpha-1)/4}}\|Z\|_{4/(1-2\alpha)}$$

where $\|Z\|_p$ denotes the L_p norm of a $\mathcal{N}(0, 1)$ variable Z.

Proof. Since the increments of the Brownian motion are Gaussian, we have for every $p > 0$

$$\mathbb{E}[|B_t - B_s|^{2p}] = K_p|t - s|^p$$

with $K_p = \sqrt{2\pi}^{-1}\int_{-\infty}^{+\infty}|x|^{2p}e^{-x^2/2}dx$. Choose p such that $\alpha < (p-1)/2p$, then

$$\left(\mathbb{E}\|X\|_\alpha^{2p}\right)^{1/2p} \le \frac{2^{1+\alpha}}{1-2^{\alpha-1/2+1/2p}}(2K_p)^{1/2p}.$$

A suitable choice is $1/p = 1/2 - \alpha$, and the right-hand side becomes

$$C_\alpha = \frac{2^{1+\alpha}}{1-2^{(\alpha-1/2)/2}}(2G_\alpha)^{(1/2-\alpha)/2}$$

with $G_\alpha = \sqrt{2\pi}^{-1} \int_{-\infty}^{+\infty} |x|^{4/(1-2\alpha)} e^{-x^2/2} dx$. By Hölder's inequality, the result follows.

$\square$

Corollary C.3. *Let X_t be the solution on $[0, T]$ of*

$$dX_t = \sigma(X_t)dB_t + b(X_t)dt$$

with $\sigma, b : \mathbb{R} \to \mathbb{R}$ locally Lipschitz and satisfying the following hypotheses :

- $\sup_x |\sqrt{tr\sigma(x)^t\sigma(x)}| \le A,$
- $\sup_x |b(x)| \le B.$

Then for $\alpha < 1/2$, for p such that $\alpha < (p-1)/2p$, there exists $C < +\infty$ depending explicitely on A, B, T, α ,p such that

$$\mathbb{E}\|X\|_\alpha^p \le C.$$

Proof. We first apply Itô's formula to the function $|X_t - X_s|^2$: this yields

$$\mathbb{E}|X_t - X_s|^2 \le 2B \int_s^t \mathbb{E}|X_u - X_s|du + A|t - s|.$$

Using the elementary inequality $x \le 1/2(1+x^2)$, we get

$$\mathbb{E}|X_t - X_s|^2 \le B \int_s^t \mathbb{E}|X_u - X_s|^2 du + (A+B)|t - s|.$$

Gronwall's lemma entails

$$\mathbb{E}|X_t - X_s|^2 \le (A+B)e^{BT}|t - s|$$

Likewise, applying Itô's formula to $|X_t - X_s|^4$, we get

$$\mathbb{E}|X_t - X_s|^4 \le 4B \int_s^t \mathbb{E}|X_u - X_s|^3 ds + 6A \int_s^t \mathbb{E}|X_u - X_s|^2 du$$

$$\le (6A + 2B) \int_s^t \mathbb{E}|X_u - X_s|^2 du + 2B \int_s^t \mathbb{E}|X_u - X_s|^4 du$$

$$\le \frac{1}{2}(6A + 2B)(A + B)e^{BT}|t - s|^2 + 2B \int_s^t \mathbb{E}|X_u - X_s|^4 du$$

and by Gronwall's lemma $\mathbb{E}|X_t - X_s|^4 \le \frac{1}{2}(6A + 2B)(A + B)e^{3BT}|t - s|^2$. By an easy recurrence, following the above, one may show that

$$\mathbb{E}|X_t - X_s|^{2p} \le C(A, B, T, p)|t - s|^p.$$

To conclude it suffices to call on Theorem C.1.

$\square$

D Transportation inequalities for Gaussian measures on a Banach space

The next proposition is a bound on the size of the tail of a Gaussian measure. A proof may be found e.g. in [22], Chapter 4.

Lemma D.1. *Let (E, μ) be a Gaussian Banach space, and define $m = \int \|x\| \mu(dx)$. Also let σ^2 denote the weak variance of μ. The tail of μ is bounded as follows : for all $R \ge 0$,*

$$\mu\{x \in E, \|x\| \ge m + R\} \le e^{-R^2/2\sigma^2}.$$

Finally we collect some (loose) results on the Wiener measure on the Banach space $(C([0, 1], \mathbb{R}), \|.\|_\infty)$.

Lemma D.2. *The Wiener measure satisfies a $\mathbf{T}_2(8)$ inequality (and therefore a $\mathbf{T}_1(8)$ inequality).*

Proof. The Wiener measure satisfies the $\mathbf{T}_2(2\sigma^2)$ inequality, where

$$\sigma^2 = \sup_\mu \mathbb{E}\left(\int_0^1 B_s d\mu(s)\right)^2$$

and the supremum runs over all Radon measures on $[0, 1]$ with total variation bounded by 1. Note that the weak variance σ^2 is bounded by the variance s^2 defined as $s^2 = \mathbb{E}(\sup_t |B_t|)^2$ (here and hereafter $\sup_t |B_t|$ refers to the supremum on $[0, 1]$). In turn we can give a (quite crude) bound on s : write $\sup_t |B_t| \le \sup_t B_t - \inf_t B_t$, thus $(\sup_t |B_t|)^2 \le (\sup_t B_t - \inf_t B_t)^2 \le 2(\sup_t B_t)^2 + 2(-\inf_t B_t)^2$. Remember the well-known fact that $\sup_t B_t$, $-\inf_t B_t$ and $|B_1|$ have the same law, so that

$$\mathbb{E}(\sup_t |B_t|)^2 \leq 4\mathbb{E}|B_1|^2 = 4.$$

$\square$

Lemma D.3. *Let γ denote the Wiener measure. For $a = \sqrt{2\log 2}/3$, we have*

$$\int e^{a\|x\|_\infty} \gamma(dx) \leq 2$$

.

Proof. We have

$$
\begin{aligned}
\int e^{a\|x\|_\infty} \gamma(dx) &= \int_0^{+\infty} \mathbb{P}(e^{a\|x\|_\infty} \geq t)dt \\
&= \int_0^{+\infty} \mathbb{P}(\|x\|_\infty \geq u)ae^{au}du \\
&= \int_0^{+\infty} \mathbb{P}(\tau_u \leq 1)ae^{au}du
\end{aligned}
$$

where τ_u is the stopping time $\inf\{t, |B_t| = u\}$. It is a simple exercise to compute

$$\mathbb{E}e^{-\lambda^2 \tau_u/2} = 1/\cosh(\lambda u) \leq 2e^{-\lambda u}.$$

This yields

$$\int e^{a\|x\|_\infty} \gamma(dx) \leq 2ae^{\lambda^2/2} \int_0^{+\infty} e^{(a-\lambda)u}du = \frac{2ae^{\lambda^2/2}}{\lambda - a}.$$

We can choose $\lambda = 3a$ to get $\int e^{a\|x\|_\infty} \gamma(dx) \leq e^{9a^2/2}$. In turn it implies the desired result.

$\square$

References

[1] M. Ajtai, J. Komlos, and G. Tusnády. On optimal matchings. *Combinatorica*, 4(4):259–264, 1984. MR0779885

[2] P. Baldi, G. Ben Arous, and G. Kerkyacharian. Large deviations and the Strassen theorem in Hölder norm. *Stochastic processes and their applications*, 42(1):171–180, 1992. MR1172514

[3] F. Barthe and C. Bordenave. Combinatorial optimization over two random point sets. *Arxiv preprint arXiv:1103.2734*, 2011.

[4] S. Bobkov and M. Ledoux. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probability Theory and Related Fields*, 107(3):383–400, 1997. MR1440138

[5] S.G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 163(1):1–28, 1999. MR1682772

[6] E. Boissard and T.L. Gouic. On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. *Arxiv preprint arXiv:1105.5263*, 2011.

[7] F. Bolley. Quantitative concentration inequalities on sample path space for mean field interaction. *ESAIM: PS, July 2010, Vol. 14, p. 192-209*, 2010. MR2741965

[8] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137:541–593, 2007. MR2280433

[9] F. Bolley and C. Villani. Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. In *Annales Faculté des Sciences Toulouse Mathématiques*, volume 14, page 331. Université Paul Sabatier, 2005. MR2172583

[10] E. Del Barrio, E. Giné, and C. Matrán. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Annals of Probability*, 27(2):1009–1071, 1999. MR1698999

[11] A. Dembo and O. Zeitouni. *Large Deviations Techniques Amd Applications*. Jones and Bartlett, 1993. MR1202429

[12] S. Dereich, F. Fehringer, A. Matoussi, and M. Scheutzow. On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces. *Journal of Theoretical Probability*, 16(1):249–265, 2003. MR1956830

[13] V. Dobric and J.E. Yukich. Exact asymptotics for transportation cost in high dimensions. *J. Theoretical Prob*, pages 97–118, 1995. MR1308672

[14] R.M. Dudley. The speed of mean Glivenko-Cantelli convergence. *The Annals of Mathematical Statistics*, 40(1):40–50, 1969. MR0236977

[15] F. Fehringer. *Kodierung von Gaußmaßen*. PhD thesis, 2001.

[16] N. Gozlan and C. Léonard. A large deviation approach to some transportation cost inequalities. *Probability Theory and Related Fields*, 139:235–283, 2007. MR2322697

[17] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Processes and Related Fields 16 (2010) 635-736*, 2010.

[18] S. Graf, H. Luschgy, and G. Pages. Functional quantization and small ball probabilities for Gaussian processes. *Journal of Theoretical Probability*, 16(4):1047–1062, 2003. MR2033197

[19] A. Guillin H. Djellout and L. Wu. Transportation cost-information inequalities for random dynamical systems and diffusions. *Annals of Probability*, 32:2702–2732, 2004. MR2078555

[20] A. Joulin and Y. Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *The Annals of Probability*, 38(6):2418–2442, 2010. MR2683634

[21] J. Kuelbs and W.V. Li. Metric entropy and the small ball problem for Gaussian measures. *Journal of Functional Analysis*, 116(1):133–157, 1993. MR1237989

[22] M. Ledoux. Isoperimetry and Gaussian analysis. *Lectures on probability theory and statistics*, pages 165–294, 1996. MR1600888

[23] M. Ledoux. The concentration of measure phenomenon, volume 89 of Mathematical Surveys and Monographs. *American Mathematical Society, Providence, RI*, 208:2005–2006, 2001. MR1849347

[24] W.V. Li and W. Linde. Approximation, metric entropy and small ball estimates for Gaussian measures. *The Annals of Probability*, 27(3):1556–1578, 1999. MR1733160

[25] K. Marton. Bounding $\bar{d}$-distance by informational divergence: a method to prove measure concentration. *The Annals of Probability*, 24(2):857–866, 1996. MR1404531

[26] P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, 18(3):1269–1283, 1990. MR1062069

[27] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. MR2319879

[28] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810–864, 2009. MR2484937

[29] M. M. Rao and Z. D. Ren. *Theory of Orlicz spaces*, volume 146 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York, 1991. MR1113700

[30] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999. MR1725357

[31] M. Talagrand. Matching random samples in many dimensions. *The Annals of Applied Probability*, 2(4):846–856, 1992. MR1189420

[32] A.W. Van der Vaart and J.A. Wellner. *Weak convergence and empirical processes*. Springer Verlag, 1996. MR1385671

[33] V.S. Varadarajan. On the convergence of sample probability distributions. *Sankhyā: The Indian Journal of Statistics*, 19(1):23–26, 1958. MR0094839

[34] C. Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. MR2459454

[35] R. Wang, X. Wang, and L. Wu. Sanov's theorem in the Wasserstein distance: A necessary and sufficient condition. *Statistics & Probability Letters*, 80(5-6):505–512, 2010. MR2593592