ON THE RATE OF CONVERGENCE IN WASSERSTEIN DISTANCE OF THE EMPIRICAL MEASURE

NICOLAS FOURNIER AND ARNAUD GUILLIN

Abstract. Let μ_N be the empirical measure associated to a N-sample of a given probability distribution μ on $\mathbb{R}^d$. We are interested in the rate of convergence of μ_N to μ, when measured in the Wasserstein distance of order $p > 0$. We provide some satisfying non-asymptotic L^p-bounds and concentration inequalities, for any values of $p > 0$ and $d \geq 1$. We extend also the non asymptotic L^p-bounds to stationary ρ-mixing sequences, Markov chains, and to some interacting particle systems.

Mathematics Subject Classification (2010): 60F25, 60F10, 65C05, 60E15, 65D32.

Keywords: Empirical measure, Sequence of i.i.d. random variables, Wasserstein distance, Concentration inequalities, Quantization, Markov chains, ρ-mixing sequences, Mc Kean-Vlasov particles system.

1. Introduction and results

1.1. **Notation.** Let $d \geq 1$ and $\mathcal{P}(\mathbb{R}^d)$ stand for the set of all probability measures on $\mathbb{R}^d$. For $\mu \in \mathcal{P}(\mathbb{R}^d)$, we consider an i.i.d. sequence $(X_k)_{k\geq 1}$ of μ-distributed random variables and, for $N \geq 1$, the empirical measure

$$\mu_N := \frac{1}{N}\sum_{k=1}^N \delta_{X_k}.$$

As is well-known, by Glivenko-Cantelli's theorem, μ_N tends weakly to μ as $N \to \infty$ (for example in probability, see Van der Vaart-Wellner [40] for details and various modes of convergence). The aim of the paper is to quantify this convergence, when the error is measured in some Wasserstein distance. Let us set, for $p \geq 1$ and μ, ν in $\mathcal{P}(\mathbb{R}^d)$,

$$\mathcal{T}_p(\mu,\nu) = \inf\left\{ \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \xi(dx,dy)\right) \ : \ \xi \in \mathcal{H}(\mu,\nu)\right\},$$

where $\mathcal{H}(\mu,\nu)$ is the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν. See Villani [41] for a detailed study of $\mathcal{T}_p$. The Wasserstein distance $\mathcal{W}_p$ on $\mathcal{P}(\mathbb{R}^d)$ is defined by $\mathcal{W}_p(\mu,\nu) = \mathcal{T}_p(\mu,\nu)$ if $p \in (0,1]$ and $\mathcal{W}_p(\mu,\nu) = (\mathcal{T}_p(\mu,\nu))^{1/p}$ if $p > 1$.

The present paper studies the rate of convergence to zero of $\mathcal{T}_p(\mu_N,\mu)$. This can be done in an asymptotic way, finding e.g. a sequence $\alpha(N) \to 0$ such that $\limsup_N \alpha(N)^{-1}\mathcal{T}_p(\mu_N,\mu) < \infty$ a.s. or $\limsup_N \alpha(N)^{-1}\mathbb{E}(\mathcal{T}_p(\mu_N,\mu)) < \infty$. Here we will rather derive some non-asymptotic moment estimates such as

$$\mathbb{E}(\mathcal{T}_p(\mu_N,\mu)) \leq \alpha(N) \quad \text{for all } N \geq 1$$

as well as some non-asymptotic concentration estimates (also often called deviation inequalities)

$$\Pr(\mathcal{T}_p(\mu_N,\mu) \geq x) \leq \alpha(N,x) \quad \text{for all } N \geq 1, \text{ all } x > 0.$$

They are naturally related to moment (or exponential moment) conditions on the law μ and we hope to derive an interesting interplay between the dimension $d \geq 1$, the cost parameter $p > 0$ and these moment conditions. Let us introduce precisely these moment conditions. For $q > 0$, $\alpha > 0$, $\gamma > 0$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$, we define

$$M_q(\mu) := \int_{\mathbb{R}^d} |x|^q \mu(dx) \quad \text{and} \quad \mathcal{E}_{\alpha,\gamma}(\mu) := \int_{\mathbb{R}^d} e^{\gamma |x|^\alpha} \mu(dx).$$

We now present our main estimates, the comparison with the existing results and methods will be developped after this presentation. Let us however mention at once that our paper relies on some recent ideas of Dereich-Scheutzow-Schottstedt [16].

1.2. **Moment estimates.** We first give some L^p bounds.

Theorem 1. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and let $p > 0$. Assume that $M_q(\mu) < \infty$ for some $q > p$. There exists a constant C depending only on p, d, q such that, for all $N \geq 1$,*

$$\mathbb{E}\left(\mathcal{T}_p(\mu_N, \mu)\right) \leq C M_q^{p/q}(\mu) \begin{cases} N^{-1/2} + N^{-(q-p)/q} & \text{if } p > d/2 \text{ and } q \neq 2p, \\ N^{-1/2} \log(1+N) + N^{-(q-p)/q} & \text{if } p = d/2 \text{ and } q \neq 2p, \\ N^{-p/d} + N^{-(q-p)/q} & \text{if } p \in (0, d/2) \text{ and } q \neq d/(d-p). \end{cases}$$

Observe that when μ has sufficiently many moments (namely if $q > 2p$ when $p \geq d/2$ and $q > dp/(d-p)$ when $p \in (0, d/2)$), the term $N^{-(q-p)/q}$ is small and can be removed. We could easily treat, for example, the case $p > d/2$ and $q = 2p$ but this would lead to some logarithmic terms and the paper is technical enough.

This generalizes [16], in which only the case $p \in [1, d/2)$ (whence $d \geq 3$) and $q > dp/(d-p)$ was treated. The argument is also slightly simplified.

To show that Theorem 1 is really sharp, let us give examples where lower bounds can be derived quite precisely.

(a) If $a \neq b \in \mathbb{R}^d$ and $\mu = (\delta_a + \delta_b)/2$, one easily checks (see e.g. [16, Remark 1]) that $\mathbb{E}(\mathcal{T}_p(\mu_N, \mu)) \geq c N^{-1/2}$ for all $p \geq 1$. Indeed, we have $\mu_N = Z_N \delta_a + (1 - Z_N)\delta_b$ with $Z_N = N^{-1} \sum_1^N \mathbf{1}_{\{X_i = a\}}$, so that $\mathcal{T}_p(\mu_N, \mu) = |a-b|^p |Z_N - 1/2|$, of which the expectation is of order $N^{-1/2}$.

(b) Such a lower bound in $N^{-1/2}$ can easily be extended to any μ (possibly very smooth) of which the support is of the form $A \cup B$ with $d(A, B) > 0$ (simply note that $\mathcal{T}_p(\mu_N, \mu) \geq d^p(A, B) |Z_N - \mu(A)|$, where $Z_N = N^{-1} \sum_1^N \mathbf{1}_{\{X_i \in A\}}$).

(c) If μ is the uniform distribution on $[-1, 1]^d$, it is well-known and not difficult to prove that for $p > 0$, $\mathbb{E}(\mathcal{T}_p(\mu_N, \mu)) \geq c N^{-p/d}$. Indeed, consider a partition of $[-1, 1]^d$ into (roughly) N cubes with length $N^{-1/d}$. A quick comptation shows that with probability greater than some $c > 0$ (uniformly in N), half of these cubes will not be charged by μ_N. But on this event, we clearly have $\mathcal{T}_p(\mu_N, \mu) \geq a N^{-1/d}$ for some $a > 0$, because each time a cube is not charged by μ_N, a (fixed) proportion of the mass of μ (in this cube) is at distance at least $N^{-1/d}/2$ of the support of μ_N. One easily concludes.

(d) When $p = d/2 = 1$, it has been shown by Ajtai-Komlós-Tusnády [2] that for μ the uniform measure on $[-1, 1]^d$, $\mathcal{T}_1(\mu_N, \mu) \simeq c(\log N/N)^{1/2}$ with high probability, implying that $\mathbb{E}(\mathcal{T}_1(\mu_N, \mu)) \geq c(\log N/N)^{1/2}$.

(e) Let $\mu(dx) = c|x|^{-q-d}\mathbf{1}_{\{|x|\geq 1\}}dx$ for some $q > 0$. Then $M_r(\mu) < \infty$ for all $r \in (0, q)$ and for all $p \geq 1$, $\mathbb{E}(\mathcal{T}_p(\mu_N, \mu)) \geq cN^{-(q-p)/q}$. Indeed, $\mathbb{P}(\mu_N(\{|x| \geq N^{1/q}\}) = 0) = (\mu(\{|x| < N^{1/q}\}))^N = (1 - c/N)^N \geq c > 0$ and $\mu(\{|x| \geq 2N^{1/q}\}) \geq c/N$. One easily gets convinced that $\mathcal{T}_p(\mu_N, \mu) \geq N^{p/q}\mathbf{1}_{\{\mu_N(\{|x|\geq N^{1/q}\})=0\}}\mu(\{|x| \geq 2N^{1/q}\})$, from which the claim follows.

As far as *general laws* are concerned, Theorem 1 is really sharp: the only possible improvements are the following. The first one, quite interesting, would be to replace $\log(1 + N)$ by something like $\sqrt{\log(1 + N)}$ when $p = d/2$ (see point (d) above). It is however not clear it is feasible in full generality. The second one, which should be a mere (and not very interesting) refinement, would be to sharpen the bound in $N^{-(q-p)/q}$ when $M_q(\mu) < \infty$: point (e) only shows that there is μ with $M_q(\mu) < \infty$ for which we have a lowerbound in $N^{-(q-p)/q-\varepsilon}$ for all $\varepsilon > 0$.

However, some improvements are possible when restricting the class of laws μ. First, when μ is the uniform distribution in $[-1, 1]^d$, the results of Talagrand [38, 39] strongly suggest that when $d \geq 3$, $\mathbb{E}(\mathcal{T}_p(\mu_N, \mu)) \simeq N^{-p/d}$ for all $p > 0$, and this is much better than $N^{-1/2}$ when p is large. Such a result would of course immediately extend to any distribution $\mu = \lambda \circ F^{-1}$, for λ the uniform distribution in $[-1, 1]^d$ and $F : [-1, 1]^d \mapsto \mathbb{R}^d$ Lipschitz continuous. In any case, a smoothness assumption for μ cannot be sufficient, see point (b) above.

Second, for irregular laws, the convergence can be much faster that $N^{-p/d}$ when $p < d/2$, see point (a) above where, in an extreme case, we get $N^{-1/2}$ for all values of $p > 0$. It is shown by Dereich-Scheutzow-Schottstedt [16] (see also Barthe-Bordenave [3]) that indeed, for a singular law, $\lim_N N^{-p/d}\mathbb{E}(\mathcal{T}_p(\mu_N, \mu)) = 0$.

1.3. **Concentration inequalities.** We next state some concentration inequalities.

Theorem 2. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and let $p > 0$. Assume one of the three following conditions:*

$$\exists\, \alpha > p,\ \exists\, \gamma > 0,\ \mathcal{E}_{\alpha,\gamma}(\mu) < \infty, \tag{1}$$

$$\text{or}\quad \exists\, \alpha \in (0, p),\ \exists\, \gamma > 0,\ \mathcal{E}_{\alpha,\gamma}(\mu) < \infty, \tag{2}$$

$$\text{or}\quad \exists\, q > 2p,\ M_q(\mu) < \infty. \tag{3}$$

Then for all $N \geq 1$, all $x \in (0, \infty)$,

$$\mathbb{P}(\mathcal{T}_p(\mu^N, \mu) \geq x) \leq a(N, x)\mathbf{1}_{\{x\leq 1\}} + b(N, x),$$

where

$$a(N, x) = C\begin{cases} \exp(-cNx^2) & \text{if } p > d/2, \\ \exp(-cN(x/\log(2 + 1/x))^2) & \text{if } p = d/2, \\ \exp(-cNx^{d/p}) & \text{if } p \in [1, d/2) \end{cases}$$

and

$$b(N, x) = C\begin{cases} \exp(-cNx^{\alpha/p})\mathbf{1}_{\{x>1\}} & \text{under (1)}, \\ \exp(-c(Nx)^{(\alpha-\varepsilon)/p})\mathbf{1}_{\{x\leq 1\}} + \exp(-c(Nx)^{\alpha/p})\mathbf{1}_{\{x>1\}} & \forall\, \varepsilon \in (0, \alpha) \text{ under (2)}, \\ N(Nx)^{-(q-\varepsilon)/p} & \forall\, \varepsilon \in (0, q) \text{ under (3)}. \end{cases}$$

The positive constants C and c depend only on p, d and either on $\alpha, \gamma, \mathcal{E}_{\alpha,\gamma}(\mu)$ (under (1)) or on $\alpha, \gamma, \mathcal{E}_{\alpha,\gamma}(\mu), \varepsilon$ (under (2)) or on $q, M_q(\mu), \varepsilon$ (under (3)).

We could also treat the *critical* case where $\mathcal{E}_{\alpha,\gamma}(\mu) < \infty$ with $\alpha = p$, but the result we could obtain is slightly more intricate and not very satisfying for small value of x (even if good for large ones).

Remark 3. *When assuming* (2) *with* $\alpha \in (0,p)$*, we actually also prove that*

$$b(N,x) \le C\exp(-cNx^2(\log(1+N))^{-\delta}) + C\exp(-c(Nx)^{\alpha/p}),$$

with $\delta = 2p/\alpha - 1$*, see Step 5 of the proof of Lemma 13 below. This allows us to extend the inequality* $b(N,x) \le C\exp(-c(Nx)^{\alpha/p})$ *to all values of* $x \ge x_N$*, for some (rather small)* x_N *depending on* N,α,p*. But for very small values of* $x > 0$*, this formula is less interesting than that of Theorem 2. Despite much effort, we have not been able to get rid of the logarithmic term.*

We believe that these estimates are quite satisfying. To get convinced, first observe that the *scales* seem to be the good ones. Recall that $\mathbb{E}(\mathcal{T}_p(\mu_N,\mu)) = \int_0^\infty \mathbb{P}(\mathcal{T}_p(\mu^N,\mu) \ge x)dx$.

(a) One easily checks that $\int_0^\infty a(N,x)dx \le CN^{-p/d}$ if $p < d/2$, $CN^{-1/2}\log(1+N)$ if $p = d/2$, and $CN^{-1/2}$ if $p > d/2$, as in Theorem 1.

(b) When integrating $b(N,x)$ (or rather $b(N,x)\wedge 1$), we find $N^{-(q-\varepsilon-p)/(q-\varepsilon)}$ under (3) and something smaller under (1) or (2). Since we can take $q-\varepsilon > 2p$, this is less than $N^{-1/2}$ (and thus also less than $N^{-p/d}$ if $p < d/2$ and than $N^{-1/2}\log(1+N)$ if $p = d/2$).

The *rates of decrease* are also satisfying in most cases. Recall that in deviation estimates, we never get something better than $\exp(-Ng(x))$ for some function g. Hence $a(N,x)$ is probably optimal. Next, for $\bar{Y}_N$ the empirical mean of a family of centered i.i.d. random variables, it is well-known that the *good* deviation inequalities are the following.

(a) If $\mathbb{E}[\exp(a|Y_1|^\beta)] < \infty$ with $\beta \ge 1$, then $\Pr[|\bar{Y}_N| \ge x] \le Ce^{-cNx^2}\mathbf{1}_{\{x\le 1\}} + Ce^{-cNx^\beta}\mathbf{1}_{\{x>1\}}$, see for example Djellout-Guillin-Wu [18], Gozlan [24] or Ledoux [27], using transportation cost inequalities.

(b) If $\mathbb{E}[\exp(a|Y_1|^\beta)] < \infty$ with $\beta < 1$, then $\Pr[|\bar{Y}_N| \ge x] \le Ce^{-cNx^2} + Ce^{-c(Nx)^\beta}$, see Merlevède-Peligrad-Rio [31, Formula (1.4)] which is based on results by Borovkov [8].

(c) If $\mathbb{E}[|Y_1|^r] < \infty$ for some $r > 2$, then $\Pr[|\bar{Y}_N| \ge x] \le Ce^{-cNx^2} + CN(Nx)^{-r}$, see Fuk-Nagaev [23], using usual truncation arguments.

Our result is in perfect adequation with these facts (up to some arbitratry small loss due to ε under (2) and (3)) since $\mathcal{T}_p(\mu_N,\mu)$ should behave very roughly as the mean of the $|X_i|^p$'s, which e.g. has an exponential moment with power $\beta := \alpha/p$ under (1) and (2).

1.4. **Comments.** The control of the distance between the empirical measure of an i.i.d. sample and its true distribution is of course a long standing problem central both in probability, statistics and informatics with a wide number of applications: quantization (see Delattre-Graf-Luschgy-Pagès [14] and Pagès-Wilbertz [33] for recent results), optimal matching (see Ajtai-Komlós-Tusnády [2], Dobrić-Yukich [19], Talagrand [39], Barthe-Bordenave [3]), density estimation, clustering (see Biau-Devroye-Lugosi [5] and Laloë [26]), MCMC methods (see [36] for bounds on ergodic averages), particle systems and approximations of partial differential equations (see Bolley-Guillin-Villani [11] and Fournier-Mischler [22]). We refer to these papers for an extensive introduction on this vast topic.

If many distances can be used to consider the problem, the Wasserstein distance is quite natural, in particular in quantization or for particle approximations of P.D.E.'s. However the depth of the problem was discovered only recently by Ajtai-Komlós-Tusnády [2], who considered the uniform measure on the square, investigated thoroughly by Talagrand [39]. As a review of the litterature is somewhat impossible, et us just say that the methods involved were focused on two methods inherited by the definitions of the Wasserstein distance: the construction of a coupling or by duality to control a particular empirical process.

Concerning moment estimates (as in Theorem 1), some results can be found in Horowitz-Karandikar [25], Rachev-Rüschendorf [35] and Mischler-Mouhot [32]. But theses results are far from optimal, even when assuming that μ is compactly supported. Very recently, strickingly clever alternatives were considered by Boissard-Le Gouic [7] and by Dereich-Scheutzow-Schottstedt [16]. Unfortunately, the construction of Boissard-Le Gouic, based on iterative trees, was a little too complicated to yield sharp rates. On the contrary, the method of [16], exposed in details in the next section, is extremely simple, robust, and leads to the almost optimal results exposed here. Some sharp moment estimates were already obtained in [16] for a limited range of parameters.

Concerning concentration estimates, only few results are available. Let us mention the work of Bolley-Guillin-Villani [11] and very recently by Boissard [6], on which we considerably improve. Our assumptions are often much weaker (the reference measure μ was often assumed to satisfy some functional inequalities, which may be difficult to verify and usually include more "structure" than mere integrability conditions) and $\Pr[\mathcal{T}_p(\mu_N,\mu) \geq x]$ was estimated only for rather large values of x. In particular, when integrating the concentration estimates of [11], one does never find the *good* moment estimates, meaning that the scales are not the good ones.

Moreover, the approach of [16] is robust enough so that we can also give some good moment bounds for the Wasserstein distance between the empirical measure of a Markov chain and its invariant distribution (under some conditions). This could be useful for MCMC methods because our results are non asymptotic. We can also study very easily some ρ-mixing sequences (see Doukhan [20]), for which only very few results exist, see Biau-Devroye-Lugosi [7]. Finally, we show on an example how to use Theorem 1 to study some particle systems. For all these problems, we might also obtain some concentration inequalities, but this would need further refinements which are out of the scope of the present paper, somewhat already technical enough, and left for further works.

1.5. **Plan of the paper.** In the next section, we state some general upper bounds of $\mathcal{T}_p(\mu,\nu)$, for any $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$, essentially taken from [16]. Section 3 is devoted to the proof of Theorem 1. Theorem 2 is proved in three steps: in Section 4 we study the case where μ is compactly supported and where N is replaced by a Poisson(N)-distributed random variable, which yields some pleasant independance properties. We show how to remove the randomization in Section 5, concluding the case where μ is compactly supported. The non compact case is studied in Section 6. The final Section 7 is devoted to dependent random variables: ρ-mixing sequences, Markov chains and a particular particle system.

2. Coupling

The following notion of distance, essentially taken from [16], is the main ingredient of the paper.

Notation 4. *(a) For $\ell \geq 0$, we denote by $\mathcal{P}_\ell$ the natural partition of $(-1,1]^d$ into $2^{d\ell}$ translations of $(-2^{-\ell},2^{-\ell}]^d$. For two probability measures μ,ν on $(-1,1]^d$ and for $p>0$, we introduce*

$$\mathcal{D}_p(\mu,\nu) := \frac{2^p-1}{2} \sum_{\ell \geq 1} 2^{-p\ell} \sum_{F \in \mathcal{P}_\ell} |\mu(F)-\nu(F)|,$$

which obviously defines a distance on $\mathcal{P}((-1,1]^d)$, always bounded by 1.

(b) We introduce $B_0 := (-1,1]^d$ and, for $n \geq 1$, $B_n := (-2^n,2^n]^d \setminus (-2^{n-1},2^{n-1}]^d$. For $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $n \geq 0$, we denote by $\mathcal{R}_{B_n}\mu$ the probability measure on $(-1,1]^d$ defined as the image of $\mu|_{B_n}/\mu(B_n)$ by the map $x \mapsto x/2^n$. For two probability measures μ,ν on $\mathbb{R}^d$ and for $p>0$, we

introduce

$$\mathcal{D}_p(\mu,\nu) := \sum_{n\geq 0} 2^{pn}\big(|\mu(B_n)-\nu(B_n)| + (\mu(B_n)\wedge\nu(B_n))\mathcal{D}_p(\mathcal{R}_{B_n}\mu,\mathcal{R}_{B_n}\nu)\big).$$

A little study, using that $\mathcal{D}_p \leq 1$ on $\mathcal{P}((-1,1]^d)$, shows that this defines a distance on $\mathcal{P}(\mathbb{R}^d)$.

Having a look at $\mathcal{D}_p$ in the compact case, one sees that in some sense, it measures distance of the two probability measures simultaneously at all the scales. The optimization procedure can be made for all scales and outperforms the approach based on a fixed diameter covering of the state space (which is more or less the approach of Horowitz-Karandikar [25]). Moreover one sees that the principal control is on $|\pi(F)-\mu(F)|$ which is a quite simple quantity. The next results are slightly modified versions of estimates found in [16], see [16, Lemma 2] for the compact case and [16, proof of Theorem 3] for the non compact case. It contains the crucial remark that $\mathcal{D}_p$ is an upper bound (up to constant) of the Wasserstein distance.

Lemma 5. *Let $d \geq 1$ and $p > 0$. For all pairs of probability measures μ,ν on $\mathbb{R}^d$, $\mathcal{T}_p(\mu,\nu) \leq \kappa_{p,d}\mathcal{D}_p(\mu,\nu)$, with $\kappa_{p,d} := 2^{p(1+d/2)}(2^p+1)/(2^p-1)$.*

Proof. We separate the proof into two steps.

Step 1. We first assume that μ and ν are supported in $(-1,1]^d$. We infer from [16, Lemma 2], in which the conditions $p \geq 1$ and $d \geq 3$ are clearly not used, that, since the diameter of $(-1,1]^d$ is $2^{1+d/2}$,

$$\mathcal{T}_p(\mu,\nu) \leq 2^{p(1+d/2)-1}\sum_{\ell\geq 0} 2^{-p\ell}\sum_{F\in\mathcal{P}_\ell}\mu(F)\sum_{C\ child\ of\ F}\left|\frac{\mu(C)}{\mu(F)}-\frac{\nu(C)}{\nu(F)}\right|,$$

where "C child of F" means that $C \in \mathcal{P}_{\ell+1}$ and $C \subset F$. Consequently,

$$\begin{aligned}
\mathcal{T}_p(\mu,\nu) \leq& 2^{p(1+d/2)-1}\sum_{\ell\geq 0} 2^{-p\ell}\sum_{F\in\mathcal{P}_\ell}\sum_{C\ child\ of\ F}\left(\frac{\nu(C)}{\nu(F)}|\mu(F)-\nu(F)| + |\mu(C)-\nu(C)|\right)\\
\leq& 2^{p(1+d/2)-1}\sum_{\ell\geq 0} 2^{-p\ell}\left(\sum_{F\in\mathcal{P}_\ell}|\mu(F)-\nu(F)| + \sum_{C\in\mathcal{P}_{\ell+1}}|\mu(C)-\nu(C)|\right)\\
\leq& 2^{p(1+d/2)-1}(1+2^p)\sum_{\ell\geq 1} 2^{-p\ell}\sum_{F\in\mathcal{P}_\ell}|\mu(F)-\nu(F)|,
\end{aligned}$$

which is nothing but $\kappa_{p,d}\mathcal{D}_p(\mu,\nu)$. We used that $\sum_{F\in\mathcal{P}_0}|\mu(F)-\nu(F)| = 0$.

In [16]*, Dereich-Scheutzow-Schottstedt use directly the formula with the children to study the rate of convergence of empirical measures. This leads to some (small) technical complications, and does not seem to improve the estimates.*

Step 2. We next consider the general case. We consider, for each $n \geq 1$, the optimal coupling $\pi_n(dx,dy)$ between $\mathcal{R}_{B_n}\mu$ and $\mathcal{R}_{B_n}\nu$ for $\mathcal{T}_p$. We define $\xi_n(dx,dy)$ as the image of π_n by the map $(x,y) \mapsto (2^n x, 2^n y)$, which clearly belongs to $\mathcal{H}(\mu|_{B_n}/\mu(B_n), \nu|_{B_n}/\nu(B_n))$ and satisfies $\iint |x-y|^p\xi_n(dx,dy) = 2^{np}\iint|x-y|^p\pi_n(dx,dy) = 2^{np}\mathcal{T}_p(\mathcal{R}_{B_n}\mu,\mathcal{R}_{B_n}\nu)$.

Next, we introduce $q := \frac{1}{2}\sum_{n\geq 0}|\nu(B_n)-\mu(B_n)|$ and we define

$$\xi(dx,dy) = \sum_{n\geq 0}(\mu(B_n)\wedge\nu(B_n))\xi_n(dx,dy) + \frac{\alpha(dx)\beta(dy)}{q},$$

where

$$\alpha(dx) := \sum_{n\geq 0}(\mu(B_n)-\nu(B_n))_+ \frac{\mu|_{B_n}(dx)}{\mu(B_n)} \text{ and } \beta(dy) := \sum_{n\geq 0}(\nu(B_n)-\mu(B_n))_+ \frac{\nu|_{B_n}(dy)}{\nu(B_n)}.$$

Using that

$$q = \sum_{n\geq 0}(\nu(B_n)-\mu(B_n))_+ = \sum_{n\geq 0}(\mu(B_n)-\nu(B_n))_+ = 1-\sum_{n\geq 0}(\nu(B_n)\wedge\mu(B_n)),$$

it is easily checked that $\xi \in \mathcal{H}(\mu,\nu)$. Furthermore, we have, setting $c_p = 1$ if $p \in (0,1]$ and $c_p = 2^{p-1}$ if $p > 1$,

$$\begin{aligned}\iint |x-y|^p \frac{\alpha(dx)\beta(dy)}{q} \leq& \frac{1}{q}\iint c_p(|x|^p+|y|^p)\alpha(dx)\beta(dy)\\ =& c_p\int |x|^p\alpha(dx) + c_p\int |y|^p\beta(dy)\\ \leq& c_p\sum_{n\geq 0} 2^{pn}[(\mu(B_n)-\nu(B_n))_+ + (\nu(B_n)-\mu(B_n))_+]\\ =& c_p\sum_{n\geq 0} 2^{pn}|\mu(B_n)-\nu(B_n)|.\end{aligned}$$

Recalling that $\iint |x-y|^p\xi_n(dx,dy) \leq 2^{np}\mathcal{T}_p(\mathcal{R}_{B_n}\mu,\mathcal{R}_{B_n}\nu)$, we deduce that

$$\begin{aligned}\mathcal{T}_p(\mu,\nu) \leq& \iint |x-y|^p\xi(dx,dy)\\ \leq& \sum_{n\geq 0} 2^{np}\left(c_p|\mu(B_n)-\nu(B_n)| + (\mu(B_n)\wedge\nu(B_n))\mathcal{T}_p(\mathcal{R}_{B_n}\mu,\mathcal{R}_{B_n}\nu)\right).\end{aligned}$$

We conclude using Step 1 and that $c_p \leq \kappa_{p,d}$. □

When proving the concentration inequalities, which is very technical, it will be good to break the proof into several steps to separate the difficulties and we will first treat the compact case. On the contrary, when dealing with moment estimates, the following formula will be easier to work with.

Lemma 6. *Let $p > 0$ and $d \geq 1$. There is a constant C, depending only on p, d, such that for all $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$,*

$$\mathcal{D}_p(\mu,\nu) \leq C\sum_{n\geq 0} 2^{pn}\sum_{\ell\geq 0} 2^{-p\ell}\sum_{F\in\mathcal{P}_\ell} |\mu(2^nF\cap B_n)-\nu(2^nF\cap B_n)|$$

with the notation $2^nF = \{2^nx \,:\, x\in F\}$.

Proof. For all $n \geq 1$, we have $|\mu(B_n)-\nu(B_n)| = \sum_{F\in\mathcal{P}_0}|\mu(2^nF\cap B_n)-\nu(2^nF\cap B_n)|$ and $(\mu(B_n)\wedge\nu(B_n))\mathcal{D}_p(\mathcal{R}_{B_n}\mu,\mathcal{R}_{B_n}\nu)$ is smaller than

$$\begin{aligned}&(\mu(B_n)\wedge\nu(B_n))\mathcal{D}_p(\mathcal{R}_{B_n}\mu,\mathcal{R}_{B_n}\nu)\\ \leq& \mu(B_n)\sum_{\ell\geq 1}2^{-p\ell}\sum_{F\in\mathcal{P}_\ell}\left|\frac{\mu(2^nF\cap B_n)}{\mu(B_n)} - \frac{\nu(2^nF\cap B_n)}{\nu(B_n)}\right|\\ \leq& \sum_{\ell\geq 1}2^{-p\ell}\sum_{F\in\mathcal{P}_\ell}|\mu(2^nF\cap B_n)-\nu(2^nF\cap B_n)| + \left|1-\frac{\mu(B_n)}{\nu(B_n)}\right|\sum_{\ell\geq 1}2^{-p\ell}\sum_{F\in\mathcal{P}_\ell}\nu(2^nF\cap B_n).\end{aligned}$$

This last term is smaller than $2^{-p}|\mu(B_n)-\nu(B_n)|/(1-2^{-p})$ and this ends the proof. □

3. Moment estimates

The aim of this section is to give the

Proof of Theorem 1. We thus assume that $\mu \in \mathcal{P}(\mathbb{R}^d)$ and that $M_q(\mu) < \infty$ for some $q > p$. By a scaling argument, we may assume that $M_q(\mu) = 1$. This implies that $\mu(B_n) \le 2^{-q(n-1)}$ for all $n \ge 0$. By Lemma 5, we have $\mathcal{T}_p(\mu_N, \mu) \le \kappa_{p,d} \mathcal{D}_p(\mu_N, \mu)$, so that it suffices to study $\mathbb{E}(\mathcal{D}_p(\mu_N, \mu))$.

For a Borel subset $A \subset \mathbb{R}^d$, since $N\mu_N(A)$ is Binomial$(N, \mu(A))$-distributed, we have

$$\mathbb{E}(|\mu_N(A) - \mu(A)|) \le \min\left\{2\mu(A), \sqrt{\mu(A)/N}\right\}.$$

Using the Cauchy-Scharz inequality and that $\#(\mathcal{P}_\ell) = 2^{d\ell}$, we deduce that for all $n \ge 0$, all $\ell \ge 0$,

$$\sum_{F \in \mathcal{P}_\ell} \mathbb{E}(|\mu_N(2^n F \cap B_n) - \mu(2^n F \cap B_n)|) \le \min\left\{2\mu(B_n), 2^{d\ell/2}(\mu(B_n)/N)^{1/2}\right\}.$$

Using finally Lemma 6 and that $\mu(B_n) \le 2^{-q(n-1)}$, we find

$$\mathbb{E}(\mathcal{D}_p(\mu_N, \mu)) \le C \sum_{n \ge 0} 2^{pn} \sum_{\ell \ge 0} 2^{-p\ell} \min\left\{2^{-qn}, 2^{d\ell/2}(2^{-qn}/N)^{1/2}\right\}. \tag{4}$$

Step 1. Here we show that for all $\varepsilon \in (0,1)$, all $N \ge 1$,

$$\sum_{\ell \ge 0} 2^{-p\ell} \min\left\{\varepsilon, 2^{d\ell/2}(\varepsilon/N)^{1/2}\right\} \le C \begin{cases} \min\{\varepsilon, (\varepsilon/N)^{1/2}\} & \text{if} \quad p > d/2, \\ \min\{\varepsilon, (\varepsilon/N)^{1/2}\log(2+\varepsilon N)\} & \text{if} \quad p = d/2, \\ \min\{\varepsilon, \varepsilon(\varepsilon N)^{-p/d}\} & \text{if} \quad p \in (0, d/2). \end{cases}$$

First of all, the bound by $C\varepsilon$ is obvious in all cases (because $p > 0$). Next, the case $p > d/2$ is immediate. If $p \le d/2$, we introduce $\ell_{N,\varepsilon} := \lfloor \log(2+\varepsilon N)/(d \log 2) \rfloor$, for which $2^{d\ell_{N,\varepsilon}} \simeq 2 + \varepsilon N$ and get an upper bound in

$$(\varepsilon/N)^{1/2} \sum_{\ell \le \ell_{N,\varepsilon}} 2^{(d/2-p)\ell} + \varepsilon \sum_{\ell \ge \ell_{N,\varepsilon}} 2^{-p\ell}.$$

If $p = d/2$, we find an upper bound in

$$(\varepsilon/N)^{1/2} \ell_{N,\varepsilon} + C\varepsilon 2^{-p\ell_{N,\varepsilon}} \le C(\varepsilon/N)^{1/2} \log(2+\varepsilon N) + C\varepsilon(1+\varepsilon N)^{-1/2} \le C(\varepsilon/N)^{1/2}\log(2+\varepsilon N)$$

as desired. If $p \in (0, d/2)$, we get an upper bound in

$$C(\varepsilon/N)^{1/2} 2^{(d/2-p)\ell_{N,\varepsilon}} + C\varepsilon 2^{-p\ell_{N,\varepsilon}} \le C(\varepsilon/N)^{1/2}(2+\varepsilon N)^{1/2-p/d} + C\varepsilon(2+\varepsilon N)^{-p/d}.$$

If $\varepsilon N \ge 1$, then $(2+\varepsilon N)^{1/2-p/d} \le (3\varepsilon N)^{1/2-p/d}$ and the conclusion follows. If now $\varepsilon N \in (0,1)$, the result is obvious because $\min\{\varepsilon, \varepsilon(\varepsilon N)^{-p/d}\} = \varepsilon$.

Step 2: $p > d/2$. By (4) and Step 1 (with $\varepsilon = 2^{-qn}$), we find

$$\mathbb{E}(\mathcal{D}_p(\mu_N, \mu)) \le C \sum_{n \ge 0} 2^{pn} \min\left\{2^{-qn}, (2^{-qn}/N)^{1/2}\right\} \le C \begin{cases} N^{-1/2} & \text{if} \quad q > 2p, \\ N^{-(q-p)/q} & \text{if} \quad q \in (p, 2p). \end{cases}$$

Indeed, this is obvious if $q > 2p$, while the case $q \in (p, 2p)$ requires to separate the sum in two parts $n \le n_N$ and $n > n_N$ with $n_N = \lfloor \log N/(q \log 2) \rfloor$. This ends the proof when $p > d/2$.

Step 3: $p = d/2$. By (4) and Step 1 (with $\varepsilon = 2^{-qn}$), we find

$$\mathbb{E}(\mathcal{D}_p(\mu_N, \mu)) \le C \sum_{n \ge 0} 2^{pn} \min\left\{2^{-qn}, (2^{-qn}/N)^{1/2} \log(2 + 2^{-qn} N)\right\}.$$

If $q > 2p$, we immediately get a bound in

$$\mathbb{E}(\mathcal{D}_p(\mu_N,\mu)) \le C\sum_{n\ge 0} 2^{(p-q/2)n} N^{-1/2}\log(2+N) \le C\log(2+N)N^{-1/2},$$

which ends the proof (when $p = d/2$ and $q > 2p$).

If $q \in (p, 2p)$, we easily obtain, using that $\log(2+x) \le 2\log x$ for all $x \ge 2$, an upper bound in

$$\begin{aligned}\mathbb{E}(\mathcal{D}_p(\mu_N,\mu)) \le& C\sum_{n\ge 0} \mathbf{1}_{\{N<2.2^{nq}\}} 2^{(p-q)n} + C\sum_{n\ge 0} \mathbf{1}_{\{N\ge 2.2^{nq}\}} 2^{(p-q/2)n} N^{-1/2}\log(N2^{-nq})\\ \le& CN^{-(q-p)/q} + CN^{-1/2}\sum_{n=0}^{n_N} 2^{(p-q/2)n}(\log N - nq\log 2)\\ =:& CN^{-(q-p)/q} + CN^{-1/2}K_N,\end{aligned}$$

where $n_N = \lfloor \log(N/2)/(q\log 2)\rfloor$. A tedious exact computation shows that

$$\begin{aligned}K_N =& \log N \frac{2^{(p-q/2)(n_N+1)}-1}{2^{(p-q/2)}-1}\\ &- q\log 2\left[(n_N+1)\frac{2^{(p-q/2)(n_N+1)}-1}{2^{(p-q/2)}-1} + \frac{n_N+1}{2^{(p-q/2)}-1} - \frac{2^{(p-q/2)(n_N+2)}-2^{(p-q/2)}}{(2^{(p-q/2)}-1)^2}\right].\end{aligned}$$

Using that the contribution of the middle term of the second line is negative and the inequality $\log N - (n_N+1)q\log 2 \le \log 2$ (because $(n_N+1)q\log 2 \ge \log(N/2)$), we find

$$K_N \le C2^{(p-q/2)n_N} \le CN^{p/q-1/2}.$$

We finally have checked that $\mathbb{E}(\mathcal{D}_p(\mu_N,\mu)) \le CN^{-(q-p)/q} + CN^{-1/2}N^{p/q-1/2} \le CN^{-(q-p)/q}$, which ends the proof when $p = d/2$.

Step 4: $p \in (0, d/2)$. We then have, by (4) and Step 1,

$$\mathbb{E}(\mathcal{D}_p(\mu_N,\mu)) \le C\sum_{n\ge 0} 2^{pn}\min\left\{2^{-qn}, 2^{-qn(1-p/d)}N^{-p/d}\right\}.$$

If $q > dp/(d-p)$, which implies that $q(1-p/d) > p$, we immediately get an upper bound by $CN^{-p/d}$, which ends the proof when $p < d/2$ and $q > dp/(d-p)$

If finally $q \in (p, dp/(d-p))$, we separate the sum in two parts $n \le n_N$ and $n > n_N$ with $n_N = \lfloor \log N/(q\log 2)\rfloor$ and we find a bound in $CN^{-(q-p)/q}$ as desired. □

4. Concentration inequalities in the compact poissonized case

It is technically advantageous to first consider the case where the size of the sampling is Poisson distributed, which implies some independence properties. If we replace N (large) by a Poisson(N)-distributed random variable, this should not change much the problem, because a Poisson(N)-distributed random variable is close to N with high probability.

Notation 7. *We introduce the functions f and g defined on $(0,\infty)$ by*

$$f(x) = (1+x)\log(1+x) - x \quad \text{and} \quad g(x) = (x\log x - x + 1)\mathbf{1}_{\{x\ge 1\}}.$$

Observe that f is increasing, nonnegative, equivalent to x^2 at 0 and to $x\log x$ at infinity. The function g is positive and increasing on $(1,\infty)$.

The goal of this section is to check the following.

Proposition 8. *Assume that μ is supported in $(-1,1]^d$. Let Π_N be a Poisson measure on $\mathbb{R}^d$ with intensity measure $N\mu$ and introduce the associated empirical measure $\Psi_N = (\Pi_N(\mathbb{R}^d))^{-1}\Pi_N$. Let $p \geq 1$ and $d \geq 1$. There are some positive constants C, c (depending only on d, p) such that for all $N \geq 1$, all $x \in (0,\infty)$,*

$$\mathbb{P}\left(\Pi_N(\mathbb{R}^d)\mathcal{D}_p(\Psi_N,\mu) \geq Nx\right) \leq C \begin{cases} \exp(-Nf(cx)) & \text{if } p > d/2, \\ \exp\left(-Nf(cx/\log(2+1/x))\right) & \text{if } p = d/2, \\ \exp\left(-Nf(cx)\right) + \exp\left(-cNx^{d/p}\right) & \text{if } p \in [1, d/2). \end{cases}$$

We start with some easy and well-known concentration inequalities for the Poisson distribution.

Lemma 9. *For $\lambda > 0$ and X a Poisson(λ)-distributed random variable, we have*
(a) $E(\exp(\theta X)) = \exp(\lambda(e^\theta - 1))$ for all $\theta \in \mathbb{R}$;
(b) $E(\exp(\theta|X-\lambda|)) \leq 2\exp(\lambda(e^\theta - 1 - \theta))$ for all $\theta > 0$;
(c) $\mathbb{P}(X > \lambda x) \leq \exp(-\lambda g(x))$ for all $x > 0$;
(d) $\mathbb{P}(|X-\lambda| > \lambda x) \leq 2\exp(-\lambda f(x))$ for all $x > 0$;
(e) $\mathbb{P}(X > \lambda x) \leq \lambda$ for all $x > 0$.

Proof. Point (a) is straightforward. For point (b), write $E(\exp(\theta|X-\lambda|)) \leq e^{\theta\lambda}\mathbb{E}(\exp(-\theta X)) + e^{-\theta\lambda}\mathbb{E}(\exp(\theta X))$, use (a) and that $\lambda(e^{-\theta} - 1 + \theta) \leq \lambda(e^\theta - 1 - \theta)$. For point (c), write $\mathbb{P}(X > \lambda x) \leq e^{-\theta\lambda x}\mathbb{E}[\exp(\theta X)]$, use (a) and optimize in θ. Use the same scheme to deduce (d) from (b). Finally, for $x > 0$, $\mathbb{P}(X > \lambda x) \leq \mathbb{P}(X > 0) = 1 - e^{-\lambda} \leq \lambda$. □

We can now give the

Proof of Proposition 8. We fix $x > 0$ for the whole proof. Recalling Notation 4-(a), we have

$$\begin{aligned} \Pi_N(\mathbb{R}^d)\mathcal{D}_p(\Psi_N,\mu) &= C\sum_{\ell\geq 1} 2^{-p\ell} \sum_{F\in\mathcal{P}_\ell} |\Pi_N(F) - \Pi_N(\mathbb{R}^d)\mu(F)| \\ &\leq C|\Pi_N(\mathbb{R}^d) - N| + C\sum_{\ell\geq 1} 2^{-p\ell} \sum_{F\in\mathcal{P}_\ell} |\Pi_N(F) - N\mu(F)| \\ &\leq C|\Pi_N(\mathbb{R}^d) - N| + C(N + \Pi_N(\mathbb{R}^d))2^{-p\ell_0} + C\sum_{\ell=1}^{\ell_0} 2^{-p\ell} \sum_{F\in\mathcal{P}_\ell} |\Pi_N(F) - N\mu(F)| \end{aligned}$$

for any choice of $\ell_0 \in \mathbb{N}$. We will choose ℓ_0 later, depending on the value of x. For any nonnegative family r_ℓ such that $\sum_1^{\ell_0} r_\ell \leq 1$, we thus have

$$\begin{aligned} \varepsilon(N,x) :=& \mathbb{P}\left(\Pi_N(\mathbb{R}^d)\mathcal{D}_p(\Psi_N,\mu) \geq Nx\right) \\ \leq& \mathbb{P}\left(|\Pi_N(\mathbb{R}^d) - N| \geq cNx\right) + \mathbb{P}\left(\Pi_N(\mathbb{R}^d) \geq N(cx2^{p\ell_0} - 1)\right) \\ &+ \sum_{\ell=1}^{\ell_0} \mathbb{P}\left(\sum_{F\in\mathcal{P}_\ell} |\Pi_N(F) - N\mu(F)| \geq cNx2^{p\ell} r_\ell\right). \end{aligned}$$

By Lemma 9-(c)-(d), since $\Pi_N(\mathbb{R}^d)$ is Poisson(N)-distributed, $\mathbb{P}(\Pi_N(\mathbb{R}^d) \geq N(cx2^{p\ell_0} - 1)) \leq \exp(-Ng(cx2^{p\ell_0} - 1))$ and $\mathbb{P}(|\Pi_N(\mathbb{R}^d) - N| \geq cNx) \leq 2\exp(-Nf(cx))$. Next, using that the family $(\Pi_N(F))_{F\in\mathcal{P}_\ell}$ is independent, with $\Pi_N(F)$ Poisson($N\mu(F)$)-distributed, we use Lemma 9-(a) and that $\#(\mathcal{P}_\ell) = 2^{\ell d}$ to obtain, for any $\theta > 0$,

$$\mathbb{E}\left(\exp\left(\theta \sum_{F\in\mathcal{P}_\ell} |\Pi_N(F) - N\mu(F)|\right)\right) \leq \prod_{F\in\mathcal{P}_\ell} 2e^{N\mu(F)(e^\theta - \theta - 1)} \leq 2^{2^{d\ell}} e^{N(e^\theta - \theta - 1)}.$$

Hence

$$\mathbb{P}\left(\sum_{F\in\mathcal{P}_\ell}|\Pi_N(F)-N\mu(F)|\geq cNx2^{p\ell}r_\ell\right)\leq\exp\left(-c\theta Nx2^{p\ell}r_\ell\right)2^{2^{d\ell}}\exp\left(N(e^\theta-\theta-1)\right).$$

Choosing $\theta=\log(1+cx2^{p\ell}r_\ell)$, we find

$$\mathbb{P}\left(\sum_{F\in\mathcal{P}_\ell}|\Pi_N(F)-N\mu(F)|\geq cNx2^{p\ell}r_\ell\right)\leq 2^{2^{d\ell}}\exp(-Nf(cx2^{p\ell}r_\ell)).$$

We have checked that

$$\varepsilon(N,x)\leq 2\exp(-Nf(cx))+\exp(-Ng(cx2^{p\ell_0}-1))+\sum_{\ell=1}^{\ell_0}2^{2^{d\ell}}\exp(-Nf(cx2^{p\ell}r_\ell)).$$

At this point, the value of $c>0$ is not allowed to vary anymore. We introduce some other positive constants a whose value may change from line to line.

Case 1: $cx>2$**.** Then we choose $\ell_0=1$ and $r_1=1$. We have $cx2^{p\ell_0}-1=2^pcx-1\geq(2^p-1)cx+1$ whence $g(cx2^{p\ell_0}-1)\geq g((2^p-1)cx+1)=f((2^p-1)cx)$. We also have $\sum_{\ell=1}^{\ell_0}2^{2^{d\ell}}\exp(-Nf(cx2^{p\ell}r_\ell))=2^{2^d}\exp(-Nf(2^pcx))$. We finally get $\varepsilon(N,x)\leq C\exp(-Nf(ax))$, which proves the statement (in the three cases, when $cx>2$).

Case 2: $cx\leq 2$**.** We choose ℓ_0 so that $(1+2/(cx))\leq 2^{p\ell_0}\leq 2^p(1+2/(cx))$, i.e.

$$\ell_0:=\lfloor\log(1+2/(cx))/(p\log 2)\rfloor+1.$$

This implies that $cx2^{p\ell_0}\geq 2+cx$. Hence $g(cx2^{p\ell_0}-1)\geq g(1+cx)=f(cx)$. Furthermore, we have $cx2^{p\ell}r_\ell\leq cx2^{p\ell_0}\leq 2^p(2+cx)\leq 2^{p+2}$ for all $\ell\leq\ell_0$, whence $f(cx2^{p\ell}r_\ell)\geq ax^2 2^{2p\ell}r_\ell^2$ (because $f(x)\geq ax^2$ for all $x\in[0,2^{p+2}]$). We thus end up with (we use that $2^{2^{d\ell}}\leq\exp(2^{d\ell})$)

$$\varepsilon(N,x)\leq 3\exp(-Nf(cx))+\sum_{\ell=1}^{\ell_0}\exp(2^{d\ell}-Nax^2 2^{2p\ell}r_\ell^2).$$

Now the value of $a>0$ is not allowed to vary anymore, and we introduce $a'>0$, whose value may change from line to line.

Case 1.1: $p>d/2$. We take $r_\ell:=(1-2^{-\eta})2^{-\eta\ell}$ for some $\eta>0$ such that $2(p-\eta)>d$. If $Nx^2\geq 1$, we easily get

$$\begin{aligned}\varepsilon(N,x)\leq&3\exp(-Nf(cx))+\sum_{\ell=1}^{\ell_0}\exp(2^{d\ell}-Na'x^2 2^{2(p-\eta)\ell})\\ \leq&3\exp(-Nf(cx))+C\exp(-a'Nx^2)\\ \leq&C\exp(-Nf(a'x)).\end{aligned}$$

The last inequality uses that $y^2\geq f(y)$ for all $y>0$. If finally $Nx^2\leq 1$, we obviously have

$$\varepsilon(N,x)\leq 1\leq\exp(1-Nx^2)\leq C\exp(-Nx^2)\leq C\exp(-Nf(x)).$$

We thus always have $\varepsilon(N,x)\leq C\exp(-Nf(a'x))$ as desired.

Case 2.2: $p = d/2$. We choose $r_\ell := 1/\ell_0$. Thus, if $aN(x/\ell_0)^2 \geq 2$, we easily find

$$\begin{aligned}\varepsilon(N,x) \leq & 3\exp(-Nf(cx)) + \sum_{\ell=1}^{\ell_0} \exp(2^{d\ell}(1 - aN(x/\ell_0)^2) \\ \leq & 3\exp(-Nf(cx)) + C\exp(-a'N(x/\ell_0)^2) \\ \leq & 3\exp(-Nf(cx)) + C\exp(-Nf(a'x/\ell_0)) \\ \leq & C\exp(-Nf(a'x/\ell_0))\end{aligned}$$

because $\ell_0 \geq 1$ and f is increasing. If now $aN(x/\ell_0)^2 < 2$, we just write

$$\varepsilon(N,x) \leq 1 \leq \exp(2 - aN(x/\ell_0)^2) \leq C\exp(-aN(x/\ell_0)^2) \leq C\exp(-Nf(ax/\ell_0)).$$

We thus always have $\varepsilon(N,x) \leq C\exp(-Nf(a'x/\ell_0))$. Using that $\ell_0 \leq C\log(2+1/x)$, we immediately conclude that $\varepsilon(N,x) \leq C\exp(-Nf(a'x/\log(2+1/x)))$ as desired.

Case 2.3: $p \in [1, d/2)$. We choose $r_\ell := \kappa 2^{(d/2-p)(\ell-\ell_0)}$ with $\kappa = 1/(1-2^{p-d/2})$. For all $\ell \leq \ell_0$,

$$\begin{aligned}2^{dl} - aNx^2 2^{2p\ell} r_\ell^2 = & -a\kappa^2 N x^{d/p} 2^{2p\ell} \left[2^{(d-2p)(\ell-\ell_0)} x^{2-d/p} - 2^{(d-2p)\ell}/(Nax^{d/p})\right] \\ \leq & -a\kappa^2 N x^{d/p} 2^{2p\ell} \left[b 2^{(d-2p)\ell} - 2^{(d-2p)\ell}/(Na\kappa^2 x^{d/p})\right]\end{aligned}$$

where the constant $b > 0$ is such that $2^{-(d-2p)\ell_0} \geq b x^{d/p-2}$ (the existence of b is easily checked). Hence if $Na\kappa^2 x^{d/p} \geq 2/b$, we find

$$2^{dl} - aNx^2 2^{2p\ell} r_\ell^2 \leq -ab\kappa^2 N x^{d/p} 2^{d\ell} b/2$$

and thus, still using that $Nx^{d/p} \geq 2/(ab\kappa^2)$,

$$\sum_{\ell=1}^{\ell_0} \exp(2^{d\ell} - Nc^2x^2 2^{2p\ell} r_\ell^2) \leq C\exp(-a'Nx^{d/p}).$$

Consequently, we have $\varepsilon(N,x) \leq 3\exp(-Nf(cx)) + C\exp(-a'Nx^{d/p})$ if $Na\kappa^2 x^{d/p} \geq 2/b$. As usual, the case where $Na\kappa^2 x^{d/p} \leq 2/b$ is trivial, since then

$$\varepsilon(N,x) \leq 1 \leq \exp(2/b - Na\kappa^2 x^{d/p}) \leq C\exp(-a'Nx^{d/p}).$$

This ends the proof. □

5. Depoissonization in the compact case

We next check the following compact version of Theorem 2.

Proposition 10. *Assume that μ is supported in $(-1,1]^d$. Let $p > 0$ and $d \geq 1$ be fixed. There are some positive constants C and c (depending only on p, d) such that for all $N \geq 1$, all $x \in (0,\infty)$,*

$$\mathbb{P}\left[\mathcal{D}_p(\mu_N, \mu) \geq x\right] \leq \mathbf{1}_{\{x\leq 1\}} C \begin{cases} \exp(-cNx^2) & \text{if } p > d/2; \\ \exp\left(-cN(x/\log(2+1/x))^2\right) & \text{if } p = d/2; \\ \exp\left(-cNx^{d/p}\right) & \text{if } p \in (0, d/2). \end{cases}$$

We will need the following easy remark.

Lemma 11. *For all $N \geq 1$, for X Poisson(N)-distributed, for all $k \in \{0, \dots, \lfloor\sqrt{N}\rfloor\}$,*

$$\mathbb{P}[X = N + k] \geq \kappa_0 N^{-1/2} \text{ where } \kappa_0 = e^{-2}/\sqrt{2}.$$

Proof. By Perrin [34], we have $N! \le e\sqrt{N}(N/e)^N$. Thus

$$\mathbb{P}[X = N+k] = e^{-N}\frac{N^{N+k}}{(N+k)!} \ge e^{-N-1}\frac{N^{N+k}}{\sqrt{N+k}((N+k)/e)^{N+k}} \ge \frac{1}{\sqrt{2N}}\left(\frac{N}{N+k}\right)^{N+k} e^{k-1}.$$

Since $\log(1+x) \le x$ on $(0,1)$, we have $((N+k)/N)^{N+k} \le \exp(k+k^2/N) \le \exp(k+1)$, so that $\mathbb{P}[X = N+k] \ge e^{-2}/\sqrt{2N}$. □

Proof of Proposition 10. The probability indeed vanishes if $x > 1$, since $\mathcal{D}_p$ is smaller than 1 when restricted to probability measures on $(-1,1]^d$.

Step 1. We introduce a Poisson measure Π_N on $\mathbb{R}^d$ with intensity measure $N\mu$ and the associated empirical measure $\Psi_N = \Pi_N/\Pi_N(\mathbb{R}^d)$. Conditionally on $\{\Pi_N(\mathbb{R}^d) = n\}$, Ψ_N has the same law as μ_n (the empirical measure of n i.i.d. random variables with law μ). Consequently,

$$\mathbb{P}\left[\Pi_N(\mathbb{R}^d)\mathcal{D}_p(\Psi_N,\mu) \ge Nx\right] = \sum_{n\ge 0}\mathbb{P}[\Pi_N(\mathbb{R}^d) = n]\mathbb{P}\left[n\mathcal{D}_p(\mu_n,\mu) \ge Nx\right].$$

By Lemma 11 (since $\Pi_N(\mathbb{R}^d)$ is Poisson(N)-distributed),

$$\frac{1}{\sqrt{N}}\sum_{k=0}^{\lfloor\sqrt{N}\rfloor}\mathbb{P}\left[(N+k)\mathcal{D}_p(\mu_{N+k},\mu) \ge Nx\right] \le \kappa_0^{-1}\mathbb{P}\left[\Pi_N(\mathbb{R}^d)\mathcal{D}_p(\Psi_N,\mu) \ge Nx\right],$$

which of course implies that (for all $N \ge 1$, all $x > 0$),

$$\frac{1}{\sqrt{N}}\sum_{k=0}^{\lfloor\sqrt{N}\rfloor}\mathbb{P}\left[\mathcal{D}_p(\mu_{N+k},\mu) \ge x\right] \le \kappa_0^{-1}\mathbb{P}\left[\Pi_N(\mathbb{R}^d)\mathcal{D}_p(\Psi_N,\mu) \ge Nx\right].$$

Step 2. Here we prove that there is a constant $A > 0$ such that for any $N \ge 1$, any $k \in \{0,\dots,\lfloor\sqrt{N}\rfloor\}$, any $x > AN^{-1/2}$,

$$\mathbb{P}\left[\mathcal{D}_p(\mu_N,\mu) \ge x\right] \le \mathbb{P}\left[\mathcal{D}_p(\mu_{N+k},\mu) \ge x/2\right].$$

Build μ_n for all values of $n \ge 1$ with the same i.i.d. family of μ-distributed random variables $(X_k)_{k\ge 1}$. Then a.s.,

$$|\mu_{N+k} - \mu_N|_{TV} \le \left|\frac{k}{N(N+k)}\sum_1^N \delta_{X_j}\right|_{TV} + \left|\frac{1}{N+k}\sum_{N+1}^{N+k}\delta_{X_j}\right|_{TV} \le \frac{k}{N+k} \le \frac{1}{\sqrt{N}}.$$

This obviously implies (recall Notation 4-(a)) that $\mathcal{D}_p(\mu_N,\mu_{N+k}) \le CN^{-1/2}$ a.s. (where C depends only on p). By the triangular inequality, $\mathcal{D}_p(\mu_N,\mu) \le \mathcal{D}_p(\mu_{N+k},\mu) + CN^{-1/2}$, whence

$$\mathbb{P}\left[\mathcal{D}_p(\mu_N,\mu) \ge x\right] \le \mathbb{P}\left[\mathcal{D}_p(\mu_{N+k},\mu) \ge x - CN^{-1/2}\right] \le \mathbb{P}\left[\mathcal{D}_p(\mu_{N+k},\mu) \ge x/2\right]$$

if $x - CN^{-1/2} \ge x/2$, i.e. $x \ge 2CN^{-1/2}$.

Step 3. Gathering Steps 1 and 2, we deduce that for all $N \ge 1$, all $x > AN^{-1/2}$,

$$\mathbb{P}\left[\mathcal{D}_p(\mu_N,\mu) \ge x\right] \le \frac{1}{\sqrt{N}}\sum_{k=0}^{\lfloor\sqrt{N}\rfloor}\mathbb{P}\left[\mathcal{D}_p(\mu_{N+k},\mu) \ge x/2\right] \le C\mathbb{P}\left[\Pi_N(\mathbb{R}^d)\mathcal{D}_p(\Psi_N,\mu) \ge Nx/2\right].$$

We next apply Proposition 8. Observing that, for $x \in (0,1]$,

(i) $\exp(-Nf(cx/2)) \le \exp(-cNx^2)$ (case $p > d/2$),

(ii) $\exp(-Nf(cx/2\log(2+2/x))) \le \exp(-cN(x/\log(2+1/x)^2)$ (case $p = d/2$),

(iii) $\exp(-Nf(cx/2)) + \exp(cN(x/2)^{d/p}) \le \exp(-cNx^{d/p})$ (case $p \in (0, d/2)$)
concludes the proof when $x > AN^{-1/2}$. But the other case is trivial, because for $x \le AN^{-1/2}$,

$$\mathbb{P}[\mathcal{D}_p(\mu_N, \mu) \ge x] \le 1 \le \exp(A^2 - Nx^2) \le C\exp(-Nx^2),$$

which is also smaller than $C\exp(-N(x/\log(2+1/x))^2)$ and than $C\exp(-Nx^{d/p})$ (if $d > 2p$). □

6. Concentration inequalities in the non compact case

Here we conclude the proof of Theorem 2. We will need some concentration estimates for the Binomial distribution.

Lemma 12. *Let X be Binomial(N,p)-distributed. Recall that f was defined in Notation 7.*
(a) $\mathbb{P}[|X - Np| \ge Npz] \le (\mathbf{1}_{\{p(1+z)\le 1\}} + \mathbf{1}_{\{z\le 1\}})\exp(-Npf(z))$ for all $z > 0$.
(b) $\mathbb{P}[|X - Np| \ge Npz] \le Np$ for all $z > 1$.
(c) $\mathbb{E}(\exp(-\theta X)) = (1 - p + pe^{-\theta})^N \le \exp(-Np(1 - e^{-\theta}))$ for $\theta > 0$.

Proof. Point (c) is straightforward. Point (b) follows from the fact that for $z > 1$, $\mathbb{P}[|X - Np| \ge Npz] = \mathbb{P}[X \ge Np(1+z)] \le \mathbb{P}[X \ne 0] = 1 - (1-p)^N \le pN$. For point (a), we use Bennett's inequality [4], see Devroye-Lugosi [17, Exercise 2.2 page 11], together with the obvious facts that $\mathbb{P}[X - Np \ge Npz] = 0$ if $p(1+z) > 1$ and $\mathbb{P}[X - Np \le -Npz] = 0$ if $z > 1$. The following elementary tedious computations also works: write $\mathbb{P}[|X - Np| \ge Npz] = \mathbb{P}(X \ge Np(1+z)) + \mathbb{P}(N - X \ge N(1-p+zp)) =: \Delta(p,z) + \Delta(1-p, zp/(1-p))$, observe that $N - X \sim \text{Binomial}(N, 1-p)$. Use that $\Delta(p,z) \le \mathbf{1}_{\{p(1+z)\le 1\}}\exp(-\theta Np(1+z))(1-p+pe^{\theta})^N$ and choose $\theta = \log((1-p)(1+z)/(1-p-pz))$, this gives $\Delta(p,z) \le \mathbf{1}_{\{p(1+z)\le 1\}}\exp(-N[p(1+z)\log(1+z) + (1-p-pz)\log((1-p-pz)/(1-p))])$. A tedious study shows that $\Delta(p,z) \le \mathbf{1}_{\{p(1+z)\le 1\}}\exp(-Npf(z))$ and that $\Delta(1-p, zp/(1-p)) \le \mathbf{1}_{\{z\le 1\}}\exp(-Npf(z))$. □

We next estimate the first term when computing $\mathcal{D}_p(\mu_N, \mu)$.

Lemma 13. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $p > 0$. Assume (1), (2) or (3). Recall Notation 4 and put $Z_N^p := \sum_{n\ge 0} 2^{pn}|\mu_N(B_n) - \mu(B_n)|$. Let x_0 be fixed. For all $x > 0$,*

$$\mathbb{P}[Z_N^p \ge x] \le C\exp(-cNx^2)\mathbf{1}_{\{x\le x_0\}} + C\begin{cases} \exp(-cNx^{\alpha/p})\mathbf{1}_{\{x>x_0\}} & \text{under (1),} \\ \exp(-c(Nx)^{(\alpha-\varepsilon)/p})\mathbf{1}_{\{x\le x_0\}} + \exp(-c(Nx)^{\alpha/p})\mathbf{1}_{\{x>x_0\}} & \forall\, \varepsilon \in (0,\alpha) \text{ under (2),} \\ N(Nx)^{-(q-\varepsilon)/p} & \forall\, \varepsilon \in (0,q) \text{ under (3).} \end{cases}$$

Proof. Under (1) or (2), we assume that $\gamma = 1$ without loss of generality (by scaling), whence $\mathcal{E}_{\alpha,1}(\mu) < \infty$ and thus $\mu(B_n) \le Ce^{-2^{(n-1)\alpha}}$ for all $n \ge 0$. Under (3), we have $\mu(B_n) \le C2^{-qn}$ for all $n \ge 0$. For $\eta > 0$ to be chosen later (observe that $\sum_{n\ge 0}(1-2^{-\eta})2^{-\eta n} = 1$), putting $c := 1 - 2^{-\eta}$

and $z_n := cx2^{-(p+\eta)n}/\mu(B_n)$,

$$\begin{aligned}\mathbb{P}\left(Z_N^p \geq x\right) &\leq \left(\sum_{n\geq 0} \mathbf{1}_{\{z_n\leq 2\}}\mathbb{P}\left[|N\mu_N(B_n) - N\mu(B_n)| \geq N\mu(B_n)z_n\right]\right) \wedge 1 \\ &\quad + \left(\sum_{n\geq 0} \mathbf{1}_{\{z_n> 2\}}\mathbb{P}\left[|N\mu_N(B_n) - N\mu(B_n)| \geq N\mu(B_n)z_n\right]\right) \wedge 1 \\ &=: \left(\sum_{n\geq 0} I_n(N,x)\right) \wedge 1 + \left(\sum_{n\geq 0} J_n(N,x)\right) \wedge 1.\end{aligned}$$

From now on, the value of $c > 0$ is not allowed to vary anymore. We introduce another positive constant $a > 0$ whose value may change from line to line.

Step 1: bound of I_n. Here we show that under (3) (which is of course implied by (1) or (2)), if $\eta \in (0, q/2 - p)$, there is $A_0 > 0$ such that

$$\sum_{n\geq 0} I_n(N,x) \leq C\exp(-aNx^2)\mathbf{1}_{\{x\leq A_0\}} \quad \text{if } Nx^2 \geq 1.$$

This will obviously imply that for all $N \geq 1$, all $x > 0$,

$$\left(\sum_{n\geq 0} I_n(N,x)\right) \wedge 1 \leq C\exp(-aNx^2)\mathbf{1}_{\{x\leq A_0\}}.$$

First, $\sum_{n\geq 0} I_n(N,x) = 0$ if $z_n > 2$ for all $n \geq 0$. Recalling that $\mu(B_n) \leq C2^{-qn}$, this is the case if $x \geq (2C/c)\sup_{n\geq 0} 2^{(p+\eta-q)n} = (2C/c) := A_0$. Next, since $N\mu_N(B_n) \sim \text{Binomial}(N, \mu(B_n))$, Lemma 12-(a) leads us to

$$I_n(N,x) \leq 2\mathbf{1}_{\{z_n\leq 2\}}\exp(-N\mu(B_n)f(z_n)) \leq 2\exp(-N\mu(B_n)z_n^2/4)),$$

because $f(x) \geq x^2/4$ for $x \in [0,2]$. Since finally $\mu(B_n)z_n^2/4 \geq ax^2 2^{(q-2p-2\eta)n}$, we easily conclude, since $q - 2p - 2\eta > 0$ and since $Nx^2 \geq 1$, that

$$\sum_{n\geq 0} I_n(N,x) \leq C\sum_{n\geq 0}\exp(-aNx^2 2^{(q-2p-2\eta)n})\mathbf{1}_{\{x\leq A_0\}} \leq C\exp(-aNx^2)\mathbf{1}_{\{x\leq A_0\}}.$$

Step 2: bound of J_n under (1) or (2) when $x \leq A$. Here we fix $A > 0$ and prove that if $\eta > 0$ is small enough, for all $x \in (0, A]$ such that $Nx^2 \geq 1$,

$$\sum_{n\geq 0} J_n(N,x) \leq C\begin{cases} \exp(-aNx^2) & \text{under (1)}, \\ \exp(-aNx^2) + \exp(-a(Nx)^{(\alpha-\varepsilon)/p}) & \forall\, \varepsilon \in (0,\alpha) \text{ under (2)}. \end{cases}$$

This will imply, as usual, that for all $N \geq 1$, all $x > 0$,

$$\left(\sum_{n\geq 0} J_n(N,x)\right) \wedge 1 \leq C\begin{cases} \exp(-aNx^2) & \text{under (1)}, \\ \exp(-aNx^2) + \exp(-a(Nx)^{(\alpha-\varepsilon)/p}) & \forall\, \varepsilon \in (0,\alpha) \text{ under (2)}. \end{cases}$$

By Lemma 12-(a)-(b) (since $z_n > 2$ implies $\mathbf{1}_{\{\mu(B_n)(1+z_n)\leq 1\}} + \mathbf{1}_{\{z_n\leq 1\}} \leq \mathbf{1}_{\{z_n\leq 1/\mu(B_n)\}}$),

$$\begin{aligned}J_n(N,x) &\leq \mathbf{1}_{\{2<z_n\leq 1/\mu(B_n)\}}\min\left\{\exp(-N\mu(B_n)f(z_n)), N\mu(B_n)\right\} \\ &\leq \mathbf{1}_{\{z_n\mu(B_n)\leq 1\}}\min\left\{\exp\left(-aN\mu(B_n)z_n\log[2\vee z_n]\right), N\mu(B_n)\right\}\end{aligned}$$

because $f(y) \geq ay \log y \geq ay \log[2 \vee y]$ for $y > 2$. Since $\mu(B_n) \leq Ce^{-2^{(n-1)\alpha}}$, we get

$$J_n(N,x) \leq C \min\{\exp(-aNx2^{-(p+\eta)n} \log[2 \vee (ax2^{-(p+\eta)n} e^{2^{(n-1)\alpha}})]), Ne^{2^{-(n-1)\alpha}}\}.$$

A straightforward computation shows that there is a constant K such that for $n \geq n_1 := \lfloor K(1 + \log\log(K/x))\rfloor$, we have $\log(ax2^{-(p+\eta)n} e^{2^{(n-1)\alpha}}) \geq 2^{(n-1)\alpha}/2$. Consequently,

$$\begin{aligned}\sum_{n\geq 0} J_n(N,x) \leq & Cn_1 \exp(-aNx2^{-(p+\eta)n_1}) + C \sum_{n>n_1} \min\left\{\exp(-aNx2^{(\alpha-p-\eta)n}), e^{-2^{(n-1)\alpha}}\right\} \\ = & CJ^1(N,x) + CJ^2(N,x).\end{aligned}$$

We first show that $J^1(N,x) \leq Ce^{-aNx^2}$ (here we actually could get something much better). First, since $n_1 = \lfloor K + K \log\log(K/x)\rfloor$ and $x \in [0,A]$, we clearly have e.g. $x2^{-(p+\eta)n_1} \geq ax^{3/2}$. Next, $Nx^2 \geq 1$ implies that $1/x \leq (Nx^{3/2})^2$. Thus

$$J^1(N,x) \leq C(1 + \log\log(C(Nx^{3/2})^2)) \exp(-aNx^{3/2}) \leq C\exp(-aNx^{3/2}) \leq \exp(-aNx^2).$$

We now treat $J^2(N,x)$.

Step 2.1. Under (1), we immediately get, if $\eta \in (0, \alpha - p)$ (recall that $x \in [0,A]$),

$$J^2(N,x) \leq \sum_{n\geq 0} \exp(-aNx2^{(\alpha-p-\eta)n}) \leq C\exp(-aNx) \leq C\exp(-aNx^2),$$

where we used that $x \leq A$ and $Nx^2 \geq 1$ (whence $Nx \geq 1/A$).

Step 2.2. Under (2), we first write

$$J^2(N,x) \leq \sum_{n\geq 0} \min\left\{\exp(-aNx2^{(\alpha-p-\eta)n}), e^{-2^{(n-1)\alpha}}\right\} \leq n_2 \exp(-cNx2^{(\alpha-p-\eta)n_2}) + Ne^{-2^{(n_2-1)\alpha}}.$$

We choose $n_2 := \lfloor \log(Nx)/((p+\eta)\log 2)\rfloor$, which yields us to $2^{(n_2-1)\alpha} \geq (Nx)^{\alpha/(p+\eta)}/2^{2\alpha}$ and $(Nx)2^{(\alpha-p-\eta)n_2} \leq (Nx)^{\alpha/(p+\eta)}$. Consequently (recall that $x \in (0,A]$),

$$J^2(N,x) \leq C(1 + \log(Nx) + N)\exp(-a(Nx)^{\alpha/(p+\eta)}) \leq C(1+N)\exp(-a(Nx)^{\alpha/(p+\eta)}).$$

For any fixed $\varepsilon \in (0,\alpha)$, we choose $\eta > 0$ small enough so that $\alpha/(p+\eta) \geq (\alpha-\varepsilon)/p$ and we conclude that (recall that $Nx \geq 1/A$ because $Nx^2 \geq 1$ and $x \leq A$)

$$J^2(N,x) \leq C(1+N)\exp(-a(Nx)^{(\alpha-\varepsilon)/p}) \leq C\exp(-a(Nx)^{(\alpha-\varepsilon)/p}).$$

The last inequality is easily checked, using that $Nx^2 \geq 1$ implies that $N \leq (Nx)^2$.

Step 3: bound of J_n under (3). Here we show that for all $\varepsilon \in (0,q)$, if $\eta > 0$ is small enough,

$$\sum_{n\geq 0} J_n(N,x) \leq CN\left(\frac{1}{Nx}\right)^{(q-\varepsilon)/p} \quad \text{if } Nx \geq 1.$$

As usual, this will imply that for all $x > 0$, all $N \geq 1$,

$$\left(\sum_{n\geq 0} J_n(N,x)\right) \wedge 1 \leq CN\left(\frac{1}{Nx}\right)^{(q-\varepsilon)/p}.$$

Exactly as in Step 2, we get from Lemma 12-(a)-(b) that

$$J_n(N,x) \leq \min\left\{\exp\left(-aN\mu(B_n)z_n \log[2 \vee z_n]\right), N\mu(B_n)\right\}.$$

Hence for n_3 to be chosen later, since $aN\mu(B_n)z_n = aNx2^{-(p+\eta)n}$,

$$\begin{aligned}\sum_{n\geq 0} J_n(N,x) \leq & C\sum_{n=0}^{n_3} \exp(-aNx2^{-(p+\eta)n}) + CN\sum_{n>n_3} 2^{-qn} \\ \leq & Cn_3\exp(-aNx2^{-(p+\eta)n_3}) + CN2^{-qn_3}.\end{aligned}$$

We choose $n_3 := \lfloor (q-\varepsilon)\log(Nx)/(pq\log 2)\rfloor$, which implies that $2^{-qn_3} \leq 2^q(Nx)^{-(q-\varepsilon)/p}$ and that $2^{-(p+\eta)n_3} \geq (Nx)^{-(q-\varepsilon)(p+\eta)/(pq)}$. Hence

$$\sum_{n\geq 0} J_n(N,x) \leq C\log(Nx)\exp(-a(Nx)^{1-(q-\varepsilon)(p+\eta)/(pq)}) + CN(Nx)^{-(q-\varepsilon)/p}.$$

If $\eta \in (0, p\varepsilon/(q-\varepsilon))$, then $1-(q-\varepsilon)(p+\eta)/(pq) > 0$, and thus

$$\log(Nx)\exp(-a(Nx)^{1-(q-\varepsilon)(p+\eta)/(pq)}) \leq C(Nx)^{-(q-\varepsilon)/p}.$$

This ends the step.

Step 4. We next assume (1) and prove that for all $x \geq A_1 := 2^p[M_p(\mu) + (2\log \mathcal{E}_{\alpha,1}(\mu))^{p/\alpha}]$,

$$\Pr[Z_N^p \geq x] \leq C\exp(-aNx^{\alpha/p}).$$

A simple computation shows that for any $\nu \in \mathcal{P}(\mathbb{R}^d)$, $\sum_{n\geq 0} 2^{pn}\nu(B_n) \leq 2^p M_p(\nu)$, whence $Z_N^p \leq 2^p M_p(\mu) + 2^p N^{-1}\sum_1^N |X_i|^p \leq 2^p M_p(\mu) + 2^p[N^{-1}\sum_1^N |X_i|^\alpha]^{p/\alpha}$. Thus

$$\Pr[Z_N^p \geq x] \leq \Pr\left[N^{-1}\sum_1^N |X_i|^\alpha \geq [x2^{-p} - M_p(\mu)]^{\alpha/p}\right].$$

Next, we note that for $y \geq 2\log \mathcal{E}_{\alpha,1}(\mu)$,

$$\Pr\left[N^{-1}\sum_1^N |X_i|^\alpha \geq y\right] \leq \exp(-Ny + N\log \mathcal{E}_{\alpha,1}(\mu)) \leq \exp(-Ny/2).$$

The conclusion easily follows, since $x \geq A_1$ implies that $y := [x2^{-p} - M_p(\mu)]^{\alpha/p} \geq 2\log \mathcal{E}_{\alpha,1}(\mu)$ and since $y \geq [x2^{-p-1}]^{\alpha/p} - [M_p(\mu)]^{\alpha/p}$.

Step 5. Assume (2) and put $\delta := 2p/\alpha - 1$. Here we show that for all $x > 0$, $N \geq 1$,

$$\Pr[Z_N^p \geq x] \leq C\exp(-a(Nx)^{\alpha/p}) + C\exp(-aNx^2(\log(1+N))^{-\delta}).$$

Step 5.1. For $R > 0$ (large) to be chosen later, we introduce the probability measure μ^R as the law of $X\mathbf{1}_{\{|X|\leq R\}}$. We also denote by μ_N^R the corresponding empirical measure (coupled with μ_N in that the X_i's are used for μ_N and the $X_i\mathbf{1}_{\{|X_i|\leq R\}}$'s are chosen for μ_N^R). We set $Z_N^{p,R} := \sum_{n\geq 0} 2^{pn}|\mu_N^R(B_n) - \mu^R(B_n)|$ and first observe that $|Z_N^p - Z_N^{p,R}| \leq 2^p N^{-1}\sum_1^N |X_i|^p\mathbf{1}_{\{|X_i|>R\}} + 2^p\int_{\{|x|>R\}} |x|^p\mu(dx)$. On the one hand, $\int_{\{|x|>R\}} |x|^p\mu(dx) \leq \exp(-R^\alpha/2)\int |x|^p e^{|x|^\alpha/2}\mu(dx) \leq C\exp(-R^\alpha/2)$ by (2) (with $\gamma = 1$). On the other hand, since $\alpha \in (0,p]$, $\sum_1^N |X_i|^p\mathbf{1}_{\{|X_i|>R\}} \leq$

$(\sum_1^N |X_i|^\alpha \mathbf{1}_{\{|X_i|>R\}})^{p/\alpha}$. Hence if $x \geq A\exp(-R^\alpha/2)$, where $A := 2^{p+1}C$,

$$\begin{aligned}\Pr\left(|Z_N^p - Z_N^{p,R}| \geq x\right) &\leq \Pr\left(N^{-1}\sum_1^N |X_i|^p \mathbf{1}_{\{|X_i|>R\}} \geq x 2^{-p-1}\right)\\ &\leq \Pr\left(\sum_1^N |X_i|^\alpha \mathbf{1}_{\{|X_i|>R\}} \geq (Nx2^{-p-1})^{\alpha/p}\right)\\ &\leq \exp(-(Nx2^{-p-1})^{\alpha/p}/2)\mathbb{E}\left[\exp\left(|X_1|^\alpha \mathbf{1}_{\{|X_1|>R\}}/2\right)\right]^N.\end{aligned}$$

Observing that $\mathbb{E}[\exp(|X_1|^\alpha \mathbf{1}_{\{|X_1|>R\}}/2)] \leq 1 + \mathbb{E}[\exp(|X_1|^\alpha/2)\mathbf{1}_{\{|X_1|>R\}}] \leq 1 + C\exp(-R^\alpha/2)$ by (2) and using that $\log(1+u) \leq u$, we deduce that for all $x \geq 2^{p+1}C\exp(-R^\alpha/2)$,

$$\Pr\left(|Z_N^p - Z_N^{p,R}| \geq x\right) \leq \exp\left(-(Nx2^{-p-1})^{\alpha/p}/2 + CN\exp(-R^\alpha/2)\right).$$

With the choice

$$R := (2\log(1+N))^{1/\alpha}, \tag{5}$$

we finally find

$$\Pr\left(|Z_N^p - Z_N^{p,R}| \geq x\right) \leq \exp\left(-(Nx2^{-p-1})^{\alpha/p}/2 + C\right) \leq C\exp\left(-a(Nx)^{\alpha/p}\right)$$

provided $x \geq A\exp(-R^\alpha/2)$, i.e. $(N+1)x \geq A$. As usual, this immediately extends to any value of $x > 0$.

Step 5.2. To study $Z_N^{p,R}$, we first observe that since $\mu^R(B_n) = 0$ if $2^{n-1} \geq R$, we have $2^{pn}\mu^R(B_n) \leq (2R)^{p-\alpha/2}2^{\alpha n/2}\mu^R(B_n)$ for all $n \geq 0$. Hence $Z_N^{p,R} \leq (2R)^{p-\alpha/2}Z_N^{\alpha/2,R}$. But μ^R satisfies $\int_{\mathbb{R}^d}\exp(|x|^\alpha)\mu^R(dx) < \infty$ uniformly in R, so that we may use Steps 1, 2 and 4 (with $p = \alpha/2 < \alpha$) to deduce that for all $x > 0$, $\Pr(Z_N^{\alpha/2,R} \geq x) \leq C\exp(-aNx^2)$. Consequently, $\Pr(Z_N^{p,R} \geq x) \leq C\exp(-aN(x/R^{p-\alpha/2})^2)$. Recalling (5) and that $\delta := 2p/\alpha - 1$, we see that that $\Pr(Z_N^{p,R} \geq x) \leq C\exp(-aNx^2(\log(1+N))^{-\delta})$. This ends the step.

Conclusion. Recall that $x_0 > 0$ is fixed.

First assume (1). By Step 4, $\Pr[Z_N^p \geq x] \leq C\exp(-aNx^{\alpha/p})$ for all $x \geq A_1$. We deduce from Steps 1 and 2 that for $x \in (0, A_1)$, $\Pr[Z_N^p \geq x] \leq C\exp(-aNx^2)$. We easily conclude that for all $x > 0$, $\Pr[Z_N^p \geq x] \leq C\exp(-aNx^2)\mathbf{1}_{\{x\leq x_0\}} + C\exp(-aNx^{\alpha/p})\mathbf{1}_{\{x>x_0\}}$ as desired.

Assume next (2). By Step 5, $\Pr[Z_N^p \geq x] \leq C\exp(-aNx^2(\log(1+N))^{-\delta}) + C\exp(-a(Nx)^{\alpha/p})$. But if $x \geq x_0$, we clearly have $(Nx)^{\alpha/p} \leq CNx^2(\log(1+N))^{-\delta}$ because $\alpha < p$, so that $\Pr[Z_N^p \geq x] \leq C\exp(-a(Nx)^{\alpha/p})$. If now $x \leq x_0$, we use Steps 1 and 2 to write $\Pr[Z_N^p \geq x] \leq C\exp(-aNx^2) + C\exp(-a(Nx)^{(\alpha-\varepsilon)/p})$.

Assume finally (3). By Steps 1 and 3, $\Pr[Z_N^p \geq x] \leq C\exp(-aNx^2) + CN(Nx)^{-(q-\varepsilon)/q}$ for all $x > 0$. But if $x \geq x_0$, $\exp(-aNx^2) \leq \exp(-aNx) \leq C(Nx)^{-(q-\varepsilon)/q} \leq CN(Nx)^{-(q-\varepsilon)/q}$. We conclude that for all $x > 0$, $\Pr[Z_N^p \geq x] \leq C\exp(-aNx^2)\mathbf{1}_{\{x\leq x_0\}} + CN(Nx)^{-(q-\varepsilon)/q}$ as desired. □

We can now give the

Proof of Theorem 2. Using Lemma 5, we write

$$\begin{aligned}\mathcal{T}_p(\mu_N,\mu) \leq& \kappa_{p,d}\mathcal{D}_p(\mu_N,\mu)\\ \leq& \kappa_{p,d}\sum_{n\geq 0} 2^{pn}|\mu_N(B_n)-\mu(B_n)| + \kappa_{p,d}\sum_{n\geq 0} 2^{pn}\mu(B_n)\mathcal{D}_p(\mathcal{R}_{B_n}\mu_N,\mathcal{R}_{B_n}\mu)\\ =:& \kappa_{p,d}(Z_N^p+V_N^p).\end{aligned}$$

Hence

$$\Pr(\mathcal{T}_p(\mu_N,\mu)\geq x) \leq \Pr(Z_N^p \geq x/(2\kappa_{p,d})) + \Pr(V_N^p \geq x/(2\kappa_{p,d})).$$

By Lemma 13 (choosing $x_0 := 1/(2\kappa_{p,d})$), we easily find $\Pr(Z_N^p \geq x/(2\kappa_{p,d})) \leq Ce^{-cNx^2}\mathbf{1}_{\{x\leq 1\}} + b(N,x) \leq a(N,x)\mathbf{1}_{\{x\leq 1\}} + b(N,x)$, these quantities being defined in the statement of Theorem 2. We now check that there is $A>0$ such that for all $x>0$,

$$\Pr[V_N^p \geq x/(2\kappa_{p,d})] \leq a(N,x)\mathbf{1}_{\{x\leq A\}}. \tag{6}$$

This will end the proof, since one easily checks that $a(N,x)\mathbf{1}_{\{x\leq A\}} \leq a(N,x)\mathbf{1}_{\{x\leq 1\}} + b(N,x)$ (when allowing the values of the constants to change).

Let us thus check (6). For $\eta>0$ to be chosen later, we set $c := (1-2^{-\eta})/(2\kappa_{p,d})$ and $z_n := cx2^{-(p+\eta)n}/\mu(B_n)$. Observing that $\sum_{n\geq 0}(1-2^{-\eta})2^{-\eta n}=1$), we write

$$\mathbb{P}\left(V_N^p \geq x/(2\kappa_{p,d})\right) \leq \left(\sum_{n\geq 0}\mathbb{P}\left[\mathcal{D}_p(\mathcal{R}_{B_n}\mu_N,\mathcal{R}_{B_n}\mu)\geq z_n\right]\right)\wedge 1 =: \left(\sum_{n\geq 0} K_n(N,x)\right)\wedge 1.$$

From now on, the value of $c>0$ is not allowed to vary anymore. We introduce another positive constant $a>0$ whose value may change from line to line. We only assume (3) (which is implied by (1) or (2)). We now show that if $\eta>0$ is small enough,

$$\sum_{n\geq 0} K_n(N,x) \leq C\exp(-aNh(x))\mathbf{1}_{\{x\leq A\}} \quad \text{if } Nh(x)\geq 1, \tag{7}$$

where $h(x)=x^2$ if $p>d/2$, $h(x)=(x/\log(2+1/x))^2$ if $p=d/2$ and $h(x)=x^{d/p}$ if $p<d/2$. This will obviously imply as usual that for all $x>0$,

$$\left(\sum_{n\geq 0} K_n(N,x)\right)\wedge 1 \leq C\exp(-aNh(x))\mathbf{1}_{\{x\leq A\}}$$

and thus conclude the proof of (6). We thus only have to prove (7).

Conditionally on $\mu_N(B_n)$, $\mathcal{R}_{B_n}\mu_N$ is the empirical measure of $N\mu_N(B_n)$ points which are $\mathcal{R}_{B_n}\mu$-distributed. Since $\mathcal{R}_{B_n}\mu$ is supported in $(-1,1]^d$, we may apply Proposition 10 and obtain

$$K_n(N,x) \leq C\mathbb{E}\left[\mathbf{1}_{\{z_n\leq 1\}}\exp\left(-aN\mu_N(B_n)h(z_n)\right)\right] \leq C\mathbf{1}_{\{z_n\leq 1\}}\exp(-N\mu(B_n)(1-e^{-ah(z_n)}))$$

by Lemma 12-(c). But the condition $z_n\leq 1$ implies that $h(z_n)$ is bounded (by a constant depending only on p and d), whence

$$K_n(N,x) \leq C\mathbf{1}_{\{z_n\leq 1\}}\exp(-aN\mu(B_n)h(z_n)).$$

By (3), we have $\mu(B_n)\leq C2^{-qn}$. Hence if $x>A:=C/c$, we have $z_n \geq (c/C)x2^{(q-p-\eta)n} > 1$ for all $n\geq 1$ (if $\eta\in(0,q-p)$) and thus $\sum_{n\geq 0}K_n(N,x)=0$ as desired.

Next, we see that $\theta \mapsto \theta h(x/\theta)$ is decreasing, whence for all $x \le A$,

$$K_n(N,x) \le C\exp(-aN2^{-qn}h(cx2^{(q-p-\eta)n}/C)) \le C\exp(-aN2^{-qn}h(x2^{(q-p-\eta)n})).$$

We now treat separately the three cases.

Step 1: case $p > d/2$. Since $h(x) = x^2$, we have, if $\eta \in (0, q/2-p)$,

$$\sum_{n\ge 0} K_n(N,x) \le C\sum_{n\ge 0}\exp(-aNx^2 2^{n(q-2p-2\eta)}) \le C\exp(-aNx^2)$$

if $Nx^2 \ge 1$.

Step 2: case $p = d/2$. Since $h(x) = (x/\log(2+1/x))^2$, we have, if $\eta \in (0, q/2-p)$,

$$\begin{aligned}\sum_{n\ge 0} K_n(N,x) \le& C\sum_{n\ge 0}\exp\left(-aNx^2 2^{(q-2p-2\eta)n}/\log^2[2+1/(x2^{(q-p-\eta)n})]\right)\\ \le& C\sum_{n\ge 0}\exp(-aNh(x)2^{n(q-2p-2\eta)})\\ \le& C\exp(-aNh(x))\end{aligned}$$

if $Nh(x) \ge 1$. The third inequality only uses that $\log^2(2+1/(x2^{n(q-p-\eta)})) \le \log^2(2+1/x)$.

Step 3: case $p < d/2$. Here $h(x) = x^{d/p}$. Since $p < d/2$ and $q > 2p$, it holds that $q(1-p/d)-p > 0$. We thus may take $\eta \in (0, q(1-p/d)-p)$ (so that $q(d/p-1)-d-d\eta/p > 0$) and we get

$$\sum_{n\ge 0} K_n(N,x) \le C\sum_{n\ge 0}\exp(-aNx^{d/p}2^{n(q(d/p-1)-d-d\eta/p)}) \le C\exp(-aNx^{d/p})$$

if $Nx^{d/p} \ge 1$. □

7. The dependent case

We finally study a few classes of dependent sequences of random variables. We only give some moment estimates. Concentration inequalities might be obtained, but this should be much more complicated.

7.1. ρ-mixing stationary sequences.

A stationary sequence of random variables $(X_n)_{n\ge 1}$ with common law μ is said to be ρ-mixing, for some $\rho : \mathbb{N} \to \mathbb{R}^+$ with $\rho_n \to 0$, if for all $f, g \in L^2(\mu)$ and all $i, j \ge 1$

$$\mathbb{C}\mathrm{ov}\left(f(X_i), g(X_j)\right) \le \rho_{|i-j|}\sqrt{\mathbb{V}\mathrm{ar}\left(f(X_i)\right)\mathbb{V}\mathrm{ar}\left(g(X_j)\right)}.$$

We refer for example to Rio [37], Doukhan [20] or Bradley [10].

Theorem 14. *Consider a stationary sequence of random variables $(X_n)_{n\ge 1}$ with common law μ and set $\mu_N := N^{-1}\sum_1^N \delta_{X_i}$. Assume that this sequence is ρ-mixing, for some $\rho : \mathbb{N} \to \mathbb{R}^+$ satisfying $\sum_{n\ge 0}\rho_n < \infty$. Let $p > 0$ and assume that $\mu \in M_q(\mathbb{R}^d)$ for some $p > q$. There exists a constant C depending only on $p, d, q, M_q(\mu), \rho$ such that, for all $N \ge 1$,*

$$\mathbb{E}\left(\mathcal{T}_p(\mu_N, \mu)\right) \le C\begin{cases} N^{-1/2} + N^{-(q-p)/q} & \text{if } p > d/2 \text{ and } q \ne 2p,\\ N^{-1/2}\log(1+N) + N^{-(q-p)/q} & \text{if } p = d/2 \text{ and } q \ne 2p,\\ N^{-p/d} + N^{-(q-p)/q} & \text{if } p \in (0, d/2) \text{ and } q \ne d/(d-p).\end{cases}$$

This is very satisfying: we get the same estimate as in the independent case. The case $\sum_{n\geq 0}\rho_n = \infty$ can also be treated (but then the upper bounds will be less good and depend on the rate of decrease of ρ). Actually, the ρ-mixing condition is slightly too strong (we only need the covariance inequality when $f = g$ is an indicator function), but it is best adapted notion of mixing we found in the litterature.

Proof. We first check that for any Borel subset $A \subset \mathbb{R}^d$,

$$\mathbb{E}[|\mu_N(A) - \mu(A)|] \leq \min\{2\mu(A), C\mu(A)N^{-1/2}\}.$$

But this is immediate: $\mathbb{E}[\mu_N(A)] = \mu(A)$ (whence $\mathbb{E}[|\mu_N(A) - \mu(A)|] \leq 2\mu(A)$) and

$$\begin{aligned}\mathbb{V}\mathrm{ar}\,\mu_N(A) =& \frac{1}{N^2}\sum_{i,j\leq N}\mathbb{C}\mathrm{ov}\left(\mathbf{1}_A(X_i), \mathbf{1}_A(X_i)\right)\\ \leq& \frac{1}{N^2}\sum_{i,j\leq N}\rho_{|i-j|}\mathbb{V}\mathrm{ar}\left(\mathbf{1}_A(X_1)\right)\\ \leq& \frac{\mu(A)(1-\mu(A))}{N^2}\sum_{i,j\leq N}\rho_{|i-j|}.\end{aligned}$$

This is smaller than $C\mu(A)/N$ as desired, since $\sum_{i,j\leq N}\rho_{|i-j|} \leq N\sum_{k\geq 0}\rho_k = CN$. Once this is done, it suffices to copy (without any change) the proof of Theorem 1. □

7.2. **Markov chains.** Here we consider a $\mathbb{R}^d$-valued Markov chain $(X_n)_{n\geq 1}$ with transition kernel P and initial distribution $\nu \in \mathcal{P}(\mathbb{R}^d)$ and we set $\mu_N := N^{-1}\sum_1^N \delta_{X_n}$. We assume that it admits a unique invariant probability measure π and the following L^2-decay property (usually related to a Poincaré inequality)

$$\forall\, n \geq 1,\ \forall\, f \in L^2(\pi), \quad \|P^n f - \pi(f)\|_{L^2(\pi)} \leq C\rho_n \|f - \pi(f)\|_{L^2(\pi)} \tag{8}$$

for some sequence $(\rho_n)_{n\geq 1}$ decreasing to 0.

Theorem 15. *Let $p \geq 1$, $d \geq 1$ and $r > 2$ be fixed. Assume that our Markov chain $(X_n)_{n\geq 0}$ satisfies (8) with a sequence $(\rho_n)_{n\geq 1}$ satisfying $\sum_{n\geq 1}\rho_n < \infty$. Assume also that the initial distribution ν is absolutely continuous with respect to π and satisfies $\|d\nu/d\pi\|_{L^r(\pi)} < \infty$. Assume finally that $M_q(\pi) < \infty$ for some $q > pr/(r-1)$. Setting $q_r := q(r-1)/r$ and $d_r = d(r+1)/r$, there is a constant C such that for all $N \geq 1$,*

$$\mathbb{E}_\nu\left(\mathcal{T}_p(\mu_N, \pi)\right) \leq C\begin{cases} N^{-1/2} + N^{-(q_r-p)/q_r} & \text{if } p > d_r/2r \text{ and } q_r \neq 2p,\\ N^{-1/2}\log(1+N) + N^{-(q_r-p)/q_r} & \text{if } p = d_r/2r \text{ and } q_r \neq 2p,\\ N^{-p/d} + N^{-(q_r-p)/q_r} & \text{if } p \in (0, d_r/2) \text{ and } q_r \neq d_r/(d_r-p).\end{cases}$$

Once again, we might adapt the proof to get a complete picture corresponding to other decay than L^2-L^2 and to slower mixing rates $(\rho_n)_{n\geq 1}$.

Proof. We only have to show that for any $\ell \geq 0$, any $n \geq 0$,

$$\begin{aligned}\Delta_{n,\ell}^N :=& \sum_{F\in\mathcal{P}_\ell}\mathbb{E}_\nu\left(|\mu_N(2^nF \cap B_n) - \pi(2^nF \cap B_n)|\right)\\ \leq& C\min\left\{(\pi(B_n))^{(r-1)/r}, [2^{d_r\ell}(\pi(B_n))^{(r-1)/r}/N]^{1/2}\right\}.\end{aligned}$$

Since $M_q(\pi) < \infty$ (whence $\pi(B_n) \le C2^{-qn}$), we will deduce that

$$\Delta_{n,\ell}^N \le C \min\left\{2^{-q_r n}, 2^{d_r \ell/2}(2^{-q_r n}/N)^{1/2}\right\}.$$

Then the rest of the proof is exactly the same as that of Theorem 1, replacing everywhere q and d by q_r and d_r.

We first check that $\Delta_{n,\ell}^N \le C(\pi(B_n))^{(r-1)/r}$. Using that $\|d\nu/d\pi\|_{L^r(\pi)} < \infty$, we write

$$\mathbb{E}_\nu(\mu_N(B_n)) = \frac{1}{N}\sum_{i=1}^N \mathbb{E}_\pi\left[\frac{d\nu}{d\pi}(X_0)1_{\{X_i\in B_n\}}\right] \le \|d\nu/d\pi\|_{L^r(\pi)}\pi(B_n)^{(r-1)/r}.$$

We next consider a Borel subset A of $\mathbb{R}^d$ and check that

$$\mathbb{E}_\nu(|\mu_N(A) - \pi(A)|) \le C(\pi(A))^{(r-1)/(2r)}N^{-1/2}.$$

To do so, as is usual when working with Markov chains or covariance properties (see [7]), we introduce $f = 1_A - \pi(A)$ and write

$$\mathbb{E}_\nu(|\mu_N(A) - \pi(A)|) = \frac{1}{N}\mathbb{E}_\nu\left(\left|\sum_{i=1}^N f(X_i)\right|\right) \le \frac{1}{N}\left(\sum_{i,j=1}^N \mathbb{E}_\nu(f(X_i)f(X_j))\right)^{1/2}.$$

For $j \ge i$, it holds that

$$\mathbb{E}_\nu(f(X_i)f(X_j)) = \mathbb{E}_\nu[f(X_i)P^{j-i}f(X_i)] = \mathbb{E}_\pi\left[\frac{d\nu}{d\pi}(X_0)f(X_i).P^{j-i}f(X_i)\right].$$

Using the Hölder inequality (recall that $\|d\nu/d\pi\|_{L^r(\pi)} < \infty$ with $r > 2$) and (8), we get

$$\mathbb{E}_\nu(f(X_i)f(X_j)) \le \|d\nu/d\pi\|_{L^r(\pi)}\|f\|_{L^{2r/(r-2)}(\pi)}\|P^{j-i}f\|_{L^2(\pi)} \le C\rho_{j-i}\|f\|_{L^{2r/(r-2)}(\pi)}\|f\|_{L^2(\pi)}.$$

But for $s > 1$, $\|f\|_{L^s(\pi)} \le C_s(\pi(A) + (\pi(A))^s)^{1/s} \le C_s(\pi(A))^{1/s}$, we find $\mathbb{E}_\nu(f(X_i)f(X_j)) \le C\rho_{j-i}(\pi(A))^{(r-1)/r}$ and thus

$$\mathbb{E}_\nu(|\mu_N(F) - \pi(F)|) \le \frac{C}{N}\left(\sum_{i,j=1}^N \rho_{|i-j|}(\pi(F))^{(r-1)/2r}\right)^{1/2} \le C(\pi(F))^{(r-1)/(2r)}N^{-1/2}$$

as desired. We used that $\sum_{i,j=1}^N \rho_{|i-j|} \le CN$.

We can finally conclude that

$$\Delta_{n,\ell}^N \le CN^{-1/2}\sum_{F\in\mathcal{P}_\ell}(\pi(2^nF\cap B_n))^{(r-1)/(2r)} \le CN^{-1/2}2^{d_r\ell/2}(\pi(B_n))^{(r-1)/(2r)}$$

by the Hölder inequality (and because $\#\mathcal{P}_\ell = 2^{d\ell}$), where $d_r = d(r+1)/r$ as in the statement. □

7.3. **Mc Kean-Vlasov particles systems.** Particle approximation of nonlinear equations has attracted a lot of attention in the past thirty years. We will focus here on the following $\mathbb{R}^d$-valued nonlinear S.D.E.

$$dX_t = \sqrt{2}dB_t - \nabla V(X_t)dt - \nabla W * u_t(X_t)dt, \qquad X_0 = x$$

where $u_t = Law(X_t)$ and (B_t) is and $\mathbb{R}^d$-Brownian motion. This is a probabilistic representation of the so-called Mc Kean-Vlasov equation, which has been studied in particular by Carillo-Mac Cann-Villani [12], Malrieu [28] and Cattiaux-Guillin-Malrieu [13] to which we refer for further motivations and existence and uniqueness of solutions. We will mainly consider here the case

where V and W are convex (and if $V = 0$ the center of mass is fixed) and W is even. To fix the ideas, let us consider only two cases:

(a) $Hess\, V \geq \beta Id > 0$, $Hess\, W \geq 0$.

(b) $V(x) = |x|^\alpha$ for $\alpha > 2$, $Hess\, W \geq 0$.

The particle system introduced to approximate the nonlinear equation is the following. Let $(B_t^i)_{t\geq 0}$ be N independent Brownian motion. For $i = 1, \ldots, N$, set $X_0^{i,N} = x$ and

$$dX_t^{i,N} = \sqrt{2} dB_t^i - \nabla V(X_t^{i,N})dt - \frac{1}{N}\sum_j \nabla W(X_t^{i,N} - X_t^{j,N})dt.$$

Usual propagation of chaos property is usually concerned with control of

$$\mathcal{W}_2(Law(X_t^{1,N}), u_t)$$

uniformly (or not) in time. It is however very natural to consider rather a control of

$$\mathcal{W}_2(\hat{u}_t^N, u_t)$$

where $\hat{u}_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}}$, as in Bolley-Guillin-Villani [11].

To do so, and inspired by the usual proof of propagation of chaos, let us consider nonlinear independent particles

$$dX_t^i = \sqrt{2} dB_t^i - \nabla V(X_t^i)dt - \nabla W * u_t(X_t^i)dt, \qquad X_0^i = x$$

(driven by the same Brownian motions as the particle system) and the corresponding empirical measure $u_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$. We then have

$$\mathcal{W}_2(\hat{u}_t^N, u_t) \leq \mathcal{W}_2(\hat{u}_t^N, u_t^N) + \mathcal{W}_2(u_t^N, u_t).$$

Then following [28] in case (a) and [13] in case (b), one easily gets (for some time-independent constant C)

$$\mathbb{E}(\mathcal{W}_2^2(\hat{u}_t^N, u_t^N)) \leq \frac{1}{N}\mathbb{E}\left(\sum_{i=1}^N |X_t^{i,N} - X_t^i|^2\right) \leq C\alpha(N)$$

where $\alpha(n) = N^{-1}$ in case (a) and $\alpha(N) = N^{-1/(\alpha-1)}$ in case (b). It is not hard to prove here that the nonlinear particles have infinitely many moments (uniformly in time) so that combining Theorem 1 with the previous estimates gives

$$\sup_{t\geq 0} \mathbb{E}(\mathcal{W}_2(\hat{u}_t^N, u_t)) \leq C(\alpha(N) + \beta(N))$$

where $\beta(N) = N^{-1/2}$ if $d = 1$, $\beta(N) = N^{-1/2}\log(1 + N)$ if $d = 2$ and $\beta(N) = N^{-1/d}$ if $d \geq 3$.

References

[1] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* 13 (2008), 1000–1034.

[2] M. Ajtai, J. Komlós, G. Tusnády. On optimal matchings, *Combinatorica*, 4 (1984), 259–264.

[3] F. Barthe, C. Bordenave. Combinatorial optimization over two random point sets. *Séminaire de Probabilités* XLV, Pages 483–535, Lecture Notes in Mathematics 2078, Springer, 2013.

[4] G. Bennett. Probability Inequalities for the Sum of Independent Random Variables. *J. Amer. Statist. Assoc.* 57 (1962), 33–45.

[5] G. Biau, L. Devroye, G. Lugosi. On the performance of clustering in Hilbert spaces. *IEEE Trans. Information Theory.* 54 (2008), 781–790.

[6] E. Boissard. Simple Bounds for the Convergence of Empirical and Occupation Measures in 1-Wasserstein Distance. *Elect. Journ. Prob.* 16 (2011), 2296–2333.

[7] E. Boissard, T. Le Gouic. On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. arXiv:1105.5263.
[8] A.A. Borovkov. Estimates for the distribution of sums and maxima of sums of random variables when the Cramér condition is not satisfied. *Siberian Math. J.* 41 (2000), 811–848.
[9] R.C. Bradley. A central limit theorem for stationary ρ-mixing sequences with infinite variance. *Ann. Prob.* 16 (1988), 313–332.
[10] R.C. Bradly. *Introduction to strong mixing conditions Vol 1,2,3.* Kendrick Press, Heber City, 2007.
[11] F. Bolley, A. Guillin, C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields* 137 (2007), 541–593.
[12] J-A. Carrillo, R. Mac Cann, C. Villani. Kinetic equilibration rates fro granular media and related equations: entropy dissipation and mass transportation. *Rev. Mat. Iberoamericana* 19 (2003), 971–1018.
[13] P. Cattiaux, A. Guillin, F. Malrieu. Probabilistic approach for granular media equations in the non uniformly convex case. *Probab. Theory and Related Fields* 140 (2008), 19–40.
[14] S. Delattre, S. Graf, H. Luschgy, G. Pagès. Quantization of probability distributions under norm-based distortion measures. *Stat. Decisions* 22 (2004), 261–282.
[15] S. Dereich. Asymptotic formulae for coding problems and intermediate optimization problems: a review. In *Trends in Stochastic Analysis*. Cambridge University Press. (2009), 187–232.
[16] S. Dereich, M. Scheutzow, R. Schottstedt, Constructive quantization: approximation by empirical measures. *Ann. Inst. Henri Poincar Probab. Stat.* 49 (2013), 1183–1203.
[17] L. Devroye, G. Lugosi. *Combinatorial methods in density estimation.* Springer, 2001.
[18] H. Djellout, A. Guillin, L. Wu. Transportation Cost-information inequalities and applications ro random dynamical systems and diffusions. *Ann. Prob.* 32 (2004), 2702–2732.
[19] V. Dobrić, J. E. Yukich. Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* 8 (1995), 97–118.
[20] P. Doukhan. *Mixing: properties and examples*. Springer-Verlag, New-York 1995.
[21] R.M. Dudley. Central limit theorems for empirical measures. *Ann. Probab.* 6 (1978), 899–929.
[22] N. Fournier, S. Mischler. Rate of convergence of the Nanbu particle system for hard potentials. arXiv:1302.5810.
[23] D.H. Fuk, S.V. and Nagaev. Probability inequalities for sums of independent random variables. *Theory Probab. Appl.* 16 (1971), 660–675.
[24] N. Gozlan. Integral criteria for transportation cost inequalities. *Elec. Com. Prob.* 11 (2006), 64–77.
[25] J. Horowitz, R.L. Karandikar. Mean rates of convergence of empirical measures in the Wasserstein metric. *J. Comput. Appl. Math.* 55 (1994), 261–273.
[26] T. Laloë. L_1-quantization and clustering in Banach spaces. *Math. Meth. Stat.* 19 (2009), 136–150.
[27] M. Ledoux. *The concentration of measure phenomenon*. Mathematical Surveys and Monographs 89, American Math. Society, Providence, 2001.
[28] F. Malrieu. Convergence to equlibrium for granular media equations. *Ann. Appl. Prob.* 13 (2003), 540–560.
[29] P. Massart. Concentration Inequalities and Model Selection: *Ecole d'Été de Probabilités de Saint-Flour XXXIII. Springer.* 2003.
[30] F. Merlevède, M. Peligrad. Rosenthal-type inequalities for the maximum of partial sums of stationary processes and examples. *Ann. Prob.* 41 (2013), 914–960.
[31] F. Merlevède, M. Peligrad, E. Rio. A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory and Related Fields* 151 (2011), 435–474.
[32] S. Mischler, C. Mouhot, Kac's programm in kinetic theory, *Invent. Math.* 193 (2013), 1–147.
[33] G. Pagès, B. Wilbertz. Optimal Delaunay and Voronoi quantization schemes for pricing american style options. In *Numerical methods in Finance,* R. Carmona, P. Hu, P. Del Moral, N. Oudjane eds, Springer, 171-217, 2012.
[34] D. Perrin. Une variante de la formule de Stirling. Available at www.math.u-psud.fr/∼perrin/CAPES /analyse/Suites/Stirling.pdf.
[35] S.T. Rachev, L. Rüschendorf. *Mass transportation problems. Vol. I. and II.* Probability and its Applications. Springer-Verlag, 1998.
[36] G. Roberts, J-S. Rosenthal. Shift-coupling and convergence rates of ergodic averages. *Comm. Stat. Stoch. Models*, 13 (1996), 147–165.
[37] E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*, Mathématiques et Applications 31, Springer, Paris, 2000.
[38] M. Talagrand. Matching random samples in many dimensions. *Ann. Appl. Probab.,* 2 (1992), 846–856.
[39] M. Talagrand. The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *Ann. Prob.,* 22 (1994), 919–959.

[40] A. Van der Vaart, J.A. Wellner. *Weak convergence of empirical processes.* Springer-Verlag, 1996.
[41] C. Villani, *Topics in optimal transportation.* Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003.

Nicolas Fournier, Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, UPMC, Case 188, 4 pl. Jussieu, F-75252 Paris Cedex 5, France, E-mail: nicolas.fournier@upmc.fr

Arnaud Guillin, Laboratoire de Mathématiques, UMR 6620, Université Blaise Pascal, Av. des landais, 63177 Aubiere cedex, E-mail: arnaud.guillin@math.univ-bpclermont.fr