Some results on the optimal matching problem for the Jacobi model

Jie-Xiang Zhu

Fudan University, People’s Republic of China,
and University of Toulouse, France

Abstract

We establish some exact asymptotic results for a matching problem with respect to a family of beta distributions. Let $X_1, \ldots, X_n$ be independent random variables with common distribution the symmetric Jacobi measure $d\mu(x) = C_d(1-x^2)^{\frac{d}{2}-1}dx$ with dimension $d \geq 1$ on $[-1, 1]$, and let $\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$ be the associated empirical measure. We show that

$$\lim_{n \to \infty} nE[W_2^2(\mu_n, \mu)] = \sum_{k=1}^{\infty} \frac{1}{k(k+d-1)},$$

where $W_2$ is the quadratic Kantorovich distance with respect to the intrinsic cost $\rho(x, y) = |\arccos(x) - \arccos(y)|$, $(x, y) \in [-1, 1]^2$, associated to the model. When $\mu$ is the product of two Jacobi measures with dimensions $d$ and $d'$ respectively, then

$$E[W_2^2(\mu^n, \mu)] \approx \frac{\log n}{n}.$$ 

In the particular case $d = d' = 1$ (corresponding to the product of arcsine laws),

$$\lim_{n \to \infty} \frac{n}{\log n} E[W_2^2(\mu^n, \mu)] = \frac{\pi}{4}.$$ 

Similar results do hold for non-symmetric Jacobi distributions. The proofs are based on the recent PDE and mass transportation approach developed by L. Ambrosio, F. Stra and D. Trevisan.

Key words and phrases: Empirical measure, Kantorovich distance, Markov triple, Jacobi semigroup

Mathematics Subject Classification (2010): 60D05, 60F25, 60H15, 49J55, 58J35

1Research supported by China Scholarship Council (No. 201806100097) and partially supported by NSF of China (Grants No. 11625102 and No. 11571077)
1 Introduction and main results

Optimal matching problems are very classical in computer science, physics and mathematics, and have been widely investigated from different viewpoints. One general formulation, of particular interest in probability theory and mathematical statistics, is expressed in terms of Kantorovich (Wasserstein) distances. Let \((M, \rho)\) be a (complete separable) metric space. Given \(p \geq 1\), the Kantorovich distance (see e.g. [18]) between two probability measures \(\nu\) and \(\mu\) on the Borel sets of \(M\) with a finite \(p\)-th moment is defined by

\[
W_p(\nu, \mu) = \inf \left( \int_{M \times M} \rho^p(x, y) d\pi(x, y) \right)^{\frac{1}{p}}
\]

where the infimum is taken over all couplings \(\pi\) on \(M \times M\) with respective marginals \(\nu\) and \(\mu\).

In this work, we mainly focus on the quadratic one \((p = 2)\).

Given then a probability measure \(\mu\) on the Borel sets of \((M, \rho)\), denote by \(X_1, \ldots, X_n, \ldots\) independent random variables with common distribution \(\mu\), and let \(\mu^n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}, n \geq 1\), be the empirical measure on the sample \((X_1, \ldots, X_n)\). One version of the optimal matching problem, at a first order, is to estimate the order of growth, and possibly the renormalized limit, of

\[
\mathbb{E} \left[ W_2^2(\mu^n, \mu) \right]
\]
as \(n \to \infty\).

A reasonable answer to this question involves conditions on the nature of the metric space \((M, \rho)\) and on the distribution \(\mu\). In the past three decades, several authors made contributions to this problem, and we mention here a few relevant ones. One first well-known result is the Ajtai-Komlós-Tusnády theorem [1] expressing that for \(\mu\) the uniform distribution on the unit cube \([0, 1]^k\) in \(\mathbb{R}^k, k \geq 1\),

\[
\mathbb{E} \left[ W_2^p(\mu^n, \mu) \right] \approx \begin{cases} 
\frac{1}{n^{p/2}}, & k = 1, \\
\left( \frac{\log n}{n} \right)^{p/2}, & k = 2, \\
\frac{1}{n^{p/k}}, & k \geq 3.
\end{cases}
\]

(1.1)

In this paper we use the notation \(A \lesssim B\) if there exists a constant \(C > 0\) (depending on the underlying model but not on \(n\)) such that \(A \leq CB\). We also write \(A \approx B\), if \(A \lesssim B\) and \(B \lesssim A\). (It will also be implicit throughout the paper that \(n \geq 1\) is large enough so that all the relevant quantities involving \(n\) are well defined.) This first result was completed, refined and extended to more general situations and examples in several works. We refer to the monograph [17] by M. Talagrand for an account and relevant references on this result and its developments (see also [19, 13]). Additional recent advances include [8, 6, 10, 12, 14] to cite a few. For an extensive discussion of the one-dimensional optimal matching problem, see [7].

It should be emphasized that in (1.1) the most delicate case is \(k = 2\) since the rate does not reflect the uniform spacings in \([0, 1]^2\) and both local and global discrepancies have to be combined (cf. [17]). This is also the case for \(k = 1\), but here it is classical that the optimal
matching is achieved by monotone rearrangement. In particular, order statistics may be used to produce the exact expression

$$\mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] = \frac{1}{6n}$$

for \( \mu \) the uniform distribution on \([0, 1]\) (cf. e.g. [7]). Order statistics can also be applied to estimate the mean Kantorovich distance for log-concave distributions on the real line (see Section 6 of [7]). For the purpose of this work, let us observe in particular that if

$$d\mu(x) = C_d(1-x^2)^{\frac{d}{2}-1}dx, \ d \geq 1,$$

is the symmetric Jacobi (beta) distribution on \([-1, 1]\) (for a detailed definition, see Section 2), then

$$\mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] \approx \frac{1}{n}.$$

Recently, a major achievement was achieved by L. Ambrosio, F. Stra and D. Trevisan [3] who answered rigorously a conjecture put forward in [9]. They indeed provided the exact value of the limiting behaviour of \( \mathbb{E}[W_2^2(\mu^n, \mu)] \) in the Ajtai-Komlós-Tusnády theorem for \( k = 2 \), namely

$$\lim_{n \to \infty} \frac{n}{\log n} \mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] = \frac{1}{4\pi}.$$ (1.4)

This is actually one of the very few explicit limits known in this setting, even in dimension one actually (some unknown limits are achieved via subadditivity in higher dimensions in [11, 8, 6, 10]).

More generally, based on the PDE ansatz of [9], the authors of [3] proved that the same limit (1.4) holds true for \( \mu \) the normalized uniform distribution on a compact 2-dimensional Riemannian manifold \((M, g)\) without boundary. The validity of the PDE ansatz of [9] is based on a fine analysis of a linearized Monge-Ampère equation and mass transportation tools, which highly depend on properties of the associated heat kernel \( p_t \). It is expected that the method can be further developed for the case \( d \geq 3 \), but this is mostly conjectural at this point.

Relying on the methodology emphasized in [3], the purpose of this work is to provide some further examples, in (topological) dimensions 1 and 2, where the exact asymptotic behaviour of \( \mathbb{E}[W_2^2(\mu^n, \mu)] \) may be achieved. These examples are the aforementioned symmetric Jacobi measures

$$d\mu(x) = C_d(1-x^2)^{\frac{d}{2}-1}dx \text{ on } [-1, 1],$$

(also called beta laws). Note that when \( d = 2 \), this is the uniform measure on \([-1, 1]\) while when \( d = 1 \), \( \mu \) is the famous arcsine law. However, to fully develop the PDE and optimal transport approach, the interval \([-1, 1]\), and thus the Kantorovich distance \( W_2 \), will be equipped with the intrinsic distance \( \rho(x, y) = |\arccos(x) - \arccos(y)|, \ (x, y) \in [-1, 1]^2 \), of the underlying Jacobi model (see Section 2 below). As indeed described below, the intrinsic distance is inherited from the geodesic metric on the sphere (Jacobi measures are projections of the spherical uniform measure on a diameter when \( d \) is an integer), and compares to the Euclidean one, but for the
limit to hold via the PDE approach, the intrinsic distance is in force. In particular, Theorem 1 below for \( d = 2 \) (for which the limit is 1) is not equivalent to (1.2).

In this setting, the main results of this work read as follows.

**Theorem 1.** If \( \mu \) is the symmetric Jacobi measure with \( d \geq 1 \), then

\[
\lim_{n \to \infty} n \mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] = \sum_{k=1}^{\infty} \frac{1}{k(k + d - 1)}. 
\]

Note that when \( d \geq 2 \) is an integer,

\[
\sum_{k=1}^{\infty} \frac{1}{k(k + d - 1)} = \frac{1}{d-1} \sum_{k=1}^{d-1} \frac{1}{k} 
\]

which is of the order \( \frac{\log d}{d} \) as \( d \to \infty \).

**Theorem 2.** If \( \mu \) is the product measure of two symmetric Jacobi measures with dimensions \( d, d' \geq 1 \) respectively, then

\[
\mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] \approx \frac{\log n}{n} 
\]

(where \( W_2 \) is defined via the product distance of the intrinsic metric \( \rho \)).

**Theorem 3.** If \( \mu \) is the product measure of two symmetric Jacobi measures with dimensions \( d = d' = 1 \) (i.e. arcsine laws), then

\[
\lim_{n \to \infty} \frac{n}{\log n} \mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] = \frac{\pi}{4}. 
\]

It is likely that Theorem 3 holds true when \( d + d' = 2 \), but the probabilistic tail estimates require \( d + d' \leq 2 \) while at the same time \( d, d' \geq 1 \) is requested for the heat kernel bounds. If \( \mu \) is the product measure of \( k \geq 3 \) symmetric Jacobi measures with dimensions \( d_j \geq 1, j = 1, \ldots, k \) respectively, Theorem 2 extends in the form

\[
\mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] \approx \frac{1}{n^{2/k}} 
\]

in accordance with (1.1).

The results may also be formulated in the bipartite form, corresponding really to the optimal matching problem. If \( X_1, \ldots, X_n \) and \( Y_1, \ldots, Y_n \) independent random variables with common distribution \( \mu \), let \( \mu^n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \) and \( \nu^n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i} \) be the corresponding empirical measures. Then, if \( \mu \) is the symmetric Jacobi measure with \( d \geq 1 \),

\[
\lim_{n \to \infty} n \mathbb{E} \left[ W_2^2(\mu^n, \nu^n) \right] = \sum_{k=1}^{\infty} \frac{2}{k(k + d - 1)} 
\]

and if \( \mu \) is the product measure of two symmetric Jacobi measures with dimensions \( d = d' = 1 \),

\[
\lim_{n \to \infty} \frac{n}{\log n} \mathbb{E} \left[ W_2^2(\mu^n, \nu^n) \right] = \frac{\pi}{2}. 
\]
As discussed during the proofs below, the main conclusions above actually also apply to non-symmetric Jacobi distributions

\[ d\mu(x) = C_{\alpha,\beta}(1-x)^{\alpha-1}(1+x)^{\beta-1}dx \]

with \( \alpha, \beta \geq \frac{1}{2} \). In this case, Theorem 1 takes the same form with \( d = \alpha + \beta \). In Theorem 2, \( \mu \) may be taken as the product of two such non-symmetric Jacobi measures with \( \alpha, \beta \geq \frac{1}{2}, \alpha', \beta' \geq \frac{1}{2} \). Theorem 3 is unchanged since the restrictions \( \alpha, \beta \geq \frac{1}{2} \) imply that we are necessarily in the symmetric case.

As mentioned above, the symmetric Jacobi model when \( d \) is an integer can be seen as a quotient of the \( d \)-dimensional sphere \( S^d \), and as a Markov Triple (see [5] and Section 2 below) satisfies a curvature-dimension condition \( CD(d-1,d) \) as the sphere. The scaling and limit in Theorem 1 however do not reflect this property. This is due to the fact that the PDE approach is a tradeoff between the (small time) heat kernel behaviour

\[ p_s(x,x) \leq Cs^{-\frac{d}{2}}, \quad 0 < s \leq 1, \]

which is alike \( d \)-dimensional for the Jacobi semigroup, and the trace estimate

\[ \int_{-1}^{1} p_s(x,x)d\mu(x) \leq Cs^{-\frac{1}{2}}, \quad 0 < s \leq 1, \tag{1.8} \]

which actually reflects a one-dimensional feature. While the heat kernel and trace small time asymptotics are of the same order on a compact Riemannian manifold, the Jacobi model is intermediate in this respect.

This paper is therefore a contribution towards further examples where exact limiting behaviours in the optimal matching problem can be achieved, following the conjectures in [9]. At this point very few results are available on this rather delicate problem. The methodology developed to this task is the PDE and optimal transport approach of [3], that is followed quite closely. In particular, some intermediate results and arguments of [3] may be immediately adapted to the Jacobi setting, so that we only quote them without detailed proofs. In addition, the investigation also benefits from the recent contribution [2], in particular at the level of a strengthened regularization step. At the same time, the full organization of the proof presented here produces several simplified arguments and steps, which may be used in return to streamline the proofs in [3, 2]. In particular, Riesz transform bounds are not used.

Turning to the content of the paper, Section 2 presents the fundamental Markov Triple structure from [3], the example of the Jacobi model, and several properties of the associated (Jacobi) semigroup. From Section 3 to Section 5, the main technical tools are introduced, including density fluctuation bounds, energy estimates and a refined contractivity estimate, following [3] and [2]. In Section 6, we prove Theorem 1 by treating upper and lower bounds respectively, according to the same outline as in [3]. The proof of Theorem 3 is similar. In Section 7, we prove Theorem 2 by means of Proposition 3 of [15] and a simple comparison argument.


2 Markov Triple, Jacobi model and properties of the semigroup

This section is a brief exposition of the Jacobi model within the setting of Markov Triples as put forward in [5] (see in particular therein Chapters 1 to 3). To start with, we briefly recall the basic definition of a Markov Triple \((E, \mu, \Gamma)\) referring to [5] for the complete picture.

On a measurable space \((E, \mathcal{F})\), a Markov semigroup \((P_t)_{t \geq 0}\) is a family of operators defined on some subset of real-valued measurable functions on \((E, \mathcal{F})\) satisfying the positivity and mass conservation properties. A \(\sigma\)-finite measure \(\mu\) is said to be invariant for the semigroup \((P_t)_{t \geq 0}\) if for every bounded positive measurable function \(f\) on \(E\) and \(t \geq 0\),

\[
\int_E P_t f \, d\mu = \int_E f \, d\mu.
\]

Denote the infinitesimal generator of \((P_t)_{t \geq 0}\) by \(L\). The associated carré du champ operator \(\Gamma\) is defined by

\[
\Gamma(f, g) = \frac{1}{2} \left[ L(fg) - fL(g) - gL(f) \right],
\]

where \(f\) and \(g\) are elements of a suitable algebra \(\mathcal{A}\) of (smooth) functions. The integration by parts formula expresses in this context that

\[
\int_E gL(f) \, d\mu = -\int_E \Gamma(f, g) \, d\mu.
\]

In sequel, \(\Gamma(f, f)\) is abbreviated as \(\Gamma(f)\). The family \((E, \mu, \Gamma)\) is then called a Markov Triple.

The Markov Triple structure induces in addition a notion of curvature-dimension condition \(CD(K, N)\), \(K \in \mathbb{R}, N \geq 1\), which extends the Ricci curvature lower bounds in Riemannian manifolds.

As an example, let \(E\) be \(\mathbb{R}^k\), and let \(\mu\) be Lebesgue measure. For the carré du champ operator \(\Gamma(f, g) = \sum_{i=1}^k \partial_i f \partial_i g\) on smooth functions \(f, g : \mathbb{R}^k \to \mathbb{R}\), \((\mathbb{R}^k, \mu, \Gamma)\) is a Markov Triple, the generator \(L\) being the standard Laplace operator \(\Delta\) on \(\mathbb{R}^k\). More generally, a (smooth complete) Riemannian manifold with the Riemannian volume element and the Riemannian length squared \(\Gamma(f) = |\nabla f|^2\) of the gradient of a smooth function \(f\) defines similarly a Markov Triple. An \(k\)-dimensional manifold \((M, g)\) with Ricci curvature bounded below by \(K\) satisfies the curvature-dimension condition \(CD(K, k)\).

The Markov Triple definition \((E, \mu, \Gamma)\) also induces an intrinsic metric structure on \(E\) by means of the distance

\[
\rho(x, y) = \text{esssup}[f(x) - f(y)], \quad (x, y) \in E \times E,
\]

where the essential supremum is taken over all \(1\)-Lipschitz functions \(f\) on \(E\) such that \(\Gamma(f) \leq 1\) \(\mu\)-almost everywhere. For example, on the preceding example \((\mathbb{R}^k, \mu, \Gamma)\), \(\rho\) is simply the Euclidean metric on \(\mathbb{R}^k\). When dealing with the Kantorovich distance \(W_2\) on a Markov Triple \((E, \mu, \Gamma)\), the underlying distance will be the intrinsic distance \(\rho\) as just defined.
Markov Triples are extensively discussed and studied in the monograph [5]. The Jacobi model is one family of examples. On the interval \( E = [-1, 1] \), the Jacobi measure with parameters \( \alpha, \beta > 0 \) is defined by

\[
d\mu_{\alpha,\beta}(x) = C_{\alpha,\beta}(1-x)^{\alpha-1}(1+x)^{\beta-1}dx
\]

where \( C_{\alpha,\beta} > 0 \) is a normalization constant and \( dx \) is Lebesgue measure. (While for \( 0 < \alpha < 1 \) or \( 0 < \beta < 1 \), \( \mu_{\alpha,\beta} \) is formally only defined on \((-1,1)\), the relevant Markov Triple properties emphasized below do hold similarly on the closed interval \( E = [-1, 1] \) that we thus keep for simplicity in the exposition.) Throughout the investigation, it will be assumed that \( \alpha, \beta \geq \frac{1}{2} \) to ensure the necessary analytic properties towards the main conclusions – specifically the ultracontractivity property (UC), see below.

The associated Jacobi operator is acting on smooth functions \( f \) as

\[
L_{\alpha,\beta}f = (1 - x^2)f'' - [(\alpha + \beta)x + \alpha - \beta]f',
\]

and the carré du champ operator is given by \( \Gamma(f) = (1 - x^2)f'^2 \). The family \(([[-1,1], \mu_{\alpha,\beta}, \Gamma])\) is therefore a Markov Triple, called the Jacobi model (of parameters \( \alpha, \beta \)). It is easy to check that the intrinsic distance \( \rho \) between \( x \) and \( y \in [-1,1] \) is given by

\[
\rho(x,y) = \left| \int_x^y \frac{1}{\sqrt{1-u^2}}du \right| = |\arccos(x) - \arccos(y)|.
\]

When \( \alpha = \beta = \frac{d}{2} \) for some \( d \geq 1 \), we speak of the symmetric Jacobi model of dimension \( d \), with thus the symmetric Jacobi distribution \( d\mu_{\frac{d}{2}, \frac{d}{2}}(x) = C_d(1-x^2)^{\frac{d}{2}-1}dx \). The “dimension” terminology has a geometric flavour. Indeed, when \( d \geq 1 \) is an integer, the symmetric Jacobi model can be seen as a quotient of the sphere \( S^d \) in \( \mathbb{R}^{d+1} \) via the projection operator \( T : x = (x_1, \ldots, x_{d+1}) \in S^d \mapsto x_1 \in [-1,1] \). If \( X \) is distributed uniformly on \( S^d \), then the law of \( T(X) \in [-1,1] \) is precisely \( \mu_{\frac{d}{2}, \frac{d}{2}} \). Moreover, the Jacobi operator \( L_{\frac{d}{2}, \frac{d}{2}} \) can be seen as an image of the Laplace-Beltrami operator \( \Delta_{S^d} \) on \( S^d \) via its action on functions depending on only one coordinate in \( \mathbb{R}^{d+1} \). According to this geometric description, the Markov Triple \(([[-1,1], \mu_{\frac{d}{2}, \frac{d}{2}}, \Gamma])\) satisfies the curvature-dimension condition \( CD(d-1,d) \) as the sphere \( S^d \), a property which extends to non-integer \( d \).

In Theorems 2 and 3, we also consider the product Markov Triple generated by two or more Jacobi models. On the square \( E = [-1,1]^2 \), the product measure \( \mu \) of two Jacobi measures with parameters \( \alpha, \beta > 0, \alpha', \beta' > 0 \) respectively, is defined by

\[
d\mu(x,y) = C(1-x)^{\alpha-1}(1+x)^{\beta-1}(1-y)^{\alpha'-1}(1+y)^{\beta'-1}dxdy
\]

where \( C > 0 \) is a normalization constant and \( dxdy \) is Lebesgue measure. The associated carré du champ operator is

\[
\Gamma(f)(x,y) = (1-x^2)(\partial_x f)^2 + (1-y^2)(\partial_y f)^2.
\]
Then the intrinsic distance $\rho$ between $x = (x_1, x_2)$ and $y = (y_1, y_2) \in [-1, 1]^2$ satisfies

$$\rho(x, y) = \sqrt{\rho_1^2(x_1, y_1) + \rho_2^2(x_2, y_2)},$$

where $\rho_i, i = 1, 2$, are the intrinsic distances of the respective Jacobi Triples. When $\alpha = \beta = \frac{d}{2}$ and $\alpha' = \beta' = \frac{d'}{2}$, the product Jacobi model is of curvature-dimension $CD(\min(d, d') - 1, d + d')$. Similar properties hold for the Markov triple generated by more than two Jacobi models.

To streamline the exposition and emphasize the main ideas, we concentrate on the symmetric Jacobi model throughout the proofs and the arguments. The necessary modifications to cover the non-symmetric case will be mentioned at the appropriate places. Accordingly, in the sequel, we drop for simplicity the subscripts $\frac{d}{2}$ in $\mu_{\frac{d}{2}, \frac{d}{2}}$, and omit the word “symmetric”.

Given then a Jacobi Triple $([-1, 1], \mu, \Gamma)$ (where $\mu = \mu_{\frac{d}{2}, \frac{d}{2}}$ with $d \geq 1$ fixed), denote by $(P_t)_{t \geq 0}$ the symmetric Markov semigroup with infinitesimal generator $L$. In following sections, we will investigate properties of the solution of the (Poisson) equation $Lf = u$. To this task, note that, formally, $(-L)^{-1}$ is described as

$$(-L)^{-1} = \int_0^\infty P_t dt,$$

acting on mean zero functions in the suitable domain. On the other hand, it is known that for the Jacobi model, the eigenvalues of $-L$ are given by the sequence $\lambda_k = k(k + d - 1)$, $k \in \mathbb{N}$. The corresponding eigenvectors $J_k$ are the Jacobi polynomials of degree $k$ which form an orthogonal basis of $L^2(\mu)$. In this spectral description, $(-L)^{-1}$ can then be represented by

$$(-L)^{-1}f = \sum_{k=0}^\infty \frac{1}{\lambda_k} f_k J_k,$$

where $f_k = \int_{-1}^1 f J_k d\mu$.

The operators $P_t$ of the Markov semigroup are represented by kernels

$$P_t f(x) = \int_{-1}^1 f(y)p_t(x, y)d\mu(y), \quad t > 0, \ x \in [-1, 1],$$

with respect to the invariant measure $\mu$, called the heat kernel. In the preceding spectral representation,

$$P_t f = \sum_{k=0}^\infty e^{-\lambda_k t} f_k J_k,$$

and the trace formula is expressed by

$$\int_{-1}^1 p_t(x, x)d\mu(x) = \sum_{k=0}^\infty e^{-t\lambda_k}.$$

The PDE and optimal transport approach developed in [3] is based on several quantitative properties of the Markov semigroup $(P_t)_{t \geq 0}$ and the associated heat kernel $p_t(x, y)$, which we gather in the following list.
There exist constants \( C > 0 \) and \( K \geq 0 \), possibly changing from line to line, such that:

(SG) Spectral gap: \( \|P_t f\|_2 \leq e^{-Ct}\|f\|_2 \) for any \( f \) with \( \int_{-1}^1 f d\mu = 0 \) and \( t \geq 0 \).

(UC) Ultracontractivity: \( p_t(x, y) \leq Ct^{-d/2} \) for any \( x, y \in [-1, 1] \) and \( 0 < t \leq 1 \).

(GE) Gradient estimate: \( \text{Lip}(p_t(x, \cdot)) \leq Ct^{-(d+1)/2} \) for any \( x \in [-1, 1] \) and \( 0 < t \leq 1 \).

(DR) Dispersion rate: \( \int_{-1}^1 \int_{-1}^1 \rho^2(x, y)p_t(x, y)d\mu(x)d\mu(y) \leq Ct, \ t > 0. \)

(SGB) Strong gradient bound: \( \sqrt{\Gamma(P_t f)} \leq e^{-Kt}P_t(\sqrt{\Gamma(f)}) \) for any smooth \( f \) and \( t \geq 0 \).

These properties are presented and detailed in [3] in the context of a compact Riemannian manifold (without boundary). We provide here the necessary arguments in case of the Jacobi model under investigation, referring to [5] for the relevant properties needed to this task.

The spectral gap condition (SG) may be seen as a consequence of the spectral decomposition. Namely, mean zero functions are orthogonal to constants, so that the exponential decay (SG) holds true with \( C = \lambda_1 > 0 \). The other four properties actually derive from the curvature-dimension condition \( CD(d-1, d) \) of the Jacobi model. Under this condition, a Sobolev inequality (of dimension \( d \)) holds, yielding equivalently the uniform heat kernel bound (UC). (The case \( d = 1 \) is a bit particular here since only \( CD(0, 1) \) is then available while the proof of the Sobolev inequality under a curvature-dimension condition requires positive curvature. We may nevertheless rely then on the direct, and more precise, heat kernel bounds developed in [16]. This reference is also used to justify the restriction \( d \geq 1 \) in this investigation since the heat kernel bounds do not seem to have been clearly put forward in the range \( 0 < d < 1 \).) The strong gradient bound (SGB), actually equivalent to \( CD(K, \infty) \), therefore holds for the Jacobi model with \( K = d - 1 \geq 0 \).

Some more care has to be taken with (GE) and (DR). First, a curvature condition \( CD(0, \infty) \) implies the local Poincaré inequality

\[
\Gamma(P_t f) \leq \frac{1}{2t} \left[ P_t (f^2) - (P_t f)^2 \right]
\]

for any (smooth) \( f \) and \( t > 0 \). For \( s > 0 \) and \( x \in [-1, 1] \) fixed, the choice of \( f = p_s(x, \cdot) \) together with the semigroup property imply that

\[
\|\Gamma(p_{t+s}(x, \cdot))\|_{\infty} \leq \frac{1}{2t} \|P_t(f^2)\|_{\infty} \leq \frac{1}{2t} \|P_t\|_{1 \to \infty} \int_{-1}^1 f^2 d\mu = \frac{1}{2t} \|P_t\|_{1 \to \infty} p_{2s}(x, x).
\]

Now (UC) tells us that \( \|P_t\|_{1 \to \infty} \) and \( p_{2s}(x, x) \) are bounded from above by \( \frac{C}{2^{d/2}} \) and \( \frac{C}{2^{d/2}} \) respectively, so that (GE) follows by choosing \( s = t \).

Next we turn to (DR). To this task we investigate, for any 1-Lipschitz function \( f \), the integral

\[
\int_{-1}^1 \int_{-1}^1 (f(x) - f(y))^2 p_t(x, y) d\mu(x) d\mu(y).
\]
By the semigroup properties,
\[
\int_{-1}^{1} \int_{-1}^{1} (f(x) - f(y))^2 p_t(x, y) d\mu(x) d\mu(y) \\
= 2 \left( \int_{-1}^{1} \int_{-1}^{1} f^2(x) p_t(x, y) d\mu(x) d\mu(y) - \int_{-1}^{1} \int_{-1}^{1} f(x) f(y) p_t(x, y) d\mu(x) d\mu(y) \right) \\
= 2 \left( \int_{-1}^{1} P_t(f^2)(y) d\mu(y) - \int_{-1}^{1} f(y) P_t(f)(y) d\mu(y) \right) \\
= 2 \left( \int_{-1}^{1} f^2(y) d\mu(x) d\mu(y) - \int_{-1}^{1} P_{\frac{t}{2}}(f^2)(y) d\mu(y) \right).
\]

Then we make use of \( \Gamma \)-calculus to obtain that
\[
\int_{-1}^{1} f^2(y) d\mu(x) d\mu(y) - \int_{-1}^{1} P_{\frac{t}{2}}(f^2)(y) d\mu(y) \\
= - \int_{0}^{\frac{t}{2}} \frac{d}{ds} \left( \int_{-1}^{1} P_s(f)^2(y) d\mu(y) \right) ds \\
= -2 \int_{0}^{\frac{t}{2}} \left( \int_{-1}^{1} P_s(f)(y) L P_s(f)(y) d\mu(y) \right) ds \\
= 2 \int_{0}^{\frac{t}{2}} \left( \int_{-1}^{1} \Gamma(P_s(f))(y) d\mu(y) \right) ds \\
\leq 2 \int_{0}^{\frac{t}{2}} \left( \int_{-1}^{1} e^{-2Ks} P_s(\Gamma(f))(y) d\mu(y) \right) ds \leq 2 \int_{0}^{\frac{t}{2}} e^{-2Ks} ds.
\]

The last line comes from (SGB) and the fact that \( \Gamma(f) \leq 1 \). Hence for all \( t > 0 \),
\[
\int_{-1}^{1} \int_{-1}^{1} (f(x) - f(y))^2 p_t(x, y) d\mu(x) d\mu(y) \leq \frac{2}{K} (1 - e^{-Kt}) \leq 2t. \quad (2.1)
\]

The preceding actually holds in a general context. But for the specific Jacobi model, we may choose \( f(x) = \text{arccos}(x) \) in (2.1) for which \( \Gamma(\text{arccos}(\cdot)) = 1 \). Since \( \rho(x, y) = |\text{arccos}(x) - \text{arccos}(y)| \), this is exactly (DR).

**Remark 4.** Using the sharp estimates on the Jacobi kernel \( p_t \) from \[16\], it is possible to reach a stronger version of (DR) in the sense that for some \( C > 0 \), for all \( x \in [-1, 1] \) and \( t > 0 \),
\[
\int_{-1}^{1} \rho^2(x, y) p_t(x, y) d\mu(y) \leq C t.
\]

As observed in \[3\], we record for further purposes that (SGB) and (SG) also imply (with the same proof as in \[3\]) that there exists a positive constant \( C \) such that for every \( f \in L^2(\mu) \) with \( Lf \in L^\infty(\mu) \),
\[
\| \sqrt{\Gamma(f)} \|_\infty \leq C \| Lf \|_\infty. \quad (2.2)
\]

Moreover, the preceding heat kernel properties are stable under product, and therefore do hold similarly for the product of two Jacobi Triples, the dimension \( d \) being replaced by the sum \( d + d' \) of the respective dimensions.
We quote finally the contraction property in the Kantorovich metric $W_2$

$$W_2(P^*_t \nu, P^*_t \mu) \leq e^{-Kt}W_2(\nu, \mu), \quad (2.3)$$

where $P^*_t$ is the heat semigroup acting (by duality) on measures, as a consequence again of a curvature-dimension condition $CD(K, \infty)$ (cf. [5]).

Before proceeding to the proof of the main results in the next sections, we mention here the corresponding properties in the non-symmetric case with parameters $\alpha, \beta \geq \frac{1}{2}$. Actually, all the above quantitative properties still hold for the non-symmetric Jacobi model with some minor modifications. Particularly, the heat kernel bounds developed in [16] imply that

$$p_t(x, x) \leq C t^{-\max\{\alpha, \beta\}}, \quad x \in [-1, 1], \quad 0 < t \leq 1. \quad (2.4)$$

This conclusion also follows from a suitable Sobolev inequality as developed in [4]. On the other hand, when $\alpha, \beta \geq \frac{1}{2}$, the curvature-dimension condition $CD(\frac{\alpha + \beta - 1}{2}, 2(\alpha + \beta) - 1)$ holds. Note that this condition does not coincide with $CD(d - 1, d)$ when $\alpha = \beta = \frac{d}{2}$ (it is weaker), but is good enough to ensure all the curvature lower bounds necessary for the preceding semigroup properties. Taking these properties for granted, the arguments developed below in the symmetric case extend similarly to the non-symmetric Jacobi model.

To conclude this section, it is meaningful to make a comparison between the main results emphasized in the introduction and what is known on the sphere, or the unit interval equipped with the Euclidean metric. Recall the projection operator

$$T : x = (x_1, \ldots, x_{d+1}) \in \mathbb{S}^d \mapsto x_1 \in [-1, 1]$$

and denote by $\mu_{\mathbb{S}^d}$ the uniform distribution on $\mathbb{S}^d$, and by $X$ a random variable with distribution $\mu_{\mathbb{S}^d}$. Then it holds true that, with the obvious notation,

$$W_2^n(\mu^n, \mu) \leq W_2^n(\mu_{\mathbb{S}^d}^n, \mu_{\mathbb{S}^d}).$$

To prove this assertion, notice that for any coupling $\pi$ on $\mathbb{S}^d \times \mathbb{S}^d$ with marginals $\mu_{\mathbb{S}^d}^n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ and $\mu_{\mathbb{S}^d}$, the push-forward of $\pi$ by $T \otimes T$ is a coupling on $[-1, 1] \times [-1, 1]$ with marginals $\mu^n = \frac{1}{n} \sum_{i=1}^n \delta_T(X_i)$ and $\mu$. Then, using the spherical coordinate system, it is easy to see that for all $x, y \in \mathbb{S}^d$,

$$\rho(T(x), T(y)) \leq \rho_{\mathbb{S}^d}(x, y)$$

where $\rho_{\mathbb{S}^d}$ is the geodesic distance on $\mathbb{S}^d$. Therefore by the definition of $W_2$, the claim is proved since $T(X_i)$ has distribution $\mu$.

The known upper bounds for the sphere model (see e.g. [15])

$$\mathbb{E} \left[ W_2^2(\mu_{\mathbb{S}^d}^n, \mu_{\mathbb{S}^d}) \right] \lesssim \begin{cases} \frac{1}{n}, & d = 1, \\ \log \frac{n}{n}, & d = 2, \\ \frac{1}{n^{d/2}}, & d \geq 3, \end{cases}$$

thus transfer to the Jacobi model, at least when $d$ is an integer. However, unless $d = 1$, the rate given by Theorem 1 is smaller than the rates on the sphere $\mathbb{S}^d$. 

11
Remark 5. It may be observed that the PDE proof of the lower bound in the Ajtai-Komlós-Tusnády provided in Section 5.2 of [3] may be adapted, together with the suitable version of Lusin’s approximation of Sobolev functions, to show that the rate $\frac{1}{n^{2/d}}$ is optimal on $S^d$ with $d \geq 3$ (and similarly on compact manifolds satisfying the volume doubling property).

In another direction, note that for $x, y \in [-1, 1]$, $\rho(x, y) \geq |x - y|$. Hence by the definition of $W_2$, $W_2^2(\mu^n, \mu) \geq \tilde{W}_2^2(\mu^n, \mu)$, where $\tilde{W}_2$ is the Kantorovich distance defined via the standard metric $|\cdot|$ on the line. Combining the relevant result obtained by ordered statistics (1.3), the order of growth of $E[W_2^2(\mu^n, \mu)]$ is at least $\frac{1}{n}$, which is compatible with Theorem 4.

3 Density fluctuation bounds

With this section, we start addressing the proofs of the main results, following the approach of [3]. The following objects and notation are taken from [3]. For $t \geq 0$, define

$$r^n = \sqrt{n}(\mu^n - \mu), \quad \mu^{n,t} = P_t^n \mu^n, \quad r^{n,t} \mu = P_t^n r^n = \sqrt{n}(\mu^{n,t} - \mu).$$

In particular therefore

$$r^{n,t}(y) = \int_E (p_t(\cdot, y) - 1)dr^n$$

with $y \in E = [-1, 1]$ or $[-1, 1]^2$.

The heuristics here is that the empirical measure $\mu_n$ converges to $\mu$ as $n \to \infty$ in both the weak and $W_2$ topologies (see e.g. [7]). On the other hand, by definition of the heat kernel regularization, for fixed $n$, $\mu^{n,t}$ is close to $\mu^n$ as $t \to 0$. The choice of the normalization of $r^n$ derives from the central limit theorem, and it is therefore reasonable that if $t = t(n) \to 0$ is chosen properly, $\frac{r^{n,t}}{\sqrt{n}}$ will tend to 0 as $n \to \infty$.

The following probabilistic statements quantify this heuristics, and describe the relationship between the rate of $\frac{r^{n,t}}{\sqrt{n}}$ converging to 0 and $n, t$ in the probability sense. They are immediate versions of the corresponding Propositions 3.9 and 3.10 of [3], following the analogous semigroup properties described in the previous section. Uniformity in $y$ is achieved as in [3] by a covering argument. Namely, for $\delta > 0$, if $N_E(\delta)$ denotes the smallest cardinality of a $\delta$-net of $E$, then for $E = [-1, 1]$, $N_E(\delta) \lesssim \max\{1, \delta^{-1}\}$ while for $E = [-1, 1]^2$, $N_E(\delta) \lesssim \max\{1, \delta^{-2}\}$.

Proposition 6 (Fluctuation bounds for the Jacobi model with $d \geq 1$). There exist constants $C, C' > 0$ with the following property: for all $\eta \in (0, 1)$, $\gamma \in (0, \frac{2}{d})$ and $\frac{1}{\eta n^{1-d/4}} \leq t \leq C'$, we have

$$\mathbb{P}\left(\sup_{y \in [-1,1]} \frac{|r^{n,t}(y)|}{\sqrt{n}} > \eta\right) \leq C \exp\left(-\theta n^{1-\frac{d}{4}}\right)$$

with $\theta = \theta(\eta) > 0$.

In parallel, the following proposition is valid on the product Jacobi model with $d = d' = 1$. It is in particular in this result that the constraint $d + d' \leq 2$ appears, restricting the scope of
Theorem 3. Indeed, the product Jacobi model is of curvature-dimension $CD(\min(d, d') - 1, d + d')$ yielding the ultracontractivity property (UC) with decay $t^{-(d+d')/2}$ for the small values of $t > 0$. According to Proposition 3.10 in [3], when $t$ is of the order of $\frac{\log n}{n}$, the probabilistic tail estimate requires $d+d' \leq 2$. On the other hand, $d, d' \geq 1$ is needed for the suitable heat kernel bounds.

Proposition 7 (Fluctuation bounds for the product Jacobi model). On the product Jacobi model $E = [-1,1]^2$ with $d = d' = 1$, there exist constants $C, C' > 0$ with the following property: for all $\eta \in (0, 1)$ and $C' \geq t \geq \frac{1}{n} \log n$ where $\gamma = \gamma(\eta) > 0$ is sufficiently large, we have

$$\mathbb{P} \left( \sup_{y \in E} \frac{|r^{n,t}(y)|}{\sqrt{n}} > \eta \right) \leq \frac{C}{n^2}.$$  

4 Energy estimates

This section is devoted to the energy or trace estimates. It is here that the one-dimensional feature of the Jacobi model is reflected, in contrast with the case of a compact manifold such as the sphere.

Recall that the eigenvalues of the $d$-dimensional Jacobi operator $-L$ on $[-1,1]$ are given by the sequence $\lambda_k = k(k + d - 1), k \in \mathbb{N}$. By the spectral representation of the heat kernel $p_s(x,y), s > 0, x,y \in [-1,1]$,

$$p_s(x,y) = \sum_{k=0}^{\infty} e^{-sk(k+d-1)}J_k(x)J_k(y),$$

and the trace formula may be expressed in the form

$$\int_{-1}^{1} p_s(x,x)d\mu(x) = \sum_{k=0}^{\infty} e^{-sk(k+d-1)}$$

which will be essential for the validity of the next statement.

Lemma 8 (Asymptotics for the trace of the Jacobi model with $d \geq 1$).

$$\int_{-1}^{1} p_s(x,x)d\mu(x) = \frac{1}{\sqrt{s}} \left( \frac{\sqrt{\pi}}{2} + o(1) \right) \quad \text{as} \ s \to 0.$$

Proof. Let $I(s) = \int_{0}^{\infty} e^{-su(u+d-1)}du, s > 0$. Since

$$e^{-s(k+1)(k+d)} \leq e^{-su(u+d-1)} \leq e^{-sk(k+d-1)}$$

for $k \leq u \leq k + 1$, it follows that

$$I(s) \leq \sum_{k=0}^{\infty} e^{-sk(k+d-1)} \leq I(s) + 1.$$  

13
Now
\[ \sqrt{s}I(s) = \sqrt{s} \int_0^\infty e^{-su+d-1} du = e^{\frac{(d-1)^2s}{4}} \sqrt{s} \int_0^\infty e^{-s(u+d-1)^2} du = e^{\frac{(d-1)^2s}{4}} \int_{(d-1)s}^\infty e^{-t^2} dt \]
so that
\[ \lim_{s \to 0} \sqrt{s}I(s) = \int_0^\infty e^{-t^2} dt = \frac{\sqrt{\pi}}{2}. \]
Combining with (4.2) and (4.1), the estimate is proved.

On a product, the previous trace estimate immediately yields the following conclusion.

**Corollary 9** (Asymptotics for the trace of the product Jacobi model). For the product Jacobi model \( E = [-1, 1]^k \) with \( d_j \geq 1, j = 1, \ldots, k \),
\[ \int_E p_s(x, x) d\mu(x) = \frac{1}{s^{\frac{k}{2}}} \left( \frac{\pi^{\frac{k}{2}}}{2} + o(1) \right) \quad \text{as } s \to 0. \]

**Remark 10.** It should be noted that the order of the trace as \( s \to 0 \) is independent of \( d \) (and \( d_j \)).

Let \( f^{n,t} \) be the solutions of the PDE \( Lf^{n,t} = r^{n,t} \) in \( E = [-1, 1] \), whose means are zero. The following results are consequences of the preceding trace asymptotics. Again, the arguments follow the investigation [3].

**Lemma 11** (Energy estimates for the Jacobi model with \( d \geq 1 \)). If \( t = t(n) \to 0 \) as \( n \to \infty \), then
\[ \lim_{n \to \infty} \mathbb{E} \left[ \int_{-1}^1 \Gamma(f^{n,t}) d\mu \right] = \int_0^\infty \left( \int_{-1}^1 p_s(x, x) d\mu(x) - 1 \right) ds = \sum_{k=1}^\infty \frac{1}{k(k + d - 1)}, \quad (4.3) \]
Moreover, if \( 1 \leq d < \frac{4}{3} \),
\[ \limsup_{n \to \infty} \mathbb{E} \left[ \left( \int_{-1}^1 \Gamma(f^{n,t}) d\mu \right)^2 \right] < \infty. \quad (4.4) \]
When \( d \geq \frac{4}{3} \), if \( t = t(n) \to 0 \) as \( n \to \infty \) and \( t \geq \frac{c}{n} \) for some \( c > 0 \), where \( \gamma \in (0, \frac{2}{3d-4}) \), then (4.4) also holds.

**Proof.** The proof of (4.3) is an easy consequence of the trace formula, and can be found in [3]. We only prove (4.4). Using Lemma 3.16 of [3],
\[ \mathbb{E} \left[ \left( \int_{-1}^1 \Gamma(f^{n,t}) d\mu \right)^2 \right] = \mathbb{E} \left[ \left( 2 \int_t^\infty \int_{-1}^1 (P_s r^n)^2 d\mu dss \right)^2 \right] \]
\[ \leq 3 \left[ \int_{2t}^\infty \left( \int_{-1}^1 p_s(x, x) d\mu(x) - 1 \right) ds \right]^2 + \frac{1}{n} \left( 2 \int_t^\infty \int_{-1}^1 [p_s(\cdot, y)]^2 d\mu(y) ds \right)^2 \]
\[ = 3 \left( \sum_{k=1}^\infty \frac{1}{k(k + d - 1)} \right)^2 + \frac{1}{n} \left( 2 \int_t^\infty \int_{-1}^1 [p_s(\cdot, y)]^2 d\mu(y) ds \right)^2, \]
\[ \]
\[ 14 \]
where \([p_s(\cdot, y)]_4 = \left( \int_{-1}^{1} (p_s(x, y) - 1)^4 d\mu(x) \right)^{\frac{1}{4}}\).

It is sufficient to estimate the limsup of the second term. Using the properties (UC) and (SG), we know that for any \(y \in [-1, 1],\)

\[\|p_s(\cdot, y)\|_4^4 \lesssim \begin{cases} s^{-\frac{3d}{2}}, & s \in (0, 1), \\ e^{-2Cs}, & s \in (1, \infty). \end{cases}\]

Thus for \(t \in (0, 1),\)

\[\int_{t}^{\infty} \|p_s(\cdot, y)\|_4^2 ds \lesssim \begin{cases} 1, & 1 \leq d < \frac{4}{3}, \\ \log(t), & d = \frac{4}{3}, \\ t^{-\frac{4d}{3}+1}, & d > \frac{4}{3}. \end{cases}\]

Then it is easy to conclude. For example, if \(d > \frac{4}{3}\) and \(t \geq \frac{c}{n^2}\) for some \(c > 0,\)

\[\frac{1}{n} \left( 2 \int_{t}^{\infty} \int_{-1}^{1} \|p_s(\cdot, y)\|_4^2 d\mu(y) ds \right)^2 \leq \frac{C}{n^{1-(\frac{3d}{2} - 2)\gamma}},\]

which is bounded if \(\gamma \in (0, \frac{2}{3d-4}).\) The proof of the lemma is complete.

In the same way, the following energy estimates hold true for the product Jacobi model with \(d = d' = 1.\)

**Lemma 12** (Energy estimates for the product Jacobi model). *On the product Jacobi model \(E = [-1, 1]^2\) with \(d = d' = 1,\) if \(t = t(n) \to 0\) as \(n \to \infty,\) then

\[\lim_{n \to \infty} \frac{1}{|\log t|} E \left[ \int_E \Gamma(f^{n,t}) d\mu \right] = \frac{\pi}{4}.\]  \hspace{1cm} (4.5)

Furthermore, assuming that \(t = t(n) > \frac{C}{n}\) for some \(C > 0,\) we have

\[\limsup_{n \to \infty} \frac{1}{(\log t)^2} E \left[ \left( \int_E \Gamma(f^{n,t}) d\mu \right)^2 \right] < \infty.\]  \hspace{1cm} (4.6)

**Proof.** Recall that

\[E \left[ \int_E \Gamma(f^{n,t}) d\mu \right] = \int_{2t}^{\infty} \left( \int_E p_s(x, x) d\mu(x) - 1 \right) ds.\]

Combining Corollary 9 and (SG),

\[\int_E p_s(x, x) d\mu(x) - 1 \lesssim \begin{cases} s^{-1}, & s \in (0, 1), \\ e^{-Cs}, & s \in (1, \infty). \end{cases}\]

Thus, we obtain

\[E \left[ \int_E \Gamma(f^{n,t}) d\mu \right] \lesssim |\log t| + 1,\]
from which

\[ \limsup_{n \to \infty} \frac{1}{\log t} \mathbb{E} \left[ \int_E \Gamma(f^{n,t})d\mu \right] < \infty. \]

But what is actually really shown here is that

\[ \frac{1}{\log t} \mathbb{E} \left[ \int_E \Gamma(f^{n,t})d\mu \right] - \frac{1}{\log |t|} \int_2^1 \left( \int_E p_s(x,x)d\mu(x) - 1 \right) ds \]

is infinitesimal as \( n \to \infty \). Combining this with Corollary 9, we obtain (4.5).

Now we turn to (4.6). Recall, as in the previous lemma, that

\[ \mathbb{E} \left[ \left( \int_E \Gamma(f^{n,t})d\mu \right)^2 \right] = \mathbb{E} \left[ \left( 2 \int t \int E (P_{s^n})^2 d\mu ds \right)^2 \right] \]

\[ \leq 3 \left[ \int_2^1 \left( \int E p_s(x,x)d\mu(x) - 1 \right) ds \right]^2 + \frac{1}{n} \left( 2 \int t \int E [p_s(\cdot,y)]^2 d\mu(y) ds \right)^2. \]

By the preceding,

\[ \int_2^1 \left( \int E p_s(x,x)d\mu(x) - 1 \right) ds \lesssim |\log t|. \]

Thus it is sufficient to estimate the limsup of the second term. Using the properties of \( p_t \), and again that \( d = d' = 1 \), we know that for any \( y \in E \),

\[ [p_s(\cdot,y)]^4 \lesssim \begin{cases} s^{-3}, & s \in (0,1), \\ e^{-Cs}, & s \in (1,\infty). \end{cases} \]

Then it is easy to verify that if \( t = t(n) > \frac{C}{n} \) for some \( C > 0 \),

\[ \frac{1}{n} \left( 2 \int t \int -1 [p_s(\cdot,y)]^2 d\mu(y) ds \right)^2 \lesssim \frac{1}{nt}, \]

which is bounded. The proof is therefore complete. \( \square \)

5 Regularization estimates

In this section, we evaluate the cost of the regularization by the heat kernel. We make use here of some recent advances from [2] which provide stronger regularization error bounds when dealing with empirical measures.

To start with, we state a general result from [3] (see also [2, 15]) on the connection between the solution of the Poisson equation \( Lf = u \) and the Kantorovich distance. The original proof is developed in the setting of Riemannian manifolds, but works similarly here. The following theorem is stated for \( E = [-1,1] \) and \( E = [-1,1]^2 \). The use of the integration by parts formula in the proof requires to add some Neumann boundary conditions to the solution of the Poisson equation \( Lf = u \). But for example, for 1-dimensional Jacobi models, \( \Gamma(f)(x) = (1-x^2)f^2 = 0 \) holds naturally when \( x = \pm 1 \) for all smooth \( f \).
Theorem 13. Given smooth, positive density functions $u_0, u_1$ in $E$, let $f$ be the unique solution of $Lf = u_0 - u_1$ with mean zero. Then

$$W_2^2(u_0 \mu, u_1 \mu) \leq 4 \int_E \frac{\Gamma(f)}{u_0} d\mu$$

and

$$W_2^2(u_0 \mu, u_1 \mu) \leq \int_E \Gamma(f) \frac{\log(u_1) - \log(u_0)}{u_1 - u_0} d\mu.$$

In the following, we start the study of the regularization procedure on $E = [-1, 1]$. First, since $\mu = \int_{-1}^1 \delta_x d\mu(x)$ and

$$W_2^2(P_t^* \delta_x, \delta_x) \leq \int_{-1}^1 \rho^2(x, y) p_t(x, y) d\mu(x) d\mu(y),$$

from the joint convexity of $W_2^2$ and (DR), for every probability $\mu$ (on $E = [-1, 1]$),

$$\mathbb{E} \left[ W_2^2(P_t^* \mu, \mu) \right] \leq \int_{-1}^1 \int_{-1}^1 \rho^2(x, y) p_t(x, y) d\mu(x) d\mu(y) \leq Ct.$$

Hence

$$\mathbb{E} \left[ W_2^2(\mu^n, \mu^{n,t}) \right] \leq Ct. \quad (5.1)$$

This estimate is good enough for the case $1 \leq d < 2$. But for the case $d \geq 2$, we need the following more refined estimate. The idea is based on Theorem 5.2 of the recent [2].

Theorem 14. Given $n \geq 1$, let the event

$$A_{1/n}^n = \left\{ \sup_{y \in [-1, 1]} \frac{|r^{n,t}(y)|}{\sqrt{n}} \leq \frac{1}{2} \right\}.$$

Then for the Jacobi model with $d \geq 1$, for $t \in (0, 1)$, we have

$$\mathbb{E} \left[ W_2^2(\mu^n, \mu^{n,t}) \right] \lesssim \mathbb{P}(A_{1/n}^n) + \frac{\sqrt{t}}{n}.$$

Proof. Fix a time $t_0 \in (0, t)$. Applying (5.1) at time $t_0$, it follows from the triangle inequality that

$$\mathbb{E} \left[ W_2^2(\mu^n, \mu^{n,t}) \right] \leq 2 \mathbb{E} \left[ W_2^2(\mu^n, \mu^{n,t_0}) \right] + 2 \mathbb{E} \left[ W_2^2(\mu^{n,t_0}, \mu^{n,t}) \right] \lesssim t_0 + \mathbb{E} \left[ W_2^2(\mu^{n,t_0}, \mu^{n,t}) \right].$$

In order to estimate $\mathbb{E}[W_2^2(\mu^{n,t_0}, \mu^{n,t})]$, let $f$ be the solution of $Lf = u^{n,t} - u^{n,t_0}$, where $u^{n,t_0}$ and $u^{n,t}$ are the density functions of $\mu^{n,t_0}$ and $\mu^{n,t}$ with respect to $\mu$, respectively. By definition,

$$|u^{n,t} - 1| = \frac{|r^{n,t}(y)|}{\sqrt{n}}.$$
Therefore in the event $A_{\frac{1}{2}, n}$, using Theorem 13 we obtain

$$W_2^2(\mu^{n,t_0}, \mu^{n,t}) \leq 4 \int_{-1}^{1} \frac{\Gamma(f)}{u^{s,t}} \, d\mu \leq 8 \int_{-1}^{1} \Gamma(f) \, d\mu.$$ 

Hence, due to the independence of the variables $X_i$,

$$\mathbb{E} \left[ W_2^2(\mu^{n,t_0}, \mu^{n,t}) \chi_{A_{\frac{1}{2}, n}} \right] \leq \mathbb{E} \left[ \int_{-1}^{1} \Gamma(f) \, d\mu \right] = \mathbb{E} \left[ \int_{-1}^{1} (-Lf) \, f \, d\mu \right] = \mathbb{E} \left[ \int_{-1}^{1} (u^{n,t_0} - u^{n,t}) \left( \int_{t_0}^{t} u^{n,s} \, ds \right) \, d\mu \right] = \frac{1}{n} \int_{t_0}^{t} \mathbb{E} \left[ \int_{-1}^{1} (p_t(X, y) - p_t(X, y)) p_s(X, y) \, d\mu(y) \right] \, ds = \frac{1}{n} \int_{t_0}^{t} \int_{-1}^{1} (p_{t+s}(x, x) - 1) \, d\mu(x) \, ds \leq \frac{1}{n} \int_{2t_0}^{t} \int_{-1}^{1} (p_s(x, x) - 1) \, d\mu(x) \, ds.
$$

Lemma $\blacktriangle$ tells us that

$$\mathbb{E} \left[ W_2^2(\mu^{n,t_0}, \mu^{n,t}) \chi_{A_{\frac{1}{2}, n}} \right] \leq \frac{1}{n} \int_{2t_0}^{t+2t_0} \frac{1}{\sqrt{s}} \, ds \leq \frac{1}{n} \frac{t - t_0}{\sqrt{t}}.$$

On the other hand $W_2^2(\mu^n, \mu) \leq \pi^2$. Therefore

$$\mathbb{E} \left[ W_2^2(\mu^n, \mu^{n,t}) \right] \lesssim t_0 + \mathbb{E} \left[ W_2^2(\mu^{n,t_0}, \mu^{n,t}) \chi_{A_{\frac{1}{2}, n}} \right] + \mathbb{E} \left[ W_2^2(\mu^{n,t_0}, \mu^{n,t}) \chi_{A_{\frac{1}{2}, n}} \right] \lesssim t_0 + \mathbb{P}(A_{\frac{1}{2}, n}) + \frac{1}{n} \frac{t - t_0}{\sqrt{t}}.$$

Letting $t_0 \to 0$, the theorem is proved.

If we choose $t = t(n) \to 0$ as $n \to \infty$ properly, then Proposition 6 implies that the term $\mathbb{P}(A_{\frac{1}{2}, n})$ is of much lower order than $\frac{\sqrt{t}}{n}$. Therefore we obtain

**Corollary 15** (Refined regularization estimate for the Jacobi model with $d \geq 1$). Assume that $d \geq 1$. If $t = t(n) \to 0$ as $n \to \infty$ and $t \geq \frac{c}{n^\gamma}$ for some $c > 0$, where $\gamma \in (0, \frac{2}{d})$,

$$\lim_{n \to \infty} n \mathbb{E} \left[ W_2^2(\mu^n, \mu^{n,t}) \right] = 0.$$

With the same arguments, and Proposition 7 we deduce a corresponding result for the product Jacobi model with $d = d' = 1$.

**Theorem 16** (Refined regularization estimate for the product Jacobi model). On the product Jacobi model $E = [-1, 1]^2$ with $d = d' = 1$, there exists a constant $C > 0$, such that for $t = t(n) = \frac{\gamma(n)}{n} \to 0$ with $\gamma(n) \geq C \log n$

$$\mathbb{E} \left[ W_2^2(\mu^n, \mu^{n,t}) \right] \lesssim \frac{\log \gamma(n)}{n}.$$
6 Proofs of Theorems 1 and 3

We first address the proof of Theorem 1 which is split in an upper bound and a lower bound. The upper bound, developed in the next Proposition 17, simplifies the corresponding step in 3 via the improved regularization Theorem 14 and Corollary 15 (and in particular avoids the use of Riesz transform bounds).

**Proposition 17.** Assume that \( d \geq 1 \), then

\[
\limsup_{n \to \infty} n \mathbb{E} \left[ W_2^2(\mu, \mu) \right] \leq \sum_{k=1}^{\infty} \frac{1}{k(k + d - 1)}.
\]

**Proof.** Fix \( \gamma \in (0, \frac{2}{d}) \) and let \( t = t(n) = n^{-\gamma} \). For \( \eta \in (0, 1) \) consider the event

\[
A_\eta = A_{\eta, n} = \left\{ \sup_{y \in [-1, 1]} \frac{|y^{n,t}(y)|}{\sqrt{n}} \leq \eta \right\}.
\]

Since \( W_2^2(\mu, \mu) \leq \pi^2 \), for \( n \) large enough using Proposition 6 we have

\[
n \mathbb{E} \left[ W_2^2(\mu, \mu) \right] \leq n \mathbb{E} \left[ W_2^2(\mu, \mu) \chi_{A_\eta} \right] + n \mathbb{E} \left[ W_2^2(\mu, \mu) \chi_{A_{\eta, n}^c} \right] \leq C n \exp \left( \theta n^{1-\frac{d\gamma}{2}} \right)
\]

with \( \theta = \theta(\eta) > 0 \). Therefore it is sufficient to estimate \( \limsup_{n \to \infty} n \mathbb{E} \left[ W_2^2(\mu, \mu) \chi_{A_\eta} \right] \).

Notice that for \( \varepsilon > 0 \),

\[
W_2^2(\mu, \mu) \leq (1 + \varepsilon)W_2^2(\mu, \mu) + (1 + \varepsilon^{-1})W_2^2(\mu, \mu, t).
\]

Corollary 15 tells us that \( \lim_{n \to \infty} n \mathbb{E} \left[ W_2^2(\mu, \mu, t) \right] = 0 \). So we only need to estimate

\[
\limsup_{n \to \infty} n \mathbb{E} \left[ W_2^2(\mu, \mu, t) \chi_{A_\eta} \right].
\]

On the event \( A_\eta \), using Theorem 13, we obtain

\[
W_2^2(\mu, \mu, t) \leq \frac{1}{n} \int_{-1}^{1} \Gamma(f) \frac{\log(u)}{u^{n,t} - 1} \, d\mu \leq \frac{1}{n \sqrt{1-\eta}} \int_{-1}^{1} \Gamma(f) \, d\mu,
\]

where \( u^{n,t} = \frac{u^{n,t}}{\sqrt{n}} \) is the density function of \( \mu^{n,t} \) with respect to \( \mu \) (since \( \frac{\log u}{|u|} \leq \frac{1}{\sqrt{1-\eta}} \) whenever \( u > 0, |u - 1| \leq \eta, 0 < \eta < 1 \)). Hence Lemma 11 implies that

\[
\limsup_{n \to \infty} n \mathbb{E} \left[ W_2^2(\mu, \mu) \chi_{A_\eta} \right] \leq \lim_{n \to \infty} \frac{1}{\sqrt{1-\eta}} \mathbb{E} \left[ \int_{-1}^{1} \Gamma(f) \, d\mu \right] = \frac{1}{\sqrt{1-\eta}} \sum_{k=1}^{\infty} \frac{1}{k(k + d - 1)}.
\]

At last we get that

\[
\limsup_{n \to \infty} n \mathbb{E} \left[ W_2^2(\mu, \mu) \right] \leq \frac{1 + \varepsilon}{\sqrt{1-\eta}} \sum_{k=1}^{\infty} \frac{1}{k(k + d - 1)}.
\]

Letting \( \varepsilon, \eta \to 0 \), the proposition is established. \( \square \)
Next, we prove the lower bound part of Theorem 1 in the form of the following statement.

**Proposition 18.** Assume that \( d \geq 1 \), then

\[
\liminf_{n \to \infty} n \mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] \geq \sum_{k=1}^{\infty} \frac{1}{k(k + d - 1)}.
\]

**Proof.** First, by the heat flow contraction in Wasserstein space (2.3), for every \( t \geq 0 \),

\[
W_2^2(\mu^n, \mu) \geq e^{2(d-1)t} W_2^2(\mu^{n,t}, \mu) \geq W_2^2(\mu^{n,t}, \mu)
\]

so that we need only concentrate on \( W_2^2(\mu^{n,t}, \mu) \). Next, by the Kantorovich duality (cf. [18]),

\[
\frac{1}{2} W_2^2(\mu^{n,t}, \mu) \geq \sup \left\{ \int_{-1}^{1} f d\mu^{n,t} + \int_{-1}^{1} g d\mu ; \ f(x) + g(y) \leq \frac{\rho(x,y)^2}{2} \right\}
\]

\[
= \sup \left\{ \int_{-1}^{1} f \sqrt{n} d\mu^{n,t} + \int_{-1}^{1} (f + g) d\mu ; \ f(x) + g(y) \leq \frac{\rho(x,y)^2}{2} \right\}.
\]

Fix \( \eta \in (0,1) \) and let \( t = t(n) = \eta^{-1} n^{-\gamma} \), where

\[
\gamma \in \begin{cases} 
(0, \frac{2}{d}) , & 1 \leq d \leq \frac{4}{3}, \\
(0, \min\{\frac{2}{d}, \frac{2}{3d-4}\}) , & d > \frac{4}{3}. 
\end{cases}
\]

As in the proof of Proposition 17, consider the event \( A_\eta \). Define \( f = -\frac{f^{n,t}}{\sqrt{n}} \), where \( f^{n,t} \) is defined in Section 3. Therefore, on the event \( A_\eta \), \( \|Lf\|_\infty \leq \eta \). Moreover, using (2.2),

\[
\|Lf\|_\infty + e^{2(d-1)t} - 1\|\Gamma(f)\|_\infty \leq \omega(\eta)
\]

with \( \omega(\eta) \to 0 \) as \( \eta \to 0 \). To this \( f \), we associate

\[
g(x) = Q_1(-f)(x) = \inf_{y \in [-1,1]} [-f(y) + \frac{1}{2} \rho^2(x,y)],
\]

so that on \( A_\eta \), by Corollary 3.3 of [3],

\[
\frac{1}{2} W_2^2(\mu^{n,t}, \mu) \geq \left( 1 - \frac{e^{\omega(\eta)}}{2} \right) \frac{1}{n} \int_{-1}^{1} \Gamma(f^{n,t}) d\mu.
\]

Next

\[
\frac{1}{2 - e^{\omega(\eta)}} \liminf_{n \to \infty} n \mathbb{E} \left[ W_2^2(\mu^{n,t}, \mu) \chi_{A_\eta} \right]
\]

\[
\geq \liminf_{n \to \infty} \mathbb{E} \left[ \chi_{A_\eta} \int_{-1}^{1} \Gamma(f^{n,t}) d\mu \right]
\]

\[
\geq \lim_{n \to \infty} \mathbb{E} \left[ \int_{-1}^{1} \Gamma(f^{n,t}) d\mu \right] - \limsup_{n \to \infty} \mathbb{E} \left[ \chi_{A_\eta} \int_{-1}^{1} \Gamma(f^{n,t}) d\mu \right].
\]
By Lemma \[11\] for our choice of \( t = t(n) \),
\[
\limsup_{n \to \infty} \mathbb{E} \left[ \left( \int_{-1}^{1} \Gamma(f^n,t)d\mu \right)^2 \right] < \infty.
\]

Hence
\[
\limsup_{n \to \infty} \mathbb{E} \left[ \chi_{A_n} \int_{-1}^{1} \Gamma(f^n,t)d\mu \right] \leq \limsup_{n \to \infty} \mathbb{P}(A_{n,n}^c)^{1/2} \left( \mathbb{E} \left[ \left( \int_{-1}^{1} \Gamma(f^n,t)d\mu \right)^2 \right] \right)^{1/2} = 0.
\]

Therefore
\[
\frac{1}{2 - e^{\omega(t)}} \liminf_{n \to \infty} n \mathbb{E} \left[ W_2^2(\mu^n, \mu_{A_n}) \right] \geq \lim_{n \to \infty} \mathbb{E} \left[ \int_{-1}^{1} \Gamma(f^n,t)d\mu \right] = \sum_{k=1}^{\infty} \frac{1}{k(k + d - 1)}
\]
from which the proposition follows by letting \( \eta \to 0 \).

The modifications to address in the same way Theorem \[3\] are minor. For Theorem \[3\] Proposition \[6\] is replaced by Proposition \[7\] and Lemma \[11\] by Lemma \[12\]. We only need to take \( t(n) = \frac{\log n}{n} \), where \( \gamma > 0 \) is a sufficiently large constant. Then, the regularization procedure is here taken into account by Theorem \[16\] which indicates that
\[
\frac{n}{\log n} \mathbb{E} \left[ W_2^2(\mu^n, \mu_{A_n}) \right] \lesssim \frac{\log \log n}{\log n}
\]
so that in particular \( \lim_{n \to \infty} \frac{n}{\log n} \mathbb{E} \left[ W_2^2(\mu^n, \mu_{A_n}) \right] = 0 \).

By the semigroup properties discussed in Section \[2\] the corresponding results of Proposition \[6\], Lemma \[11\] and Corollary \[15\] are valid for non-symmetric Jacobi models, with the value of \( d \) therein replaced by \( 2(\alpha + \beta) - 1 \) due to the curvature-dimension \( CD(\frac{\alpha + \beta - 1}{2}, 2(\alpha + \beta) - 1) \) condition. All the other steps are then basically similar to the proof of Theorem \[1\] and are thus not repeated.

In the same way, the proofs of the limits \((1.6)\) and \((1.7)\) in the bipartite case are immediately adapted from Propositions 4.9 and 4.10 of \[3\].

7 Proof of Theorem \[2\]

We cover at the same time Theorem \[2\] and its extension \((1.5)\) to \( k \geq 3 \). Let therefore \( E = [-1,1]^k \) and \( \mu \) be the product Jacobi measure with \( d_j \geq 1, j = 1, \ldots, k \) \((k \geq 2)\). The arguments also apply to non-symmetric Jacobi measures (with the conditions \( \alpha_j, \beta_j \geq \frac{1}{2} \)).

The following lemma is a general heat kernel upper bound of \( \mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] \) (for its proof, see \[15\])

Lemma 19. In the prescribed setting and notation, for every \( t > 0 \),
\[
\mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] \leq 2 \int_E \int_E r^2(x, y)p_t(x, y)d\mu(x)d\mu(y) + \frac{8}{n} \int_{2t}^{\infty} \int_E (p_s(x, x) - 1)d\mu(x)ds.
\]
Combining (DR) and Corollary 2 we know that on the product Jacobi model, for every $t \in (0, 1)$,

$$
\mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] \lesssim t + \begin{cases} 
\frac{\log(\frac{1}{t})}{n^{1/(2k-1)}}, & k = 2, \\
\frac{1}{n^{2/k}}, & k \geq 3.
\end{cases}
$$

After optimization in $t \in (0, 1)$, we obtain that

$$
\mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] \lesssim \begin{cases} 
\frac{1}{n}, & k = 2, \\
\frac{1}{n^{2/k}}, & k \geq 3,
\end{cases}
$$

which amounts to the first half of Theorem 2.

In order to prove the lower bound part of Theorem 2, we use a simple comparison argument. Denote by $\lambda$ the uniform distribution on $[0, 1]^k$, let $U_1, \ldots, U_n$ be independent and distributed as $\lambda$ and write $\nu_n = \frac{1}{n} \sum_{i=1}^n \delta_{U_i}$. Define functions $\Phi_j : [-1, 1] \to [0, 1], j = 1, \ldots, k$, as follows:

$$
\Phi_j(x) = C_{d_j} \int_{-1}^x (1 - u^2)^{d_j - 1} du,
$$

where $C_{d_j}$ are normalization constants. Write $\Phi = \Phi_1 \otimes \cdots \otimes \Phi_k : E \to [0, 1]^k$. Therefore if $X_1, \ldots, X_n$ are independent distributed as $\mu$, then $\Phi(X_1), \ldots, \Phi(X_n)$ are independent distributed as $\lambda$. The Kantorovich distance $W_1$ in this context is given by

$$
W_1(\mu^n, \mu) = \sup_{\Gamma(f) \leq 1} \left[ \frac{1}{n} \sum_{i=1}^n f(X_i) - \int_E f d\mu \right]. \quad (7.1)
$$

Notice that for any smooth function $g$ defined on $[0, 1]^k$, $g \circ \Phi$ is a function defined on $E$ and

$$
\Gamma(g \circ \Phi) = \sum_{j=1}^k (1 - x_j^2) \Phi_j''(x_j)(\partial_{x_j}g)^2 = \sum_{j=1}^k C_{d_j}^2 (1 - x_j^2)^{d_j - 1}(\partial_{x_j}g)^2.
$$

Therefore if $d_j \geq 1, j = 1, \ldots, k$ (or $\alpha_j, \beta_j \geq \frac{1}{2}$), there exists a constant $C > 0$ such that

$$
\Gamma(g \circ \Phi) \leq C^2 \|\nabla g\|_\infty^2. \quad (7.2)
$$

Combining with (7.1), we obtain that

$$
W_1(\mu^n, \mu) \geq \sup_{\Gamma(g \circ \Phi) \leq 1} \left[ \frac{1}{n} \sum_{i=1}^n g \circ \Phi(X_i) - \int_E g \circ \Phi d\mu \right]
$$

$$
\geq \sup_{\|\nabla g\|_\infty \leq \frac{1}{C}} \left[ \frac{1}{n} \sum_{i=1}^n g \circ \Phi(X_i) - \int_E g \circ \Phi d\mu \right]
$$

$$
= \frac{1}{C} \sup_{\|\nabla g\|_\infty \leq 1} \left[ \frac{1}{n} \sum_{i=1}^n g(\Phi(X_i)) - \int_{[0,1]^k} g dx dy \right]
$$

$$
= \frac{1}{C} W_1(\nu^n, \lambda).
$$
Therefore there exists a constant $C > 0$,

$$\mathbb{E} \left[ W_2^2(\mu^n, \mu) \right] \geq \mathbb{E} \left[ W_1(\mu^n, \mu) \right]^2 \geq \frac{1}{C} \mathbb{E} \left[ W_1(\nu^n, \lambda) \right]^2.$$  

From (1.1),

$$\mathbb{E} [W_1(\nu^n, \lambda)] \approx \begin{cases} \sqrt{\log n}, & k = 2, \\ \frac{\log n}{n^{1/k}}, & k \geq 3. \end{cases}$$

Theorem 2 is proved.

Acknowledgements. I thank M. Ledoux for useful indications and comments during the preparation of this work, and the referee for several helpful suggestions for improvements.

References

[1] M. Ajtai, J. Komlós, G. Tusnády. On optimal matchings. *Combinatorica* 4, 259–264 (1984).
[2] L. Ambrosio, F. Glaudo, Finer estimates on the 2-dimensional matching problem (2018).
[3] L. Ambrosio, F. Stra, D. Trevisan. A PDE approach to a 2-dimensional matching problem (2016). *Probab. Theory Related Fields* 173, 433-478 (2019).
[4] D. Bakry. Remarques sur les semigroupes de Jacobi. *Hommage à P. A. Meyer et J. Neveu. Astérisque* 236, 23–39. Société Mathématique de France (1996).
[5] D. Bakry, I. Gentil, M. Ledoux. *Analysis and geometry of Markov diffusion operators*. Grundlehren der mathematischen Wissenschaften 348. Springer (2014).
[6] F. Barthe, C. Bordenave. Combinatorial optimization over two random point sets. *Séminaire de Probabilités XLV, Lecture Notes in Math.* 2078, 483–535. Springer (2013).
[7] S. Bobkov, M. Ledoux. One-dimensional empirical measures, order statistics, and Kantorovich transport distances (2016). To appear in *Memoirs Amer. Math. Soc.*
[8] J. Boutet de Monvel, O. Martin. Almost sure convergence of the minimum bipartite matching functional in Euclidean space. *Combinatorica* 22, 523–530 (2002).
[9] S. Caracciolo, C. Lucibello, G. Parisi, G. Sicuro. Scaling hypothesis for the Euclidean bipartite matching problem. *Physical Review E* 90, 012118 (2014).
[10] S. Dereich, M. Scheutzow, R. Schottstedt. Constructive quantization: approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* 49, 1183–1203 (2013).
[11] V. Dobrić, J. Yukich. Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* 8, 97–118 (1995).
[12] N. Fournier, A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* 162, 707–738 (2015).

[13] M. Hahn, Y. Shao. An exposition of Talagrand’s mini-course on matching theorems. *Probability in Banach spaces* 8, Progr. Probab. 30, 3–38. Birkhäuser (1992).

[14] N. Holden, Y. Peres, A. Zhai. Gravitational allocation for uniform points on the sphere. *Proc. Nat. Acad. Sciences* 115, 9666–9671 (2018).

[15] M. Ledoux. On optimal matching of Gaussian samples. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* 457, Veroyatnost’ i Statistika. 25, 226–264 (2017).

[16] A. Nowak, P. Sjögren. Sharp estimates of the Jacobi heat kernel. *Studia Mathematica* 218, 219–244 (2013).

[17] M. Talagrand. *Upper and lower bounds of stochastic processes*. Modern Surveys in Mathematics 60. Springer-Verlag (2014).

[18] C. Villani. *Optimal transport. Old and new*. Grundlehren der mathematischen Wissenschaften 338. Springer (2009).

[19] J. Yukich. *Probability theory of classical Euclidean optimization problems*. Lecture Notes in Mathematics 1675. Springer (1998).

School of Mathematical Sciences
Fudan University, Shanghai 200433, People’s Republic of China
15110840006@fudan.edu.cn
and
Institut de Mathématiques de Toulouse
Université de Toulouse – Paul-Sabatier, F-31062 Toulouse, France
zhu@math.univ-toulouse.fr