Available online at www.sciencedirect.com
SciVerse ScienceDirect
Stochastic Processes and their Applications 122 (2012) 305–333
stochastic processes and their applications
www.elsevier.com/locate/spa

Asymptotic analysis of the optimal cost in some transportation problems with random locations

Giovanni Luca Torrisi*

Istituto per le Applicazioni del Calcolo "Mauro Picone", CNR, Via dei Taurini 19, I-00185 Roma, Italy

Received 13 April 2011; received in revised form 3 August 2011; accepted 23 August 2011
Available online 1 September 2011

Abstract

In this paper we provide an asymptotic analysis of the optimal transport cost in some matching problems with random locations. More precisely, under various assumptions on the distribution of the locations and the cost function, we prove almost sure convergence, and large and moderate deviation principles. In general, the rate functions are given in terms of infinite-dimensional variational problems. For a suitable one-dimensional transportation problem, we provide the expression of the large deviation rate function in terms of a one-dimensional optimization problem, which allows the numerical estimation of the rate function. Finally, for certain one-dimensional transportation problems, we prove a central limit theorem.

Keywords: Calculus of variations; Central limit theorem; Large deviations; Matching problem; Moderate deviations; Monge–Kantorovich problem; Optimal transport

1. Introduction

Let (E_1, d_1) and (E_2, d_2) be Polish spaces (i.e. complete and separable metric spaces) and $c : E_1 \times E_2 \to [0, \infty)$ a measurable cost function, i.e. the quantity $c(x, y)$ describes the cost of moving an object located at $x \in E_1$ to location $y \in E_2$. For a fixed integer $n \geq 1$, let $\mathcal{P}_n$ be the set of permutations of $\{1, \dots, n\}$. If we want to transport n objects located at $x_1, \dots, x_n \in E_1$ to locations $y_1, \dots, y_n \in E_2$, the optimal transport cost is

$$k_n = k_n(x_1, \dots, x_n, y_1, \dots, y_n) := \inf_{\sigma \in \mathcal{P}_n} \sum_{i=1}^{n} c(x_i, y_{\sigma(i)}), \quad n \geq 1.$$

* Tel.: +39 0649270963; fax: +39 0672594699.
E-mail address: torrisi@iac.rm.cnr.it.

0304-4149/$ - see front matter
doi:10.1016/j.spa.2011.08.009

In mass transportation literature (see e.g. [25,26]), the permutations σ are called transference plans, and those achieving the infimum, optimal transference plans. Adding the randomness to this problem, let $X_1, \dots, X_n$ be E_1-valued random variables (rv's), $Y_1, \dots, Y_n$ be E_2-valued rv's and consider the optimal transport cost $K_n := k_n(X_1, \dots, X_n, Y_1, \dots, Y_n)$, $n \geq 1$.

In this paper, under various assumptions on the distribution of the rv's $\{X_n\}_{n\geq 1}$, $\{Y_n\}_{n\geq 1}$ and the cost function c, we provide almost sure convergence (ASC), large deviation principles (LDP), moderate deviation principles (MDP) and central limit theorems (CLT) for $\{K_n\}_{n\geq 1}$ (properly normalized).

Most of our results refer to the so-called two-sample transportation problem (TSTP) and grid transportation problem (GTP). We speak of a TSTP if

$$\{X_n\}_{n\geq 1} \text{ is a sequence of independent and identically distributed (iid) rv's with law } \ell_1 \tag{1}$$

and

$$\{Y_n\}_{n\geq 1} \text{ is a sequence of iid rv's with law } \ell_2. \tag{2}$$

If moreover

$$\{X_n\}_{n\geq 1} \text{ and } \{Y_n\}_{n\geq 1} \text{ are independent sequences,} \tag{3}$$

then we speak of a TSTP with completely independent locations (CIL). A TSTP is called a k-dimensional TSTP (kDTSTP) if E_2 is a subset of $\mathbb{R}^k$. We speak of a GTP if $\{X_n\}_{n\geq 1}$ satisfies (1), $\{g_n\}_{n\geq 1} \subset E_2$ is a countable subset of E_2 and $Y_n = g_n$, $n \geq 1$. In such a case, K_n represents the optimal transport cost for transporting n objects located at $X_1, \dots, X_n$ to the grid locations $g_1, \dots, g_n$. A GTP is called a k-dimensional GTP (kDGTP) if E_2 is a subset of $\mathbb{R}^k$. Let $m, k \geq 1$ be integers, $n = m^k$, and $\{C_1, \dots, C_n\}$ a partition of $[0, 1]^k$ formed by cubes $C_i \subset [0, 1]^k$ with Lebesgue measure $1/n$. If $E_2 = [0, 1]^k$ and the grid locations $g_1, \dots, g_n$ are regularly spaced, i.e. $g_i \in C_i$, then we speak of a k-dimensional regular grid transportation problem (kDRGTP). Throughout this paper we say that the TSTP or the GTP is over a compact metric space (E, d) if (E, d) is a compact metric space, $E_1 = E_2 = E$ and $d_1 = d_2 = d$. The TSTP or the GTP is said to be over a compact metric space (E, d) with distance cost function (DCF) if it is over a compact metric space (E, d) and $d_1 = d_2 = c = d$.

Before describing our results in more detail, we recall some related literature. Ajtai et al. [2] studied the TSTP over $E_1 = E_2 = [0, 1]^2$ with Euclidean DCF, CIL and $\ell_1 = \ell_2 = \ell$, the Lebesgue measure. In particular, they proved the existence of a constant $\kappa > 0$ such that $\kappa^{-1}\sqrt{n \log n} \leq K_n \leq \kappa\sqrt{n \log n}$, with high probability. Refinements of this result were obtained by Talagrand [22]. Ganesh and O'Connell [12] proved a LDP for $\{K_n/n\}_{n\geq 1}$ for the case where K_n is the optimal transport cost of the TSTP over a compact metric space (E, d) with DCF, CIL and $\ell_1 = \ell_2$. The rate function is given in variational form and, although they did not solve explicitly the related optimization problem, they characterized its solution. Moreover, in [12] a MDP is proved for $\{K_n/n\}_{n\geq 1}$ in the case when K_n is the optimal transport cost of the TSTP described above with $E = [0, 1]^2$, d equal to the Euclidean distance and $\ell_1 = \ell_2 = \ell$. The expression for the rate function is provided up to a multiplicative constant. Recently, Barthe and O'Connell [3] noticed that the MDP stated in [12] may be extended to $\{K_n/n\}_{n\geq 1}$ where K_n is the optimal transport cost of the TSTP over a compact metric space (E, d) with $E \subset \mathbb{R}^k$, Euclidean DCF, CIL and $\ell_1 = \ell_2$. As a main result, Barthe and O'Connell [3] proved that if E is the unit ball of $\mathbb{R}^k$ and $\ell_1 = \ell_2$ is the uniform distribution, then the moderate deviation rate function is equal to $(k + 2)x^2/4$. A related paper is Gozlan and Léonard [13] where new transportation cost inequalities are derived by means of elementary large deviation reasonings. We refer the reader to [14] for a survey on recent developments in the area of transport inequalities.

This paper is organized as follows. In Section 2 we give some preliminaries on the Monge–Kantorovich optimal transport problem, and recall a representation formula for the normalized optimal transport cost in terms of the empirical measures of locations. In Section 3, under general assumptions on the distribution of the rv's $\{X_n\}_{n\geq 1}$, $\{Y_n\}_{n\geq 1}$ and the cost function c, we provide ASC and LDP for $\{K_n/n\}_{n\geq 1}$. The large deviation rate functions are given in terms of infinite-dimensional variational problems. In Section 4 we give more insight into the expressions for the large deviation rate functions in some specific situations. First, we consider the TSTP over a compact metric space (E, d) with DCF, CIL, and $\ell_1 = \ell_2$ and provide lower bounds for the large deviation rate function of $\{K_n/n\}_{n\geq 1}$. A similar result is derived for the GTP. Second, we investigate the relation between Maurey's τ-property and the large deviation rate function of $\{K_n/n\}_{n\geq 1}$ concerning the TSTP with $E_1 = E_2 = E$, $d_1 = d_2 = d$, $\ell_1 = \ell_2$ and CIL. Third, we consider the 1DRGTP over the compact interval [0, 1] with Euclidean DCF and $\ell_1 = \ell$, and provide the large deviation rate function of $\{K_n/n\}_{n\geq 1}$ in terms of a one-dimensional optimization problem. This allows the numerical estimation of the rate function. In Section 5 we prove MDP for $\{K_n/n\}_{n\geq 1}$ when K_n is the optimal transport cost of the GTP over a compact metric space with DCF. In Section 6 we briefly discuss possible extensions of the previous results to non-compact spaces. Finally, in Section 7 we consider the 1DTSTP over a compact interval with Euclidean DCF, CIL, and $\ell_1 = \ell_2$, and we prove that $\left\{K_n/\sqrt{n}\right\}_{n\geq 1}$ converges weakly to a rv whose tail is asymptotically equivalent to the tail of the modulus of a Gaussian rv. We show also a CLT for a 1DGTP.

2. The Monge–Kantorovich optimal transport problem

Let $\mathrm{P}(E_i)$ $(i = 1, 2)$ be the space of Borel probability measures on E_i, equipped with the topology of weak convergence. We recall that, since (E_i, d_i) is a Polish space, then the topology of weak convergence is metrizable by Prohorov's metric, say $d_P^{(i)}$, and $(\mathrm{P}(E_i), d_P^{(i)})$ is in turn a Polish space.

For $\mu \in \mathrm{P}(E_1)$ and $\nu \in \mathrm{P}(E_2)$, denote by $\Pi(\mu, \nu)$ the space of Borel probability measures π on $E_1 \times E_2$ with fixed marginals $\mu(\cdot) = \pi(\cdot \times E_2)$ and $\nu(\cdot) = \pi(E_1 \times \cdot)$. Formally, the Monge–Kantorovich optimal transport problem (see [15,16], and more recently [25,26]) is the minimization problem described by

$$C(\mu, \nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{E_1\times E_2} c(x, y)\, \pi(\mathrm{d}x\, \mathrm{d}y), \quad (\mu, \nu) \in \mathrm{P}(E_1) \times \mathrm{P}(E_2).$$

Following the standard interpretation, $c(x, y)$ is the work needed to move one unit of mass from location x to location y, and $C(\mu, \nu)$ is the value of the optimal transport cost of transport between μ and ν. For a Polish space (E, d) and $p \in [1, \infty)$, define the functional

$$W_p(\mu, \nu) := \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{E\times E} d(x, y)^p\, \pi(\mathrm{d}x\, \mathrm{d}y) \right)^{1/p}, \quad (\mu, \nu) \in \mathrm{P}(E) \times \mathrm{P}(E)$$

and denote by $\mathrm{P}_p(E)$ the subset of $\mathrm{P}(E)$ formed by the probability measures μ such that, for some $x_0 \in E$, $\int_E d(x_0, x)^p \mu(\mathrm{d}x) < \infty$. Note that if d is a bounded distance, i.e. $\|d\|_\infty := \sup_{(x,y)\in E\times E} d(x, y) < \infty$, then $\mathrm{P}_p(E) = \mathrm{P}(E)$, for all $p \geq 1$. It is well-known that W_p is a distance on $\mathrm{P}_p(E)$ (see e.g. Theorem 7.3 in [25]), called the Wasserstein distance for $p \geq 2$ and the Kantorovich–Rubinshtein distance for $p = 1$. Thus, if $E_1 = E_2 = E$, $d_1 = d_2 = d$, and the cost function is of the form $c = d^p$, then, for any $\alpha_1, \alpha_2 \in \mathrm{P}_p(E)$, $C(\alpha_1, \alpha_2) = 0$ if and only if $\alpha_1 = \alpha_2$. By the Kantorovich duality (see e.g. Theorem 1.14 in [25])

$$W_1(\mu, \nu) = \sup \left(\int_E \varphi(x)\, \mu(\mathrm{d}x) - \int_E \varphi(x)\, \nu(\mathrm{d}x) \right), \quad \forall \mu, \nu \in \mathrm{P}_1(E),$$

where the supremum is taken over all the Lipschitz functions $\varphi : E \to \mathbb{R}$ with Lipschitz constant

$$\|\varphi\|_{\mathrm{Lip}} := \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x, y)} \tag{4}$$

less than or equal to 1. If d is a bounded distance, then the set where the supremum is taken may be restricted, i.e.

$$W_1(\mu, \nu) = \sup_{\varphi \in \mathcal{L}(E)} \left(\int_E \varphi(x)\, \mu(\mathrm{d}x) - \int_E \varphi(x)\, \nu(\mathrm{d}x) \right), \quad \forall \mu, \nu \in \mathrm{P}(E), \tag{5}$$

where $\mathcal{L}(E) := \{\varphi : E \to \mathbb{R} : \text{ is Lipschitz}, 0 \leq \varphi \leq \|d\|_\infty, \ \|\varphi\|_{\mathrm{Lip}} \leq 1\}$ (see Remark 1.15(i), p. 34, and 7.5(i), p. 207, in [25] for details). Let μ and ν be probability measures on $\mathbb{R}$, with distribution functions F_1 and F_2, respectively. We mention that, if d is the Euclidean distance then, for any Borel set $E \subset \mathbb{R}$, we have

$$W_1(\mu, \nu) = \int_E |F_1(x) - F_2(x)|\, \mathrm{d}x. \tag{6}$$

Now, we recall a fundamental expression for k_n. For $n \geq 1$, $x_1, \ldots, x_n \in E_1$ and $y_1, \ldots, y_n \in E_2$, define $l_n := (\sum_{i=1}^n \delta_{x_i})/n$ and $m_n := (\sum_{i=1}^n \delta_{y_i})/n$, where δ_x denotes the Dirac measure at x.

Lemma 2.1. *$C(l_n, m_n) = k_n/n$, for any $n \geq 1$.*

For $n \geq 1$, define the random empirical measures $L_n := (\sum_{i=1}^n \delta_{X_i})/n$ and $M_n := (\sum_{i=1}^n \delta_{Y_i})/n$; then by Lemma 2.1 we have almost surely (a.s.)

$$C(L_n, M_n) = K_n/n, \quad \forall n \geq 1. \tag{7}$$

The proof of Lemma 2.1 is similar to that of Lemma 3 in [20] or Lemma 2.1 in [12], and therefore omitted.

We conclude this section by introducing some more notation. All the random quantities are defined on a complete probability space $(\Omega, \mathcal{F}, P)$ and we adopt the standard conventions that the infimum over an empty set is equal to $+\infty$ and $0 \log 0 = 0$. For a Polish space (E, d), we denote by $\mathrm{M}_b(E)$ the set of signed Borel measures of finite variation on E, equipped with the topology of weak convergence. Here again, the topology of weak convergence is metrizable by Prohorov's metric, say d_P, and $(\mathrm{M}_b(E), d_P)$ is a Polish space. Finally, for each $\nu \in \mathrm{M}_b(E)$ and measurable function φ, we set $\nu(\varphi) := \int_E \varphi(x)\, \nu(\mathrm{d}x)$.

3. ASC and LDP

We say that a sequence of rv's $\{\xi_n\}_{n \geq 1}$, taking values on a metric space (S, d_S), converges almost surely on S to a limit ξ if ξ is an S-valued rv and, with probability 1, $\xi_n \to \xi$ (wrt the metric d_S). Recall that $\{\xi_n\}_{n \geq 1}$ obeys a LDP on S with speed s_n and rate function I if $\{s_n\}_{n \geq 1}$ is a strictly increasing sequence of positive numbers diverging to $+\infty$, $I : S \to [0, \infty]$ is a lower semi-continuous (lsc) function, and the following inequalities hold for every Borel set B:

$$-\inf_{z\in\overset{\circ}{B}} I(z) \le \liminf_{n\to\infty} \frac{1}{s_n} \log P(\xi_n \in \overset{\circ}{B}) \le \limsup_{n\to\infty} \frac{1}{s_n} \log P(\xi_n \in \overline{B}) \le -\inf_{z\in\overline{B}} I(z),$$

where $\overset{\circ}{B}$ denotes the interior of B and $\overline{B}$ denotes the closure of B. We point out that the lower semi-continuity of I means that its level sets $\{z \in S : I(z) \le a\}$ are closed for all $a \ge 0$; when the level sets are compact, the rate function I is said to be good. For more insight into the large deviations theory see e.g. [9].

We start with the following ASC on general Polish spaces. In the next part, we denote by $C_b(E_1 \times E_2)$ the space of bounded and continuous functions from $E_1 \times E_2$ to $\mathbb{R}$.

Theorem 3.1. *Assume*

$$c \in C_b(E_1 \times E_2), \tag{8}$$

$\{L_n\}_{n\ge1}$ *converges almost surely on* $\mathrm{P}(E_1)$ *to* ℓ_1, *and* $\{M_n\}_{n\ge1}$ *converges almost surely on* $\mathrm{P}(E_2)$ *to* ℓ_2. *Then* $\{K_n/n\}_{n\ge1}$ *converges almost surely on* $\mathbb{R}$ *to* $C(\ell_1, \ell_2)$.

Proof. The claim follows if we show that the map $C : \mathrm{P}(E_1) \times \mathrm{P}(E_2) \to [0, \infty)$ is continuous wrt the product weak topology. It is known that C is lsc if the cost function c is lsc (see Remark 6.12 p. 97 in [26]). So, we only need to show the upper semi-continuity (usc) of C. This in turn follows by (8), the compactness of $\Pi(\mu, \nu)$ (see the proof of Theorem 4.1, pp. 44–45, in [26]) and Theorem 5.20, p. 77, in [26]. □

The following generalizations of Theorem 3.1 in [12] hold.

Theorem 3.2. (i) *Assume* (8) *and*

$$\{(L_n, M_n)\}_{n\ge1} \text{ obeys a LDP on } \mathrm{P}(E_1) \times \mathrm{P}(E_2) \text{ with speed } n \text{ and good rate function } K. \tag{9}$$

Then $\{K_n/n\}_{n\ge1}$ *obeys a LDP on* $\mathbb{R}$ *with speed* n *and good rate function*

$$I(z) = \inf_{(\mu,\nu)\in\mathrm{P}(E_1)\times\mathrm{P}(E_2):\, C(\mu,\nu)=z} K(\mu, \nu).$$

(ii) *Assume* (3), (8),

$$\{L_n\}_{n\ge1} \text{ obeys a LDP on } \mathrm{P}(E_1) \text{ with speed } n \text{ and good rate function } K^{(1)} \tag{10}$$

and

$$\{M_n\}_{n\ge1} \text{ obeys a LDP on } \mathrm{P}(E_2) \text{ with speed } n \text{ and good rate function} K^{(2)}. \tag{11}$$

Then $\{K_n/n\}_{n\ge1}$ *obeys a LDP on* $\mathbb{R}$ *with speed* n *and good rate function*

$$I(z) = \inf_{(\mu,\nu)\in\mathrm{P}(E_1)\times\mathrm{P}(E_2):\, C(\mu,\nu)=z} K^{(1)}(\mu) + K^{(2)}(\nu).$$

Proof of (i). The claim follows by relation (7), assumption (9), the continuity of C wrt the product weak topology (see the proof of Theorem 3.1) and the Contraction Principle (see e.g. Theorem 4.2.1 in [9]). □

Proof of (ii). Since $(\mathrm{P}(E_i), d_P^{(i)})$ is separable $(i = 1, 2)$, $\{(L_n, M_n)\}_{n\ge1}$ obeys a LDP on $\mathrm{P}(E_1) \times \mathrm{P}(E_2)$ with speed n and good rate function $K(\mu, \nu) = K^{(1)}(\mu) + K^{(2)}(\nu)$ (see e.g. Exercises 4.2.7 and 1.2.19 in [9]). The claim follows by part (i). □

The next Corollaries 3.3 and 3.4 are simple consequences of Theorems 3.1 and 3.2, and provide ASC and LDP for the two-sample and the grid transportation problems. Given two probability measures α_1, α_2 on a metric space (S, d_S), we consider the relative entropy of α_1 wrt α_2:

$$H(\alpha_1 \mid \alpha_2) = \begin{cases} \int_{S\times S} \frac{\mathrm{d}\alpha_1}{\mathrm{d}\alpha_2}(x) \log \frac{\mathrm{d}\alpha_1}{\mathrm{d}\alpha_2}(x)\,\mathrm{d}\alpha_2 & \text{if } \alpha_1 \ll \alpha_2 \\ +\infty & \text{otherwise} \end{cases}$$

where the symbol $\ll$ denotes absolute continuity between measures. In Corollary 3.3 K_n denotes the optimal transport cost of the TSTP, while in Corollary 3.4 K_n denotes the optimal transport cost of the GTP and $\{\gamma_n\}_{n\geq 1}$ the sequence of empirical measures defined by $\gamma_n := \sum_{i=1}^n \delta_{g_i}/n$, where the g's represent the grid locations.

Corollary 3.3. *Assume (8). Then:*

(i) $\{K_n/n\}_{n\geq 1}$ *converges almost surely on* $\mathbb{R}$ *to* $C(\ell_1, \ell_2)$.
(ii) *If in addition (3) holds, then* $\{K_n/n\}_{n\geq 1}$ *obeys a LDP on* $\mathbb{R}$ *with speed* n *and good rate function*

$$I_{\mathrm{TSTP}}(z) = \inf_{(\mu,\nu)\in \mathrm{P}(E_1)\times \mathrm{P}(E_2):\, C(\mu,\nu)=z} H(\mu \mid \ell_1) + H(\nu \mid \ell_2). \tag{12}$$

Corollary 3.4. *Assume (8) and*

$$\gamma_n \text{ converges weakly to } \ell^* \in \mathrm{P}(E_2), \quad \text{as } n \to \infty. \tag{13}$$

Then:

(i) $\{K_n/n\}_{n\geq 1}$ *converges almost surely on* $\mathbb{R}$ *to* $C(\ell_1, \ell^*)$.
(ii) $\{K_n/n\}_{n\geq 1}$ *obeys a LDP on* $\mathbb{R}$ *with speed* n *and good rate function*

$$I_{\mathrm{GTP}}(z) = \inf_{\mu\in \mathrm{P}(E_1):\, C(\mu,\ell^*)=z} H(\mu \mid \ell_1). \tag{14}$$

Before proving these results we give an example.

Example 1. (1) Assumption (13) is satisfied in the case of the kDRGTP with $\ell^* = \ell$, the Lebesgue measure on $[0, 1]^k$. Indeed, for any $\varphi \in C_b([0, 1]^k)$, $\int_{[0,1]^k} \varphi(x)\gamma_n(\mathrm{d}x) = (\sum_{i=1}^n \varphi(g_i))/n$ and the term in the right-hand side can be bounded from above and from below by Riemann's sums of φ on $[0, 1]^k$, wrt the partition $\{C_1, \ldots, C_n\}$. Since the Riemann sums converge to $\ell(\varphi)$, we have that $\gamma_n \to \ell$ weakly.

(2) Suppose that the grid points are not all distinct, i.e. $\{g_1, g_2, \ldots\} = \{x_1, \ldots, x_m\}$ for some positive integer m, where $x_1, \ldots, x_m$ are distinct points of E_2. For $n \geq 1$ and $j = 1, \ldots, m$, define $n_j(n) := \sharp\{i \in \{1, \ldots, n\} : g_i = x_j\}$. In addition, suppose that there exist $p_1, \ldots, p_m \in [0, 1]$ such that $\sum_j p_j = 1$ and $n_j(n)/(np_j) \to 1$, as $n \to \infty$. Then assumption (13) is satisfied with $\ell^* := \sum_j p_j \delta_{x_j}$. A possible choice of the integers $n_j(n)$ is the following. Let $[x]$ denote the integer part of $x \in \mathbb{R}$; since $[x] \leq x \leq [x] + 1$ and $\sum_{j=1}^m p_j = 1$, we have $\sum_{j=1}^m ([np_j] + 1) \geq n \geq \sum_{j=1}^m [np_j]$. Therefore, there exist $h_1, \ldots, h_m \in \{0, 1\}$ such that $\sum_{j=1}^m ([np_j] + h_j) = n$, and we may take $n_j(n) = [np_j] + h_j$.

Proof of Corollary 3.3(i). By (1), (2) and the ASC for empirical laws we have that $\{L_n\}_{n\geq 1}$ converges almost surely on $\mathrm{P}(E_1)$ to ℓ_1 and $\{M_n\}_{n\geq 1}$ converges almost surely on $\mathrm{P}(E_2)$ to ℓ_2. The claim follows by Theorem 3.1. □

Proof of Corollary 3.3(ii). It is a simple consequence of Theorem 3.2(ii) and Sanov's theorem (see e.g. Theorem 6.2.10 in [9]). □

Proof of Corollary 3.4(i). As noticed in the proof of Corollary 3.3(i), $\{L_n\}_{n\geq 1}$ converges almost surely on $\mathrm{P}(E_1)$ to ℓ_1. The claim follows by (13) and Theorem 3.1. □

Proof of Corollary 3.4(ii). By Sanov's theorem, $\{L_n\}_{n\geq 1}$ obeys a LDP on $\mathrm{P}(E_1)$ with speed n and good rate function $K^{(1)}(\cdot) = H(\cdot \mid \ell_1)$. By (13) we deduce that $\{\gamma_n\}_{n\geq 1}$ obeys a LDP on $\mathrm{P}(E_2)$ with speed n and good rate function $K^{(2)}(\nu) = 0$ if $\nu = \ell^*$ and $K^{(2)}(\nu) = +\infty$ if $\nu \neq \ell^*$. The conclusion follows by Theorem 3.2(ii). □

Note that if the TSTP is specified by $E_1 = E_2 = E$, $d_1 = d_2 = d$, where d is a bounded distance, and $c = d^p$, $p \geq 1$, then $C = W_p^p$, i.e. C is the power of a particular distance. So, by Corollary 3.3 we deduce that the scaling $1/n$ is optimal for K_n, i.e. $\{K_n/n\}_{n\geq 1}$ converges almost surely on $\mathbb{R}$ to a strictly positive limit if and only if $\ell_1 \neq \ell_2$. In particular, for the 1DTSTP over a compact interval E with Euclidean DCF, CIL and $\ell_1 = \ell_2$, the scaling $1/n$ is not optimal for K_n. We shall prove that (in the sense of convergence in law) the optimal scaling is given by $1/\sqrt{n}$ (see Theorem 7.1). Similarly, consider the GTP specified by $E_1 = E_2 = E$, $d_1 = d_2 = d$, where d is a bounded distance, and $c = d^p$, $p \geq 1$. If (13) holds, then by Corollary 3.4 we have that the scaling $1/n$ is optimal for K_n if and only if $\ell_1 \neq \ell^*$. In particular, if (13) holds with $\ell^* = \ell_1$ and $\{\gamma_n\}_{n\geq 1}$ satisfies a suitable assumption then, for the 1DGTP specified by $E_1 = E_2 = E \subset \mathbb{R}$ and $d_1 = d_2 = c = d$, where E is a bounded Borel set and d is the Euclidean distance, we shall prove that the optimal scaling is given by $1/\sqrt{n}$ (see Theorem 7.4).

4. On the expression of the large deviation rate functions

In this section we provide some sharp lower bounds for the large deviation rate function. Moreover, for a specific 1DRGTP over the compact interval [0, 1] we express the large deviation rate function of $\{K_n/n\}_{n\geq 1}$ in terms of a one-dimensional optimization problem.

4.1. Some lower bounds for the large deviation rate functions

Let (E, d) be a compact metric space, and assume $E_1 = E_2 = E$, $d_1 = d_2 = c = d$.

For $\ell_1 = \ell_2$, a characterization of the large deviation rate function I_{TSTP} in (12) was given in [12] where it is proved (see Theorem 3.3 therein)

$$I_{\mathrm{TSTP}}(z) = \begin{cases} \inf\limits_{\varphi \in \mathcal{L}^0(E)} \mathrm{L}_\varphi^*(z) & \text{if } z \geq 0 \\ +\infty & \text{if } z < 0, \end{cases} \tag{15}$$

where $\mathcal{L}^0(E) := \{\varphi : E \to \mathbb{R} : \varphi \text{ is Lipschitz}, \|\varphi\|_{\mathrm{Lip}} = 1, \ell_1(\varphi) = 0\}$ and, for $\theta, z \in \mathbb{R}$,

$$\mathrm{L}_\varphi^*(z) := \sup_{\theta \in \mathbb{R}} (\theta z - \mathrm{L}_\varphi(\theta)), \qquad \mathrm{L}_\varphi(\theta) := \Lambda_\varphi(\theta) + \Lambda_\varphi(-\theta),$$

$$\Lambda_\varphi(\theta) := \log \int_E \mathrm{e}^{\theta\varphi(x)}\, \ell_1(\mathrm{d}x).$$

For $\ell^* = \ell_1$, a characterization of the large deviation rate function I_{GTP} in (14) may be proved similarly. In particular, one has

$$I_{\mathrm{GTP}}(z) = \begin{cases} \inf_{\varphi\in\mathcal{L}^0(E)} \Lambda^*_\varphi(z) & \text{if } z \geq 0 \\ +\infty \quad \text{if } z < 0, \end{cases} \tag{16}$$

where $\Lambda^*_\varphi(z) := \sup_{\theta\in\mathbb{R}}(\theta z - \Lambda_\varphi(\theta))$.

The next result provides an explicit lower bound for the large deviation rate functions I_{TSTP} and I_{GTP}.

Theorem 4.1. *Under the foregoing assumptions we have*

$$I_{\mathrm{TSTP}}(z) \geq \lambda z^2/4 \quad \textit{and} \quad I_{\mathrm{GTP}}(z) \geq \lambda z^2/2, \quad \forall z \geq 0, \tag{17}$$

where

$$\lambda := \frac{1}{2\left\{\inf_{x_0\in E,\alpha>0}\left[(2\alpha)^{-1}\left(1+\log\int_E \mathrm{e}^{\alpha d(x_0,x)^2}\,\ell_1(\mathrm{d}x)\right)\right]^{1/2}\right\}^2}. \tag{18}$$

Proof. We only show the first inequality in (17). The second one can be proved similarly. By the results in [6] (see Corollary 4 therein) we have that ℓ_1 satisfies Talagrand's $T_1(\lambda)$ inequality

$$W_1(\nu, \ell_1) \leq \sqrt{\frac{2}{\lambda}H(\nu \mid \ell_1)}, \quad \forall \nu \in \mathrm{P}(E). \tag{19}$$

By Bobkov and Götze's [4] characterization of the $T_1(\lambda)$ inequality we have, for all $\theta \geq 0$ and all φ Lipschitz with $\|\varphi\|_{\mathrm{Lip}} \leq 1$, $\log\int_E \mathrm{e}^{\theta(\varphi(x)-\ell_1(\varphi))}\,\ell_1(\mathrm{d}x) \leq \theta^2/(2\lambda)$. In particular, for all $\theta \geq 0$ and $\varphi \in \mathcal{L}^0(E)$, $\Lambda_\varphi(\pm\theta) \leq \theta^2/(2\lambda)$. So, for all $z \geq 0$,

$$\begin{aligned} I_{\mathrm{TSTP}}(z) = \inf_{\varphi\in\mathcal{L}^0(E)} \mathrm{L}^*_\varphi(z) &\geq \sup_{\theta\in\mathbb{R}}\inf_{\varphi\in\mathcal{L}^0(E)}(\theta z - \Lambda_\varphi(\theta) - \Lambda_\varphi(-\theta)) \\ &\geq \sup_{\theta\geq 0}\left(\theta z - \frac{\theta^2}{\lambda}\right) = \lambda z^2/4. \quad \square \end{aligned}$$

Let (E, d) be a Polish space. Finally, we investigate the relation between Maurey's τ-property [19] and the large deviation rate function of the TSTP with $E_1 = E_2 = E$, $d_1 = d_2 = d$, $\ell_1 = \ell_2$ and CIL. Let $c \in C_b(E^2)$. For all bounded Borel functions $\varphi : E \to \mathbb{R}$, define $Q_c\varphi(x) := \inf_{y\in E}(\varphi(y)+c(x, y))$, $x \in E$. We say that the couple (ℓ_1, c) satisfies the τ-property if $\ell_1(Q_c\varphi)\ell_1(-\varphi) \leq 1$, for any φ.

Theorem 4.2. *For the* TSTP *specified above we have*

$$I_{\mathrm{TSTP}}(z) \geq z, \quad \forall z \geq 0 \quad \textit{if and only if} \quad \textit{the couple } (\ell_1, c) \textit{ has the } \tau\textit{-property}.$$

Before proving this result we give an example.

Example 2. Take $E = \mathbb{R}$, $d(x, y) = |x - y|/(1 + |x - y|)$ for any $x, y \in \mathbb{R}$, and ℓ_1 equal to the standard Gaussian measure on $\mathbb{R}$. By Talagrand's T_2 inequality we deduce

$$\inf_{\pi\in\Pi(\mu,\ell_1)}\int_{\mathbb{R}\times\mathbb{R}} |x - y|^2\,\pi(\mathrm{d}x\,\mathrm{d}y) \leq 2H(\mu \mid \ell_1), \quad \forall \mu \in \mathrm{P}(\mathbb{R}).$$

Therefore, defining the cost function $c(x, y) := d(x, y)^2/2$, we have the transport–entropy inequality: $C(\mu, \ell_1) \leq H(\mu \mid \ell_1)$, for any $\mu \in \mathrm{P}(\mathbb{R})$. So by e.g. Proposition 8.3 in [14] we have that the couple $(\ell_1, \widetilde{c})$, where $\widetilde{c}(x, y) := d(x, y)^2/4$, satisfies the τ-property. Now consider the TSTP with $E_1 = E_2 = \mathbb{R}$, $d_1 = d_2 = d$, cost function $\widetilde{c}$ and completely independent locations distributed according to the standard Gaussian measure. By Theorem 4.2 we deduce $I_{\mathrm{TSTP}}(z) \geq z$, for any $z \in \mathbb{R}$.

Proof of Theorem 4.2. Assume that the τ-property holds. By Proposition 8.2 in [14] we have $H(\mu \mid \ell_1) + H(\nu \mid \ell_1) \geq C(\mu, \nu)$, for any $\mu, \nu \in \mathrm{P}(E)$. Therefore, for any $z \geq 0$,

$$I_{\mathrm{TSTP}}(z) = \inf_{(\mu,\nu)\in \mathrm{P}(E)\times \mathrm{P}(E):\, C(\mu,\nu)=z} H(\mu \mid \ell_1) + H(\nu \mid \ell_1) \geq z.$$

Now, assume $I_{\mathrm{TSTP}}(z) \geq z$, for all $z \geq 0$. Let Φ_c be the set of all bounded continuous functions such that $\phi_1(x) + \phi_2(y) \leq c(x, y)$, for any $x, y \in E$. By the Kantorovich duality, $C(\mu, \nu) = \sup_{(\phi_1,\phi_2)\in\Phi_c}(\mu(\phi_1) + \nu(\phi_2))$. So, setting $T_n^{\phi} := L_n(\phi_1) + M_n(\phi_2)$, $\phi = (\phi_1, \phi_2) \in \Phi_c$, we have, for any $t \geq 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log P(T_n^{\phi} \geq t) \leq \limsup_{n\to\infty} \frac{1}{n} \log P(K_n/n \geq t) \leq -\inf_{z\geq t} I_{\mathrm{TSTP}}(z) \leq -t.$$

The claim follows using the techniques in the proof of Theorem 3.7 in [13] (see the equivalence $(a) \Leftrightarrow (c)$ therein).

4.2. On the large deviation rate function of a particular 1DRGTP

Let K_n denote the optimal transport cost of the 1DRGTP over the compact interval $[0, 1]$ with Euclidean DCF and $\ell_1 = \ell$, the Lebesgue measure. The following theorem holds.

Theorem 4.3. *$\{K_n/n\}_{n\geq 1}$ obeys a LDP on $\mathbb{R}$ with speed n and good rate function*

$$I_{\mathrm{GTP}}(z) = \begin{cases} 0 & \text{if } z = 0 \\ \inf_{a\in(0,1)} \gamma(a, z) & \text{if } z \in (0, 1/4] \\ \inf_{a\in(0,\sqrt{1/2-z})\cup(2-\sqrt{1/2-z},1)} \gamma(a, z) & \text{if } z \in (1/4, 1/2) \\ +\infty & \text{otherwise} \end{cases}$$

where $\gamma(a, z) := -\gamma^{(1)}(a, z) - \gamma^{(2)}(a, z)(z + a^2 + 1/2)$,

$$\gamma^{(1)}(a, z) := \log\left(-\frac{1}{\gamma^{(2)}(a, z)}\left(\mathrm{e}^{-2a\gamma^{(2)}(a,z)} - 2\mathrm{e}^{-a\gamma^{(2)}(a,z)} + \mathrm{e}^{-\gamma^{(2)}(a,z)}\right)\right)$$

and $\gamma^{(2)}(a, z)$ is solution on $(-\infty, 0)$ of the following equation in θ:

$$\frac{-\mathrm{e}^{-\theta}\left(\frac{1}{\theta} + \frac{1}{\theta^2}\right) + \mathrm{e}^{-a\theta}\left(\frac{2a}{\theta} + \frac{2}{\theta^2}\right) - \mathrm{e}^{-2a\theta}\left(\frac{2a}{\theta} + \frac{1}{\theta^2}\right)}{z + a^2 + 1/2} + \frac{1}{\theta}\left(\mathrm{e}^{-2a\theta} - 2\mathrm{e}^{-a\theta} + \mathrm{e}^{-\theta}\right)$$
$$= 0.$$

Using standard notation, for positive functions f and g such that $f(z), g(z) \to 0$, as $z \to 0$, we write $f(z) = O(g(z))$ if there exist positive constants $m, M > 0$ such that

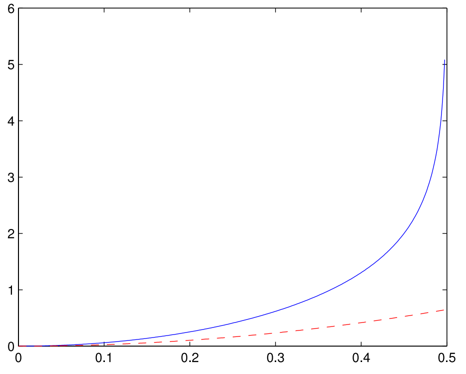

Fig. 1. Approximate graph of the rate function $I_{\mathrm{GTP}}(z)$, $0 \le z < 1/2$ (solid line), and approximate graph of the corresponding theoretical lower bound $\lambda z^2/2$, $0 \le z \le 1/2$, $\lambda \simeq 5.20$ (dotted line).

$m \le f(z)/g(z) \le M$, as $z \to 0$. As a consequence of Theorems 4.1 and 4.3, we have the following corollary whose proof will be given at the end of this section.

Corollary 4.4. *Let I_{GTP} be the rate function in Theorem 4.3. It holds that $I_{\mathrm{GTP}}(z) = O(z^2)$ as $z \to 0$.*

Using the expression of the rate function given in Theorem 4.3, we may numerically estimate I_{GTP}. In Fig. 1 we report the approximate graph of the rate function $I_{\mathrm{GTP}}(z)$, $0 \le z < 1/2$ (solid line), and the approximate graph of the corresponding theoretical lower bound $\lambda z^2/2$, $0 \le z \le 1/2$ (dotted line), provided by Theorem 4.1. In particular, for the graph of the lower bound, we evaluated numerically λ, resulting $\lambda \simeq 5.20$.

The proof of Theorem 4.3 is based on the following Lemma 4.5. Let $\mathcal{F}([0,1])$ be the set of probability densities on $[0,1]$, and define

$$\mathcal{B}_{a,z} := \left\{ f \in \mathcal{F}([0,1]) : \int_0^1 \sigma_a(x)(f(x)-1)\,\mathrm{d}x = z \right\}, \quad a \in (0,1), z \ge 0,$$

where $\sigma_a(x) := \max\{2a - x, x\}$.

Lemma 4.5. *For $a \in (0,1)$ and $z \in [0, \max\{1, 2a\} - a^2 - 1/2)$, the variational problem*

$$\inf_{f \in \mathcal{B}_{a,z}} \int_0^1 f(x) \log f(x)\,\mathrm{d}x \tag{20}$$

has a unique solution $f^*_{a,z} \in \mathcal{B}_{a,z}$ given by $f^*_{a,z}(x) := \mathrm{e}^{-\gamma^{(1)}(a,z) - \sigma_a(x)\gamma^{(2)}(a,z)}\mathbf{1}\{x \in [0,1]\}$, where $\gamma^{(1)}(a,z)$ and $\gamma^{(2)}(a,z)$ are defined in the statement of Theorem 4.3.

Proof of Theorem 4.3. A simple computation shows that the claim is equivalent to $I_{\mathrm{GTP}}(z) = \inf_{a \in \mathrm{A}_z} \gamma(a,z)$, $z \in \mathbb{R}$, where $\mathrm{A}_z := \{x \in (0,1) : z \in [0, \max\{1, 2x\} - x^2 - 1/2)\}$. Clearly, $I_{\mathrm{GTP}}(z) = +\infty$ for $z < 0$ and by the ASC, $I_{\mathrm{GTP}}(0) = 0$. So, in the rest of the proof, we consider $z > 0$. By (6) we have $W_1(\mu,\nu) = \int_0^1 |\mu([0,x]) - \nu([0,x])|\,\mathrm{d}x$; therefore $I_{\mathrm{GTP}}(z) = \inf_{f \in \mathcal{A}_z} \int_0^1 f(x)\log f(x)\,\mathrm{d}x$, where $\mathcal{A}_z := \left\{ f \in \mathcal{F}([0,1]) : \int_0^1 |F(x) - x|\,\mathrm{d}x = z \right\}$ and $F(x) := \int_0^x f(y)\,\mathrm{d}y$. We divide the proof in four steps.

Step 1: *a preliminary inequality*. For $a \in (0,1)$ and $z \geq 0$, denote by $\mathcal{A}_{a,z}$ the set of probability densities $f \in \mathcal{F}([0,1])$ such that

$$\begin{aligned} &F(x) - x \geq 0, \quad \text{for } \ell\text{-almost all } x \in (0,a);\\ &F(x) - x < 0, \quad \text{for } \ell\text{-almost all } x \in (a,1) \end{aligned} \tag{21}$$

and $\int_0^1 |F(x) - x|\, \mathrm{d}x = z$. If f satisfies (21), then integration by parts yields

$$\begin{aligned} \int_0^1 |F(x) - x|\, \mathrm{d}x &= \int_0^a F(x)\, \mathrm{d}x - \int_a^1 F(x)\, \mathrm{d}x - a^2 + \frac{1}{2}\\ &= 2aF(a) - \int_0^a x f(x)\, \mathrm{d}x + \int_a^1 x f(x)\, \mathrm{d}x - a^2 - \frac{1}{2}\\ &= \int_0^1 \sigma_a(x)(f(x) - 1)\, \mathrm{d}x. \end{aligned}$$

So $\mathcal{A}_{a,z} \subseteq \mathcal{B}_{a,z}$, and therefore

$$\inf_{f \in \mathcal{B}_{a,z}} \int_0^1 f(x) \log f(x)\, \mathrm{d}x \leq \inf_{f \in \mathcal{A}_{a,z}} \int_0^1 f(x) \log f(x)\, \mathrm{d}x. \tag{22}$$

Step 2: *a related variational problem*. Since $\mathcal{B}_{a,z} = \emptyset$ for $z \geq \max\{1, 2a\} - a^2 - 1/2$, by (22) we have

$$\inf_{f \in \mathcal{A}_{a,z}} \int_0^1 f(x) \log f(x)\, \mathrm{d}x = +\infty, \quad \forall a \in (0,1) \text{ and } z \geq \max\{1, 2a\} - a^2 - 1/2.$$

In the following we show that, for $a \in (0,1)$ and $z \in [0, \max\{1, 2a\} - a^2 - 1/2)$, the unique solution $f^*_{a,z}$ of (20) solves even $\inf_{f \in \mathcal{A}_{a,z}} \int_0^1 f(x) \log f(x)\, \mathrm{d}x$. Due to (22), this claim follows if we prove $f^*_{a,z} \in \mathcal{A}_{a,z}$. We reason by contradiction and assume the existence of a Borel set $C \subseteq (0,a)$ with positive Lebesgue measure and such that $F^*_{a,z}(x) := \int_0^x f^*_{a,z}(y)\, \mathrm{d}y < x$ on C (the same argument to follow can be easily adapted to the case when $F^*_{z,a}(x) \geq x$ on some Borel set $C \subseteq (a,1)$ of positive Lebesgue measure). By continuity,

$$F^*_{a,z}(x) < x \quad \text{on some interval } (u,v) \subseteq (0,a). \tag{23}$$

Without loss of generality we may assume

$$F^*_{a,z}(u) = u, \qquad F^*_{a,z}(v) = v. \tag{24}$$

Consider the distribution function $G^*_{a,z}(x) := x\mathbf{1}\{x \in (u,v)\} + F^*_{a,z}(x)\mathbf{1}\{x \in (0,1) \setminus (u,v)\}$ and let $g^*_{a,z}(x) := \mathbf{1}\{x \in (u,v)\} + f^*_{a,z}(x)\mathbf{1}\{x \in (0,1) \setminus (u,v)\}$ be its density wrt ℓ. We have

$$\int_0^1 \sigma_a(x)(g^*_{a,z}(x) - 1)\, \mathrm{d}x > \int_0^1 \sigma_a(x)(f^*_{a,z}(x) - 1)\, \mathrm{d}x = z. \tag{25}$$

Indeed, the equality follows since $f^*_{a,z} \in \mathcal{B}_{a,z}$ and the strict inequality is a consequence of the following argument. First note that since $g^*_{a,z} \equiv f^*_{a,z}$ on $(a,1)$, it is equivalent to

$$2aG^*_{a,z}(a) - \int_0^a x g^*_{a,z}(x)\, \mathrm{d}x > 2aF^*_{a,z}(a) - \int_0^a x f^*_{a,z}(x)\, \mathrm{d}x.$$

Using $g^*_{a,z} \equiv f^*_{a,z}$ on $(0,a) \setminus (u,v)$, $g^*_{a,z} \equiv 1$ on (u,v), (24) and integration by parts, this inequality is in turn equivalent to

$$\int_u^v x(1 - f^*_{a,z}(x))\,\mathrm{d}x = \int_u^v (F^*_{a,z}(x) - x)\,\mathrm{d}x < 0$$

and so the inequality in (25) is a consequence of (23). Define the functions $h^{(\alpha)}_{a,z}(x) := \alpha + (1-\alpha)g^*_{a,z}(x)$, for $\alpha \in (0,1)$ and $x \in (0,1)$. Clearly,

$$\int_0^1 \sigma_a(x)(h^{(1)}_{a,z}(x) - 1)\,\mathrm{d}x = 0 \le z \quad \text{and (by (25))} \quad \int_0^1 \sigma_a(x)(h^{(0)}_{a,z}(x) - 1)\,\mathrm{d}x > z.$$

Since $h^{(\alpha)}_{a,z} \in \mathcal{F}([0,1])\ \forall \alpha \in (0,1)$, by the continuity of the map $\alpha \mapsto \int_0^1 \sigma_a(x)(h^{(\alpha)}_{a,z}(x) - 1)\,\mathrm{d}x$, we deduce the existence of $\overline{\alpha} \in (0,1)$ such that $h^{(\overline{\alpha})}_{a,z} \in \mathcal{B}_{a,z}$. We get a contradiction if we show

$$\int_0^1 h^{(\overline{\alpha})}_{a,z}(x) \log h^{(\overline{\alpha})}_{a,z}(x)\,\mathrm{d}x < \int_0^1 f^*_{a,z}(x) \log f^*_{a,z}(x)\,\mathrm{d}x. \tag{26}$$

By the strict convexity on [0, 1] of $x \mapsto x \log x$, we have

$$\int_0^1 h^{(\overline{\alpha})}_{a,z}(x) \log h^{(\overline{\alpha})}_{a,z}(x)\,\mathrm{d}x < (1-\overline{\alpha}) \int_0^1 g^*_{a,z}(x) \log g^*_{a,z}(x)\,\mathrm{d}x,$$

so (26) is proved if we check

$$\int_u^v f^*_{a,z}(x) \log f^*_{a,z}(x)\,\mathrm{d}x \ge 0. \tag{27}$$

For $x \in (0,1)$, define the function $\varphi(x) := f^*_{a,z}(u + (v-u)x)$. By (24) we have $\ell(\varphi) = (v-u)^{-1} \int_u^v f^*_{a,z}(x)\,\mathrm{d}x = 1$. The inequality (27) follows since

$$\int_u^v f^*_{a,z}(x) \log f^*_{a,z}(x)\,\mathrm{d}x = (v-u) \int_0^1 \varphi(x) \log \varphi(x)\,\mathrm{d}x \ge 0.$$

Step 3: *monotonicity of* I_{GTP} *on* $[0,\infty)$. Note that, for the grid transportation problem we are considering, it holds that $\ell^* = \ell_1 = \ell$. So, by (16) we have that I_{GTP} is non-decreasing on $[0,\infty)$ if, for any fixed $\varphi \in \mathcal{L}^0([0,1])$, the Legendre transform Λ^*_φ is also such. This is immediate from the fact that Λ^*_φ is non-negative, convex and $\Lambda^*_\varphi(0) = 0$.

Step 4: *conclusion of the proof*. Since $\mathcal{A}_{a,z} \subseteq \mathcal{A}_z$, we have, for any $a \in (0,1)$ and $z \ge 0$,

$$I_{\mathrm{GTP}}(z) \le I(a,z) := \inf_{f \in \mathcal{A}_{a,z}} \int_0^1 f(x) \log f(x)\,\mathrm{d}x.$$

Note that by Lemma 4.5 and Step 2, we deduce $I(a,z) = \gamma(a,z)$ if $z \in [0, \max\{1, 2a\} - a^2 - 1/2)$ and equal to $+\infty$ otherwise. Taking the infimum over $a \in (0,1)$, we get

$$I_{\mathrm{GTP}}(z) \le \inf_{a \in \mathrm{A}_z} I(a,z) = \inf_{a \in \mathrm{A}_z} [-\gamma^{(1)}(a,z) - \gamma^{(2)}(a,z)(z + a^2 + 1/2)], \quad \forall z \ge 0. \tag{28}$$

The reversed inequality follows if we show that, for fixed $z \ge 0$ and $f \in \mathcal{A}_z$,

$$\exists z^* \ge z, \overline{a} \in \mathrm{A}_{z^*}, g_{\overline{a}} \in \mathcal{A}_{\overline{a},z^*}$$
$$\text{such that} \int_0^1 f(x) \log f(x)\,\mathrm{d}x = \int_0^1 g_{\overline{a}}(x) \log g_{\overline{a}}(x)\,\mathrm{d}x. \tag{29}$$

Indeed, (29) yields $I_{\mathrm{GTP}}(z) \geq \inf_{a\in \mathrm{A}_{z^*}} I(a, z^*)$ and, combining this inequality with Step 3 and (28), we have $I_{\mathrm{GTP}}(z) = \inf_{a\in \mathrm{A}_{z^*}} I(a, z^*) \geq \inf_{a\in \mathrm{A}_z} \gamma(a, z)$. It remains to prove (29). Take $z \geq 0$, $f \in \mathcal{A}_z$ and, to fix the ideas, suppose $F(x) - x \geq 0$ in a right neighborhood of 0. Denote by $J_k^{(+)} := (r_k, s_k) \subseteq (0, 1)$, $k \geq 1$, $r_1 = 0$, the disjoint intervals of maximal length such that $F(x) - x \geq 0$ and by $J_k^{(-)} := (s_k, r_{k+1}) \subseteq (0, 1)$, $k \geq 1$, the disjoint intervals of maximal length such that $F(x) - x < 0$. If there are only N intervals $J_k^{(+)}$ and $N - 1$ intervals $J_k^{(-)}$, then we set $s_N := 1$ and $r_k = s_k := 1$ $\forall k \geq N + 1$. If there are only N intervals $J_k^{(+)}$ and $J_k^{(-)}$, then we set $r_k = s_k := 1$ $\forall k \geq N + 1$. Define

$$z^* := z + 2\sum_{k\geq 1}(r_{k+1} - s_k) \sum_{j\geq k+1} (s_j - r_j) \quad \text{and} \quad \overline{a} := \sum_{k\geq 1}(s_k - r_k)$$

(note that, due to the above definitions, if the number of intervals $J_k^{(+)}$ is equal to N, then the sums $\sum_{k\geq 1}$ and $\sum_{j\geq k+1}$, in the definition of z^*, have to be replaced by $\sum_{k=1}^{N-1}$ and $\sum_{j=k+1}^{N}$, respectively, and the sum $\sum_{k\geq 1}$, in the definition of $\overline{a}$, by $\sum_{k=1}^{N}$), and consider the distribution function G defined as follows:

$$\text{for } n \geq 0, \quad G\left(\sum_{k=1}^{n}(s_k - r_k) + x\right) := F(r_{n+1} + x) \quad \text{if } x \in (0, s_{n+1} - r_{n+1})$$

and

$$\text{for } n \geq 0, \quad G\left(\overline{a} + \sum_{k=1}^{n}(r_{k+1} - s_k) + x\right) := F(s_{n+1} + x) \quad \text{if } x \in (0, r_{n+2} - s_{n+1}).$$

Let $g_{\overline{a}}$ be the density of G wrt ℓ. The claim follows if we check that the triplet $(z^*, \overline{a}, g_{\overline{a}})$ has the properties summarized in (29). In the rest of the proof we accomplish this task in the case where there are exactly $N \geq 1$ intervals $J_k^{(+)}$ and $J_k^{(-)}$. The other case may be checked similarly. We have

$$z^* = z + 2\sum_{k=1}^{N}(r_{k+1} - s_k) \sum_{j=k+1}^{N} (s_j - r_j) \quad \text{and} \quad \overline{a} = \sum_{k=1}^{N}(s_k - r_k)$$

and G defined by

$$\text{for } n = 0, \ldots, N-1, \quad G\left(\sum_{k=1}^{n}(s_k - r_k) + x\right) := F(r_{n+1} + x)$$
$$\text{if } x \in (0, s_{n+1} - r_{n+1})$$

and

$$\text{for } n = 0, \ldots, N-1, \quad G\left(\overline{a} + \sum_{k=1}^{n}(r_{k+1} - s_k) + x\right) := F(s_{n+1} + x)$$
$$\text{if } x \in (0, r_{n+2} - s_{n+1}).$$

We start by checking that

$$G(x) - x \geq 0 \quad \text{on } (0, \overline{a}) \quad \text{and} \quad G(x) - x < 0 \quad \text{on } (\overline{a}, 1). \tag{30}$$

By construction, for $n = 0, \ldots, N-1$, on $(0, s_{n+1} - r_{n+1})$ we have

$$G\left(\sum_{k=1}^{n}(s_k - r_k) + x\right) - \left(\sum_{k=1}^{n}(s_k - r_k) + x\right) \geq F(r_{n+1} + x) - (r_{n+1} + x) \geq 0$$

and therefore $G(x) - x \geq 0$ on $(0, \overline{a})$. Similarly, for $n = 0, \ldots, N-1$, on $(0, r_{n+2} - s_{n+1})$ we have

$$G\left(\overline{a} + \sum_{k=1}^{n}(r_{k+1} - s_k) + x\right) - \left(\overline{a} + \sum_{k=1}^{n}(r_{k+1} - s_k) + x\right) \leq F(s_{n+1} + x) - (s_{n+1} + x) < 0$$

and therefore $G(x) - x < 0$ on $(\overline{a}, 1)$. Thus, (30) is proved, and by the computations at the beginning of Step 1 we deduce

$$\int_0^1 |G(x) - x| \, \mathrm{d}x = \int_0^1 \sigma_{\overline{a}}(x)(g_{\overline{a}}(x) - 1) \, \mathrm{d}x. \tag{31}$$

A straightforward computation yields

$$\begin{aligned}\int_0^1 |F(x) - x| \, \mathrm{d}x &= \sum_{k=1}^{N} \int_{r_k}^{s_k} (F(x) - x) \, \mathrm{d}x - \sum_{k=1}^{N} \int_{s_k}^{r_{k+1}} (F(x) - x) \, \mathrm{d}x \\ &= \sum_{k=1}^{N} \int_0^{s_k - r_k} \left[G\left(\sum_{j=1}^{k-1}(s_j - r_j) + x\right) - (x + r_k) \right] \mathrm{d}x \\ &\quad - \sum_{k=1}^{N} \int_0^{r_{k+1} - s_k} \left[G\left(\overline{a} + \sum_{j=1}^{k-1}(r_{j+1} - s_j) + x\right) - (x + s_k) \right] \mathrm{d}x.\end{aligned}$$

Note that

$$\begin{aligned}&\int_0^{s_k - r_k} \left[G\left(\sum_{j=1}^{k-1}(s_j - r_j) + x\right) - (x + r_k) \right] \mathrm{d}x \\ &= \int_0^{s_k - r_k} \left[G\left(\sum_{j=1}^{k-1}(s_j - r_j) + x\right) - \left(\sum_{j=1}^{k-1}(s_j - r_j) + x\right) \right] \mathrm{d}x \\ &\quad - \left[r_k - \sum_{j=1}^{k-1}(s_j - r_j) \right] (s_k - r_k),\end{aligned}$$

$$\begin{aligned}&\int_0^{r_{k+1} - s_k} \left[G\left(\overline{a} + \sum_{j=1}^{k-1}(r_{j+1} - s_j) + x\right) - (x + s_k) \right] \mathrm{d}x \\ &= \int_0^{r_{k+1} - s_k} \left[G\left(\overline{a} + \sum_{j=1}^{k-1}(r_{j+1} - s_j) + x\right) - \left(x + \overline{a} + \sum_{j=1}^{k-1}(r_{j+1} - s_j)\right) \right] \mathrm{d}x \\ &\quad + (r_{k+1} - s_k) \sum_{j=k+1}^{N} (s_j - r_j)\end{aligned}$$

and

$$\sum_{k=1}^{N}\left[r_k-\sum_{j=1}^{k-1}(s_j-r_j)\right](s_k-r_k)=\sum_{k=1}^{N}(r_{k+1}-s_k)\sum_{j=k+1}^{N}(s_j-r_j).$$

Therefore

$$\begin{aligned}&\int_0^1 |F(x)-x|\,\mathrm{d}x\\&\quad=\int_0^{\overline{a}}(G(x)-x)\,\mathrm{d}x-\int_{\overline{a}}^1(G(x)-x)\,\mathrm{d}x-2\sum_{k=1}^{N}(r_{k+1}-s_k)\sum_{j=k+1}^{N}(s_j-r_j).\end{aligned}$$

By (30), $f\in\mathcal{A}_z$ and the definition of z^*, we have $\int_0^1|G(x)-x|\,\mathrm{d}x=z^*$. Consequently, $g_{\overline{a}}\in\mathcal{A}_{\overline{a},z^*}$. By (31) we deduce $z^*=\int_0^1\sigma_{\overline{a}}(x)g_{\overline{a}}(x)\,\mathrm{d}x-\overline{a}^2-1/2<\max\{2\overline{a},1\}-\overline{a}^2-1/2$, and so $\overline{a}\in \mathrm{A}_{z^*}$. Finally

$$\begin{aligned}\int_0^1 f(x)\log f(x)\,\mathrm{d}x&=\sum_{k=1}^{N}\int_0^{s_k-r_k}g_{\overline{a}}\left(\sum_{j=1}^{k-1}(s_j-r_j)+x\right)\log g_{\overline{a}}\left(\sum_{j=1}^{k-1}(s_j-r_j)+x\right)\mathrm{d}x\\&\quad+\sum_{k=1}^{N}\int_0^{r_{k+1}-s_k}g_{\overline{a}}\left(\overline{a}+\sum_{j=1}^{k-1}(r_{j+1}-s_j)+x\right)\\&\quad\times\log g_{\overline{a}}\left(\overline{a}+\sum_{j=1}^{k-1}(r_{j+1}-s_j)+x\right)\mathrm{d}x\\&=\int_0^1 g_{\overline{a}}(x)\log g_{\overline{a}}(x)\,\mathrm{d}x.\quad\square\end{aligned}$$

Proof of Lemma 4.5. For fixed $a\in(0,1)$ and $z\in[0,\max\{1,2a\}-a^2-1/2)$, the set $\mathcal{B}_{a,z}$ is convex. So, if it is not empty, due to the strict convexity of the relative entropy, (20) has a unique solution, say $\overline{f}_{a,z}$. We compute $\overline{f}_{a,z}$ and check retrospectively that $\mathcal{B}_{a,z}$ is not empty. For the sake of clarity, we divide the proof in four steps.

Step 1: *the Lagrange multipliers method*. Set $A_{a,z}:=\{x\in[0,1]:\overline{f}_{a,z}(x)>0\}$, and consider the Lagrangian

$$\begin{aligned}&\mathcal{L}(f,\lambda_1(A_{a,z}),\lambda_2(A_{a,z}))(x)\\&\quad:=f(x)\log f(x)+\lambda_1(A_{a,z})(f(x)-1)+\lambda_2(A_{a,z})(\sigma_a(x)(f(x)-1)-z)\end{aligned}$$

where $\lambda_1(A_{a,z})$ and $\lambda_2(A_{a,z})$ are the Lagrange multipliers. By the Euler equation (see e.g. [7]) we have $\frac{\partial\mathcal{L}}{\partial f}|_{f=\overline{f}_{a,z}}=0$, i.e. $\overline{f}_{a,z}(x)=\mathrm{e}^{-(1+\lambda_1(A_{a,z}))-\sigma_a(x)\lambda_2(A_{a,z})}\mathbf{1}\{x\in A_{a,z}\}$. By the constraints we deduce that the Lagrange multipliers $\lambda_1(A_{a,z})$ and $\lambda_2(A_{a,z})$ are solutions of

$$\int_{A_{a,z}}\mathrm{e}^{-\sigma_a(x)\lambda_2(A_{a,z})}\,\mathrm{d}x=\mathrm{e}^{1+\lambda_1(A_{a,z})}\tag{32}$$

and

$$\int_{A_{a,z}}\sigma_a(x)\mathrm{e}^{-\sigma_a(x)\lambda_2(A_{a,z})}\,\mathrm{d}x=\mathrm{e}^{1+\lambda_1(A_{a,z})}(z+a^2+1/2).\tag{33}$$

Step 2: *a suitable set function*. For a fixed Borel set $U \subseteq [0, 1]$ with positive Lebesgue measure and $u_1, u_2 \in \mathbb{R}$ define the functions

$$m_a^{(0)}(U, u_1, u_2) := \int_U \mathrm{e}^{-(1+u_1)-\sigma_a(x)u_2}\,\mathrm{d}x,$$
$$m_a^{(1)}(U, u_1, u_2) := \int_U \sigma_a(x)\mathrm{e}^{-(1+u_1)-\sigma_a(x)u_2}\,\mathrm{d}x$$

and the set function $\nu_{a,z}(U) := \sup_{u_1,u_2\in\mathbb{R}}[-m_a^{(0)}(U, u_1, u_2) - u_1 - (z + a^2 + 1/2)u_2]$. If there exist $\lambda_1(U)$ and $\lambda_2(U)$, solutions of (32) and (33) with $A_{a,z} = U$, then $\nu_{a,z}(U) = -(1 + \lambda_1(U)) - (z + a^2 + 1/2)\lambda_2(U)$. Indeed,

$$\begin{cases} \dfrac{\partial}{\partial u_1}(-m_a^{(0)}(U, u_1, u_2) - u_1 - (z + a^2 + 1/2)u_2) = m_a^{(0)}(U, u_1, u_2) - 1 \\ \dfrac{\partial}{\partial u_2}(-m_a^{(0)}(U, u_1, u_2) - u_1 - (z + a^2 + 1/2)u_2) = m_a^{(1)}(U, u_1, u_2) - (z + a^2 + 1/2). \end{cases}$$

Note that the set function $\nu_{a,z}$ is non-increasing wrt the set inclusion, so

$$\nu_{a,z}([0, 1]) = \inf\{\nu_{a,z}(U) : U \subseteq [0, 1] \text{ Borel set such that } \ell(U) > 0\}. \tag{34}$$

Step 3: *end of the proof*. In the next step we shall show that $\lambda_1(a, z) := \gamma^{(1)}(a, z) - 1$ and $\lambda_2(a, z) := \gamma^{(2)}(a, z)$, where $\gamma^{(1)}$ and $\gamma^{(2)}$ are defined in the statement of the lemma, are solutions of (32) and (33) with $A_{a,z} = [0, 1]$. So $\nu_{a,z}([0, 1]) = \gamma(a, z)$ and therefore $\inf_{f\in\mathcal{B}_{a,z}} \int_0^1 f(x)\log f(x)\,\mathrm{d}x \le \nu_{a,z}([0, 1])$. The claim follows if we prove that this latter inequality is indeed an equality. This is guaranteed by the following computation:

$$\begin{aligned}\inf_{f\in\mathcal{B}_{a,z}} \int_0^1 f(x)\log f(x)\,\mathrm{d}x &= -(1 + \lambda_1(A_{a,z})) - (z + a^2 + 1/2)\lambda_2(A_{a,z}) = \nu_{a,z}(A_{a,z}) \\ &\ge \inf\{\nu_{a,z}(U) : U \subseteq [0, 1] \text{ Borel set such that } \ell(U) > 0\} = \nu_{a,z}([0, 1])\end{aligned}$$

where the latter equality is given by (34).

Step 4: *computation of the Lagrange multipliers*. We shall check that $\lambda_1 := \lambda_1(a, z)$ and $\lambda_2 := \lambda_2(a, z)$ are solutions of (32) and (33) with $A_{a,z} = [0, 1]$. Consider the system

$$\begin{cases} \displaystyle\int_0^1 \mathrm{e}^{-\sigma_a(x)\lambda_2}\,\mathrm{d}x = \mathrm{e}^{1+\lambda_1} \\ \displaystyle\int_0^1 \sigma_a(x)\mathrm{e}^{-\sigma_a(x)\lambda_2}\,\mathrm{d}x = \mathrm{e}^{1+\lambda_1}(z + a^2 + 1/2) \end{cases}$$

and define the function

$$G_{a,z}(\theta) := \frac{\int_0^1 \sigma_a(x)\mathrm{e}^{-\sigma_a(x)\theta}\,\mathrm{d}x}{z + a^2 + 1/2} - \int_0^1 \mathrm{e}^{-\sigma_a(x)\theta}\,\mathrm{d}x. \tag{35}$$

Note that

$$\lim_{\theta\to 0} G_{a,z}(\theta) = \frac{a^2 + 1/2}{z + a^2 + 1/2} - 1 \le 0. \tag{36}$$

Computing the integrals the system reduces to

$$\begin{cases} -\dfrac{1}{\lambda_2}\left(\mathrm{e}^{-2a\lambda_2} - 2\mathrm{e}^{-a\lambda_2} + \mathrm{e}^{-\lambda_2}\right) = \mathrm{e}^{1+\lambda_1} \\ -\mathrm{e}^{-\lambda_2}\left(\dfrac{1}{\lambda_2} + \dfrac{1}{\lambda_2^2}\right) + \mathrm{e}^{-a\lambda_2}\left(\dfrac{2a}{\lambda_2} + \dfrac{2}{\lambda_2^2}\right) - \mathrm{e}^{-2a\lambda_2}\left(\dfrac{2a}{\lambda_2} + \dfrac{1}{\lambda_2^2}\right) \\ \quad = \mathrm{e}^{1+\lambda_1}\left(z + a^2 + \dfrac{1}{2}\right), \end{cases}$$

and

$$G_{a,z}(\theta) = \frac{-\mathrm{e}^{-\theta}\left(\frac{1}{\theta} + \frac{1}{\theta^2}\right) + \mathrm{e}^{-a\theta}\left(\frac{2a}{\theta} + \frac{2}{\theta^2}\right) - \mathrm{e}^{-2a\theta}\left(\frac{2a}{\theta} + \frac{1}{\theta^2}\right)}{z + a^2 + 1/2} + \frac{1}{\theta}\left(\mathrm{e}^{-2a\theta} - 2\mathrm{e}^{-a\theta} + \mathrm{e}^{-\theta}\right). \tag{37}$$

We have

$$\lim_{\theta \to -\infty} G_{a,z}(\theta) = +\infty. \tag{38}$$

Indeed, $G_{a,z}(\theta)$ can be rewritten as

$$\frac{\mathrm{e}^{-\theta}}{\theta}\Bigg[1 + \mathrm{e}^{\theta(1-2a)}\left(1 - \kappa(z,a)\left(2a + \frac{1}{\theta}\right)\right) - 2\mathrm{e}^{\theta(1-a)}\left(1 - \kappa(z,a)\left(a + \frac{1}{\theta}\right)\right) - \kappa(z,a)\left(1 + \frac{1}{\theta}\right)\Bigg],$$

where we set $\kappa(z,a) := (z + a^2 + 1/2)^{-1}$. If $a < 1/2$, as $\theta \to -\infty$, the term inside the square brackets converges to $1 - \kappa(z,a)$, and this quantity is strictly less than zero since $z < 1/2 - a^2$. So (38) follows for $a < 1/2$. Similarly, for $a = 1/2$, as $\theta \to -\infty$, the term inside the square brackets converges to $2(1 - \kappa(z,a))$, which is again a strictly negative quantity. Finally, if $a > 1/2$, as $\theta \to -\infty$, the term inside the square brackets is asymptotically equivalent to $(1 - 2a\kappa(z,a))\mathrm{e}^{\theta(1-2a)}$. Multiplying this quantity by $\mathrm{e}^{-\theta}/\theta$ and passing to the limit as $\theta \to -\infty$ we get (38) (indeed $1 - 2a\kappa(z,a)$ is strictly less than zero since $z < 2a - a^2 - 1/2$). The conclusion follows by (36), (38) and the Intermediate Values Theorem. □

Proof of Corollary 4.4. By Theorem 4.1 the constant $\lambda/2$, where λ is defined by (18) with $E = [0,1]$, d the Euclidean distance and ℓ_1 the Lebesgue measure, is a lower bound for the ratio $I_{\mathrm{GTP}}(z)/z^2$, for any $z > 0$ (as already mentioned, a numerical evaluation of λ yields $\lambda \simeq 5.20$). So the claim follows if we show

$$\limsup_{z \to 0} \frac{I_{\mathrm{GTP}}(z)}{z^2} \le M, \tag{39}$$

for some positive constant $M > 0$. Let $G_{a,z}(\theta)$ be the function defined by (35), with $a \in [0, 1/2)$, $z \in [0, 1/4)$ and $\theta \in (-\infty, 0]$. Let $\gamma^{(2)}(a,z)$ be a solution of the equation in $\theta\, G_{a,z}(\theta) = 0$. By Dini's Implicit Function Theorem we have that $\gamma^{(2)}(0,z)$ is continuously differentiable (and strictly negative) on $(0, 1/4)$ and

$$\gamma^{(2)\prime}(0,z) := \frac{\mathrm{d}}{\mathrm{d}z}\gamma^{(2)}(0,z) = -\frac{\mathrm{d}}{\mathrm{d}z}G_{0,z}(\theta)\Big|_{\theta=\gamma^{(2)}(0,z)} \Big/ \frac{\mathrm{d}}{\mathrm{d}\theta}G_{0,z}(\theta)\Big|_{\theta=\gamma^{(2)}(0,z)}.$$

Using relation (37) (with $a = 0$), by an easy computation we deduce, for $z \in (0, 1/4)$,

$$\gamma^{(2)\prime}(0, z) = \frac{(z + 1/2)^{-2} A(z)}{A(z) + (z + 1/2)^{-1} B(z)},$$

where

$$A(z) := 1 - \mathrm{e}^{-\gamma^{(2)}(0,z)} - \gamma^{(2)}(0, z)\mathrm{e}^{-\gamma^{(2)}(0,z)}$$

and

$$B(z) := \frac{-2A(z) + (\gamma^{(2)}(0, z))^2 \mathrm{e}^{-\gamma^{(2)}(0,z)}}{\gamma^{(2)}(0, z)}.$$

By the continuity of $\gamma^{(2)}(0, z)$ on $[0, 1/4)$ we have $G_{0,0}(0) = 0 = G_{0,0}(\gamma^{(2)}(0, 0))$. Since $G'_{0,0}(\theta) < 0$ on $(-\infty, 0]$, we deduce $\gamma^{(2)}(0, 0) = 0$. By applying e.g. l'Hopital's rule we have

$$\lim_{z\to 0} \frac{A(z)}{(\gamma^{(2)}(0, z))^2} = 1/2 \quad \text{and} \quad \lim_{z\to 0} \frac{B(z)}{(\gamma^{(2)}(0, z))^2} = -1/3.$$

Therefore $\lim_{z\to 0} \gamma^{(2)\prime}(0, z) = -12$. An easy computation yields

$$\begin{aligned} \gamma^{(2)\prime\prime}(0, z) = & -2(z + 1/2)^{-3} \frac{A(z)^2}{[A(z) + (z + 1/2)^{-1} B(z)]^2} \\ & - (z + 1/2)^{-4} \frac{A(z)B(z)}{[A(z) + (z + 1/2)^{-1} B(z)]^2} \\ & + (z + 1/2)^{-3} \frac{A'(z)B(z) - A(z)B'(z)}{[A(z) + (z + 1/2)^{-1} B(z)]^2}. \end{aligned} \tag{40}$$

Now, we investigate the behavior of $A'(z)B(z) - A(z)B'(z)$, as $z \to 0$. After some manipulations, we have

$$\begin{aligned} & \frac{A'(z)B(z) - A(z)B'(z)}{(\gamma^{(2)}(0, z))^3} = \gamma^{(2)\prime}(0, z) \\ & \quad \times \frac{2A(z)^2/[\gamma^{(2)}(0, z)]^4 - \mathrm{e}^{-\gamma^{(2)}(0,z)} + (A(z)/[\gamma^{(2)}(0, z)]^2)\mathrm{e}^{-\gamma^{(2)}(0,z)}(1 - \gamma^{(2)}(0, z))}{\gamma^{(2)}(0, z)}, \end{aligned}$$

and so

$$\begin{aligned} & \lim_{z\to 0} \frac{A'(z)B(z) - A(z)B'(z)}{(\gamma^{(2)}(0, z))^3} \\ & \quad = -12 \lim_{z\to 0} \frac{2A(z)^2/[\gamma^{(2)}(0, z)]^4 - 1 + A(z)/[\gamma^{(2)}(0, z)]^2}{\gamma^{(2)}(0, z)}. \end{aligned} \tag{41}$$

Note that

$$\frac{\mathrm{d}}{\mathrm{d}z} \frac{A(z)}{\gamma^{(2)}(0, z)^2} = \frac{\gamma^{(2)\prime}(0, z)B(z)}{[\gamma^{(2)}(0, z)]^2} \to 4, \quad \text{as } z \to 0.$$

So, by l'Hopital's rule and (41) we have

$$\lim_{z\to 0}\frac{A'(z)B(z)-A(z)B'(z)}{(\gamma^{(2)}(0,z))^3}=12.$$

Therefore, by (40) it follows that

$$\lim_{z\to 0} z\gamma^{(2)''}(0,z)=\lim_{z\to 0}\left(\frac{z}{\gamma^{(2)}(0,z)}\right)\gamma^{(2)}(0,z)\gamma^{(2)''}(0,z)=-288. \tag{42}$$

Now, consider the function

$$\gamma^{(1)}(0,z):=\log\left(-\frac{\mathrm{e}^{-\gamma^{(2)}(0,z)}-1}{\gamma^{(2)}(0,z)}\right),\qquad z\in(0,1/4).$$

We have

$$\gamma^{(1)'}(0,z)=\gamma^{(2)'}(0,z)\frac{A(z)}{\gamma^{(2)}(0,z)(\mathrm{e}^{-\gamma^{(2)}(0,z)}-1)}$$

and

$$\begin{aligned}\gamma^{(1)''}(0,z)&=\gamma^{(2)''}(0,z)\frac{A(z)}{\gamma^{(2)}(0,z)(\mathrm{e}^{-\gamma^{(2)}(0,z)}-1)}\\&\quad+\gamma^{(2)'}(0,z)\frac{\mathrm{d}}{\mathrm{d}z}\frac{A(z)}{\gamma^{(2)}(0,z)(\mathrm{e}^{-\gamma^{(2)}(0,z)}-1)}.\end{aligned}$$

In particular, $\lim_{z\to 0}\gamma^{(1)'}(0,z)=6$. A straightforward computation shows

$$\lim_{z\to 0}\frac{\mathrm{d}}{\mathrm{d}z}\frac{A(z)}{\gamma^{(2)}(0,z)(\mathrm{e}^{-\gamma^{(2)}(0,z)}-1)}=-1$$

and so

$$\begin{aligned}&\lim_{z\to 0}(-\gamma^{(1)''}(0,z)-\gamma^{(2)''}(0,z)(z+1/2)-2\gamma^{(2)'}(0,z))\\&\quad=12+\lim_{z\to 0}\left[-\gamma^{(2)''}(0,z)\left(\frac{A(z)}{\gamma^{(2)}(0,z)(\mathrm{e}^{-\gamma^{(2)}(0,z)}-1)}+z+1/2\right)\right].\end{aligned} \tag{43}$$

Using l'Hopital's rule it is easily seen that

$$\lim_{z\to 0} z^{-1}\left(\frac{A(z)}{\gamma^{(2)}(0,z)(\mathrm{e}^{-\gamma^{(2)}(0,z)}-1)}+z+1/2\right)=0.$$

Therefore, by (42) and (43) we have

$$\lim_{z\to 0}(-\gamma^{(1)''}(0,z)-\gamma^{(2)''}(0,z)(z+1/2)-2\gamma^{(2)'}(0,z))=12. \tag{44}$$

By Theorem 4.3 (and the continuity on $[0,1/2)$ of the function $a\mapsto\gamma^{(2)}(a,z)$, $z\in[0,1/4)$) we have, for any $z\in(0,1/4)$,

$$I_{\mathrm{GTP}}(z)\le\gamma(0,z)=-\gamma^{(1)}(0,z)-\gamma^{(2)}(0,z)(z+1/2).$$

Applying l'Hopital's rule twice and using (44) we deduce

$$\limsup_{z\to 0}\frac{I_{\mathrm{GTP}}(z)}{z^2}\leq \lim_{z\to 0}\frac{-\gamma^{(1)\prime\prime}(0,z)-\gamma^{(2)\prime\prime}(0,z)(z+1/2)-2\gamma^{(2)\prime}(0,z)}{2}=6.$$

So the upper bound (39) holds with $M=6$, and the proof is completed. □

5. MDP for the GTP

MDP for the TSTP were given in [3,12]. In this section we provide MDP for $\{K_n/n\}_{n\geq 1}$, where K_n is the optimal transport cost of the GTP over a compact metric space (E,d) with DCF.

In this paper $\{s_n\}_{n\geq 1}$ denotes a strictly increasing sequence of positive numbers diverging to $+\infty$. If in addition $s_n/\sqrt{n}\to 0$, we say that a sequence of rv's $\{\xi_n\}_{n\geq 1}$, taking values on a metric space (S,d_S), obeys a MDP on S with speed s_n^2 and rate function J if $\left\{\xi_n\sqrt{n}/s_n\right\}_{n\geq 1}$ obeys a LDP on S with speed s_n^2 and rate function J.

Let $\mathcal{L}(E)^*$ be the space of bounded real functionals $F:\mathcal{L}(E)\to\mathbb{R}$ equipped with the norm $\|F\|_*:=\sup_{\varphi\in\mathcal{L}(E)}|F(\varphi)|$. For each $\nu\in \mathrm{M}_b(E)$, we define $\nu^*\in\mathcal{L}(E)^*$ by $\nu^*(\varphi):=\nu(\varphi)$. We denote by $N(E,r)$ the metric entropy of (E,d), i.e. the minimal number of closed balls with radius r needed to cover E. The following theorems hold.

Theorem 5.1. *Assume that*

$$\lim_{n\to\infty}\frac{s_n}{\sqrt{n}}=0,\qquad \lim_{n\to\infty}\frac{n^{\alpha/2}}{s_n^{\alpha+2}}\log\frac{\sqrt{n}}{s_n}=0,\quad \textit{for some } \alpha>0, \tag{45}$$

and

$$\lim_{n\to\infty}\sqrt{n}W_1(\gamma_n,\ell_1)/s_n=0. \tag{46}$$

If moreover

$$\textit{there exists } \kappa>0 \textit{ such that } N(E,\delta)\leq\kappa\delta^{-\alpha},\quad \forall\delta>0, \tag{47}$$

then $\{K_n/n\}_{n\geq 1}$ obeys a MDP on $\mathbb{R}$ with speed s_n^2 and good rate function

$$J_{\mathrm{GTP}}(z)=\begin{cases} z^2/\left[2\sup\limits_{\varphi\in\mathcal{L}^0(E)}\ell_1(\varphi^2)\right] & \textit{if } z\geq 0\\ +\infty \quad \textit{if } z<0.\end{cases}$$

In the particular case when E is a compact subset of $\mathbb{R}^k$ and d is the Euclidean distance, in dimensions 1 and 2 it is possible to have a wider choice for $\{s_n\}_{n\geq 1}$:

Theorem 5.2. *If E is a compact subset of the line, d is the Euclidean distance and $\{s_n\}_{n\geq 1}$ satisfies the first relation in (45) and (46), then $\{K_n/n\}_{n\geq 1}$ obeys a MDP on $\mathbb{R}$ with speed s_n^2 and good rate function J_{GTP}.*

Theorem 5.3. *Suppose that d is the Euclidean distance on the plane and $\{s_n\}_{n\geq 1}$ satisfies the first relation in (45) and (46). Moreover, assume one of the following additional conditions:*

(i) $E=[0,1]^2$, *ℓ_1 is the uniform distribution, and* $\lim_{n\to\infty}s_n/\sqrt{\log n}=+\infty$.
(ii) *E is a compact subset of the plane and* $\lim_{n\to\infty}s_n/\log n=+\infty$.

Then $\{K_n/n\}_{n\geq 1}$ obeys a MDP on $\mathbb{R}$ with speed s_n^2 and good rate function J_{GTP}.

Finally, we state the following corollary of Theorem 5.1.

Corollary 5.4. *Assume that E is a compact subset of $\mathbb{R}^k$, $k \geq 3$, d is the Euclidean distance and $\{s_n\}_{n\geq 1}$ satisfies the first relation in (45), (46) and is such that, for all n large enough,*

$$s_n \geq \kappa n^{\frac{1}{2}-\beta}, \quad \text{for some } \kappa > 0 \text{ and } \beta \in (0, 1/(k+2)). \tag{48}$$

Then $\{K_n/n\}_{n\geq 1}$ obeys a MDP on $\mathbb{R}$ with speed s_n^2 and good rate function J_{GTP}.

It was proved in [3] that if E is the Euclidean unit ball of $\mathbb{R}^k$, d is the Euclidean distance and ℓ_1 is the uniform distribution, then $\sup_{\varphi\in\mathcal{L}^0(E)} \ell_1(\varphi^2) = 1/(k+2)$. We also note that if $E = [0, 1]$, d is the Euclidean distance and $\ell_1 = \ell$ is the Lebesgue measure, then $\sup_{\varphi\in\mathcal{L}^0(E)} \ell_1(\varphi^2) = 1/12$. Indeed,

$$\sup_{\varphi\in\mathcal{L}^0(E)} \ell_1(\varphi^2) = \frac{1}{2} \sup_{\varphi\in\mathcal{L}^0(E)} \int_{E\times E} |\varphi(x) - \varphi(y)|^2 \,\mathrm{d}x\mathrm{d}y \leq \frac{1}{2} \int_{E\times E} |x - y|^2 \,\mathrm{d}x\mathrm{d}y,$$

so the supremum is attained at $\varphi^*(x) = x - \frac{1}{2}$, and the claim follows.

To clarify the choice of $\{s_n\}_{n\geq 1}$, before proving the above results, we provide an example.

Example 3. (1) Consider the kDRGTP with $C = W_1$ and $\ell_1 = \ell$, the Lebesgue measure on $E = [0, 1]^k$. Assume that $\{s_n\}_{n\geq 1}$ satisfies the first relation in (45). In addition, suppose that $\{s_n\}_{n\geq 1}$ goes to $+\infty$ faster than $\log n$ if $k = 2$, or satisfies (48) if $k \geq 3$. Then Theorem 5.2 or 5.3 or Corollary 5.4 may be applied. Indeed, let $\{g_1, \dots, g_n\} \subset [0, 1]^k$ be the grid formed by n regularly spaced points g_i, i.e. each g_i is a point of the cube C_i, $i = 1, \dots, n$. We have

$$\begin{aligned} W_1(\gamma_n, \ell_1) &= \sup_{\varphi\in\mathcal{L}(E)} \left(\int_E \varphi(x)\gamma_n(\mathrm{d}x) - \int_E \varphi(x)\,\mathrm{d}x \right) \\ &\leq \sup_{\varphi\in\mathcal{L}(E)} \sum_{i=1}^{n} \int_{C_i} |\varphi(g_i) - \varphi(x)|\,\mathrm{d}x \\ &\leq \frac{1}{n} \sum_{i=1}^{n} \sup_{x\in C_i} |g_i - x| \leq \sqrt{k} n^{-1/k}. \end{aligned}$$

So condition (46) is satisfied.

(2) Consider the kDGTP with $C = W_1$ and ℓ_1 specified below. As in Example 1 part (2), suppose that the grid points are not all distinct, i.e. $\{g_1, g_2, \dots\} = \{x_1, \dots, x_m\}$ for some positive integer m, where $x_1, \dots, x_m$ are distinct points of E. Using the same notation as in Example 1, define $p_{\min} := \min\{p_1, \dots, p_m\} > 0$, $n_j(n) := [np_j] + h_j$ and $\ell_1 := \sum_{j=1}^m p_j \delta_{x_j}$. Assume that $\{s_n\}_{n\geq 1}$ satisfies the first relation in (45). In addition, suppose that $\{s_n\}_{n\geq 1}$ goes to $+\infty$ faster than $\log n$ if $k = 2$, or satisfies (48) if $k \geq 3$. Then Theorem 5.2 or 5.3 or Corollary 5.4 may be applied. Indeed, (46) may be checked as follows. By 7.3.3 in Talagrand [23] we have

$$W_1(\gamma_n, \ell_1) = W_1\left(\frac{1}{n} \sum_{j=1}^{m} n_j(n)\delta_{x_j}, \sum_{j=1}^{m} p_j \delta_{x_j} \right) \leq \|d\|_\infty \sum_{j=1}^{m} \left| \frac{n_j(n)}{n} - p_j \right|.$$

So (46) is a consequence of $\sum_{j=1}^m |([np_j] + h_j)/(np_j) - 1| \leq 2m/(np_{\min})$. Alternatively, (46) may be checked using a classical transportation (or T_1) inequality (see Theorem 22.10, p. 575, in [26]).

Let d_2 be the L^2-metric on $\mathrm{L}^2(E, \ell_1)$, the space of square-integrable functions on E, wrt ℓ_1. The following lemmas are used to prove Theorem 5.1. The proof of Lemma 5.5 is omitted. It may be reproduced following Step 2 of the proof of Theorem 2.1 in [5]. The proof of Lemma 5.6 is postponed to the end of this section.

Lemma 5.5. *For all $n \geq 1$, $x \geq 0$ and $\delta \in (0, 2\|d\|_\infty)$,*

$$P\left(W_1(L_n, \ell_1) \geq x\right) \leq \left(\frac{16\mathrm{e}\|d\|_\infty}{\delta}\right)^{N(E,\delta/4)} \exp\left\{-n\frac{\lambda}{2}\max\{x-\delta, 0\}^2\right\},$$

where λ is defined in (18).

Lemma 5.6. *The metric space $(\mathcal{L}(E), \mathrm{d}_2)$ is totally bounded.*

Proof of Theorem 5.1. We divide the proof in two steps.

Step 1: *MDP for the Kantorovich–Rubinshtein distance*. In this step we shall prove that $\{W_1(L_n, \ell_1)\}_{n\geq 1}$ obeys a MDP on $\mathbb{R}$ with speed s_n^2 and good rate function J_{GTP}. For each $n \geq 1$, the process $\{(L_n - \ell_1)(\varphi)\}_{\varphi\in\mathcal{L}(E)}$ is continuous in probability. So, due to the separability of $(\mathcal{L}(E), \mathrm{d}_2)$, there exist separable versions of the processes $\{(L_n - \ell_1)(\varphi)\}_{\varphi\in\mathcal{L}(E)}$, $n \geq 1$. Throughout this proof we always refer to these separable versions so that $\sup_{\varphi\in\mathcal{L}(E)}(L_n - \ell_1)(\varphi) = \sup_{\varphi\in\mathcal{L}_{\mathrm{c}}(E)}(L_n - \ell_1)(\varphi)$ a.s., for some countable subset $\mathcal{L}_{\mathrm{c}}(E)$ of $\mathcal{L}(E)$ (see e.g. [18], p. 44, for the notion of separable process). We start by proving that $\{(L_n - \ell_1)^*\}_{n\geq 1}$ obeys a MDP on $\mathcal{L}(E)^*$ with speed s_n^2 and good rate function

$$J_{\mathcal{L}(E)^*}(F) = \inf\left\{\frac{1}{2}\ell_1\left[\left(\frac{\mathrm{d}\nu}{\mathrm{d}\ell_1}\right)^2\right] : \nu \ll \ell_1, \nu(E) = 0, \nu^* = F\right\}.$$

By Lemma 5.6, Theorem 2 in [27] (the implication (iii) ⇒ (i)) and (5), we only need to verify

$$\lim_{n\to\infty} P\left(W_1(L_n, \ell_1) > \varepsilon s_n/\sqrt{n}\right) = 0, \quad \forall \varepsilon > 0. \tag{49}$$

By Lemma 5.5 and (47) we deduce, for all $n \geq 1$, $x, \delta \in (0, 2\|d\|_\infty)$,

$$P(W_1(L_n, \ell_1) > x) \leq \left(\frac{16\mathrm{e}\|d\|_\infty}{\delta}\right)^{(4^\alpha\kappa)\delta^{-\alpha}} \exp\left\{-n\frac{\lambda\max\{x-\delta, 0\}^2}{2}\right\}.$$

Choosing, for n large enough and $\varepsilon > \varepsilon' > 0$, $x = \varepsilon s_n/\sqrt{n}$ and $\delta = \varepsilon' s_n/\sqrt{n}$, we get

$$P\left(W_1(L_n, \ell_1) > \varepsilon s_n/\sqrt{n}\right) \leq \left(\kappa_1\frac{\sqrt{n}}{s_n}\right)^{\kappa_2(s_n/\sqrt{n})^{-\alpha}} \exp\left\{-\kappa_3 s_n^2\right\}, \tag{50}$$

where $\kappa_1 := 16\mathrm{e}\|d\|_\infty/\varepsilon'$, $\kappa_2 := (4/\varepsilon')^\alpha\kappa$, $\kappa_3 := \lambda(\varepsilon - \varepsilon')^2/2$. By (45) it follows that the right-hand side of (50) goes to zero, as $n \to \infty$, and so $\{(L_n - \ell_1)^*\}_{n\geq 1}$ obeys a MDP on $\mathcal{L}(E)^*$ with speed s_n^2 and good rate function $J_{\mathcal{L}(E)^*}$. By (5) we deduce

$$\frac{\sqrt{n}}{s_n}W_1(L_n, \ell_1) = \left\|\left[\frac{\sqrt{n}}{s_n}(L_n - \ell_1)\right]^*\right\|_*.$$

So, by the Contraction Principle, $\{W_1(L_n, \ell_1)\}_{n\geq 1}$ obeys a MDP on $\mathbb{R}$ with speed s_n^2 and good rate function

$$J_{\mathrm{GTP}}(z) = \inf\left\{\frac{1}{2}\ell_1\left[\left(\frac{\mathrm{d}\nu}{\mathrm{d}\ell_1}\right)^2\right] : \nu \ll \ell_1, \nu(E) = 0, \sup_{\varphi\in\mathcal{L}(E)}\nu(\varphi) = z\right\}.$$

Arguing similarly to in the proof of Lemma 3.4 in [12] one can deduce that J_{GTP} has the desired expression.

Step 2: *exponential equivalence*. By the triangular inequality and the symmetry, we deduce

$$|W_1(\mu,\nu_1) - W_1(\mu,\nu_2)| \le W_1(\nu_1,\nu_2), \quad \text{for all } \mu,\nu_1,\nu_2 \in \mathrm{P}(E). \tag{51}$$

Combining this with (46) we have that for any $\delta > 0$ there exists n_δ such that, for all $n \ge n_\delta$,

$$P\left(\frac{\sqrt{n}}{s_n}|K_n/n - W_1(L_n,\ell_1)| > \delta\right) \le P\left(W_1(\gamma_n,\ell_1) > \frac{s_n}{\sqrt{n}}\,\delta\right) = 0.$$

Therefore, for a fixed $\delta > 0$, we have

$$\lim_{n\to\infty}\frac{1}{s_n^2}\log P\left(\frac{\sqrt{n}}{s_n}|K_n/n - W_1(L_n,\ell_1)| > \delta\right) = -\infty.$$

So, $\{K_n/(s_n\sqrt{n})\}_{n\ge1}$ and $\{\sqrt{n}W_1(L_n,\ell_1)/s_n\}_{n\ge1}$ are exponentially equivalent (see e.g. Definition 4.2.10 in [9]). The conclusion follows by Step 1 and Theorem 4.2.13 in [9]. □

Proof of Theorem 5.2. To avoid trivialities we assume that E is a compact interval (otherwise $\mathcal{L}(E) = \emptyset$). Arguing as in the proof of Theorem 5.1 we only need to check (49). In Section 7, we shall show that if $E \subset \mathbb{R}$ is a compact interval, then $\mathcal{L}(E)$ is an ℓ_1-Donsker class (see Lemma 7.3). Combining this with Theorem 3.2 (the implication (ii) ⇒ (iii)) and Remark (ii) p.19 in [27], we deduce (49). □

Proof of Theorem 5.3. Here again, arguing as in the proof of Theorem 5.1, we only need to check (49). By Markov's inequality

$$P\left(W_1(L_n,\ell_1) > \varepsilon s_n/\sqrt{n}\right) \le \sqrt{n}\mathrm{E}[W_1(L_n,\ell_1)]/(\varepsilon s_n). \tag{52}$$

First, suppose that the additional conditions (i) hold. By Theorem 4.1 in [22] we have $\mathrm{E}[W_1(L_n,\ell_1)] \le \kappa\sqrt{\log n/n}$, for all $n \ge 1$ and some constant $\kappa > 0$ not depending on n. Combining this with (52) we deduce (49). Now, assume (ii). Equip $\mathcal{L}(E)$ with the norm $\|\varphi\|_{BL} := \|\varphi\|_\infty + \|\varphi\|_{\mathrm{Lip}}$. Here, $\|\cdot\|_\infty$ denotes the usual sup-norm and $\|\cdot\|_{\mathrm{Lip}}$ is defined by (4) with d being the Euclidean distance. For $\alpha \in \mathrm{M}_b(E)$, set $\|\alpha\|_{BL} := \sup_{\varphi:\,\|\varphi\|_{BL}\le1}|\alpha(\varphi)|$. The BL-metric on $\mathrm{P}(E)$ is defined by $\beta(\mu,\nu) := \|\mu-\nu\|_{BL}$. A straightforward computation shows that, for any $M \ge 1$, $\mu,\nu \in \mathrm{P}(E)$, $\sup\{|(\mu-\nu)(\varphi)| : \varphi \in \mathcal{L}(E),\ \|\varphi\|_{BL} \le M\} = M\beta(\mu,\nu)$. So $W_1(\mu,\nu) \le (\|d\|_\infty + 1)\beta(\mu,\nu)$, for all $\nu,\mu \in \mathrm{P}(E)$. Therefore, (49) follows if we show that $\sqrt{n}\mathrm{E}[\beta(L_n,\ell_1)]/s_n \to 0$. As noticed in [10], p. 44, $\mathrm{E}[\beta(L_n,\ell_1)] \le \kappa n^{-1/2}(1+\log n)$, for all $n \ge 1$ and some constant $\kappa > 0$ not depending on n. The claim is a consequence of the choice of $\{s_n\}_{n\ge1}$. □

Proof of Corollary 5.4. Note that (47) is satisfied with $\alpha = k$ and (48) implies the second relation in (45) with $\alpha = k$. The claim follows by Theorem 5.1. □

Proof of Lemma 5.6. We need to prove that, for any $\varepsilon > 0$, there exist a finite number of sets in $(\mathcal{L}(E), \mathrm{d}_2)$, with radius less than or equal to ε, whose union covers $\mathcal{L}(E)$. Equip $C_b(E)$ with the usual sup-norm $\|\cdot\|_\infty$. Clearly, $\mathcal{L}(E) \subset C_b(E)$, and so by the Ascoli–Arzelà' Theorem (see e.g. [9, Theorem C.8, p. 352]) $\mathcal{L}(E)$ is totally bounded (i.e. precompact) in $(C_b(E), \|\cdot\|_\infty)$. Indeed $(\mathcal{L}(E), \|\cdot\|_\infty)$ is bounded and $\mathcal{L}(E)$ is equicontinuous, i.e. for any $\varepsilon > 0$ and $x \in E$, there exists a neighborhood of x, say $\mathcal{N}_\varepsilon(x)$, such that $\sup_{\varphi\in\mathcal{L}(E)}\sup_{y\in\mathcal{N}_\varepsilon(x)}|\varphi(x)-\varphi(y)| < \varepsilon$. By one of the equivalent definitions of totally bounded metric space, it follows that, for a fixed

$\varepsilon > 0$, there exist a finite number of functions in $C_b(E)$, say $\varphi_1, \ldots, \varphi_n \in C_b(E)$, such that $\mathcal{L}(E) \subset \bigcup_{i=1}^n \overset{\circ}{\mathrm{B}}_\varepsilon(\varphi_i; \|\cdot\|_\infty)$, where $\overset{\circ}{\mathrm{B}}_\varepsilon(\varphi_i; \|\cdot\|_\infty) := \{\psi \in C_b(E) : \|\varphi_i - \psi\|_\infty < \varepsilon\}$. The total boundedness of $(\mathcal{L}(E), \mathrm{d}_2)$ follows on noticing that $\mathrm{B}_\varepsilon(\varphi_i; \mathrm{d}_2) \supset \overset{\circ}{\mathrm{B}}_\varepsilon(\varphi_i; \|\cdot\|_\infty)$, for any $i = 1, \ldots, n$, where $\mathrm{B}_\varepsilon(\varphi_i; \mathrm{d}_2) := \{\psi \in C_b(E) : \mathrm{d}_2(\varphi_i, \psi) \le \varepsilon\}$. □

6. Extensions to non-compact spaces: a brief discussion

The ASC and LDP in Section 3 and Theorem 4.2 hold on general Polish spaces. The proofs of (15), (16) and Lemma 5.6 rely on the compactness of (E, d), and we do not know how to prove Theorems 4.1 and 5.1 relaxing such a topological assumption. However, the exact deviation bounds in [5] may be used to get explicit large deviation upper bounds for certain TSTP and GTP on non-compact spaces.

Let d be the Euclidean distance on $E \subseteq \mathbb{R}^k$. We denote by $K_n^{(T)}$ the optimal transport cost of the kDTSTP with $E_1 = E_2 = E$, $\ell_1 = \ell_2$, $d_1 = d_2 = d$, and by $K_n^{(G)}$ the optimal transport cost of the kDGTP with $E_1 = E_2 = E$ and $d_1 = d_2 = d$. The following theorem holds.

Theorem 6.1. *Assume $c(x, y) \le d(x, y)$ for all $x, y \in E$ and $\int_E \mathrm{e}^{\alpha d(x_0,x)^2} \ell_1(\mathrm{d}x) < \infty$ for some $x_0 \in E$ and $\alpha > 0$. Let λ be defined by (18). Then*

$$\limsup_{n\to\infty} \frac{1}{n} \log P\left(K_n^{(T)}/n \in F\right) \le -\lambda \inf_{x\in F} x^2/8, \quad \forall \text{ closed } F \subseteq [0, \infty). \tag{53}$$

If in addition (13) holds with $\ell^* \in \mathrm{P}_1(E)$, then

$$\limsup_{n\to\infty} \frac{1}{n} \log P\left(K_n^{(G)}/n \in F\right) \le -\lambda \inf_{x\in F} \max\{x - W_1(\ell_1, \ell^*), 0\}^2/2,$$
$$\forall \text{ closed } F \subseteq [0, \infty). \tag{54}$$

Proof. We only prove (53); the bound (54) may be shown similarly. For ease of notation we set $K_n^{(T)} = K_n$. By Corollary 4 in [6], ℓ_1 satisfies the $T_1(\lambda)$ inequality (19). So by Theorem 2.1 in [5] we have, for any $x > 0$ and all n large enough, $P(W_1(L_n, \ell_1) > x) \le \mathrm{e}^{-n\lambda x^2/2}$. For any $n \ge 1$ and some $\sigma \in \mathcal{P}_n$ we have $K_n/n \le n^{-1} \sum_{i=1}^n c(X_i, Y_{\sigma(i)}) \le n^{-1} \sum_{i=1}^n d(X_i, Y_{\sigma(i)}) = W_1(L_n, M_n)$ a.s. By the square-exponential moment condition it follows that $\mathrm{E}[d(x_0, X_1)] < \infty$, and so $L_n, M_n, \ell_1 \in \mathrm{P}_1(E)$. Combining the above inequalities and recalling that W_1 is a distance on $\mathrm{P}_1(E)$, for all n large enough and $x \ge 0$, we have

$$P(K_n/n \ge x) \le P(W_1(L_n, M_n) \ge x) \le 2P(W_1(L_n, \ell_1) > x/2) \le 2\mathrm{e}^{-n\lambda x^2/8}.$$

Therefore

$$\limsup_{n\to\infty} \frac{1}{n} \log P\left(K_n/n \ge x\right) \le -\lambda x^2/8, \quad \forall x \ge 0. \tag{55}$$

This upper bound can be extended from closed half-intervals $[x, \infty)$ to arbitrary closed subsets $F \subseteq [0, \infty)$ as follows. Letting x denote the infimum of F, we have $x^2/8 = \inf_{y\in F} y^2/8$. The claim is a consequence of (55), on noticing that F is contained in $[x, \infty)$. □

Although Theorem 6.1 extends to some non-compact spaces the large deviation upper bounds provided by Theorem 4.1, (53) and (54) are weaker than the corresponding upper bounds provided by (17). We also remark that if the tail of ℓ_1 decays so slowly that it does not admit a finite square-exponential moment, but only polynomial or exponential moments, then explicit large deviation upper bounds for the optimal transport cost with locations on non-compact spaces may be deduced by Theorems 2.7 and 2.8 in [5].

7. CLT

We say that a sequence of rv's $\{\xi_n\}_{n\geq 1}$, taking values on a separable metric space (S, d_S), obeys a CLT on S with limit ξ if ξ is an S-valued rv and $\xi_n/\sqrt{n} \to \xi$ weakly. In this section we provide CLT for $\{K_n\}_{n\geq 1}$, concerning certain 1DTSTP and 1DGTP with Euclidean DCF (see Theorems 7.1, 7.2 and 7.4).

7.1. CLT for the 1DTSTP

Throughout this subsection, E denotes a non-empty compact interval of $\mathbb{R}$ and d the Euclidean distance. Let ℓ_1 be the common law of iid rv's $\{X_n\}_{n\geq 1}$ with values on E and $G^{(\ell_1)} \equiv \{G_\varphi^{(\ell_1)}\}_{\varphi\in\mathcal{L}(E)}$ a centered Gaussian process with covariance $\mathrm{E}[G_{\varphi_1}^{(\ell_1)} G_{\varphi_2}^{(\ell_1)}] := \ell_1(\varphi_1\varphi_2) - \ell_1(\varphi_1)\ell_1(\varphi_2)$ (it exists since the given covariance is non-negative definite; see [11, p. 92]). We endow $\mathcal{L}(E)$ with the semi-metric $\rho_{\ell_1}(\varphi_1, \varphi_2) := \mathrm{E}^{1/2}[(G_{\varphi_1}^{(\ell_1)} - G_{\varphi_2}^{(\ell_1)})^2] = \mathrm{d}_2(\varphi_1 - \ell_1(\varphi_1), \varphi_2 - \ell_1(\varphi_2))$. For each $n \geq 1$, consider the random signed measure $\sqrt{n}(L_n - \ell_1)$. Due to the continuity of W_1 and the measurability of L_n, the function $\omega \mapsto \sup_{\varphi\in\mathcal{L}(E)} |\sqrt{n}(L_n(\omega) - \ell_1)(\varphi)|$ is a rv, and each process $\left\{\sqrt{n}(L_n - \ell_1)(\varphi)\right\}_{\varphi\in\mathcal{L}(E)}$ may be viewed as a random element on $\mathcal{L}(E)^*$ through the measurable map $\omega \mapsto \left[\sqrt{n}(L_n(\omega) - \ell_1)\right]^*$. In Lemma 7.3 we shall check that $\mathcal{L}(E)$ is an ℓ_1-Donsker class (see e.g. [18, p. 404]), i.e.:

(i) $\mathcal{L}(E)$ is pre-Gaussian, i.e. the process $G^{(\ell_1)}$ admits a version such that, for each ω, the function from $\mathcal{L}(E)$ to $\mathbb{R}$ defined by $G_\omega^{(\ell_1)}(\varphi) := G_\varphi^{(\ell_1)}(\omega)$ is bounded and uniformly continuous wrt ρ_{ℓ_1}.
(ii) $\left[\sqrt{n}(L_n - \ell_1)\right]^*$ converges weakly to $G^{(\ell_1)}$ on $\mathcal{L}(E)^*$.

Thanks to the pre-Gaussianity, there exists a separable version of $G^{(\ell_1)}$ (see e.g. [18]). In the following (and in part (ii) of the above definition), we are considering a separable version, in such a way that the pointwise supremum is a rv, and the Gaussian process $G^{(\ell_1)}$ may be viewed as a random element on $\mathcal{L}(E)^*$ through the (measurable) map $\omega \mapsto G_\omega^{(\ell_1)}$.

Finally, we recall that, to take into account the measurability questions (we are dealing with random elements on the non-separable space $(\mathcal{L}(E)^*, \|\cdot\|_*)$), the weak convergence of probability measures on $\mathcal{L}(E)^*$ is defined via the upper integral (see [18, p. 404] and [11, pp. 93–94]).

Let K_n denote the optimal transport cost of the 1DTSTP over E with Euclidean DCF, CIL and $\ell_1 = \ell_2$. The following theorems hold.

Theorem 7.1. *$\{K_n\}_{n\geq 1}$ obeys a CLT on $\mathbb{R}$ with limit $\sqrt{2}\|G^{(\ell_1)}\|_*$.*

Set

$$\sigma^2 := \int_E x^2\,\ell_1(\mathrm{d}x) - \left(\int_E x\,\ell_1(\mathrm{d}x)\right)^2 \quad \text{and} \quad \Psi(x) := (2\pi)^{-1/2}\int_x^\infty \mathrm{e}^{-y^2/2}\,\mathrm{d}y.$$

Theorem 7.2. *Under the foregoing assumptions:*

(i) *$\|G^{(\ell_1)}\|_*$ has the same law of $\int_{\mathbb{R}} |B(F(t))|\,\mathrm{d}t$, where $\{B(t)\}_{0\leq t\leq 1}$ is a Brownian bridge and F is the distribution function of ℓ_1.*
(ii) *If moreover $\sigma > 0$, then $\lim_{x\to\infty} P(\|G^{(\ell_1)}\|_* > x)/[2\Psi(x/\sigma)] = 1$.*

The proofs of Theorems 7.1 and 7.2 are based on the following lemma, which we shall show later on.

Lemma 7.3. *$\mathcal{L}(E)$ is an ℓ_1-Donsker class.*

Proof of Theorem 7.1. By Lemma 7.3 and Theorem 11.1.1, p. 333, in [11] we have that $\left[\sqrt{n/2}(L_n - M_n)\right]^* \to G^{(\ell_1)}$ weakly on $\mathcal{L}(E)^*$. By the Continuous Mapping Theorem (see e.g. Theorem 3.6.7, p. 116, in [11]) $\left\|\left[\sqrt{n/2}(L_n - M_n)\right]^*\right\|_* \to \|G^{(\ell_1)}\|_*$ weakly. The claim follows on noticing that $K_n/\sqrt{2n} = \left\|\left[\sqrt{n/2}(L_n - M_n)\right]^*\right\|_*$. □

Proof of Theorem 7.2(i). By Theorem 2.1(a) in [8], $\sqrt{n}W_1(L_n, \ell_1)$ converges in law to $\int_{\mathbb{R}} |B(F(t))|\,\mathrm{d}t$. Note that

$$\sqrt{n}W_1(L_n, \ell_1) = \sup_{\varphi\in\mathcal{L}(E)} \int_E \varphi(x)\left[\sqrt{n}(L_n - \ell_1)\right](\mathrm{d}x) = \left\|\left[\sqrt{n}(L_n - \ell_1)\right]^*\right\|_*,$$

and by Lemma 7.3 $\left[\sqrt{n}(L_n - \ell_1)\right]^* \to G^{(\ell_1)}$ weakly on $\mathcal{L}(E)^*$. So by the Continuous Mapping Theorem, $\left\|\left[\sqrt{n}(L_n - \ell_1)\right]^*\right\|_* \to \|G^{(\ell_1)}\|_*$ in law. The claim follows by the uniqueness of the limit in distribution. □

Proof of Theorem 7.2(ii). We start by proving the lower bound. By the Lipschitz property of φ, for all $\varphi \in \mathcal{L}(E)$, we deduce

$$\begin{aligned}\sigma_\varphi^2 := \mathrm{E}[(G_\varphi^{(\ell_1)})^2] &= \frac{1}{2}\int_{E\times E} |\varphi(x) - \varphi(y)|^2\,\ell_1(\mathrm{d}x)\ell_1(\mathrm{d}y)\\ &\le \frac{1}{2}\int_{E\times E} |x - y|^2\,\ell_1(\mathrm{d}x)\ell_1(\mathrm{d}y)\\ &= \sigma^2 = \ell_1(\varphi_1^2) - \ell_1^2(\varphi_1) = \ell_1(\varphi_2^2) - \ell_1^2(\varphi_2),\end{aligned}$$

where $\varphi_1(x) := x - \inf E$ and $\varphi_2(x) := \sup E - x$. Since $\varphi_1, \varphi_2 \in \mathcal{L}(E)$, they are points of maximal variance. Therefore, for all $x > 0$, letting Z denote a standard normal rv, we have

$$P(\|G^{(\ell_1)}\|_* > x) \ge \sup_{\varphi\in\mathcal{L}(E)} P(|G_\varphi| > x) = \sup_{\varphi\in\mathcal{L}(E)} P(|Z| > x/\sigma_\varphi) = 2\Psi(x/\sigma),$$

and the lower bound follows. We now show the matching upper bound. A straightforward computation gives $\mathrm{E}[G_{\varphi_1}G_{\varphi_2}] = -\sigma^2$. Therefore, $G_{\varphi_1} = -G_{\varphi_2}$ a.s. Consequently,

$$\begin{aligned}P(\|G^{(\ell_1)}\|_* > x) &= P\left(\sup_{\varphi\in\mathcal{L}(E)\setminus\{\varphi_2\}} \max\{G_\varphi, -G_\varphi\} > x\right)\\ &\le P\left(\sup_{\varphi\in\mathcal{L}(E)\setminus\{\varphi_2\}} G_\varphi > x\right) + P\left(\sup_{\varphi\in\mathcal{L}(E)\setminus\{\varphi_2\}} -G_\varphi > x\right)\\ &= 2P\left(\sup_{\varphi\in\mathcal{L}(E)\setminus\{\varphi_2\}} G_\varphi > x\right),\end{aligned}$$

where the latter equality follows by the symmetry of $G^{(\ell_1)}$. The matching upper bound follows if we prove

$$\lim_{x\to\infty} \frac{P\left(\sup_{\varphi\in\mathcal{L}(E)\setminus\{\varphi_2\}} G_\varphi^{(\ell_1)} > x\right)}{\Psi(x/\sigma)} = 1. \tag{56}$$

We apply Theorem 5.5, p. 121, in [1] (see also [21]). Note that φ_1 and φ_2 are the unique points of maximal variance for $G^{(\ell_1)}$. Indeed, reasoning by contradiction, assume that there exists $\overline{\varphi} \in \mathcal{L}(E) \setminus \{\varphi_1, \varphi_2\}$ of maximal variance. Then

$$\int_{E\times E} |\overline{\varphi}(x) - \overline{\varphi}(y)|^2\, \ell_1(\mathrm{d}x)\ell_1(\mathrm{d}y) = \int_{E\times E} |x-y|^2\, \ell_1(\mathrm{d}x)\ell_1(\mathrm{d}y)$$

which implies $|\overline{\varphi}(x) - \overline{\varphi}(y)| = |x - y|$ for $\ell_1 \otimes \ell_1$-almost all $(x, y) \in E \times E$. So, ℓ_1-almost everywhere on E, $\overline{\varphi}(x) = x + c_1$ or $\overline{\varphi}(x) = -x + c_2$, where $c_1, c_2 \in E$ are two constants. Since $\overline{\varphi} \in \mathcal{L}(E)$, then $\overline{\varphi}(x) = x - \inf E$ or $\overline{\varphi}(x) = \sup E - x$, for all $x \in E$, which is impossible. Consequently, φ_1 is the unique point of maximal variance for $\{G_\varphi\}_{\varphi\in\mathcal{L}(E)\setminus\{\varphi_2\}}$. For $h > 0$, define $\mathcal{L}^{(h)}(E) := \{\varphi \in \mathcal{L}(E) \setminus \{\varphi_2\} :\ \mathrm{E}[G_\varphi^{(\ell_1)} G_{\varphi_1}^{(\ell_1)}] \ge \sigma^2 - h^2\}$. We have

$$\mathcal{L}^{(h)}(E) \subset \mathrm{B}_{h\sqrt{2}}(\varphi_1; \rho_{\ell_1}) := \left\{\varphi \in \mathcal{L}(E) :\ \rho_{\ell_1}(\varphi, \varphi_1) \le h\sqrt{2}\right\}. \tag{57}$$

Indeed, for $\varphi \in \mathcal{L}^{(h)}(E)$,

$$\begin{aligned} 2h^2 &\ge 2[\ell_1(\varphi_1^2) - \ell_1^2(\varphi_1)] - 2\mathrm{E}[G_\varphi^{(\ell_1)} G_{\varphi_1}^{(\ell_1)}] \\ &\ge \ell_1(\varphi_1^2) - \ell_1^2(\varphi_1) + \ell_1(\varphi^2) - \ell_1^2(\varphi) - 2[\ell_1(\varphi\varphi_1) - \ell_1(\varphi)\ell_1(\varphi_1)] = \rho_{\ell_1}^2(\varphi, \varphi_1). \end{aligned}$$

Next, we prove $\lim_{h\to 0} \mathrm{E}\left[\sup_{\varphi\in\mathcal{L}^{(h)}(E)} (G_\varphi^{(\ell_1)} - G_{\varphi_1}^{(\ell_1)})\right]/h = 0$. By (57) we have a.s. $0 \le \sup_{\varphi\in\mathcal{L}^{(h)}(E)} (G_\varphi^{(\ell_1)} - G_{\varphi_1}^{(\ell_1)}) \le \left[\sup_{\varphi\in\mathrm{B}_{h\sqrt{2}}(\varphi_1;\rho_{\ell_1})} G_\varphi^{(\ell_1)}\right] - G_{\varphi_1}^{(\ell_1)}$. So, it suffices to check

$$\lim_{h\to 0} \frac{\mathrm{E}\left[\sup_{\varphi\in\mathrm{B}_{h\sqrt{2}}(\varphi_1;\rho_{\ell_1})} G_\varphi^{(\ell_1)}\right]}{h} = 0. \tag{58}$$

This limit is consequence of a classical entropy bound. For $\varepsilon > 0$, let $N(\varepsilon, \mathrm{B}_{h\sqrt{2}}(\varphi_1; \rho_{\ell_1}), \rho_{\ell_1})$ denote the minimal number of ρ_{ℓ_1}-balls of radius ε needed to cover $\mathrm{B}_{h\sqrt{2}}(\varphi_1; \rho_{\ell_1})$ (this number is finite for each ε by Theorem 3.18, pp. 80–81, in [18]). Clearly, $N(\varepsilon, \mathrm{B}_{h\sqrt{2}}(\varphi_1; \rho_{\ell_1}), \rho_{\ell_1}) = 1$ for all $\varepsilon \ge h\sqrt{2}$. Therefore, by Corollary 4.15, p. 106, in [1], there exists a constant $K > 0$ such that

$$\mathrm{E}\left[\sup_{\varphi\in\mathrm{B}_{h\sqrt{2}}(\varphi_1;\rho_{\ell_1})} G_\varphi^{(\ell_1)}\right] \le K \int_0^{h\sqrt{2}} (\log N(\varepsilon, \mathrm{B}_{h\sqrt{2}}(\varphi_1; \rho_{\ell_1}), \rho_{\ell_1}))^{1/2}\, \mathrm{d}\varepsilon.$$

Combining this inequality with $N\left(h\sqrt{2}, \mathrm{B}_{h\sqrt{2}}(\varphi_1; \rho_{\ell_1}), \rho_{\ell_1}\right) = 1$ we deduce (58). So, all the assumptions of Theorem 5.5, p. 121, in [1] are verified, and the proof is complete. □

Proof of Lemma 7.3. For a fixed positive constant $M > 0$, define $\mathcal{C}_M(E) := \{\varphi : E \to \mathbb{R} : \|\varphi\|_{BL} \le M\}$, where $\|\varphi\|_{BL} := \|\varphi\|_\infty + \|\varphi\|_{\mathrm{Lip}}$ and $\|\cdot\|_{\mathrm{Lip}}$ is given by (4) with d being the Euclidean distance. For a subset $\mathcal{C} \subset C_b(E)$, we denote by $N(\varepsilon, \mathcal{C}, \|\cdot\|_\infty)$ the minimal number of balls of radius ε needed to cover $\mathcal{C}$, wrt the sup-norm. By a result in [17] (see also Proposition 1.1 in [24]) we have, for all $\varepsilon > 0$, $\log N(\varepsilon, \mathcal{C}_1(E), \|\cdot\|_\infty) \le K\varepsilon^{-1}$, where $K \in (0, \infty)$ is a positive constant, not depending on ε. Consequently, for all $\delta > 0$, $\log N(\delta, \mathcal{C}_M(E), \|\cdot\|_\infty) \le KM\,\delta^{-1}$. Note that $\mathcal{L}(E) \subset \mathcal{C}_{\|d\|_\infty+1}(E)$; therefore

$$\log N(\delta, \mathcal{L}(E), \|\cdot\|_\infty) \le K(\|d\|_\infty + 1)\,\delta^{-1}, \quad \forall \delta > 0. \tag{59}$$

Given two functions $\varphi_1, \varphi_2 \in \mathrm{L}^2(E, \ell_1)$ the bracket $[\varphi_1, \varphi_2]$ is the set of functions φ such that $\varphi_1 \le \varphi \le \varphi_2$; the size of the bracket is the quantity $\mathrm{d}_2(\varphi_1, \varphi_2)$. For a subset $\mathcal{L} \subset \mathrm{L}^2(E, \ell_1)$, we denote by $N_{[\,]}^{(2)}(\varepsilon, \mathcal{L}, \ell_1)$ the bracketing number, i.e. the minimal number of brackets of size smaller than or equal to ε needed to cover $\mathcal{L}$ (see e.g. [11, p. 234]). Let $\varphi_0 \in C_b(E)$. It is easily checked that $\{\varphi \in C_b(E) : \|\varphi - \varphi_0\|_\infty \le \delta\} \subset [\varphi_0 - \delta, \varphi_0 + \delta]$. This latter bracket is of size 2δ, and so $N_{[\,]}^{(2)}(2\delta, \mathcal{L}(E), \ell_1) \le N(\delta, \mathcal{L}(E), \|\cdot\|_\infty)$. Therefore, by (59) we deduce $\log N_{[\,]}^{(2)}(\delta, \mathcal{L}(E), \ell_1) \le 2K(\|d\|_\infty + 1)\,\delta^{-1}$, for all $\delta > 0$. Consequently,

$$\begin{aligned}\int_0^1 (\log N_{[\,]}^{(2)}(x, \mathcal{L}(E), \ell_1))^{1/2}\,\mathrm{d}x &\le [2K(\|d\|_\infty + 1)]^{1/2} \int_0^1 x^{-1/2}\,\mathrm{d}x \\ &= 2[2K(\|d\|_\infty + 1)]^{1/2} < \infty.\end{aligned}$$

The claim follows by Theorem 7.2.1 in [11]. □

7.2. *CLT for the 1DGTP*

Let K_n denote the optimal transport cost of the 1DGTP specified by $E_1 = E_2 = E \subseteq \mathbb{R}$ and $d_1 = d_2 = c = d$, where d is the Euclidean distance. We have:

Theorem 7.4. *Assume* $\int_{\mathbb{R}} |x|\ell_1(\mathrm{d}x) < \infty$ *and*

$$\lim_{n\to\infty} \sqrt{n}\,W_1(\gamma_n, \ell_1) = 0. \tag{60}$$

Then $\{K_n\}_{n\ge1}$ *obeys a CLT on* $\mathbb{R}$ *with limit* $\int_{\mathbb{R}} |B(F(t))|\,\mathrm{d}t$, *where* $\{B(t)\}_{0\le t\le1}$ *and* F *are defined in the statement of Theorem 7.2.*

Proof. By the triangular inequality and the symmetry, we deduce

$$|W_1(\mu, \nu_1) - W_1(\mu, \nu_2)| \le W_1(\nu_1, \nu_2), \quad \text{for all } \mu, \nu_1, \nu_2 \in \mathrm{P}_1(E).$$

Since $\mathrm{E}[|X_1|] < \infty$ we have $L_n, \ell_1 \in \mathrm{P}_1(E)$. So, combining the above inequality with (7) and (60), we have, with probability 1,

$$|K_n/\sqrt{n} - \sqrt{n}\,W_1(L_n, \ell_1)| = \sqrt{n}|W_1(L_n, \gamma_n) - W_1(L_n, \ell_1)| \le \sqrt{n}\,W_1(\gamma_n, \ell_1) \to 0.$$

Note that $\mathrm{E}[|X_1|] < \infty$ implies $\int_0^\infty \sqrt{P(|X_1| > t)}\,\mathrm{d}t < \infty$ and so the claimed convergence in law follows by Theorem 2.1(a) in [8]. □

In particular, note that condition (60) is satisfied in the cases considered in Example 3.

Acknowledgments

The author thanks the Editor, the Associate Editor and two anonymous referees for their insightful comments which improved a first version of the paper. The author is particularly grateful to one of the two referees who suggested stating and proving Corollary 4.4.

References

[1] R.J. Adler, An Introduction to Continuity, Extrema and Related Topics for General Gaussian Processes, in: Lecture Notes of Inst. Math. Stat., vol. 12, IMS, Hayword, 1990.
[2] M. Ajtai, J. Komlós, G. Tusnády, On optimal matchings, Combinatorica 4 (1984) 259–264.
[3] F. Barthe, N. O'Connell, Matchings and the variance of Lipschitz functions, ESAIM: Probab. Statist. 13 (2010) 400–408.
[4] S.G. Bobkov, F. Götze, Exponential integrability and transportation cost related to logarithmic Sobolev inequalities, J. Funct. Anal. 163 (1999) 1–28.
[5] F. Bolley, A. Guillin, C. Villani, Quantitative concentration inequalities for empirical measures on non-compact spaces, Probab. Theory Related Fields 137 (2007) 541–593.
[6] F. Bolley, C. Villani, Weighted Csizár–Kullback–Pinsker inequalities and applications to transportation inequalities, Ann. Fac. Sci. Toulouse Math. 14 (2005) 331–352.
[7] G. Buttazzo, M. Giaquinta, S. Hildebrandt, One-dimensional Variational Problems, in: Oxford Lecture Series in Mathematics and its Applications, vol. 15, Oxford University Press, New York, 1998.
[8] E. Del Barrio, E. Giné, C. Matrán, Central limit theorems for the Wasserstein distance between the empirical and the true distributions, Ann. Probab. 27 (1999) 1009–1071.
[9] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, 2nd ed., Springer, New York, 1998.
[10] M.R. Dudley, The speed of mean Glivenko–Cantelli convergence, Ann. Math. Statist. 40 (1969) 40–50.
[11] M.R. Dudley, Uniform Central Limit Theorems, Cambridge University Press, Cambridge, 1999.
[12] A. Ganesh, N. O'Connell, Large and moderate deviations for matching problems and empirical discrepancies, Markov Process. Related Fields 13 (2007) 85–98.
[13] N. Gozlan, C. Léonard, A large deviation approach to some transportation cost inequalities, Probab. Theory Related Fields 139 (2007) 235–283.
[14] N. Gozlan, C. Léonard, Transport inequalities. A survey, Markov Process. Related Fields 16 (2010) 635–736.
[15] L.V. Kantorovich, On the transfer of masses, Dokl. Acad. Nauk SSSR 37 (1942) 227–229.
[16] L.V. Kantorovich, On a problem of Monge, Uspekhi Mat. Nauk 3 (1948) 225–226.
[17] A.N. Kolmogorov, V.M. Tikhomirov, ε-entropy and ε-capacity of sets in function spaces, Uspekhi Mat. Nauk 14 (1959) 3–86.
[18] M. Ledoux, M. Talagrand, Probability in Banach Spaces, Springer, New York, 1991.
[19] B. Maurey, Some deviation inequalities, Geom. Funct. Anal. 1–2 (1991) 188–197.
[20] N. O'Connell, A large deviations heuristic made precise, Math. Proc. Cambridge Philos. Soc. 128 (2000) 561–569.
[21] M. Talagrand, Small tails for the supremum of a Gaussian process, Ann. Inst. Henri Poincaré Probab. Statist. 24 (1988) 307–315.
[22] M. Talagrand, Matching theorems and empirical discrepancy computations using majorizing measures, J. Amer. Math. Soc. 7 (1994) 455–537.
[23] M. Talagrand, Spin Glasses: A Challenge for Mathematicians, Springer, New York, 2003.
[24] A. van der Vaart, Bracketing smooth functions, Stochastic Process. Appl. 52 (1994) 93–105.
[25] C. Villani, Topics in Optimal Transportation, in: Graduate Studies in Mathematics, vol. 58, American Mathematical Society, 2003.
[26] C. Villani, Optimal Transport, Old and New, Springer, New York, 2009.
[27] L. Wu, Large deviations, moderate deviations and LIL for empirical processes, Ann. Probab. 22 (1994) 17–27.