Probab. Theory Relat. Fields (2018) 170:229–262
https://doi.org/10.1007/s00440-017-0756-2

Berry–Esseen bounds and Edgeworth expansions in the central limit theorem for transport distances

Sergey G. Bobkov[1]

Received: 16 February 2016 / Revised: 31 August 2016 / Published online: 19 January 2017

Abstract For sums of independent random variables $S_n = X_1 + \cdots + X_n$, Berry–Esseen-type bounds are derived for the power transport distances W_p in terms of Lyapunov coefficients L_{p+2}. In the case of identically distributed summands, the rates of convergence are refined under Cramér's condition.

Keywords Central limit theorem · Transport distances · Edgeworth expansions · Coupling

Mathematics Subject Classification 60F

1 Introduction

Let $\mathcal{F}_p$ denote the collection of all Borel probability measures on the real line $\mathbb{R}$ with finite absolute moments of order $p \geq 1$. The power transport distance of order p between two measures μ, ν in $\mathcal{F}_p$ (also called the Kantorovich or minimal distance) is defined by

$$W_p(\mu, \nu) = \inf_{\pi} \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x - y|^p \, \mathrm{d}\pi(x, y) \right)^{1/p}, \tag{1.1}$$

Partially supported by the Alexander von Humboldt Foundation and NSF Grant DMS-1612961.

✉ Sergey G. Bobkov
bobkov@math.umn.edu

[1] School of Mathematics, University of Minnesota, 127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455, USA

where the infimum runs over all probability measures π on $\mathbb{R} \times \mathbb{R}$ with marginals μ and ν. The quantity W_p represents a metric in the space $\mathcal{F}_p$, which is closely related to the topology of week convergence of probability distributions on the line.

Given independent random variables $X_1, \dots, X_n$ with zero mean and variances $\mathbb{E}X_k^2 = \sigma_k^2$ such that $\sum_{k=1}^n \sigma_k^2 = 1$, we consider the transport distances from the distribution μ_n of the sum $S_n = X_1 + \cdots + X_n$ to the standard normal law γ, i.e., to the Gaussian measure with density and distribution function

$$\varphi(x) = \frac{\mathrm{d}\gamma(x)}{\mathrm{d}x} = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \quad \Phi(x) = \gamma((-\infty, x]) = \int_{-\infty}^{x} \varphi(y)\,\mathrm{d}y \quad (x \in \mathbb{R}).$$

Of a large interest there has been the problem of rates at which μ_n may converge to γ in W_p for growing n, as well as of finding sharp upper bounds on $W_p(\mu_n, \gamma)$, which would quantify the central limit theorem for this important class of metrics.

Like in many similar problems, it is natural to impose moment conditions by involving the Lyapunov coefficients

$$L_s = \sum_{k=1}^{n} \mathbb{E}\,|X_k|^s \quad (s \geq 2).$$

In case of the identically distributed random variables $X_k = \xi_k/\sqrt{n}$ with $\mathbb{E}\xi_1 = 0$, $\mathbb{E}\xi_1^2 = 1$, these quantities have a polynomial decay with respect to the number of summands:

$$L_s = n^{-(s-2)/2}\,\beta_s, \quad \text{where } \beta_s = \mathbb{E}\,|\xi_1|^s.$$

For short, we refer to this particular model as the i.i.d. case.

For example, in presence of finite absolute moments of order $2 < s \leq 3$, for the uniform (Kolmogorov) distance there is a Berry–Esseen bound

$$\rho\,(\mu_n, \gamma) = \sup_x\,|F_n(x) - \Phi(x)| \,\leq\, cL_s, \tag{1.2}$$

where $F_n(x) = \mathbb{P}\{S_n \leq x\}$ is the distribution function of S_n, and c is a positive numerical constant. This bound is optimal in terms of L_s, but is no longer true for $s > 3$, as can be seen in the i.i.d. case with Bernoulli summands. Therefore, the critical value $s = 3$ is most popular in (1.2), since it leads to the standard rate $1/\sqrt{n}$ in the general i.i.d. case with finite 3rd absolute moment.

Similar results, the so-called "global forms of the CLT" (going back to the works of Agnew and Esseen in 1950s, [1–3,19]) remain to hold for the L^p distances between F_n and Φ. In particular, under the 3rd order moment condition, using a non-uniform Berry–Esseen bound due to Bikjalis (cf. e.g. Petrov [27], Chapter V), we have

$$\|F_n - \Phi\|_p \,=\, \left(\int_{-\infty}^{\infty} |F_n(x) - \Phi(x)|^p\,\mathrm{d}x\right)^{1/p} \,\leq\, cL_3, \quad p \geq 1.$$

In view of the well-know representation $W_1(\mu_n, \gamma) = \|F_n - \Phi\|_1$, we are still lead to the bound $W_1(\mu_n, \gamma) \leq cL_3$ for $p = 1$ similarly to (1.2).

However, for the power transport distances of order $p > 1$, the situation turns out to be somewhat different. First, one should mention the work by Sakhanenko [33], who derived a transport inequality (with a stronger cost function) implying

$$W_p(\mu_n, \gamma) \leq c_p \, L_p^{1/p}, \quad p \geq 2, \tag{1.3}$$

with some p-dependent constants c_p. In the i.i.d. case, it yields $W_p(\mu_n, \gamma) = O(n^{\frac{1}{p}-\frac{1}{2}})$, which is however worse than the standard rate. A similar bound was earlier derived by Bártfai [4], assuming that an exponential moment for ξ_1 is finite. Afterwards, an important step in this direction was made Rio [29,30], who discovered that, in order to reach the desired relation $W_p(\mu_n, \gamma) = O(\frac{1}{\sqrt{n}})$ in the i.i.d. case, we have to require that the moment β_{p+2} be finite. More precisely, he derived a general Berry–Esseen-type inequality

$$W_p(\mu_n, \gamma) \leq cL_{p+2}^{1/p}, \quad 1 \leq p \leq 2, \tag{1.4}$$

which in the i.i.d. case reads

$$W_p(\mu_n, \gamma) \leq \frac{c}{\sqrt{n}} \, \beta_{p+2}^{1/p}. \tag{1.5}$$

Moreover, the latter was shown to be optimal with respect to the absolute moment: For any $p \geq 1$ and any prescribed number $\beta_{p+2} \geq 1$, there is a sequence of i.i.d. random variables $\xi_1, \xi_2, \dots$, such that $\mathbb{E}\xi_1 = 0$, $\mathbb{E}\xi_1^2 = 1$, $\mathbb{E}\,|\xi_1|^{p+2} = \beta_{p+2}$, and with the property that

$$\liminf_{n \to \infty} \left[W_p(\mu_n, \gamma) \sqrt{n} \, \right] \geq \frac{1}{4} \, \beta_{p+2}^{1/p}$$

([30], Theorems 4.1 and 5.1). In particular, up to an absolute factor, the right-hand side of (1.4) cannot be improved as a function of Lyapunov's coefficients.

The proof of (1.4) given in [30] was based on the relating the transport distances to Zolotarev's ideal metrics and on the Poisson approximation. For the particular parameter $p = 2$, a different approach to this result was proposed in [10]; it makes use of the Talagrand transport-entropy inequality. However, whether or not the bound (1.4) is true for $p > 2$ up to some p-dependent constants remained open (which would be the best possible relation); it was known as Rio's conjecture. Recently, Bonis [15] has given an affirmative solution in the i.i.d case, by showing that we do have the standard rate $W_p(\mu_n, \gamma) = O(\frac{1}{\sqrt{n}})$, as long as the moment β_{p+2} is finite. (In fact, when $p = 2$, the new argument seems to extend this asymptotic result to the multidimensional situation.)

The main purpose of this note is to prove Rio's conjecture about the validity of the Berry–Esseen bound such as (1.4) for the whole range of the parameter p.

Theorem 1.1 *For any real $p \geq 1$, we have $W_p(\mu_n, \gamma) \leq c_p\, L_{p+2}^{1/p}$ with some constants c_p continuously depending on p.*

As was noted in [30], cf. Corollary 4.2, once L_{p+2} is finite, one may also involve stronger transport distances than W_p. More precisely, combining Theorem 1.1 with the Sakhanenko bound (1.3) and applying Hölder's inequality, we get the following more general assertion.

Corollary 1.2 *For any $p \geq 1$ and $r \in [p, p+2]$, we have $W_r(\mu_n, \gamma) \leq c_p\, L_{p+2}^{1/r}$ with some constants c_p depending on p.*

These bounds cover the i.i.d. case as well. In fact, then, under stronger moment assumptions, the bound (1.5) may be further strengthened, if we involve the Cramér condition

$$\limsup_{t \to \infty} |\,\mathbb{E}\, e^{it\xi_1}| < 1. \tag{1.6}$$

In the sequel, we denote by Z a standard normal random variable.

Theorem 1.3 *Given $p \geq 1$ and an integer $l \geq 3$, suppose that the first $l-1$ moments of ξ_1 coincide with the corresponding moments of Z with $\beta_{(l-2)p+2} < \infty$, and let (1.6) be fulfilled. Then $W_p(\mu_n, \gamma) = O(n^{-\frac{l-2}{2}})$, and moreover,*

$$\lim_{n \to \infty} \Big[n^{\frac{l-2}{2}}\, W_p\,(\mu_n, \gamma) \Big] = \frac{|\gamma_l|}{l!}\, \big(\mathbb{E}\, |H_{l-1}(Z)|^p\big)^{1/p}. \tag{1.7}$$

Here, H_{l-1} denotes the Chebyshev–Hermite polynomial of degree $l-1$, and γ_l stands for the l-th cumulant of ξ_1, which under the above moment assumptions may be defined just as the difference of the l-th moments $\mathbb{E}\,\xi_1^l - \mathbb{E}\, Z^l$.

In case $l = 3$ we return in Theorem 1.3 to the basic moment assumptions $\mathbb{E}\xi_1 = 0$, $\mathbb{E}\xi_1^2 = 1$, and then (1.7) yields an asymptotic result refining the bound with the standard rate. Namely,

$$\lim_{n \to \infty} \big[\sqrt{n}\, W_p\,(\mu_n, \gamma)\big] = \frac{|\gamma_3|}{6}\, \Big(\mathbb{E}\, \Big|Z^2 - 1\Big|^p\Big)^{1/p}, \quad \gamma_3 = \mathbb{E}\xi_1^3,$$

provided that β_{p+2} is finite. This relation was established by Rio [31] for the range $1 < p \leq 2$, under a weaker assumption that ξ_1 has a non-lattice distribution [which replaces (1.6)].

If the distribution of ξ_1 is symmetric about the origin, so that $\mathbb{E}\xi_1^3 = 0$ (that is, $l = 4$), then, under Cramer's condition, we get a stronger convergence

$$\lim_{n \to \infty} \big[\, n\, W_p\,(\mu_n, \gamma)\big] = \frac{|\gamma_4|}{24}\, \Big(\mathbb{E}\, \Big|Z^3 - 3Z\Big|^p\Big)^{1/p}, \quad \gamma_4 = \mathbb{E}\xi_1^4 - 3,$$

provided that the moment β_{2p+2} is finite.

In the context of a strong approximation, these results allow a coupling reformulation in terms of L^p-closeness of S_n to the sum $\zeta_1 + \cdots + \zeta_n$ of independent centered Gaussian random variables with $\mathbb{E}\,\zeta_k^2 = \sigma_k^2$, defined on the same underlying probability space. Assuming that this space is reach enough, e.g., a Lebesgue space in the sense of Rokhlin, let us give such a statement in the i.i.d. situation (see also Theorem 11.2 below).

Corollary 1.4 *Let $\xi_1, \dots, \xi_n$ be i.i.d. random variables such that $\mathbb{E}\xi_1 = 0$, $\mathbb{E}\xi_1^2 = 1$, and $\beta_{p+2} < \infty$ $(p \geq 1)$. Then on the same probability space there exist i.i.d. random variables $\zeta_1, \dots, \zeta_n$ with a standard normal distribution such that*

$$\left(\mathbb{E}\left|\sum_{k=1}^n \xi_k - \sum_{k=1}^n \zeta_k\right|^p\right)^{1/p} \leq c_p\, \beta_{p+2}^{1/p}.$$

Moreover, under the assumptions of Theorem 1.3, this L^p-norm $\leq c\, n^{-(l-3)/2}$ with some constant c which does not depend on n.

One natural approach to these results is relying on the employment of Edgeworth expansions. As was mentioned before, the distributions μ_n are at the distance at most L_3 from γ in the Kolmogorov metric ρ. But, if the Lyapunov coefficient L_s is finite for some integer value $s > 3$, the rate of approximation of μ_n can be made in some (different) sense much better—to be of order at most L_s, if we replace the normal law by a certain "corrected normal" signed measure ν_{s-1} on the real line. The density φ_{s-1} of this measure involves the cumulants γ_r of S_n of orders up to $s-1$ (which are just the sums of the cumulants of X_k); for example,

$$\varphi_3(x) = \varphi(x)\left(1 + \frac{\gamma_3}{3!}\,H_3(x)\right),$$
$$\varphi_4(x) = \varphi(x)\left(1 + \frac{\gamma_3}{3!}\,H_3(x) + \frac{\gamma_4}{4!}\,H_4(x) + \frac{\gamma_3^2}{2!\,3!^2}\,H_6(x)\right).$$

If s is not integer, $s = m + \alpha$ with m integer and $0 < \alpha < 1$, then, as a corresponding approximation for μ_n, one may take the measure ν_m. It represents a small oscillation of γ, and on this way one can be reduced to the study of the distance $W_p(\mu_n, \nu_m)$. However, this quantity does not make sense in (1.1), since in general the approximating measures are not positive. Therefore, we propose to redefine and extend the transport distances to the larger space of measures.

Denote by $\mathbb{M}_p$ the collection of all Borel signed measures μ on the line with total "mass" $\mu(\mathbb{R}) = 1$ and finite absolute moment of order $p \geq 1$, i.e., such that $\int_{-\infty}^{\infty} |x|^p\,|\mathrm{d}\mu(x)| < \infty$, where $|\mu|$ denotes the variation of μ (viewed as a positive measure). With every μ in $\mathbb{M}_p$ we associate the generalized "distribution function" $F(x) = \mu((-\infty, x])$, which may be an arbitrary right-continuous function of bounded variation, with $F(-\infty) = 0$, $F(\infty) = 1$, and with finite p-th absolute moment for $|\mu|$ (the monotonicity property is not required). For $\mu, \nu \in \mathbb{M}_p$ with distribution functions F and G, respectively, put

$$\widetilde{W}_p(\mu, \nu) = \sup \int_{-\infty}^{\infty} |u(F(x)) - u(G(x))| \, \mathrm{d}x, \tag{1.8}$$

where the supremum is taken over all smooth functions $u : \mathbb{R} \to \mathbb{R}$ such that

$$\|u'\|_q = \left(\int_{-\infty}^{\infty} |u'(t)|^q \, \mathrm{d}t \right)^{1/q} \leq 1, \quad \frac{1}{p} + \frac{1}{q} = 1.$$

As we will see, the quantity $\widetilde{W}_p$ represents a metric on $\mathbb{M}_p$, which coincides with W_p on $\mathcal{F}_p$ (once this extension is recognized, one may use the same notation W_p instead of $\widetilde{W}_p$). Formula (1.8) and the important triangle inequality for the extended distance will allow us to activate the analysis of transport distances with participation of corrected normal measures.

More details on the extended transport distances are given in Sects. 2, 3. Sections 5, 6 are focused on the study of asymptotic behavior and bounding of $W_p(\nu, \gamma)$ in the situation where ν behaves similarly to the approximating measures ν_m (with preliminary technical lemmas located in Sect. 4). In Sects. 7 and 9 we remind basic definitions and results related to the Edgeworth expansions, including recent ones obtained in [12] to cover the case of non-integer values of p. They are used to properly bound $W_p(\mu_n, \nu_m)$. As part of the proof of Theorem 1.1, a smoothing argument is discussed separately in Sect. 8, and final steps are made in Sect. 10. We conclude with some remarks on the "coupling" version of Theorem 1.1 in Sect. 11 (where Corollary 1.4 is also proved).

Contents:

1. Introduction.
2. Extended transport metrics.
3. Explicit representations and general bounds.
4. Normal distribution function.
5. Perturbations of the Gaussian measure.
6. Second order approximation.
7. Edgeworth-type approximation in the non-i.i.d. case.
8. Smoothing.
9. Edgeworth expansion (the i.i.d. case).
10. Proof of Theorems 1.1 and 1.3.
11. Coupling.

2 Extended transport metrics

Returning to the definition (1.8), first let us state basic properties of the functional $\widetilde{W}_p$.

Proposition 2.1 *For all $\mu, \nu, \lambda \in \mathbb{M}_p$,*

(a) $0 \leq \widetilde{W}_p(\mu, \nu) < \infty$; $\widetilde{W}_p(\mu, \nu) = 0 \Longleftrightarrow \mu = \nu$;
(b) $\widetilde{W}_p(\mu, \nu) = \widetilde{W}_p(\nu, \mu)$;
(c) $\widetilde{W}_p(\mu, \lambda) \leq \widetilde{W}_p(\mu, \nu) + \widetilde{W}_p(\nu, \lambda)$.

That is, $\widetilde{W}_p$ is a metric on $\mathbb{M}_p$. Moreover, $\widetilde{W}_p(\mu, \nu) = W_p(\mu, \nu)$, whenever $\mu, \nu \in \mathcal{F}_p$.

Proof Properties (*a*)–(*c*) are obvious except for the finiteness of $\widetilde{W}_p$. Let μ and ν have (generalized) distribution functions $F(x) = \mu((-\infty, x])$ and $G(x) = \mu((-\infty, x])$. We need to check that for every (equivalently, for some) point $x_0 \in \mathbb{R}$,

$$\sup \int_{-\infty}^{x_0} |u(F(x)) - u(G(x))| \, \mathrm{d}x < \infty, \quad \sup \int_{x_0}^{\infty} |u(F(x)) - u(G(x))| \, \mathrm{d}x < \infty,$$

where the supremum is running over all functions u as in (1.8). Since $F(-\infty) = G(-\infty) = 0$ and $F(\infty) = G(\infty) = 1$, these assertions are equivalent to each other, and one may assume additionally that $u(0) = 0$ under the first supremum. Hence, it is sufficient to see that

$$\sup_{\|u'\|_q \le 1, \ u(0)=0} \int_{-\infty}^{x_0} |u(F(x))| \, \mathrm{d}x \ < \ \infty. \tag{2.1}$$

Given a non-decreasing, bounded, right-continuous function A on the real line, such that $A(-\infty) = 0$, $A(\infty) = c$ $(0 < c < \infty)$, define the generalized inverse function

$$A^{-1}(t) = \min\{x \in \mathbb{R} : A(x) \ge t\}, \quad 0 < t < c. \tag{2.2}$$

It is left-continuous, and as a random variable under the Lebesgue measure on $(0, c)$, it is distributed according to the Lebesgue–Stieltjes measure $\mathrm{d}A(x)$ generated by A, so that

$$\int_0^c \left|A^{-1}(t)\right|^p \, \mathrm{d}t = \int_{-\infty}^{\infty} |x|^p \, \mathrm{d}A(x). \tag{2.3}$$

We apply the definition (2.2) to $A(x)$ defined to be the total variation of F on $(-\infty, x]$, which is the total variation of μ restricted to this half-axis, so that $\mathrm{d}A(x) = |\mu(\mathrm{d}x)|$. In particular, $|F(x)| \le A(x) \le c$ for all $x \in \mathbb{R}$, where $c = \|\mu\|_{\mathrm{TV}} \ge 1$. Since A^{-1} is distributed on $(0, c)$ according to $|\mu|$, the integrals in (2.3) are finite, by the moment assumption on μ.

Now, given a function u participating in the sup of (2.1), let $\psi = |u'|$, thus $\|\psi\|_q \le 1$. By Fubini's theorem,

$$\begin{aligned} \int_{-\infty}^{x_0} |u(F(x))| \, \mathrm{d}x &\le \int_{-\infty}^{x_0} \left| \int_0^{F(x)} \psi(t) \, \mathrm{d}t \right| \mathrm{d}x \\ &= \int_{-c}^{c} \psi(t)\xi(t) \, \mathrm{d}t \ \le \ \left(\int_{-c}^{c} \xi(t)^p \, \mathrm{d}t \right)^{1/p}, \end{aligned} \tag{2.4}$$

where $\xi(t) = \mathrm{mes}\{x \le x_0 : t \text{ is between } 0 \text{ and } F(x)\}$, and where we applied Hölder's inequality on the last step. In the case $0 < t < F(x)$, necessarily $A(x) > t$, hence $x \ge A^{-1}(t)$. If $F(x) < t < 0$, then $A(x) \ge -F(x) > -t$, hence $x \ge A^{-1}(-t)$. In both cases $x \ge A^{-1}(|t|)$ for any $t \in (-c, c)$, which implies that

$$\xi(t) \leq \Big(x_0 - A^{-1}(|t|)\Big)_+ \leq |x_0| + |A^{-1}(|t|)|.$$

Therefore, the last integral in (2.4) does not exceed $2^p\, c\, |x_0|^p + 2^{p-1} \int_{-c}^{c} |A^{-1}(|t|)|^p\, \mathrm{d}t$, which is finite according to (2.3). This proves (2.1) and provides the finiteness of $\widetilde{W}_p$.

To show that $\widetilde{W}_p(\mu,\nu) = W_p(\mu,\nu)$ for all probability measures μ and ν on the line with distribution functions F and G, we make use of the classical representation going back to Fréchet [20],

$$W_p(\mu,\nu) = \bigg(\int_0^1 \Big|F^{-1}(t) - G^{-1}(t)\Big|^p\, \mathrm{d}t\bigg)^{1/p}, \tag{2.5}$$

in terms of the inverse functions defined according to (2.2), cf. e.g. [35]. In analogy with the previous step, put

$$\xi(t) = \mathrm{mes}\{x \in \mathbb{R} : t \text{ is between } F(x) \text{ and } G(x)\}.$$

For any function u with $\|u'\|_q \leq 1$, putting $\psi = |u'|$, we have, by Hölder's inequality,

$$\begin{aligned}\int_{-\infty}^{\infty} |u(F(x)) - u(G(x))|\, \mathrm{d}x &\leq \int_{-\infty}^{\infty} \bigg|\int_{F(x)}^{G(x)} \psi(t)\, \mathrm{d}t\bigg|\, \mathrm{d}x \\ &= \int_0^1 \psi(t)\xi(t)\, \mathrm{d}t \leq \bigg(\int_0^1 \xi(t)^p\, \mathrm{d}t\bigg)^{1/p}. \end{aligned} \tag{2.6}$$

But the inequalities $F(x) < t < G(x)$ and $G(x) < t < F(x)$ imply respectively that

$$G^{-1}(t) \leq x \leq F^{-1}(t) \quad \text{and} \quad F^{-1}(t) \leq x \leq G^{-1}(t).$$

So, $\xi(t) \leq |F^{-1}(t) - G^{-1}(t)|$ for any $t \in (0,1)$, and we obtain from (2.5)–(2.6) and definition (1.8) that $\widetilde{W}_p(\mu,\nu) \leq W_p(\mu,\nu)$.

To derive an opposite bound, let us also show that one can remove the modulus sign from the integrand in (1.8). Suppose that $\psi \in L^q(0,1)$, $\|\psi\|_q \leq 1$. Putting $u(t) = \int_0^t \psi(s)\, \mathrm{d}s$ and integrating by parts we have

$$\int_0^1 \psi(t)\, \Big(F^{-1}(t) - G^{-1}(t)\Big)\, \mathrm{d}t = -\int_0^1 u(t)\, \mathrm{d}\Big(F^{-1}(t) - G^{-1}(t)\Big).$$

As easy to verify, $F^{-1}(t) \leq x < F^{-1}(s)$, if and only if $t \leq F(x) < s$ for all $0 < t < s < 1$. So, the map $x \to F(x)$ pushes forward the Lebesgue measure on $\mathbb{R}$ to the (positive) Borel measure $\mu_{F^{-1}}$ on $(0,1)$ associated to F^{-1} via the equality $\mu_{F^{-1}}([t,s)) = F^{-1}(s) - F^{-1}(t)$. Hence, the identity

$$\int_{-\infty}^{\infty} u(F(x))\, \mathrm{d}x = \int_0^1 u(t)\, \mathrm{d}\mu_{F^{-1}}(t) = \int_0^1 u(t)\, \mathrm{d}F^{-1}(t)$$

holds true for all indicator functions $u = 1_{[t,s)}$ and therefore for any continuous function u such that $u(0) = u(1) = 0$. Thus, for any absolutely continuous function $u : \mathbb{R} \to \mathbb{R}$ with $u(0) = u(1) = 0$ and with Radon-Nikodym derivative $u' = \psi$ such that $\|\psi\|_q \le 1$,

$$\int_0^1 \psi(t) \left(F^{-1}(t) - G^{-1}(t)\right) \mathrm{d}t = -\int_{-\infty}^{\infty} (u(F(x)) - u(G(x)))\, \mathrm{d}x. \quad (2.7)$$

But this equality also holds for all affine functions u, so the constraint $u(0) = u(1) = 0$ may be removed. It remains to take supremum in (2.7) over all smooth u with $\|u'\|_q = \|\psi\|_q \le 1$. $\square$

3 Explicit representations and general bounds

From now on, we use the notation W_p instead of $\widetilde{W}_p$. Following (1.8), one can give a more explicit representation for W_p as an L^p-norm of the function

$$\eta(t) = \mathrm{mes}\{x \in \mathbb{R} : G(x) < t \le F(x)\} - \mathrm{mes}\{x \in \mathbb{R} : F(x) < t \le G(x)\}, \quad t \in \mathbb{R}. \quad (3.1)$$

The properties $F(-\infty) = G(-\infty) = 0$ and $F(\infty) = G(\infty) = 1$ (together with the boundedness of the total variation) ensure that, for any $t \ne 0, 1$, both terms on the right of (3.1) are finite and vanishing outside a large interval, so that $\eta(t)$ is well-defined and compactly supported on the real line. In addition, $\eta(t)$ is bounded on sets $\mathbb{R}\setminus((-\varepsilon, \varepsilon) \cup (1-\varepsilon, 1+\varepsilon))$, $\varepsilon > 0$. Note also that this function appears as the limit case $\eta = \eta_\infty$ for

$$\begin{aligned}\eta_T(t) = \mathrm{mes}\{x \in [-T, T] : G(x) < t \le F(x)\}\\ -\mathrm{mes}\{x \in [-T, T] : F(x) < t \le G(x)\}.\end{aligned}$$

Proposition 3.1 *For all $\mu, \nu \in \mathbb{M}_p$ with distribution functions F and G respectively,*

$$W_p(\mu, \nu) = \left(\int_{-\infty}^{\infty} |\eta(t)|^p \, dt\right)^{1/p}. \quad (3.2)$$

Moreover, for any $T > 0$, with some $|\theta| \le 1$

$$W_p(\mu, \nu) = \left(\int_{-\infty}^{\infty} |\eta_T(t)|^p \, dt\right)^{1/p} + \theta \int_{|x|>T} |F(x) - G(x)|^{1/p} \, dx, \quad (3.3)$$

provided that the last integral is finite.

Equality (3.2) provides a natural generalization of (2.5), since in case $\mu, \nu \in \mathcal{F}_p$, we have $|\eta| = |F^{-1} - G^{-1}|$. On the other hand, formula (3.2) follows from (3.3) by letting $T \to \infty$ (which has to be justified).

Proof Any function u participating in the sup of (1.8) is Lip(α) for $\alpha = 1/p$, with Lipschitz semi-norm ≤ 1, so that

$$|u(F(x)) - u(G(x))| \leq |F(x) - G(x)|^{1/p}.$$

Integrating this inequality over the region $|x| > T$, we obtain

$$\int_{|x|>T} (u(F(x)) - u(G(x)))\, \mathrm{d}x = \theta \int_{|x|>T} |F(x) - G(x)|^{1/p}\, \mathrm{d}x$$

with some $\theta = \theta(u)$ such that $|\theta| \leq 1$. As for the interval $|x| < T$, we use the property (as was explained before) that one can remove the modulus sign from the integrand in (1.8). First write

$$u(F(x)) - u(G(x)) = \int_{-\infty}^{\infty} u'(t)\, \left(1_{\{G(x)<t\leq F(x)\}} - 1_{\{F(x)<t\leq G(x)\}}\right) \mathrm{d}t. \quad (3.4)$$

Here, the integrand is vanishing outside some finite interval. Integrating over x, we get

$$\int_{-T}^{T} (u(F(x)) - u(G(x)))\, \mathrm{d}x = \int_{-\infty}^{\infty} u'(t)\, \eta_T(t)\, \mathrm{d}t.$$

Hence

$$\int_{-\infty}^{\infty} (u(F(x)) - u(G(x)))\, \mathrm{d}x = \int_{-\infty}^{\infty} u'(t)\, \eta_T(t)\, \mathrm{d}t + \theta \int_{|x|>T} |F(x) - G(x)|^{1/p}\, \mathrm{d}x.$$

It remains to take the supremum over all admissible u, and then we obtain (3.3).

To reach the limit case (3.2), fix $N > 1$, $\varepsilon \in (0, \frac{1}{2})$, and assume that u belongs to the class $\mathcal{C}_{N,\varepsilon}$ of all smooth functions on the line whose derivative u' is supported and bounded on the set $A_{N,\varepsilon} = (-N, N) \backslash ((-\varepsilon, \varepsilon) \cup (1-\varepsilon, 1+\varepsilon))$, with $\|u'\|_q \leq 1$. Returning to (3.4), write

$$u(F(x)) - u(G(x)) = \int_{A_{N,\varepsilon}} u'(t)\, v(x,t)\, \mathrm{d}t \quad (3.5)$$

with $v(x,t) = 1_{\{G(x)<t\leq F(x)\}} - 1_{\{F(x)<t\leq G(x)\}}$. If N is large enough and $t \in A_{N,\varepsilon}$, then $v(x,t) = 0$ whenever $|x| > N$. In this case, one may freely integrate (3.5) over x, which gives

$$\int_{-\infty}^{\infty} (u(F(x)) - u(G(x)))\, \mathrm{d}x = \int_{A_{N,\varepsilon}} u'(t)\, \eta(t)\, \mathrm{d}t,$$

where the left integral may be restricted to $[-N, N]$, while the integrand on the right is a bounded function. From this,

$$\sup_{u \in \mathcal{C}_{\mathcal{N},\varepsilon}} \left| \int_{-\infty}^{\infty} (u(F(x)) - u(G(x)))\, \mathrm{d}x \right| = \left(\int_{A_{N,\varepsilon}} |\eta(t)|^p\, \mathrm{d}t \right)^{1/p}.$$

On this step, letting $N \to \infty$ and $\varepsilon \to 0$, the left supremum is extended to all u such that $\|u'\|_q \leq 1$, and then we arrive at (3.2). □

Since

$$\eta_T(t) \;\leq\; \xi_T(t) \;=\; \mathrm{mes}\{x \in [-T, T] : t \text{ is between } F(x) \text{ and } G(x)\}, \tag{3.6}$$

as an immediate consequence of Proposition 3.1, we obtain:

Corollary 3.2 *For all $\mu, \nu \in \mathbb{M}_p$ with (generalized) distribution functions F and G respectively,*

$$W_p(\mu, \nu) \leq \int_{-\infty}^{\infty} |F(x) - G(x)|^{1/p}\, dx.$$

Moreover, for any $T > 0$,

$$W_p(\mu, \nu) \;\leq\; \left(\int_{-\infty}^{\infty} \xi_T(t)^p\, dt \right)^{1/p} + \int_{|x|>T} |F(x) - G(x)|^{1/p}\, dx.$$

The last bound will be used in the proof of Theorem 1.1, while the more precise relation (3.3) is needed to study second order approximations for transport distances as in Theorem 1.3.

Remark In general, the moment assumption $\mu, \nu \in \mathbb{M}_p$ with distribution functions F and G does not guarantee the finiteness of the integral $\int_{-\infty}^{\infty} |F(x) - G(x)|^{1/p}\, \mathrm{d}x$. For a counter-example, one may take for ν any compactly supported measure with total mass one, and for μ a probability measure with distribution function such that $1 - F(x) = (x \log x)^{-p}$ for large x.

4 Normal distribution function

Here we collect a few calculus relations involving the normal distribution function. Although most of them are rather elementary and certainly known, we include some proofs for reader's convenience.

Lemma 4.1 *For all $x \leq 0$, we have $\Phi(x) \leq \frac{1}{2}\, e^{-x^2/2}$. Moreover,*

$$\frac{1}{1 + |x|}\, \varphi(x) \;\leq\; \Phi(x) \;\leq\; \frac{1}{|x|}\, \varphi(x). \tag{4.1}$$

Let us only explain the left inequality in (4.1). The function $v(x) = (1-x)\,\Phi(x) - \varphi(x)$ satisfies $v(-\infty) = 0$ and $v(0) = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} > 0$. In addition,

$$v'(x) = -\Phi(x) + \varphi(x), \quad v''(x) = -(1+x)\,\varphi(x).$$

Hence, v is convex in $x \leq -1$ and concave in $x \geq -1$. Moreover, $v'(x) \geq -\frac{1}{|x|}\,\varphi(x) + \varphi(x) \geq 0$ for $x \leq -1$, so v is also increasing on the half-axis $x \leq -1$. These properties readily imply that v is positive for all $x \leq 0$.

Lemma 4.2 *For all $d \geq 0$ and $x \leq 0$,*

$$\left(1 + |x|^d\right) e^{-x^2/2} \leq 2\pi \left(1 + |x|^{d+1}\right) \Phi(x).$$

Proof If $d = 0$, the left inequality in (4.1) is a bit sharper. In the general case, for a parameter $a > 0$, put $v(x) = \varphi(x) - (a+x)(1-\Phi(x))$, $x \geq 0$. This function is vanishing at infinity, while $v(0) \leq 0$, as long as $a \geq \sqrt{\frac{2}{\pi}}$. Using $1 - \Phi(x) \leq \frac{1}{2}\,e^{-x^2/2}$, we also get

$$v'(x) = -(1-\Phi(x)) + a\varphi(x) \geq 0, \quad a \geq \sqrt{\frac{\pi}{2}},$$

in which case v is increasing. In particular, $v(x) \leq 0$ for all $x \geq 0$, that is, $\frac{\varphi(x)}{1-\Phi(x)} \leq \sqrt{\frac{\pi}{2}} + x$, or $\frac{e^{-x^2/2}}{1-\Phi(x)} \leq \pi + \sqrt{2\pi}\,x$. Equivalently, $\frac{e^{-x^2/2}}{\Phi(x)} \leq \pi + \sqrt{2\pi}\,|x|$ for all $x \leq 0$, which gives

$$\frac{\left(1+|x|^d\right)\,e^{-x^2/2}}{\Phi(x)} \leq \pi(1+|x|)\left(1+|x|^d\right) \leq 2\pi\left(1+|x|^{d+1}\right).$$

□

Lemma 4.3 *For all $T \geq 0$ and $d \geq 0$,*

$$\int_{|x|\geq T} |x|^d\, e^{-x^2/2}\,dx \leq 4d^{d/2}\,e^{-T^2/4}. \tag{4.2}$$

Proof Denote by J_d the integral in (4.2). If $d \geq 1$, we apply the elementary inequalities $x^{d-1}e^{-x^2/4} \leq \left(\frac{2(d-1)}{e}\right)^{(d-1)/2} \leq d^{d/2}$ which give

$$\begin{aligned} J_d &= 2\int_T^\infty x^{d-1}e^{-x^2/4}\cdot xe^{-x^2/4}\,\mathrm{d}x \\ &\leq 4\left(\frac{2(d-1)}{e}\right)^{(d-1)/2} e^{-T^2/4} \leq 4d^{d/2}\,e^{-T^2/4}. \end{aligned}$$

In case $0 \le d \le 1$, one may integrate by parts and write

$$J_d = -\frac{2}{d+1}\, T^{d+1} e^{-T^2/2} + \frac{2}{d+1} \int_T^\infty x^{d+2} e^{-x^2/2}\, \mathrm{d}x \le \frac{1}{d+1}\, J_{d+2}.$$

By the previous step, the last expression is bounded by

$$\begin{aligned}\frac{4}{d+1} \left(\frac{2(d+1)}{e}\right)^{(d+1)/2} e^{-T^2/4} &= 4 \left(\frac{2}{e}\right)^{(d+1)/2} (d+1)^{(d-1)/2}\, e^{-T^2/4} \\ &\le 4 \left(\frac{2}{e}\right)^{(d+1)/2} e^{-T^2/4} \\ &= 4 \left(\frac{2}{e}\right)^{1/2} \left(\frac{2}{de}\right)^{d/2} d^{d/2}\, e^{-T^2/4}.\end{aligned}$$

Here, the quantity $\left(\frac{2}{de}\right)^{d/2}$ is maximized at $d = \frac{2}{e^2}$, at which $\left(\frac{2}{e}\right)^{1/2} \left(\frac{2}{de}\right)^{d/2} = 0.982\ldots < 1$. □

Let us now turn to the inverse function Φ^{-1}. Some of its properties can be explored by involving the so-called Gaussian profile

$$I(t) = \varphi(\Phi^{-1}(t)), \quad 0 \le t \le 1.$$

It is a (strongly) concave function on [0, 1], symmetric about the point 1/2, which behaves near zero like $t\sqrt{2\log(1/t)}$.

Lemma 4.4 *Given $\beta > 1$, the function $\Phi^{-1}(\beta t) - \Phi^{-1}(t)$ is increasing in $0 < t < 1/\beta$.*

This property follows from the (strong) concavity of I and the fact that $(\Phi^{-1}(t))' = 1/I(t)$.

Using $\varphi'(x) = -x\varphi(x)$, by the chain rule, $I'(t) = -\Phi^{-1}(t)$, which immediately implies $I''(t) = -1/I(t)$. Using the I-function, one can simplify further differentiation of the inverse normal distribution function. In particular, for all $t \in (0, 1)$,

$$\left(\Phi^{-1}\right)''(t) = \frac{\Phi^{-1}(t)}{I(t)^2}, \quad \left(\Phi^{-1}\right)'''(t) = \frac{1 + 2\left(\Phi^{-1}(t)\right)^2}{I(t)^3}.$$

Thus, the second derivative is strictly increasing. We are prepared to develop a Taylor expansion of the inverse function up to the linear and second terms.

Lemma 4.5 *If $|\varepsilon| \le \frac{1}{2(1+|x|)}$ $(x \in \mathbb{R})$, then $0 < \Phi(x) + \varepsilon\varphi(x) < 1$ and*

$$\left|\Phi^{-1}(\Phi(x) + \varepsilon\varphi(x)) - x\right| \le 2\varepsilon.$$

Moreover,

$$\left|\Phi^{-1}\left(\Phi(x)+\varepsilon\varphi(x)\right)-x-\varepsilon\right| \,\leq\, 2\,(1+|x|)\,\varepsilon^2.$$

Proof One may assume that $x<0$, so that $t=\Phi(x)<\frac{1}{2}$. The function $R(\varepsilon)=\Phi^{-1}(\Phi(x)+\varepsilon\varphi(x))$ is well-defined in the interval $|\varepsilon|\,\varphi(x)<\Phi(x)$, hence for $|\varepsilon|<\frac{1}{1+|x|}$ (by Lemma 4.1). Clearly, $R(0)=x$, $R'(0)=1$, and putting $\delta=\varepsilon\varphi(x)$, we have

$$R'(\varepsilon)=\frac{1}{I(t+\delta)}\,\varphi(x)\,=\,\frac{I(t)}{I(t+\delta)},$$
$$R''(\varepsilon)=\left(\Phi^{-1}\right)''(t+\delta)\,\varphi(x)^2\,=\,\left(\Phi^{-1}\right)''(t+\delta)\,\frac{I(t)^2}{I(t+\delta)^2}.$$

By the left inequality in (4.1), we have $|\delta|\leq\frac{\Phi(x)}{2}=\frac{t}{2}$. Since $(\Phi^{-1})''$ is increasing, we get

$$R'(\varepsilon)\leq\frac{I(t)}{I(t/2)},\quad |R''(\varepsilon)|\leq\left|\left(\Phi^{-1}\right)''(t/2)\right|\,I(t)^2=\left|\Phi^{-1}(t/2)\right|\,\frac{I(t)^2}{I(t/2)^2}. \tag{4.3}$$

But, by the concavity, $I(t/2)\geq I(t)/2$ for all $t\in[0,1]$, so $R'(\varepsilon)\leq 2$. Also, by Lemma 4.4,

$$\Phi^{-1}(t)-\Phi^{-1}(t/2)\leq\Phi^{-1}(1/2)-\Phi^{-1}(1/4)<0.7$$

for $0<t\leq\frac{1}{2}$, and therefore $|\Phi^{-1}(t/2)|<1+|\Phi^{-1}(t)|=1+|x|$. As a result, the last expression on the right-hand side of (4.3) is bounded by $4\,(1+|x|)$. It remains to apply Taylor's formula. □

5 Perturbations of the Gaussian measure

The second bound of Corollary 3.2 may be used to quantify the closeness of the Edgeworth correction to the standard normal law in terms of the extended transport distance. As a preliminary step, here we prove a more general assertion. Let $\nu\in\mathbb{M}_p$ have the (generalized) distribution function G.

Proposition 5.1 *Assume that, for some real numbers $\varepsilon>0$ and $d\geq 1$,*

$$|G(x)-\Phi(x)|\,\leq\,\varepsilon\,\left(1+|x|^d\right)\,e^{-x^2/2},\quad x\in\mathbb{R}. \tag{5.1}$$

Then

$$W_p(\nu,\gamma)\,\leq\,C_{p,d}\,\varepsilon,$$

where one may take $C_{p,d}=(Cpd)^{3(d+1)/2}$ with some absolute constant C.

First, we derive:

Lemma 5.2 *Assume that* (5.1) *is fulfilled with some* $\varepsilon \in (0, 1/e]$ *and* $d \geq 1$. *Then*

$$W_p(\nu, \gamma) \ \leq \ 13\, d^{d/2p} \sqrt{p}\, \varepsilon + \left(\int_{-\infty}^{\infty} \xi_T(t)^p \, \mathrm{d}t \right)^{1/p}, \tag{5.2}$$

where $T = 2\sqrt{p \, \log(1/\varepsilon)}$ *and*

$$\xi_T(t) = \mathrm{mes}\{x \in [-T, T] : t \ \mathit{between} \ \Phi(x) \ \mathit{and} \ G(x)\}.$$

Proof Applying $\Phi(x) \leq \frac{1}{2}\, e^{-x^2/2}$ $(x \leq 0)$ and the assumption (5.1), in case $x \leq -1$, we have

$$|G(x)| \leq \Phi(x) + \left(1 + |x|^d\right) e^{-x^2/2} \leq \frac{5}{2}\, |x|^d \, e^{-x^2/2}$$

and hence

$$|G(x)|^{1/p} \leq \frac{5}{2}\, |x|^{d/p} \, e^{-x^2/2p}.$$

A similar bound holds true for $|1 - G(x)|^{1/p}$ in case $x \geq 1$. Since $T \geq 1$, we get, by (4.2),

$$\begin{aligned}
\int_{-\infty}^{-T} |G(x)|^{1/p} \, \mathrm{d}x + \int_{T}^{\infty} |1 - G(x)|^{1/p} \, \mathrm{d}x &\leq \frac{5}{2} \int_{|x| \geq T} |x|^{d/p} \, e^{-x^2/2p} \, \mathrm{d}x \\
&= \frac{5}{2}\, p^{\frac{p+d}{2p}} \int_{|y| \geq T/\sqrt{p}} |y|^{d/p} \, e^{-y^2/2} \, \mathrm{d}y \\
&\leq \frac{5}{2}\, p^{\frac{p+d}{2p}} \cdot 4 \left(\frac{d}{p}\right)^{\frac{d}{2p}} e^{-T^2/4p} \\
&= 10\, d^{d/2p} \sqrt{p}\, \varepsilon.
\end{aligned}$$

A similar bound is also true for Φ in place of G, even with better constants, since then

$$\begin{aligned}
\int_{-\infty}^{-T} \Phi(x)^{1/p} \, \mathrm{d}x + \int_{T}^{\infty} (1 - \Phi(x))^{1/p} \, \mathrm{d}x &\leq \int_{|x| \geq T} e^{-x^2/2p} \, \mathrm{d}x \\
&= 2\sqrt{p} \int_{T/\sqrt{p}}^{\infty} e^{-y^2/2} \, \mathrm{d}y \\
&= 2\sqrt{2\pi p}\, \left(1 - \Phi(T/\sqrt{p})\right) \\
&\leq \sqrt{2\pi p}\, e^{-T^2/4p} \ = \ \sqrt{2\pi p}\, \varepsilon.
\end{aligned}$$

Combining the two inequalities and applying Corollary 3.2, we arrive at the bound (5.2). □

Proof of Proposition 5.1. We use Lemma 5.2 with $\varepsilon \le \varepsilon_0 \equiv (80\,p(d+1))^{-(d+1)}$ and in essence the linear bound of Lemma 4.5. As before, let $T = 2\sqrt{p \log(1/\varepsilon)}$. To estimate $\xi_T(t)$, one may assume that $t \le \frac{1}{2}$ (by the symmetry of the problem about the point $t = \frac{1}{2}$).

First consider the inequalities $\Phi(x) < t < G(x)$. In particular, $t > 0$ and $x < \Phi^{-1}(t) \le 0$. By the assumption (5.1) and applying Lemma 4.2, we have, for all $x \le 0$,

$$t \;<\; G(x) \;\le\; \Phi(x) + \varepsilon\left(1 + |x|^d\right) e^{-x^2/2} \;\le\; \Phi(x)\,\left(1 + 2\pi\,\left(1 + |x|^{d+1}\right)\,\varepsilon\right). \tag{5.3}$$

In the interval $|x| \le T$,

$$\begin{aligned}\left(1 + |x|^{d+1}\right)\,\varepsilon \le \left(1 + T^{d+1}\right)\,\varepsilon \;&\le\; 2T^{d+1}\varepsilon \\ &= 2\varepsilon\,(4p\,\log(1/\varepsilon))^{(d+1)/2} \;\equiv\; 2v(\varepsilon).\end{aligned}$$

Putting $\varepsilon = \exp(-s)$, we have $v(\varepsilon) = (4ps)^{(d+1)/2}\,e^{-s/2}\,e^{-s/2} \le \left(\frac{4p(d+1)}{e}\right)^{(d+1)/2}\sqrt{\varepsilon}$. Hence

$$2T^{d+1}\varepsilon \le 2\,\left(\frac{4p(d+1)}{e}\right)^{(d+1)/2}\sqrt{\varepsilon}.$$

But, since $\varepsilon \le \varepsilon_0$, we get $2T^{d+1}\varepsilon \le 2\,\left(\frac{1}{20e}\right)^{(d+1)/2} \le \frac{1}{10e}$ and thus

$$2\pi\,\left(1 + |x|^{d+1}\right)\varepsilon \le \frac{2\pi}{10\,e} < \frac{1}{4}. \tag{5.4}$$

In particular, $1 + 2\pi\,\left(1 + |x|^{d+1}\right)\varepsilon < 2$ and $t < G(x) \le 2\,\Phi(x)$ implying $|x| \le |\Phi^{-1}(t/2)|$. Hence, from (5.3),

$$t < \Phi(x)\,\left(1 + 2\pi\,\left(1 + \left|\Phi^{-1}(t/2)\right|^{d+1}\right)\,\varepsilon\right).$$

By Lemma 4.4 (as was already noted before), $|\Phi^{-1}(t/2)| \le 0.7 + |\Phi^{-1}(t)|$. Applying the latter together with Jensen's inequality, the above bound can easily be simplified to

$$t < \Phi(x)\,\left(1 + 5^{d+1}\,\left(1 + \left|\Phi^{-1}(t)\right|^{d+1}\right)\,\varepsilon\right),$$

which is solved as $x > \Phi^{-1}\left(\frac{t}{1+c(t)\varepsilon}\right)$ with

$$c(t) = 5^{d+1}\left(1 + \left|\Phi^{-1}(t)\right|^{d+1}\right). \tag{5.5}$$

Consequently, for any $t \in (0, \frac{1}{2}]$,

$$\mathrm{mes}\,\{x \in [-T, T] : \Phi(x) < t < G(x)\} \le \Phi^{-1}(t) - \Phi^{-1}\left(\frac{t}{1+c(t)\varepsilon}\right). \tag{5.6}$$

Now consider the second possibility described by the inequalities $G(x) < t < \Phi(x)$. Necessarily $x > \Phi^{-1}(t)$. By the assumption (5.1) and once more Lemma 4.2, for all x,

$$t > G(x) \ge \Phi(x)\left(1 - \varepsilon\,\frac{\left(1+|x|^d\right)\,e^{-x^2/2}}{\Phi(x)}\right) \ge \Phi(x)\left(1 - 2\pi\,\left(1+|x|^{d+1}\right)\varepsilon\right). \tag{5.7}$$

Here, for $|x| \le T$, according to (5.4), the expression on the right is positive and, moreover, it is larger than $\frac{3}{4}\,\Phi(x)$. Using $\frac{1}{1-\delta} \le 1 + \frac{4}{3}\,\delta$ for $0 \le \delta \le \frac{1}{4}$, we therefore obtain from (5.7) that

$$\Phi(x) \le t\left(1 + \frac{8\pi}{3}\,\left(1+|x|^{d+1}\right)\varepsilon\right). \tag{5.8}$$

In particular, $t > 0$. Moreover, by Lemma 4.4, $x < \Phi^{-1}\left(\frac{4}{3}\,t\right) \le \Phi^{-1}(t) + \Phi^{-1}\left(\frac{2}{3}\right) < \Phi^{-1}(t) + 0.7$. Hence, $|x| \le 0.7 + \left|\Phi^{-1}(t)\right|$, and (5.8) readily implies that

$$\Phi(x) \le t\left(1 + 5^{d+1}\left(1 + \left|\Phi^{-1}(t)\right|^{d+1}\right)\varepsilon\right).$$

It is solved as $x < \Phi^{-1}\left(t\,(1 + c(t)\varepsilon)\right)$ with the same function $c(t)$ as in (5.5).

As a result, we obtain an analog of (5.6), namely

$$\mathrm{mes}\,\{x \in [-T, T] : G(x) < t < \Phi(x)\} \le \Phi^{-1}\left(t(1 + c_1(t)\varepsilon)\right) - \Phi^{-1}(t),$$

where the left-hand side is vanishing for $t \le 0$. Combining this with (5.6), we conclude that $\xi_T(t) = 0$ for $t \le 0$, and moreover,

$$\xi_T(t) \le \Phi^{-1}\left(t(1 + c(t)\varepsilon)\right) - \Phi^{-1}\left(\frac{t}{1+c(t)\varepsilon}\right) \tag{5.9}$$

for any $0 < t \le \frac{1}{2}$ with the function $c(t)$ described in (5.5). On this stage, we need to verify that the values $t(1 + c(t)\varepsilon)$ are bounded away from 1. Putting $t = \Phi(-y)$,

$y \geq 0$, we have

$$t\,\left|\Phi^{-1}(t)\right|^{d+1} = y^{d+1}\Phi(-y) \leq \frac{1}{2}\,y^{d+1}\,e^{-y^2/2} \leq \frac{1}{2}\,(d+1)^{(d+1)/2}.$$

Hence

$$\begin{aligned} tc(t)\,\varepsilon &= 5^{d+1}\left(t + t\,\left|\Phi^{-1}(t)\right|^{d+1}\right)\varepsilon \\ &\leq 5^{d+1}\left(\frac{1}{2} + \frac{1}{2}\,(d+1)^{(d+1)/2}\right)\cdot(80\,p(d+1))^{-(d+1)} \\ &\leq 5^{d+1}\,(d+1)^{(d+1)/2}\cdot(80\,p(d+1))^{-(d+1)} \leq \frac{1}{32}, \end{aligned}$$

so that, $t(1 + c(t)\varepsilon) \leq 0.6$.

Now, to simplify the bound (5.9), we recall that the derivative of $\Phi^{-1}(s)$ is $1/I(s)$, while, by the concavity, $I(s) \geq 2I(1/2)\,s \geq \frac{1}{2}\,s$ in the interval $0 < s \leq \frac{1}{2}$. Hence, using also $I(0.6) > \frac{1}{3}$, we have, for all $0 < a < b < 0.6$ $(a \leq \frac{1}{2})$,

$$\Phi^{-1}(b) - \Phi^{-1}(a) = \int_a^b \frac{1}{I(s)}\,\mathrm{d}s \leq \frac{b-a}{\min\{I(a), I(b)\}} \leq \frac{b-a}{\min\left\{\frac{1}{2}\,a, \frac{1}{3}\right\}} \leq 2\,\frac{b-a}{a}.$$

Using this bound in (5.9) with $a = \frac{t}{1+c(t)\varepsilon}$, $b = t(1 + c(t)\varepsilon)$, first note that

$$b - a = t\varepsilon\,\frac{2c(t) + c(t)^2\varepsilon}{1 + c(t)\varepsilon} \leq t\varepsilon\,(1 + c(t))^2.$$

Using $a \geq \frac{t}{1+c(t)}$, we get $\frac{b-a}{a} \leq 2\varepsilon(1 + \varepsilon(t))^3$. Thus, we may conclude that, for all $0 < t \leq \frac{1}{2}$.

$$\xi_T(t)\ \leq\ 4\varepsilon\,(1 + c(t))^3 \tag{5.10}$$

with $\xi_T(t) = 0$ for $t \leq 0$.

But the function $c(t)$ is symmetric about $1/2$, so (5.10) remains to hold for $\frac{1}{2} \leq t < 1$ as well, and $\xi_T(t) = 0$ for $t \geq 1$. Moreover, the function on the right-hand side of (5.10) belongs to $L^p(0, 1)$ and has an L^p-norm which only depends on p and d. To derive a quantitative bound on this norm, one may use the property that $Z = \Phi^{-1}(t)$ has a standard normal distribution under the uniform measure on $(0, 1)$. First, by Jensen's inequality, $(1 + c(t))^{3p} \leq 2^{3p-1} + 2^{3p-1}c(t)^{3p}$ and $c(t)^{3p} \leq 5^{3p(d+1)}\,2^{3p-1}\,\left(1 + |Z|^{3p(d+1)}\right)$. Using also (4.2) with $T = 0$ so as to bound absolute moments of Z, we have

$$\begin{aligned} 4\int_0^1 (1 + c(t))^{3p}\,\mathrm{d}t &\leq 2^{3p+1} + 5^{3p(d+1)}\,4^{3p}\,\left(1 + \mathbb{E}\,|Z|^{3p(d+1)}\right) \\ &\leq 2^{3p+1} + 5^{3p(d+1)}\,4^{3p}\cdot 2\pi\,\left(1 + 4\,(3p(d+1))^{3p(d+1)/2}\right). \end{aligned}$$

From this and (5.10),

$$\frac{1}{\varepsilon}\left(\int_0^1 \xi_T(t)^p\,\mathrm{d}t\right)^{1/p} \le 16 + 128\,\pi\cdot 5^{3(d+1)}\left(1+4\,(3p(d+1))^{3(d+1)/2}\right)$$
$$\le (Cpd)^{3(d+1)/2}$$

with some absolute constant C. In view of (5.2), a similar bound also holds for $\frac{1}{\varepsilon}\,W_p(\nu,\gamma)$.

Finally, in case $\varepsilon \ge \varepsilon_0$, one may apply the first estimate of Corollary 3.2. Under the assumption (5.1), and using (5.2) with $T=0$, we get

$$W_p(\nu,\gamma) \le \int_{-\infty}^{\infty} |G(x)-\Phi(x)|^{1/p}\,\mathrm{d}x$$
$$\le \varepsilon^{1/p}\int_{-\infty}^{\infty}\left(1+|x|^{d/p}\right)e^{-x^2/2p} \le 4\varepsilon^{1/p}\sqrt{2p\pi}\left(1+d^{d/2p}\right).$$

Since $\varepsilon^{1/p} \le \frac{1}{\varepsilon_0}\,\varepsilon = (80\,p(d+1))^{(d+1)}\,\varepsilon$, we get an estimate $W_p(\nu,\gamma) \le (Cpd)^{d+3/2}\,\varepsilon$. □

6 Second order approximation

Under stronger assumptions on G, Proposition 5.1 may further be sharpened with the help of the quadratic bound of Lemma 4.5. Let us consider a measure $\nu \in \mathbb{M}_p$ with (generalized) distribution function of the form

$$G(x) = \Phi(x) + \varepsilon(x)\,\varphi(x), \quad x \in \mathbb{R}. \tag{6.1}$$

when $\varepsilon(x)$ is small and regular in some sense, our next aim is to show that $W_p(\nu,\gamma)$ is described by the quantity

$$I_p = \left(\int_{-\infty}^{\infty} |\varepsilon(x)|^p\,\varphi(x)\,\mathrm{d}x\right)^{1/p} = \|\varepsilon(Z)\|_p, \quad Z \sim N(0,1),$$

up to an error term which has a smaller order. To make a corresponding estimate simpler, we assume that the integral $\int_{-\infty}^{\infty} |F(x)-G(x)|^{1/p}\,\mathrm{d}x$ is finite. Below we use θ to denote a quantity such that $|\theta| \le 1$.

Proposition 6.1 *If the function $\varepsilon(x)$ is smooth and satisfies on the interval $[-T,T]$*

$$|\varepsilon(x)| \le \frac{1}{2\,(1+|x|)}, \quad \left|\varepsilon'(x)\right| \le \frac{1}{2}, \tag{6.2}$$

then

$$W_p(\nu,\gamma) = A_p + \theta\int_{|x|>T} |\varepsilon(x)|^{1/p}\,\varphi(x)^{1/p}\,dx, \tag{6.3}$$

where $A_p \geq 0$ *satisfies*

$$\begin{aligned} \left|A_p^p - I_p^p\right| \leq \int_{|x|>T} |\varepsilon(x)|^p \,\varphi(x)\,\mathrm{d}x + \int_{-T}^{T} |\varepsilon(x)|^p \,|\varepsilon'(x)|\,\varphi(x)\,\mathrm{d}x \\ + \left(p\cdot 2^{p+1}+1\right) \int_{-T}^{T} |\varepsilon(x)|^{p+1}\,(1+|x|)\,\varphi(x)\,\mathrm{d}x \;+\; 2\,(2T)^{p-1}\,\varphi(T). \end{aligned} \tag{6.4}$$

Letting $T \to \infty$ in (6.3)–(6.4), we get a simpler representation

$$\begin{aligned} W_p^p(\nu,\gamma) = I_p^p \;+\; \theta \int_{-\infty}^{\infty} |\varepsilon(x)|^p \,|\varepsilon'(x)|\,\varphi(x)\,\mathrm{d}x \\ + \theta\left(p\cdot 2^{p+1}+1\right) \int_{-\infty}^{\infty} |\varepsilon(x)|^{p+1}\,(1+|x|)\,\varphi(x)\,\mathrm{d}x, \end{aligned}$$

however, under a much stronger requirement that condition (6.2) is fulfilled on the whole real line. In further applications, we will use Proposition 6.1 with $\varepsilon(x)$ being small multiples of polynomials [and then (6.2) may only hold on bounded, although large intervals].

Proof Proposition 3.1 for the couple (Φ, G) provides (6.3) with $A_p^p = \int_{-\infty}^{\infty} |\eta_T(t)|^p\,\mathrm{d}t$, where $\eta_T(t) = \eta_T^+(t) - \eta_T^-(t)$,

$$\begin{aligned} \eta_T^+(t) &= \mathrm{mes}\{x \in [-T,T] : G(x) < t \leq \Phi(x)\}, \\ \eta_T^-(t) &= \mathrm{mes}\{x \in [-T,T] : \Phi(x) < t \leq G(x)\}. \end{aligned}$$

Put $G(-T) = t_0$, $G(T) = t_1$, and note that $0 < t_0 < t_1 < 1$ (by Lemma 4.5). In particular, $0 < G(x) < 1$ on $[-T,T]$, so, $\eta_T^+(t) = \eta_T^-(t) = 0$ outside $[0,1]$, and thus

$$A_p^p = \int_0^1 |\eta_T(t)|^p\,\mathrm{d}t.$$

If $0 < t < t_0$, then $\eta_T^+(t) = 0$ and

$$\eta_T^-(t) \leq \mathrm{mes}\{x \in [-T,T] : x < \Phi^{-1}(t)\} \leq (\Phi^{-1}(t)+T)^+ = \max\{\Phi^{-1}(t)+T, 0\}.$$

The latter expression may only be positive in the interval $\Phi(-T) \leq t \leq G(-T)$, whose length is at most $|\varepsilon(-T)|\,\varphi(-T) \leq \frac{1}{2\,(1+T)}\,\varphi(T)$, according to (6.1)-(6.2). On the other hand, $\eta_T^-(t) \leq 2T$. Hence

$$\int_0^{t_0} |\eta_T(t)|^p\,\mathrm{d}t \;\leq\; (2T)^{p-1}\,\varphi(T). \tag{6.5}$$

If $t_1 < t < 1$, then $\eta_T^-(t) = 0$ and, by similar arguments,

$$\int_{t_1}^{1} |\eta_T(t)|^p \,\mathrm{d}t \;\leq\; (2T)^{p-1}\,\varphi(T). \tag{6.6}$$

As for the intermediate interval (t_0, t_1), we first note that G has density

$$g(x) = \big(1 + \varepsilon'(x) - \varepsilon(x)\,x\big)\;\varphi(x),$$

which is positive on $[-T, T]$ under the assumption (6.2). Hence G is increasing on this interval, and moreover, for all $|x| \leq T$,

$$|g(x) - \varphi(x)| \leq \big(|\varepsilon'(x)| + |\varepsilon(x)|\,|x|\big)\;\varphi(x) < \varphi(x). \tag{6.7}$$

Denote by $G^{-1} : [t_0, t_1] \to [-T, T]$ the inverse function to G restricted to $[-T, T]$. Then, for $t_0 < t < t_1$,

$$\eta_T^+(t) \;=\; \Big(\Phi^{-1}(t) - G^{-1}(t)\Big)^+, \quad \eta_T^-(t) \;=\; \Big(G^{-1}(t) - \Phi^{-1}(t)\Big)^+,$$

so that $|\eta_T(t)| = |\Phi^{-1}(t) - G^{-1}(t)|$. Changing the variable $t = G(x)$, one may write

$$\begin{aligned}\int_{t_0}^{t_1} |\eta_T(t)|^p \,\mathrm{d}t &= \int_{t_0}^{t_1} |\Phi^{-1}(t) - G^{-1}(t)|^p \,\mathrm{d}t \;=\; \int_{-T}^{T} |\Phi^{-1}(G(x)) - x|^p\; g(x)\,\mathrm{d}x\\ &\equiv J_T.\end{aligned}$$

By Lemma 4.5,

$$\Delta(x) \;\equiv\; \Phi^{-1}(G(x)) - x \;=\; \varepsilon(x)\;\left(1 + 2\theta\,(1 + |x|)\,\varepsilon(x)\right), \quad |\theta| \leq 1.$$

Applying a simple inequality $|\,|1 + y|^p - 1| \;\leq\; p \cdot 2^{p-1}\,|y|$, $|y| \;\leq\; 1$, with $y = 2\theta\,(1 + |x|)\,\varepsilon(x)$, we get

$$|\Delta(x)|^p \;=\; |\varepsilon(x)|^p + p \cdot 2^p\;\theta\,(1 + |x|)\,|\varepsilon(x)|^{p+1}$$

and thus [using $g \leq 2\varphi$ according to (6.7)],

$$\begin{aligned}J_T &= \int_{-T}^{T} |\varepsilon(x)|^p\; g(x)\,\mathrm{d}x + \theta p\,\cdot 2^p \int_{-T}^{T} (1 + |x|)\,|\varepsilon(x)|^{p+1}\; g(x)\,\mathrm{d}x\\ &= \int_{-T}^{T} |\varepsilon(x)|^p\; g(x)\,\mathrm{d}x + \theta p \cdot 2^{p+1} \int_{-T}^{T} (1 + |x|)\,|\varepsilon(x)|^{p+1}\;\varphi(x)\,\mathrm{d}x.\end{aligned}$$

Moreover, the first integral on the right may be written as

$$\int_{-T}^{T} |\varepsilon(x)|^p\;\varphi(x)\,\mathrm{d}x + \theta \int_{-T}^{T} |\varepsilon(x)|^p\;\big(\varepsilon'(x) - \varepsilon(x)\,x\big)\;\varphi(x)\,\mathrm{d}x.$$

Therefore,

$$\begin{aligned} J_T &= \int_{-T}^{T} |\varepsilon(x)|^p \, \varphi(x) \, dx \\ &\quad + \theta \left(p \cdot 2^{p+1} + 1\right) \int_{-T}^{T} (1 + |x|) \, |\varepsilon(x)|^{p+1} \, \varphi(x) \, dx \\ &\quad + \theta \int_{-T}^{T} |\varepsilon(x)|^p \, |\varepsilon'(x)| \, \varphi(x) \, dx. \end{aligned} \tag{6.8}$$

The first integral in (6.8) is exactly I_p^p up to the summand $\int_{|x|>T} |\varepsilon(x)|^p \, \varphi(x) \, dx$. Since, by (6.5)-(6.6), $A_p^p = J_T + 2\theta \, (2T)^{p-1} \, \varphi(T)$, the relation (6.8) leads to the estimate (6.4). $\square$

7 Edgeworth-type approximation in the non-i.i.d. case

Let us return to the sum $S_n = X_1 + \cdots + X_n$ of n independent random variables, such that $\mathbb{E}X_k = 0$, $\sum_{k=1}^n \mathbb{E}X_k^2 = 1$, and with finite Lyapunov coefficient $L_s = \sum_{k=1}^n \mathbb{E}\,|X_k|^s$ $(s > 2)$. In this case, the characteristic function $f_n(t) = \mathbb{E}\, e^{itS_n}$ has $[s]$ continuous derivatives on the real line, and one may introduce the cumulants

$$\gamma_r = \gamma_r(S_n) = \frac{d^r}{i^r \, dt^r} \log f_n(t)\Big|_{t=0} = \sum_{k=1}^n \gamma_r(X_k), \quad r = 1, \dots, [s].$$

The first values are $\gamma_1 = 0$, $\gamma_2 = 1$. Each γ_r represents a polynomial in the moments of X_k up to order r; however, an explicit expression for them will not be needed for our aims.

Writing $s = m + \alpha$ with integer $m \geq 2$ and $0 < \alpha \leq 1$, the corrected normal "characteristic" function of order m for the distribution of S_n is given by the formula

$$g_m(t) = e^{-t^2/2} + e^{-t^2/2} \sum \frac{1}{k_1! \cdots k_{m-2}!} \left(\frac{\gamma_3}{3!}\right)^{k_1} \cdots \left(\frac{\gamma_m}{m!}\right)^{k_{m-2}} (it)^k,$$

where the summation is running over all collections of non-negative integers $k_1, \dots, k_{m-2}$ that are not all zero and such that $k_1 + 2k_2 + \cdots + (m-2)k_{m-2} \leq m-2$, with $k = 3k_1 + \cdots + mk_{m-2}$. The polynomial in the sum has degree at most $3(m-2)$ in the variable t. The index m for g_m indicates that the cumulants up to γ_m participate in the construction. If $s = m+1$ is integer, one may also consider the function g_{m+1}, involving the next cumulant γ_{m+1}.

The function g_m represents the Fourier–Stieltjes transform of a signed measure ν_m with density

$$\varphi_m(x) = \varphi(x) + \varphi(x) \sum \frac{1}{k_1! \cdots k_{m-2}!} \left(\frac{\gamma_3}{3!}\right)^{k_1} \cdots \left(\frac{\gamma_m}{m!}\right)^{k_{m-2}} H_k(x),$$

where the summation is as above, and where $H_k(x)$ denotes the Chebyshev–Hermite polynomial with leading term x^k. Since $H_k(x)\varphi(x) = -(H_{k-1}(x)\varphi(x))'$, the corresponding generalized distribution function

$$\Phi_m(x) = \nu_m((-\infty, x]) = \int_{-\infty}^{x} \varphi_m(y)\,\mathrm{d}y \quad (x \in \mathbb{R})$$

may explicitly be written by virtue of the analogous expression

$$\Phi_m(x) = \Phi(x) - \varphi(x) \sum \frac{1}{k_1! \cdots k_{m-2}!} \left(\frac{\gamma_3}{3!}\right)^{k_1} \cdots \left(\frac{\gamma_m}{m!}\right)^{k_{m-2}} H_{k-1}(x). \tag{7.1}$$

Let us note that, when $2 < s \le 3$, necessarily $g_2(t) = e^{-t^2/2}$, that is, ν_2 is the standard Gaussian measure γ. If $s > 3$ and L_s is small, the measure ν_m is close to γ in many senses. For example, if $L_s \le 1$, there is a simple bound on the total variation distance

$$\|\nu_m - \gamma\|_{\mathrm{TV}} \le m\sqrt{(3(m-2))!}\, L_s^{\frac{1}{s-2}}$$

(cf. [12]). Here the right-hand side may be replaced with a much smaller quantity, if we compare the corresponding "characteristic" functions. In particular, we have:

Proposition 7.1 *If $s \ge 3$, then in the interval $|t| \le \frac{1}{L_3}$, for all $r = 0, 1, \ldots, [s]$,*

$$\left| \frac{d^r}{dt^r} (f_n(t) - g_m(t)) \right| \le C_s L_s \min\left\{1, |t|^{s-r}\right\} e^{-t^2/8}, \tag{7.2}$$

where C_s depends on s only, e.g., one may take $C_s = (Cs)^{3s}$ with some absolute constant C. In case $2 < s < 3$, the same inequality holds true in the interval $|t| \le (6L_s)^{-\frac{1}{s-2}}$ for $r = 0, 1, 2$.

In the literature, inequalities similar to (7.2) can be found for integer values $s = m + 1$, often for i.i.d. summands and $r = 0$, only. In the book by Petrov [27], (7.2) is proved in the i.i.d. case without the derivative of the maximal order $p = m + 1$, and with an indefinite constant C_s (Lemma 4, p. 140). Bikjalis derived a more precise statement with explicit constants that also depend on r ([8], cf. also [7,34] on the non-i.i.d. case with $r = 0$). A variant of (7.2) can be found in the book by Bhattacharya and Ranga Rao [6], who considered multidimensional summands. Their Theorem 9.9 covers the interval of the form $|t| \le c_s L_s^{-1/(s-2)}$ for all $r \le m + 1$, although it does not specify the constants as functions of s. As easy to see, the interval $|t| \le 1/L_3$ in (7.2) is longest possible (up to an absolute factor), but the question on the worst growth of the s-dependent constants in such inequalities seems to be open. The current formulation with not necessarily integer values of s may be found in the recent paper [12].

Proposition 7.1 may be used to derive the following non-uniform bounds.

Proposition 7.2 *Let $L_s < \infty$ $(s \geq 3)$, and suppose that the characteristic function $f_n(t)$ is vanishing outside the interval $|t| \leq L_s^{-1/(s-2)}$. Then, for all $x \in \mathbb{R}$,*

$$|F_n(x) - \Phi_m(x)| \leq \frac{C_s L_s}{(1+|x|)^{m-1}}, \quad |F_n(x) - \Phi_m(x)| \leq \frac{C_s L_s/\alpha}{(1+|x|)^m}, \tag{7.3}$$

where one may take $C_s = (Cs)^{3s}$ with some absolute constant C.

Proof The function $A_r(x) = x^r\,(F_n(x) - \Phi_m(x))$ has bounded variation on the real line, and its Fourier–Stieltjes transform may be written as

$$a_r(t) = i^{-r} \frac{d^r}{dt^r} \frac{a(t)}{t} = i^{-r} \int_0^1 \Big(a^{(r)}(t) - a^{(r)}(\eta t)\Big)\, r\eta^{r-1}\, d\eta \quad (t \neq 0),$$

where $a(t) = f_n(t) - g_m(t)$ (for more details, we refer an interested reader to [11]). Hence, by the Fourier inversion formula, for all x,

$$|A_r(x)| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_0^1 \frac{\big|a^{(r)}(t)\big| + \big|a^{(r)}(\eta t)\big|}{|t|}\, r\eta^{r-1}\, d\eta\, dt = \frac{2}{\pi} \int_0^{\infty} \frac{\big|a^{(r)}(t)\big|}{t}\, dt.$$

Using the assumption on f_n and applying (7.2), we see that the last integral does not exceed

$$C_s L_s \int_0^{L_s^{-\frac{1}{s-2}}} \frac{\min\big\{1, |t|^{s-r}\big\}}{t}\, e^{-t^2/8}\, dt + \int_{L_s^{-\frac{1}{s-2}}}^{\infty} \frac{\big|g^{(r)}(t)\big|}{t}\, dt. \tag{7.4}$$

Up to an absolute constant C, the derivatives of the corrected normal characteristic function admit the bound

$$|g^{(r)}(t)| \leq (Cs)^{2s} L_s\, e^{-t^2/8}, \quad \text{for } |t| \max\Big\{L_s^{\frac{1}{s-2}}, L_s^{\frac{1}{3(s-2)}}\Big\} \geq 1, \quad r = 0, 1, \dots, [s]$$

(cf. [12], Proposition 17.1). Hence, if $L_s \leq 1$, the second integral in (7.4) is bounded by a similar quantity $(Cs)^{2s} L_s$. One can make the same conclusion about the first integral in case $r \leq m-1$, and then we are led to the first bound in (7.3). However, in case $r = m$, when integrating over $(0, 1)$, we gain an additional factor, which may only be bounded by $1/\alpha$.

Let us also note that, by Taylor's formula, $|f_n(t) - 1| \leq \frac{1}{2} t^2$, so $f_n(t)$ may not vanish in the interval $|t| < \sqrt{2}$, and thus necessarily $L_s \leq 1$. □

8 Smoothing

Keeping the same notations, the bounds in (7.3) immediately yield:

Proposition 8.1 *Let L_s be finite for $s \geq 3$ and let $1 \leq p \leq s-2$. Under the assumptions of Proposition 7.2, for the distribution μ_n of S_n we have*

$$W_p(\mu_n, \nu_m) \leq (C_s L_s)^{1/p} \tag{8.1}$$

with some s-dependent constant C_s.

Proof We apply Corollary 3.2 and the first inequality in (7.3) to get

$$W_p(\mu_n, \nu_m) \leq \int_{-\infty}^{\infty} |F_n(x) - \Phi_m(x)|^{\frac{1}{p}}\, \mathrm{d}x \leq (C_s L_s)^{\frac{1}{p}} \int_{-\infty}^{\infty} \frac{\mathrm{d}x}{(1+|x|)^{\frac{m-1}{p}}}.$$

If $\alpha \leq \frac{1}{2}$, then $\frac{m-1}{p} \geq \frac{p+1-\alpha}{p} \geq 1 + \frac{1}{2p}$. Hence, the last integral does not exceed $4p$. In case $\alpha > \frac{1}{2}$, one may use the second inequality in (7.3), which similarly gives

$$W_p(\mu_n, \nu_m) \leq 2\,(C_s L_s)^{\frac{1}{p}} \int_{-\infty}^{\infty} \frac{\mathrm{d}x}{(1+|x|)^{\frac{m}{p}}} \leq 4p\,(C_s L_s)^{\frac{1}{p}}.$$

□

Using a smoothing argument, the assumption on the support,

$$f_n(t) = 0 \ \text{ outside the interval } \ |t| \leq L_s^{-\frac{1}{s-2}}, \tag{8.2}$$

can be removed from Proposition 8.1 for the critical value $p = s-2$. As a standard choice of smoothing, we consider the probability densities of the form

$$w_r(x) = \frac{\lambda}{\Lambda} \left(\frac{\sin(\lambda x)}{\lambda x}\right)^{2r}, \quad r = 2, 3, \dots,$$

where the positive parameters $\Lambda = \Lambda_r$ and $\lambda = \lambda_r$ are defined by

$$\Lambda = \int_{-\infty}^{\infty} \left(\frac{\sin x}{x}\right)^{2r} \mathrm{d}x, \quad \lambda^2 = \frac{1}{\Lambda} \int_{-\infty}^{\infty} x^2 \left(\frac{\sin x}{x}\right)^{2r} \mathrm{d}x.$$

Here, the normalizing constant Λ is chosen to ensure that w_r is a probability density, while the choice of λ guarantees that the second moment is equal to 1. The corresponding characteristic functions $h_r(t)$ are supported on the segments $[-T_r, T_r]$, where $T_r = \frac{2r}{\lambda} > 1$; they represent normalized rescaled r-fold power convolutions of the triangle characteristic function $(1-|t|)_+$.

Let ξ be a random variable with density w_r. Clearly, it has finite r-th absolute moment $M_r = \mathbb{E}\,|\xi|^r$. Taking $r = [s]+1$, we apply Proposition 8.1 to the random vector $\widetilde{X}$ in $\mathbb{R}^{n+1}$ with components $\sqrt{1-\tau^2}\,X_1, \dots, \sqrt{1-\tau^2}\,X_n, \tau\xi$, where $\tau =$

$cL_s^{1/(s-2)} \le 1$ with a proper constant $c > 0$, and assuming that ξ is independent of all X_k. Then the sum of the components

$$S_{n+1} = \sqrt{1-\tau^2}\, S_n + \tau\xi$$

has the characteristic function $f_{n+1}(t) = f_n(t)h_r(\tau t)$ vanishing outside the interval $|t| \le T_r/\tau$. In addition, assuming that $L_s \le 1$, the Lyapunov coefficient $\widetilde{L}_s$ corresponding to $\widetilde{X}$ satisfies

$$\widetilde{L}_s \,\le\, L_s + \tau^s M_r \,\le\, L_s(1 + M_r).$$

Therefore, the requirement (8.2) is met for $\widetilde{X}$, that is, for f_{n+1} and $\widetilde{L}_s$, as long as

$$\frac{T_r}{\tau} \,\le\, (L_s(1 + M_r))^{-\frac{1}{s-2}}$$

which is the same as $c \ge c_0 \equiv T_r(1 + M_r)^{\frac{1}{s-2}}$. Choosing $c = c_0$ and assuming that $L_s \le c_0^{-(s-2)}$ (in order to guarantee that $\tau \le 1$), the inequality (8.1) will thus hold for the distribution μ_{n+1} of S_{n+1}, i.e., we have

$$W_{s-2}(\mu_{n+1}, \nu_m) \le C_s L_s^{\frac{1}{s-2}} \tag{8.3}$$

with some s-dependent constants C_s.

On the other hand, by the definition (1.1) of the power transport distance,

$$\begin{aligned} W_p\,(\mu_{n+1}, \mu_n) &\le \big(\mathbb{E}\,|S_{n+1} - S_n|^p\big)^{1/p} \\ &\le \Big(1 - \sqrt{1-\tau^2}\Big)\,(\mathbb{E}\,|S_n|^p)^{1/p} + \tau M_r^{1/p} \,\le\, \tau\,\Big[(\mathbb{E}\,|S_n|^p)^{1/p} + M_r^{1/p}\Big]. \end{aligned}$$

Here, according to Rosenthal's inequality, $\mathbb{E}\,|S_n|^p \le B_p$ with some constants B_p depending on p, only. Applying the results of [22], this inequality may be shown to hold, for example, with $B_p = (2p)^p$, cf. also [21,28]. Hence, with $p = s-2$ we get

$$W_{s-2}(\mu_{n+1}, \mu_n) \le C'_s L_s^{\frac{1}{s-2}}.$$

It remains to combine this inequality with (8.3), and then we arrive at (8.1), by applying the triangle inequality for the distance W_{s-2}. At this point, the condition $L_s \le c_0^{-(s-2)}$ may easily be removed.

Corollary 8.2 *Let $L_s < \infty$ for $s \ge 3$. For the distribution μ_n of S_n we have*

$$W_{s-2}(\mu_n, \nu_m) \,\le\, C_s L_s^{\frac{1}{s-2}} \tag{8.4}$$

with some constants C_s continuously depending on s.

9 Edgeworth expansion (the i.i.d. case)

If the random variables $X_k = \xi_k/\sqrt{n}$ are identically distributed, with $\mathbb{E}\xi_1 = 0$ and $\mathbb{E}\xi_1^2 = 1$, then $\gamma_{j+2} = n^{-j/2}\,\gamma_{j+2}(\xi_1)$ in the sum (7.1), and the sum itself may be viewed as a polynomial in $1/\sqrt{n}$ of degree at most $m-2$, namely

$$\Phi_m(x) = \Phi(x) - \varphi(x) \sum_{j=1}^{m-2} n^{-\frac{j}{2}}\, Q_j(x). \tag{9.1}$$

Here

$$Q_j(x) = \sum \frac{1}{k_1!\cdots k_{m-2}!} \left(\frac{\gamma_3(\xi_1)}{3!}\right)^{k_1} \cdots \left(\frac{\gamma_m(\xi_1)}{m!}\right)^{k_{m-2}} H_{k-1}(x)$$

with summation over all integers $k_1,\dots,k_{m-2} \geq 0$ such that $k_1 + 2k_2 + \cdots + (m-2)k_{m-2} = j$, and where $k = 3k_1 + \cdots + mk_{m-2}$. Following Esseen [18], formula (9.1) defines the Edgeworth expansion for the distribution function $F_n(x) = \mathbb{P}\{S_n \leq x\}$.

In particular, $\Phi_2(x) = \Phi(x)$ and

$$\Phi_3(x) = \Phi(x) - \frac{\gamma_3}{6\sqrt{n}}\,(x^2-1)\,\varphi(x), \quad \gamma_3 = \mathbb{E}\xi_1^3.$$

More generally, if $\gamma_3(\xi_1) = \cdots = \gamma_{l-1}(\xi_1) = 0$ $(3 \leq l \leq m)$, that is, if the first $l-1$ moments of ξ_1 coincide with those of a standard normal random variable Z, then the first $l-3$ terms in the sum (9.1) are vanishing, $\gamma_l = \mathbb{E}\xi_1^l - \mathbb{E}Z^l$, and (9.1) is simplified to

$$\Phi_m(x) = \Phi(x) - \frac{\gamma_l}{l!}\, H_{l-1}(x)\,\varphi(x)\, n^{-\frac{l-2}{2}} - \varphi(x) \sum_{r=l-2}^{m-2} n^{-\frac{r}{2}}\, Q_r(x). \tag{9.2}$$

Although Proposition 7.2 may be applied in the i.i.d. case, recall that it contains the assumption on the support of the characteristic function $f_n(t)$. Actually, this assumption may be weakened to the requirement that $f_n(t)$ is sufficiently small on larger intervals in comparison with $|t| \leq L_s^{-1/(s-2)}$. For the i.i.d. random variables as above, this is fulfilled under the Cramér condition (1.6), in which case much more is known. When $s = m+1$ is integer, Cramér proved that $F_n(x) - \Phi_{s-1}(x) = O(n^{-(s-2)/2})$ uniformly over all x, while adding another term to the Edgeworth expansion, Esseen strengthened this result to

$$\sup_x |F_n(x) - \Phi_s(x)| = o(n^{-(s-2)/2}),$$

cf. [16,17] and [18], Theorem 1, p. 49. The following important refinement, with extension to not necessarily integer values of s, is due to Osipov [25,26], see also [27], Theorem 2, p. 168, and a more general Theorem 1, p. 159.

Proposition 9.1 *Assume that* (1.6) *is fulfilled, and let* $\mathbb{E}\,|\xi_1|^s < \infty$ $(s \geq 3)$. *Then*

$$\big|F_n(x) - \Phi_{[s]}(x)\big| \leq \frac{\varepsilon_n}{(1+|x|)^s}\, n^{-(s-2)/2} \tag{9.3}$$

uniformly over all $x \in \mathbb{R}$ *with some sequence* $\varepsilon_n \to 0$ *as* $n \to \infty$.

The critical case $s = 3$ is rather special in obtaining of uniform and non-uniform bounds. For example, Esseen showed that the relation $F_n(x) - \Phi_3(x) = o(n^{-1/2})$ remains to hold under a weaker restriction that ξ_1 has a non-lattice distribution. Removing any restriction and replacing Φ_3 with the normal distribution function, there is also a non-uniform bound

$$|F_n(x) - \Phi(x)| \leq \frac{C\beta_3}{(1+|x|)^3\sqrt{n}}.$$

This result was obtained by Nagaev [23], and later Bikjalis [9] extended it to the range $2 < s \leq 3$ in terms of L_s; see also [24] for the history of the problem for this range.

In a full analogy with Proposition 8.1, using the bound (9.3) and applying Corollary 3.2, we obtain a refinement of the inequality (8.4).

Corollary 9.2 *Let* $\mathbb{E}\,|\xi_1|^s < \infty$ *for* $s \geq 3$ *and let* $1 \leq p \leq s-2$. *Under the Cramér condition* (1.6), *for the signed measure* ν_m *with distribution function* Φ_m, $m = [s]$, *we have*

$$W_p(\mu_n, \nu_m) = o\left(n^{-(s-2)/2p}\right). \tag{9.4}$$

10 Proof of Theorems 1.1 and 1.3

Due to the triangle inequality, the Edgeworth correction ν_m in Corollaries 8.2 and 9.2 may be replaced with the standard Gaussian measure γ at the expense of an additional term $W_p(\nu_m, \gamma)$. Hence, the final step in the proof Theorem 1.1 should be provided by the corresponding bound on this distance in case $p = s-2$.

Lemma 10.1 *If* $L_s \leq 1$ *for* $s \geq 3$, *then*

$$W_{s-2}(\nu_m, \gamma) \leq C_s L_s^{\frac{1}{s-2}}, \tag{10.1}$$

where one may take $C_s = (Cs)^{12s}$ *with some absolute constant* C.

Proof If a random variable X has mean zero, its cumulants admit a simple bound in terms of the absolute moments, namely

$$|\gamma_r(X)| \leq (r-1)!\,\mathbb{E}\,|X|^r$$

(Bikjalis [8], cf. also [12]). Hence, a similar relation also holds for the cumulants $\gamma_r = \gamma_r(S_n)$ of the sum S_n in terms of the Lyapunov coefficients L_r. Moreover, since the function $L_r^{1/(r-2)}$ is non-decreasing in r (in view of $\mathbb{E}S_n^2 = 1$), we have

$$|\gamma_r| \le (r-1)!\, L_r \le (r-1)!\, L_s^{\frac{r-2}{s-2}}, \quad 3 \le r \le [s].$$

Hence, for any tuple $(k_1, \dots, k_{m-2})$ participating in (7.1),

$$\left| \left(\frac{\gamma_3}{3!}\right)^{k_1} \cdots \left(\frac{\gamma_m}{m!}\right)^{k_{m-2}} \right| \le \frac{1}{3^{k_1} \dots m^{k_{m-2}}}\, L_s^{\frac{l}{s-2}},$$

where $l = k_1 + 2k_2 + \dots + (m-2)k_{m-2}$. Necessarily, $l \ge 1$, so, $L_s^{\frac{l}{s-2}} \le L_s^{\frac{1}{s-2}}$. In addition,

$$\sum \frac{1}{k_1! \cdots k_{m-2}!}\, \frac{1}{3^{k_1} \dots m^{k_{m-2}}} < e^{1/3} \cdots e^{1/m} < m.$$

Using these bounds together with a simple inequality $|H_k(x)| \le k!\,(1 + |x|^k)$ for the Chebyshev–Hermite polynomials, which is needed with $k \le 3(m-2)+1$, we find from (7.1) that

$$|\Phi_m(x) - \Phi(x)| \le m\,(3m-5)!\, L_s^{\frac{1}{s-2}}\,(1 + |x|^{3m-5})\,\varphi(x).$$

One can now apply Proposition 5.1 with $\nu = \nu_m$, $G = \Phi_m$, $p = s-2$, $d = 3m-5$ and $\varepsilon = m\,(3m-5)!\, L_s^{1/(s-2)}$. It then yields the desired conclusion with constant

$$C_s = m\,(3m-5)!\,(C(s-2)d)^{3(d+1)/2} < (C's)^{12s}.$$

□

Proof of Theorem 1.1. Combining (10.1) with (8.4), we arrive at the desired conclusion

$$W_{s-2}(\mu_n, \gamma) \le C_s L_s^{\frac{1}{s-2}},$$

assuming that $L_s \le 1$. But, in the case $L_s \ge 1$, this inequality also holds, by taking into account the general relation relying on Rosenthal's inequality: For all $p \ge 1$,

$$W_p(\mu_n, \gamma) \le (\mathbb{E}\,|S_n|^p)^{1/p} + (\mathbb{E}\,|Z|^p)^{1/p} \le (2p)^{1/p}\, L_{p^*}^{1/p} + (\mathbb{E}\,|Z|^p)^{1/p},$$

where $p^* = \max\{p, 2\}$ and Z is a standard normal random variable. It remains to note that in case $s = p+2$, necessarily $L_{p^*} \le \max\{L_{p+2}^{(p^*-2)/p}, 1\} \le L_{p+2}$. □

Let us now turn to the i.i.d. case $X_k = \xi_k/\sqrt{n}$ and derive the following refinement of Lemma 10.1 for the special situation as in Theorem 1.2. As usual, $\beta_s = \mathbb{E}\,|\xi_1|^s$.

Lemma 10.2 *Let $\beta_s < \infty$ for $s = (l-2)p + 2$ and let the first $l-1$ moments of ξ_1 ($l \geq 3$) coincide with the corresponding moments of $Z \sim N(0,1)$. Then, under the Cramér condition (1.6),*

$$W_p\left(\nu_{[s]}, \gamma\right) = c\, n^{-\frac{l-2}{2}} + O\left(n^{-\frac{l-1}{2}}\right), \tag{10.2}$$

where

$$c = \frac{|\gamma_l|}{l!} \left(\mathbb{E}\,|H_{l-1}(Z)|^p\right)^{1/p}. \tag{10.3}$$

Proof We now involve Proposition 6.1 with the (generalized) distribution function $G(x) = \Phi_{[s]}(x) = \Phi(x) + \varepsilon(x)\varphi(x)$, where, according to (9.2),

$$\varepsilon(x) = -\frac{\gamma_l}{l!}\, H_{l-1}(x)\, n^{-\frac{l-2}{2}} - \sum_{r=l-2}^{[s]-2} Q_r(x)\, n^{-\frac{r}{2}}. \tag{10.4}$$

Since the polynomials Q_r have degree at most $3(s-2)$, we have

$$\max\{|\varepsilon(x)|, |\varepsilon'(x)|\} \leq C\,(1+|x|)^{3(s-2)}\, n^{-\frac{l-2}{2}} \tag{10.5}$$

with some constant C which does not depend on n and x. Hence, condition (6.2) on the behavior of $\varepsilon(x)$ will be fulfilled on all intervals $[-T, T]$ with $T \geq 1$ such that

$$2C\,(2T)^{3(s-2)+1} \leq n^{\frac{l-2}{2}}.$$

In particular, we may choose $T = T_n = n^{\beta}$ with a sufficiently small $\beta > 0$.

From (10.5) we also obtain that

$$\int_{|x|>T_n} |\varepsilon(x)|^{1/p}\, \varphi(x)^{1/p}\, \mathrm{d}x = o(n^{-q})$$

for any $q > 0$. Hence, we get the representation $W_p\left(\nu_{[s]}, \gamma\right) = A_p + o(n^{-q})$, in which

$$\begin{aligned}|A_p^p - I_p^p| \leq{}& \int_{|x|>T_n} |\varepsilon(x)|^p\, \varphi(x)\,\mathrm{d}x + \int_{-\infty}^{\infty} |\varepsilon(x)|^p\, |\varepsilon'(x)|\, \varphi(x)\,\mathrm{d}x \\ &+ \left(p \cdot 2^{p+1} + 1\right) \int_{-\infty}^{\infty} |\varepsilon(x)|^{p+1}\,(1+|x|)\,\varphi(x)\,\mathrm{d}x + 2\,(2T_n)^{p-1}\,\varphi(T_n),\end{aligned}$$

where

$$I_p^p = \int_{-\infty}^{\infty} |\varepsilon(x)|^p\, \varphi(x)\,\mathrm{d}x = \mathbb{E}\,|\varepsilon(Z)|^p.$$

By (10.5), this gives $A_p^p - I_p^p = O\left(n^{-\frac{l-2}{2}(p+1)}\right)$. In addition, it follows from (10.4) that

$$\mathbb{E}\,|\varepsilon(Z)|^p = c^p\, n^{-\frac{(l-2)p}{2}} + O\left(n^{-\frac{(l-2)p+1}{2}}\right)$$

with constant c described in (10.3). Therefore, the same expansion is also true for A_p, so,

$$W_p^p\left(\nu_{[s]}, \gamma\right) = A_p^p + o(n^{-q}) = c^p\, n^{-\frac{(l-2)p}{2}} + O\left(n^{-\frac{(l-2)p+1}{2}}\right).$$

Raising this equality to the power $1/p$, we arrive at (10.2). □

Proof of Theorem 1.3 Again, let $s = (l-2)p + 2$. As in the proof of Theorem 1.1 (final step), one can now combine (10.2) with inequality (9.4) of Corollary 9.2. By the triangle inequality, with some $|\theta| \le 1$ this gives

$$\begin{aligned} W_p(\mu_n, \gamma) &= W_p\left(\nu_{[s]}, \gamma\right) + \theta\, W_p\left(\mu_n, \nu_{[s]}\right) \\ &= c\, n^{-\frac{l-2}{2}} + O\left(n^{-\frac{l-1}{2}}\right) + o\left(n^{-\frac{s-2}{2p}}\right) = c\, n^{-\frac{l-2}{2}} + o\left(n^{-\frac{l-2}{2}}\right). \end{aligned}$$

□

11 Coupling

The assertion in Corollary 1.3 relies upon the following general observation on the transport distances between probability measures on $\mathbb{R}^n$ for a special cost function

$$c(x, y) = \left|\sum_{k=1}^n x_k - \sum_{k=1}^n y_k\right|^p, \quad x = (x_1, \dots, x_n),\ y = (y_1, \dots, y_n) \in \mathbb{R}^n. \tag{11.1}$$

Lemma 11.1 *Let $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_n)$ be random vectors in $\mathbb{R}^n$ with finite absolute moments of order $p \ge 1$. Then, the W_p-distance between the distributions μ and ν of the sums $X_1 + \dots + X_n$ and $Y_1 + \dots + Y_n$ admits the representation*

$$W_p(\mu, \nu) = \inf_\pi \left(\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} c(x, y)\, d\pi(x, y)\right)^{1/p}, \tag{11.2}$$

where the infimum runs over all Borel probability measures π on $\mathbb{R}^n \times \mathbb{R}^n$ whose marginals are equal to the distributions of X and Y, respectively.

Proof The argument is based on the dual description of transport distances, which we apply with the cost function $c(x, y)$ as in (11.1). Let $K_n(P, Q)$ denote the p-th power of the right-hand side of (11.2) for the distributions P and Q of X and Y. Then, we have

$$K_n(P,Q) = J_n(P,Q), \tag{11.3}$$

where

$$J_n(P,Q) \;=\; \sup\left[\int_{\mathbb{R}^n} u(x)\,dP(x) + \int_{\mathbb{R}^n} v(y)\,dQ(y)\right] \tag{11.4}$$

with supremum running over all Borel measurable functions u, v on $\mathbb{R}^n$ such that

$$u(x) + v(y) \le c(x,y), \quad x, y \in \mathbb{R}^n. \tag{11.5}$$

The latter condition together with the moment assumption ensures that the integrals in (11.4) exist in the Lebesgue sense and may not take the value $+\infty$, so that $J_n(P,Q)$ is well-defined.

The identity (11.3) is rather universal; as was shown in [5], it holds in the setting of an arbitrary complete separable metric space and for an arbitrary cost function $c \ge 0$ integrable with respect to the product measure $P \otimes Q$ (cf. also [14], p. 24). Moreover, the infimum in (11.2) is always attained at some π (called an optimal transference plan, cf. [14], p. 19).

Now, restricting the sup in (11.4) to the functions of the form $u = u(x_1 + \cdots + x_n)$ and $v = v(y_1 + \cdots + y_n)$, the constraint (11.5) is simplified to $u(a) + v(b) \le |a-b|^p$ $(a, b \in \mathbb{R})$. Hence, by the one-dimensional variant of (11.3), the restricted supremum is equal to $W_p^p(\mu,\nu)$, and thus $J_n(P,Q) \ge W_p^p(\mu,\nu)$.

For an opposite direction, consider the partition of the n-space into the hyperplanes $H(a) = \{x \in \mathbb{R}^n : x_1 + \cdots + x_n = a\}$, $a \in \mathbb{R}$. There exist a family of conditional probability measures $(P_a)_{a \in \mathbb{R}}$ and $(Q_a)_{a \in \mathbb{R}}$ for P and Q, that are supported on H_a and satisfy

$$\int_{\mathbb{R}^n} u(x)\,dP(x) = \int_{-\infty}^{\infty}\left[\int_{H(a)} u\,dP_a\right] \mathrm{d}\mu(a), \tag{11.6}$$

$$\int_{\mathbb{R}^n} v(y)\,dQ(y) = \int_{-\infty}^{\infty}\left[\int_{H(b)} v\,dQ_b\right] \mathrm{d}\nu(b) \tag{11.7}$$

(cf. [13,32] for a general theory). According to (11.5), $u(x) + v(y) \le |a-b|^p$ whenever $x \in H(a)$ and $y \in H(b)$. Hence, there is an apriori weaker property $\tilde{u}(a) + \tilde{v}(b) \le |a-b|^p$ for the functions

$$\tilde{u}(a) = \int_{H(a)} u\,dP_a, \quad \tilde{v}(b) = \int_{H(b)} u\,dQ_b.$$

Again, by the one-dimensional variant of (11.3),

$$\int_{-\infty}^{\infty} \tilde{u}\,\mathrm{d}\mu + \int_{-\infty}^{\infty} \tilde{v}\,\mathrm{d}\mu \;\le\; J_1(\mu,\nu) \;=\; K_1(\mu,\nu) \;=\; W_p^p(\mu,\nu),$$

which, by (11.6)–(11.7), yields the desired bound

$$\int_{\mathbb{R}^n} u(x)\, dP(x) + \int_{\mathbb{R}^n} v(y)\, dQ(y) \;\le\; W_p^p(\mu, \nu).$$

□

Lemma 11.1 allows us to reformulate Theorem 1.1 as a statement on the closeness of product measures to the standard Gaussian measure γ_n on $\mathbb{R}^n$ in terms of transference plans on $\mathbb{R}^n \times \mathbb{R}^n$. Let P be a product probability measure on $\mathbb{R}^n$ whose k-th marginals have mean zero and variances σ_k^2 such that $\sigma_1^2 + \cdots + \sigma_n^2 = 1$. The corresponding Lyapunov coefficients may be written as

$$L_s = \int_{\mathbb{R}^n} \sum_{k=1}^n |x_k|^s \, dP(x) \quad (s \ge 2).$$

Theorem 11.2 *If L_{p+2} is finite for $p \ge 1$, then for some probability measure π on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals P and γ_n,*

$$\left[\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \left|\sum_{k=1}^n x_k - \sum_{k=1}^n y_k\right|^p \, d\pi(x,y)\right]^{1/p} \le c_p\, L_{p+2}^{1/p}$$

with some constants c_p continuously depending on p.

In the i.i.d. case we arrive at the first assertion of Corollary 1.3, while the second one is similar.

Acknowledgements The author would like to thank Emmanuel Rio for pointing to the preprint by T. Bonis, and the referees for valuable comments and additional references.

References

1. Agnew, R.P.: Global versions of the central limit theorem. Proc. Natl. Acad. Sci. USA **40**, 800–804 (1954)
2. Agnew, R.P.: Estimates for global central limit theorems. Ann. Math. Stat. **28**, 26–42 (1957)
3. Agnew, R.P.: Asymptotic expansions in global central limit theorems. Ann. Math. Stat. **30**, 721–737 (1959)
4. Bártfai, P.: Über die Entfernung der Irrfahrtswege. (German). Stud. Sci. Math. Hungar **5**, 41–49 (1970)
5. Beiglböck, M., Schachermayer, W.: Duality for Borel measurable cost functions. Trans. Am. Math. Soc. **363**(8), 4203–4224 (2011)
6. Bhattacharya, R.N., Ranga Rao, R.: Normal Approximation and Asymptotic Expansions. Wiley. 1976. Also: Soc. for Industrial and Appl. Math., Philadelphia (2010)
7. Bikjalis, A.: An asymptotic expansion for products of multidimensional characteristic functions. (Russian) Teor. Verojatnost. i Primenen **14**, 508–511 (1969)
8. Bikjalis, A.: Remainder terms in asymptotic expansions for characteristic functions and their derivatives. (Russian) Litovsk. Mat. Sb. **7**, 571–582 (1967). Selected Transl. in Math. Statistics and Probability **11**, 149–162 (1973)
9. Bikjalis, A.: The accuracy of the approximation of the distributions of sums of independent identically distributed random variables by the normal distribution. (Russian) Litovsk Mat. Sb. **11**, 237–240 (1971)

10. Bobkov, S.G.: Entropic approach to E. Rio's central limit theorem for W_2 transport distance. Stat. Probab. Lett. **83**(7), 1644–1648 (2013)
11. Bobkov, S.G.: Closeness of probability distributions in terms of Fourier–Stieltjes transforms (English translation: Russian Math. Surveys). (Russian) Uspekhi Matem. Nauk **71**(6), 37–98 (2016)
12. Bobkov, S.G.: Asymptotic expansions for products of characteristic functions under moment assumptions of non-integer order. Preprint (2016). To appear in: The IMA Volumes in Mathematics and its Applications. Concentration, Convexity and Discrete Structures
13. Bogachev, V.I.: Measure theory. Vol. I, II. Springer, Berlin (2007). Vol. I: xviii+500 pp., Vol. II: xiv+575 pp
14. Bogachev, V.I., Kolesnikov, A.V.: The Monge–Kantorovich problem: achievements, connections, and prospects. (Russian) Uspekhi Mat. Nauk **67**(5), 3–110 (2012); translation in: Russian Math. Surveys **67**(5), 785–890 (2012)
15. Bonis, T.: Stable measures and Stein's method: rates in the sentral limit theorem and diffusion approximation. Preprint (2015). arXiv:1506.06966
16. Cramér, H.: On the composition of elementary errors. Scand. Actuarial J. **11**, 13–74 (1928)
17. Cramér, H.: Random variables and probability distributions. Cambridge Tracts, no. 36 (1937)
18. Esseen, C.-G.: Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. Acta Math. **77**, 1–125 (1945)
19. Esseen, C.-G.: On mean central limit theorems. Kungl. Tekn. Högsk. Handl. Stockholm, 121 (1958)
20. Fréchet, M.: Sur la distance de deux lois de probabilité. (French) Publ. Inst. Statist. Univ. Paris 6, 183–198 (1957)
21. Ibragimov, R., Sharakhmetov, S.H.: On an exact constant for the Rosenthal inequality. (Russian) Teor. Veroyatnost. i Primenen. **42**(2), 341–350 (1997); translation in Theory Probab. Appl. **42**(2), 294–302 (1997)
22. Johnson, W.B., Schechtman, G., Zinn, J.: Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. Ann. Probab. **13**(1), 234–253 (1985)
23. Nagaev, S.V.: Some limit theorems for large deviations. Theory Probab. Appl. **10**(1), 214–235 (1965)
24. Nefedova, Y.S., Shevtsova, I.G.: On nonuniform convergence rate estimates in the central limit theorem. Theory Probab. Appl. **57**(1), 28–59 (2013)
25. Osipov, L.V.: Asymptotic expansions in the central limit theorem. (Russian) Vestnik Leningrad. Univ. **19**, 45–62 (1967)
26. Osipov, L.V.: On asymptotic expansions of the distribution function of a sum of random variables with non uniform estimates for the remainder term. (Russian) Vestnik Leningrad. Univ. **1**, 51–59 (1972)
27. Petrov, V.V.: Sums of independent random variables. Translated from the Russian by A. A. Brown. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82. Springer, New York-Heidelberg, 1975. x+346 pp. Russian ed.: Moscow, Nauka (1972)
28. Pinelis, I.F., Utev, S.A.: Estimates of moments of sums of independent random variables. (Russian) Teor. Veroyatnost. i Primenen. **29**(3), 554–557 (1984)
29. Rio, E.: Distances minimales et distances idéales. C. R. Acad. Sci. Paris **326**, 1127–1130 (1998)
30. Rio, E.: Upper bounds for minimal distances in the central limit theorem. Ann. Inst. Henri Poincaré Probab. Stat. **45**(3), 802–817 (2009)
31. Rio, E.: Asymptotic constants for minimal distance in the central limit theorem. Electron. Commun. Probab. **16**, 96–103 (2011)
32. Rokhlin, V.A.: On the fundamental ideas of measure theory. (Russian) Mat. Sbornik N.S **25**(67), 107–150 (1949)
33. Sakhanenko, A.I.: Estimates in an invariance principle. (Russian) Limit theorems of probability theory, 27–44, 175, Trudy Inst. Mat., 5, "Nauka" Sibirsk. Otdel., Novosibirsk (1985)
34. Statulevičius, V.A.: Limit theorems for densities and the asymptotic expansions for distributions of sums of independent random variables. (Russian) Teor. Verojatnost. i Primenen. **10**, 645–659 (1965)
35. van der Vaart, A.W., Wellner, J.A.: Weak convergence and empirical processes. With applications to statistics. Springer Series in Statistics. Springer, New York (1996)