ASYMPTOTICALLY OPTIMAL WASSERSTEIN COUPLINGS FOR THE SMALL-TIME STABLE DOMAIN OF ATTRACTION

JORGE GONZÁLEZ CÁZARES*, DAVID KRAMER-BANG† & ALEKSANDAR MIJATOVIĆ‡

Abstract. We develop two novel couplings between general pure-jump Lévy processes in $\mathbb{R}^d$ and apply them to obtain upper bounds on the rate of convergence in an appropriate Wasserstein distance on the path space for a wide class of Lévy processes attracted to a multidimensional stable process in the small-time regime. We also establish general lower bounds based on certain universal properties of slowly varying functions and the relationship between the Wasserstein and Toscani–Fourier distances of the marginals. Our upper and lower bounds typically have matching rates. In particular, the rate of convergence is polynomial for the domain of normal attraction and slower than a slowly varying function for the domain of non-normal attraction.

1. Introduction

Stable processes arise naturally as universal scaling limits of a vast class of stochastic processes at either small or large times. In particular, in the small-time regime, stable processes arise as weak limits of discretisation errors of widely used models in theoretical and applied probability [1, 27, 34]. Most of these models are based on Lévy processes in the small-time domain of attraction of a stable processes [7, 12, 26]. In contrast with the more classical long-time regime of Lamperti, where literature is abundant (see, e.g. [9, 24, 25, 28, 33]), the study of the convergence in the small-time regime, which is the focus of this paper, has been underdeveloped. In the long-time regime, the convergence is a consequence of heavy tails with regularly varying tail probabilities or a finite second moment. On the other hand, in the small-time regime, the convergence depends on the activity of the small jumps of the underlying Lévy process and does not depend on the behaviour of the tail probabilites [26]. However, having a heavy-tailed limit may severely deteriorate the convergence speed as uniform integrability typically fails. Quantifying such an error is a fundamental problem, crucial in a number of disparate application areas, such as controlling the bias of discretised models in mathematical finance and elsewhere (see [26] and the references therein), quantifying the model misspecification risk [8] or asserting the convergence properties of estimators for the index of variation, such as Hill's estimator, which is known to require a second order condition for the convergence to have good properties [17, p. 193–195].

The main aim of the present paper is to establish lower and upper bounds in Wasserstein distance on the convergence rate of multivariate Lévy processes attracted to a stable process in the domains of both normal and non-normal attraction (see definition in Section 2 below). Moreover, we will show that our bounds are often sharp. Our upper bounds are applicable to a large class of Lévy processes that are attracted to a multivariate α-stable process (which is Gaussian if $\alpha = 2$ and heavy-tailed

*IIMAS, UNAM, Mexico.

†Department of Mathematics, Aarhus University, Denmark.

‡Department of Statistics, University of Warwick & The Alan Turing Institute, UK.

E-mail address: *jorge.gonzalez@sigma.iimas.unam.mx, †bang@math.au.dk, ‡a.mijatovic@warwick.ac.uk.

2020 *Mathematics Subject Classification.* Primary 60F05, 60G51; Secondary 60G52, 60F25.

Key words and phrases. small-time scaling limit, stable domain of attraction, Wasserstein distance, Coupling of Lévy processes.

if $\alpha \in (0,2)$), while our lower bounds are universal within the small-time regime. To establish the upper bounds on the path supremum norm, we construct two couplings between any two arbitrary Lévy processes, inspired by the stochastic representations in [13], and bound the L^p-norm of the maximum distance between the paths of the resulting processes. The lower bounds for the domain of normal attraction are obtained by comparing the Wasserstein distance with the Toscani–Fourier distance between the marginals and in the case of non-normal attraction, using a universal property of slowly varying functions.

We show that in the domain of normal attraction to heavy-tailed laws, under suitable second order assumptions, the rate of convergence of the upper and lower bounds are polynomial and agree for the L^q-norm in the cases $q < \alpha < 1$ and $q = 1 < \alpha$, making our couplings rate-optimal in this sense. In the domain of non-normal attraction (to either Gaussian or a heavy-tailed stable law), the upper and lower bounds are both 'slow' and, in particular, the convergence is never faster than $\log^{-1-\varepsilon}(1/t)$ as $t \downarrow 0$ for any $\varepsilon > 0$. Moreover, for large subclasses of Lévy processes, the upper and lower bounds on the convergence rate agree in the case $q = 1 < \alpha$ (see e.g. Corollary 2.4). In the domain of normal attraction to the Gaussian law, our upper and lower bounds are also polynomial and dependent on the Blumenthal–Getoor index of the attracted process. The bounds on the convergence rates in this case often agree and, when they do not, the gap between them is small (see Figure 2 below). A short YouTube presentation [21] describes our results, including the ideas behind the proofs.

1.1. **Summary of our results in the heavy-tailed stable domain of attraction.** In preparation for the summary of our results in Table 1, we introduce some notation: $f(t) \lesssim g(t)$ as $t \downarrow 0$ holds for two functions $f, g \geqslant 0$ if there exists $c, t_0 > 0$ satisfying $f(t) \leqslant cg(t)$ for all $t \in (0, t_0]$. An eventually positive function G is slowly varying at infinity, $G \in \mathrm{SV}_\infty$, if $\lim_{x\to\infty} G(cx)/G(x) = 1$ for all $c > 0$.

Table 1 summarises our results on the convergence rates established here for processes in both domains of attraction of stable processes. Recall that an α-stable process has a finite q-moment if and only if $q < \alpha$. Due to this technical constraint, our upper bounds on the L^q-Wasserstein distance, defined in (1) below, always require $q < \alpha$ for the corresponding domains of attraction. We remark that, in this case, both lower and upper bounds are typically asymptotically equivalent up to a multiplicative constant, making our methods and couplings, rate optimal. Indeed, in the domain of normal attraction, this occurs for any admissible $q > 0$ if $\alpha \in (0,1)$ and for the L^1-Wasserstein distance if $\alpha \in (1,2)$. Our bounds for the domain of non-normal attraction are also seen to be rate optimal when G is sufficiently regular (see discussion following Theorem 2.3 and Corollary 2.4 below) when $q = 1$ (and hence $\alpha > 1$).

Domain of attraction	$\alpha \in (0,2)\setminus\{1\}$, $q \in (0,\alpha)\cap(0,1]$ and $t \downarrow 0$
normal	$t^{1-q/\alpha} \lesssim \mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{Z}) \lesssim t^{1-q/\alpha}\mathbb{1}_{\{\alpha<1\}} + t^{q(1-1/\alpha)}\mathbb{1}_{\{\alpha>1\}}$
non-normal	$L(t) \lesssim \max\{\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{Z}), \mathcal{W}_q(\boldsymbol{X}^{2t}, \boldsymbol{Z})\} \lesssim L(t)^q$, where $L(t) \coloneqq t^{-1}\lvert G'(t^{-1})\rvert/G(t^{-1})$ **cannot** be bounded above by any non-decreasing integrable function $\ell > 0$: if $\int_0^1 \ell(t)t^{-1}\mathrm{d}t < \infty$, then $\limsup_{t\downarrow 0} L(t)/\ell(t) = \infty$

TABLE 1. Summary of the results in Theorem 2.1 for the domain of normal attraction and Theorem 2.3 and Corollary 2.4 for the domain of non-normal attraction. Domain of normal attraction requires $\exists \lim_{t\to 0} G(1/t) \in (0,\infty)$, where $g(t) = t^{1/\alpha}G(1/t)$ is the normalising function in the scaling limit $\boldsymbol{X}_t/g(t) \xrightarrow{d} \boldsymbol{Z}_1$ and $\boldsymbol{Z}$ is the stable process of index α. Otherwise, $\boldsymbol{X}$ is in the domain of non-normal attraction. An example of an increasing function ℓ satisfying $\int_0^1 \ell(t)t^{-1}\mathrm{d}t < \infty$ is $\ell(t) = \lvert\log t\rvert^{-1-\varepsilon}$ for any $\varepsilon > 0$.

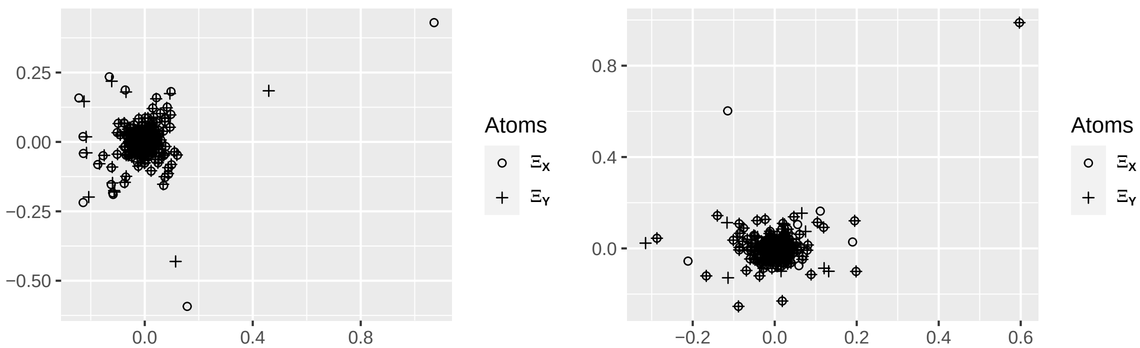
More precisely, in Table 1 we let $\boldsymbol{X} = (\boldsymbol{X}_t)_{t\in[0,1]}$ be a Lévy process in $\mathbb{R}^d$ attracted to an α-stable process $\boldsymbol{Z} = (\boldsymbol{Z}_t)_{t\in[0,1]}$ with normalising function g. That is, $\boldsymbol{X}^t = (\boldsymbol{X}^t_s)_{s\in[0,1]} := (\boldsymbol{X}_{st}/g(t))_{s\in[0,1]}$, $t \in (0,1]$, satisfies $\boldsymbol{X}^t_1 = \boldsymbol{X}_t/g(t) \xrightarrow{d} \boldsymbol{Z}_1$ as $t \downarrow 0$. The table gives asymptotic bounds on the distance $\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{Z})$ as $t \downarrow 0$ in both regimes of attraction. We let the assumptions of either Theorem 2.1 (with $p = 1$, for the domain of normal attraction) or Theorem 2.3 (for the domain of non-normal attraction) hold for $\alpha \in (0,2) \setminus \{1\}$ and pick $q \in (0,\alpha) \cap (0,1]$ satisfying $\mathbb{E}[|\boldsymbol{X}_1|^q] < \infty$. We stress that the lower bounds in Table 1 in both domains of normal and non-normal attraction require no assumptions beyond the existence of the scaling limit (see Theorem 5.1 below for the precise description of the class of Lévy processes $\boldsymbol{X}$ attracted to a stable process $\boldsymbol{Z}$). In particular, as explained in the caption of Table 1, for *any* $\mathbf{X}$ in the domain of non-normal attraction and arbitrary $\epsilon > 0$, there exists a positive increasing sequence $(t_k)_{k\in\mathbb{N}}$ tending to infinity, such that the lower bound on the L^q-Wasserstein distance satisfies $L(t_k) \geqslant |\log t_k|^{-1-\epsilon}$ for all $k \in \mathbb{N}$. In Example 3.2 below, we show that, even if the slowly varying function $G(1/t) = g(t)t^{-1/\alpha}$ in the scaling limit grows arbitrarily slowly, the lower bound L may be asymptotically equivalent to it and bounded below by $1/\log^{1+\varepsilon}(1/t)$ *for all* small $t > 0$.

Recall that the slowly varying function $t \mapsto G(1/t)$ in the scaling limit $\boldsymbol{X}_t/g(t) \xrightarrow{d} \boldsymbol{Z}_1$ (as $t \downarrow 0$) is uniquely determined up to asymptotic equivalence only. Interestingly, our results imply that the rate of convergence in the Wasserstein distance can be affected by different choices of G, see Remark 2.5(IV) below for more details. Finally, we note that the couplings yielding the upper bounds in Table 1 require some structural assumptions on the Lévy measure of $\mathbf{X}$ discussed in Sections 2 and 5 below.

1.2. **A heuristic account of our couplings of Lévy processes.** One of the main purposes of this article is to introduce two couplings between two arbitrary multivariate Lévy processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ and analyse their properties. Both coupling constructions are centered around coupling the respective Poisson jump measures $\Xi_{\boldsymbol{X}}$ and $\Xi_{\boldsymbol{Y}}$. As the Brownian components of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are coupled synchronously under both couplings, the couplings are named after the techniques involved in coupling of the Poisson jump measures $\Xi_{\boldsymbol{X}}$ and $\Xi_{\boldsymbol{Y}}$: the first is the *thinning coupling*, as it is based on Poisson thinning (see Subsection 4.1), and the second is the *comonotonic coupling*, based on the minimal transport coupling of real-valued random variables and LePage's simulation method (see Subsection 4.2). As illustrated in Figure 1, the thinning coupling aims to maximise the intersection of the Poisson jump measures, whereas the comonotonic coupling aims to establish an optimal one-to-one correspondence between the atoms of the Poisson jump measures.

The assumptions and constructions of both couplings are rather different. In the thinning coupling, we consider a common dominating Lévy measure (say, the sum of both Lévy measures) such that the Lévy measures of both processes are absolutely continuous with respect to it with a bounded density. Then, we consider a Poisson measure with mean measure given by the dominating Lévy measure and then thin the Poisson measure appropriately to produce coupled Poisson measures with mean measures given by the Lévy measures of the processes. This maximises the common jumps of both processes.

For the comonotonic coupling, we assume that both processes have a radial decomposition with their own angular measures. We then construct a radial decomposition for both with respect to a common angular measure. We use this measure and LePage's method to construct a one-to-one correspondence of jumps in which both processes jump in the same direction but with different magnitudes. Indeed, both Poisson measures are a transformation of a standard Poisson measure with independent decorations that select the direction of the jump. With our assumption, we construct such a transformation explicitly with the following properties. The decorations of both processes agree. Conditionally given a direction, the jumps of both processes, when ordered by decreasing magnitude, are in a one-to-one correspondence that mimics the relationship of real random variables under the minimal transport (or



(A) Comonotonic coupling

(B) Thinning coupling

FIGURE 1. $\Xi_{\boldsymbol{X}}$ and $\Xi_{\boldsymbol{Y}}$ are Poisson random measures of jumps of the Lévy processes $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Panel (A) depicts atoms of $\Xi_{\boldsymbol{X}}$ and $\Xi_{\boldsymbol{Y}}$ under the comonotonic coupling, where the angular component of each jump of $\boldsymbol{X}$ and $\boldsymbol{Y}$ coincides, while there magnitudes are coupled in the comonotonic fashion. In panel (B), the atoms of $\Xi_{\boldsymbol{X}}$ and $\Xi_{\boldsymbol{Y}}$ from the thinning coupling are sampled: either each jump of $\boldsymbol{X}$ coincides with a jump of $\boldsymbol{Y}$ or the two jumps are sampled independently.

comonotonic) coupling. More precisely, the magnitudes are expressed as right inverse of the radial tail Lévy measure evaluated on the epochs of a standard Poisson process.

1.3. **Comparison with the literature.** Few results identifying the small-time convergence rate exist in the multivariate setting even when the limit is Gaussian. Indeed, most results in this regime are restricted to dimension 1 and often require finite jump activity [10, 11, 14, 15]. In those situations, the limit law of the rescaled error can be identified for some functionals, leading to accurate estimates of the resulting bias and the celebrated continuity corrections [10, 15].

For heavy-tailed stable limits (i.e. non-Gaussian) and infinite activity Lévy processes attracted to the Gaussian law, again in one dimension, the literature is more scarce and only a fraction of the analogous results exist (see [7] for the convergence of certain path statistics to heavy-tailed limits). There are several complications in developing such results for small-time. First, the Berry–Esseen type bounds (see e.g. [24, 37]), commonly used to establish convergence rates to the Gaussian law, fail to give convergent upper bounds since the jump intensity is vanishing in the small-time regime. Second, the rescaled variables often either fail to be uniformly integrable or their uniform integrability is difficult to prove (see, e.g. [7]).

In [16], the authors consider estimating the density of a discretely observed Lévy process satisfying Orey's condition. Under the assumption that sufficiently large jumps are identifiable and removable in the sample, the estimation attains a minimax rate that is optimal up to a logarithmic factor if the Blumenthal–Getoor index is known. This regime is different from our situation, as the authors assume that we may remove all sufficiently large jumps. In fact, under this kind of assumption, the residual small-jump process may not be attracted to a stable process but to a Brownian motion [2].

In [30, 34], the authors introduce couplings between Lévy processes to bound the Wasserstein distance between them. The coupling in [34] is generic and pays special attention to the small jumps. However, the bounds fail to converge to zero when applied to a stable process and a Lévy process in its small-time stable domain of attraction. In contrast to the coupling used in [34], where the authors couple the big-jump components based on the magnitude of the jumps (i.e. based on a common threshold), we couple these components matching their jump intensities. Moreover, in [34] the authors couple the small jumps through an artificial Brownian motion, while we instead couple the compensated Poisson measures directly. On the other hand, the coupling in [30], based on McCann's coupling and Rogers' results on random walks, is the optimal Markovian coupling. However, the coupling requires such tight control of the infinitesimal dynamics of the processes that the coupling could only be constructed for Lévy processes with finitely many jumps on compact intervals, excluding all heavy-tailed stable processes and most processes in the small-time domain of attraction of Brownian motion.

Although the slow convergence phenomenon under the presence of a slowly varying function that does not converge to a positive constant has been observed in some specific settings such as in the case of Hill's estimator (see [17, p. 193–195]), to the best of our knowledge it was first documented rigorously in [9] in an elementary general setting. The authors in [9] lower bound the Prokhorov distance between the marginals of the limit and that of a random walk in its domain of non-normal attraction with a function $b(n)$, satisfying $\limsup_{n\to\infty} b(n)\log^{1+\varepsilon}(n)=\infty$ for any $\varepsilon>0$. However, as is often the case with lower bounds in the form of upper limits, the sparsity of the sequence of times along which the divergence holds remains unclear. The present paper extends the applicability of such a lower bound and strengthens the conclusions. In particular we show that the function analogous to $b(n)$ is typically slowly varying and provide some explicit asymptotically equivalent lower bounds.

1.4. **Organisation of the article.** In Section 2 we introduce the main results of the paper, namely, upper and lower bounds on the convergence rate for processes in the domains of normal and non-normal attraction. Subsection 2.5 explains why the existing literature cannot be directly applied to obtain the bounds presented in Section 2. We present two examples in Section 3 in which our main results are applied to tempered stable processes. The two couplings for general Lévy processes on $\mathbb{R}^d$ used to prove the upper bounds on the Wasserstein distance in Section 2 are introduced in Section 4. General upper bounds (in L^p) for each component in the Lévy–Itô decomposition of coupled Lévy processes are also established in Section 4. The upper bounds for processes in the domain of (normal and non-normal) attraction of a stable process (Gaussian and heavy-tailed) are established in Section 5, while the lower bounds are established in Section 6. The proofs of the results stated in Section 2 are given in Section 7. Section 8 concludes the paper.

2. Main results

The L^q-Wasserstein distance $\mathcal{W}_q(\boldsymbol{\xi},\boldsymbol{\zeta})$, for any $q\in(0,\infty)$, between the laws of $\mathbb{R}^d$-valued random vectors $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ equals $\inf_{\boldsymbol{\xi}'\overset{d}{=}\boldsymbol{\xi},\boldsymbol{\zeta}'\overset{d}{=}\boldsymbol{\zeta}}\mathbb{E}[|\boldsymbol{\xi}'-\boldsymbol{\zeta}'|^q]^{1/(q\vee1)}$ where the infimum is taken over all couplings $(\boldsymbol{\xi}',\boldsymbol{\zeta}')$ with $\boldsymbol{\xi}'\overset{d}{=}\boldsymbol{\xi}$ and $\boldsymbol{\zeta}'\overset{d}{=}\boldsymbol{\zeta}$ (throughout $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^d$ and $x\vee y:=\max\{x,y\}$ for $y,x\in\mathbb{R}$).[1] For $\mathbb{R}^d$-valued stochastic processes $\boldsymbol{\mathcal{X}}=(\boldsymbol{\mathcal{X}}_t)_{t\in[0,1]}$ and $\boldsymbol{\mathcal{Y}}=(\boldsymbol{\mathcal{Y}}_t)_{t\in[0,1]}$, the L^q-Wasserstein distance, based on the distance between the paths in the uniform norm, is given by:

$$\mathcal{W}_q(\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}}):=\inf_{\boldsymbol{\mathcal{X}}'\overset{d}{=}\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}}'\overset{d}{=}\boldsymbol{\mathcal{Y}}}\mathbb{E}\Big[\sup_{t\in[0,1]}|\boldsymbol{\mathcal{X}}'_t-\boldsymbol{\mathcal{Y}}'_t|^q\Big]^{1/(q\vee1)},\qquad q>0,\tag{1}$$

where the infimum is taken over all couplings $(\boldsymbol{\mathcal{X}}',\boldsymbol{\mathcal{Y}}')$ with $\boldsymbol{\mathcal{X}}'\overset{d}{=}\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}'\overset{d}{=}\boldsymbol{\mathcal{Y}}$, where $\boldsymbol{\mathcal{X}}'\overset{d}{=}\boldsymbol{\mathcal{X}}$ means that $\boldsymbol{\mathcal{X}}'$ and $\boldsymbol{\mathcal{X}}$ are equal in law as processes. The case $q\in(0,1)$ is important in our setting because the stable limit does not necessarily possess the first moment.

[1] For $q\in(0,1)$ and any $u>0$, $y\geqslant0$ we have $q(u+y)^{q-1}\leqslant qu^{q-1}$. Hence, by integrating in $u\in(0,x)$, we obtain $(x+y)^q\leqslant x^q+y^q$ for all $x,y\geqslant0$, thus implying that $\mathcal{W}_q$ is a metric.

Let $\boldsymbol{X} = (\boldsymbol{X}_t)_{t\in[0,1]}$ and $\boldsymbol{Z} = (\boldsymbol{Z}_t)_{t\in[0,1]}$ be Lévy processes in $\mathbb{R}^d$ (see [40, Ch. 1, Def. 1.6] for definition), where $\boldsymbol{Z}$ is α-stable (see Section 5.1 below for definition).[2] We say $\boldsymbol{X}$ is in the *small-time domain of attraction* of $\boldsymbol{Z}$ if $(\boldsymbol{X}_{st}/g(t))_{s\in[0,1]} \xrightarrow{d} (\boldsymbol{Z}_s)_{s\in[0,1]}$ as $t \downarrow 0$ in the Skorokhod space for some normalising positive function $g : (0,1] \to (0,\infty)$. Then, it is well known that $\boldsymbol{Z}$ is α-stable for some $\alpha \in (0,2]$ and the normalising function admits the representation $g(t) = t^{1/\alpha}G(t^{-1})$ where G is a slowly varying function at infinity (see [26, Eq. (8)]) that is asymptotically unique (see Theorem 5.1 below for the description of all Lévy processes $\boldsymbol{X}$ attracted to $\boldsymbol{Z}$). We say $\boldsymbol{X}$ is in the *domain of normal attraction* when the slowly varying function $G(x)$ converges to a positive finite constant as $x \to \infty$ (see [20, p. 181]). Otherwise, we say $\boldsymbol{X}$ is in the *domain of non-normal attraction.* Throughout the paper we denote $\boldsymbol{X}^t = (\boldsymbol{X}^t_s)_{s\in[0,1]} := (\boldsymbol{X}_{st}/g(t))_{s\in[0,1]}$ for $t \in (0,1]$.

2.1. **Heavy-tailed stable domain of normal attraction.** For a Lévy process $\boldsymbol{X}$ to be in the domain of attraction of an α-stable process $\boldsymbol{Z}$, the necessary condition in (25) of Theorem 5.1 suggests the Lévy measure of $\boldsymbol{X}$ around the origin should be "asymptotically absolutely continuous" with respect to the α-stable Lévy measure of $\boldsymbol{Z}$ (see also Remark 5.3(b) below). Assumption (T) quantifies the regularity of the corresponding density at the origin $\mathbf{0}$ via the parameter $p > 0$ (the larger p is, the more asymptotic regularity there is). Assumption (T) is required for the upper bound on the rate of convergence in the scaling limit in Theorem 2.1 and is stated in Section 5.2 below. Moreover, Assumption (T), widely satisfied in practice (e.g. in the class of tempered stable processes [39] with $p = 1$; cf. Section 3 below for specific examples), can be seen as quantifying the speed of convergence in the necessary condition (25) for $\boldsymbol{X}$ to be in the stable domain of attraction.

Throughout the paper, for positive functions f_1 and f_2, we use the notation $f_1(x) = \mathcal{O}(f_2(x))$ as $x \downarrow 0$ if $\limsup_{x\downarrow 0} f_1(x)/f_2(x) < \infty$, and $f_1(x) = \mathrm{o}(f_2(x))$ as $x \downarrow 0$ if $\lim_{x\downarrow 0} f_1(x)/f_2(x) = 0$.

Theorem 2.1. *Let $\alpha \in (0,2)\setminus\{1\}$, $\boldsymbol{Z}$ be α-stable and $\boldsymbol{X}$ be in the domain of normal attraction of $\boldsymbol{Z}$.*
(a) *Let Assumption (T) hold for $p = 1$. Then for any $q \in (0,1]\cap(0,\alpha)$ with $\mathbb{E}[|\boldsymbol{X}_1|^q] < \infty$, as $t \downarrow 0$,*

$$\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{Z}) = \begin{cases} \mathcal{O}\big(t^{1-q/\alpha}\big), & \alpha < 1, \\ \mathcal{O}\big(t^{q(1-1/\alpha)}\big), & \alpha > 1. \end{cases}$$

(b) *If $\boldsymbol{X}$ does not have the law of $\boldsymbol{Z}$, then for any $q \in (0,1]\cap(0,\alpha)$ there exists some $C_q > 0$ satisfying*

$$\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{Z}) \geqslant \mathcal{W}_q(\boldsymbol{X}^t_1, \boldsymbol{Z}_1) \geqslant C_q t^{1-q/\alpha}, \quad \textit{for all sufficiently small } t > 0.$$

Remark 2.2. (I) The upper bounds in Theorem 2.1(a) are based on the thinning coupling in Section 4.1 below. The upper bounds are asymptotically proportional to the lower bounds of Theorem 2.1(b) when either $\alpha < 1$ or $q = 1$ when $\alpha > 1$, making the thinning coupling rate-optimal with respect to these Wasserstein distances. Coincidentally, the upper bounds decay the fastest for small values of q when $\alpha < 1$ and for $q = 1$ when $\alpha > 1$. Moreover, the multiplicative constants in $\mathcal{O}$ can be made explicit and depend on the dimension d only through the characteristics of $\boldsymbol{X}$ and $\boldsymbol{Z}$. The lower bounds are based on the lower bound on the Toscani–Fourier distance, see Section 6.2 below for details.
(II) Note that $\mathbb{E}[|\boldsymbol{Z}_1|^q] < \infty$ for $q < \alpha$ and that most models in practice satisfy Assumption (T) with $p = 1$. Theorem 2.1 thus focuses on the case $p = 1$ in order to simplify the exposition, while retaining the key message of the paper. Our technical result Theorem 5.5 in Section 5 (resp. Lemma 6.4 in Subsection 6.2), used to prove part (a) (resp. part (b)) of Theorem 2.1, covers all parameters $p > 0$ and $\mathcal{W}_q$-distances with $q \in (0,1]\cap(0,\alpha)$. The statement of the corresponding general version of Theorem 2.1 is omitted for brevity. ◊

[2]Note that $\boldsymbol{Z}$ need not be isotropic: the angular component of its Lévy measure is not necessarily a uniform probability measure on the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$, see Section 5.1 for details.

2.2. **Stable domain of non-normal attraction.** Consider the case where the slowly varying function G in the scaling limit $(\boldsymbol{X}_{st}/(t^{1/\alpha}G(1/t)))_{s\in[0,1]} \xrightarrow{d} (\boldsymbol{Z}_s)_{s\in[0,1]}$, as $t\downarrow 0$, is not asymptotically equivalent to a positive constant. In this section, we show that the lower bound on the $\mathcal{W}_q$-distance cannot be upper bounded by a positive non-decreasing function ϕ satisfying $\int_0^1 \phi(t)t^{-1}\mathrm{d}t<\infty$. The lower bound requires no assumptions (beyond $\boldsymbol{X}$ being in domain of attraction), while the assumptions for the upper bounds give us multiplicative non-asymptotic control over the distance from $G(x/t)/G(1/t)$ to 1 for small $t>0$ and any $x>0$.

Assumption (S). *There exist $G_1, G_2:[0,\infty)\to[0,\infty)$, such that G_2 is bounded with $G_2(t)\to 0$ as $t\downarrow 0$, G_1 is a slowly varying function both at 0 and at infinity and*

$$|G(x/t)/G(1/t)-1|\leqslant G_1(x)G_2(t),\qquad \textit{for all } x>0 \textit{ and all sufficiently small } t>0.$$

The upper bound in Theorem 2.3(a) below require an additional technical Assumption (C), see Section 5 below. Intuitively, these assumptions require non-parametric structural properties of the Lévy measure of $\boldsymbol{X}$ that allows us to compare it to the Lévy measure of the stable limit $\boldsymbol{Z}$. Indeed, the necessary condition in (25) of Theorem 5.1 suggests the Lévy measure of $\boldsymbol{X}$ around the origin should "asymptotically admit a radial decomposition that is close to that of the stable process". Assumption (C) states precisely this and specifies the proximity of the corresponding radial decomposition to that of the stable process via the parameters $p,\delta>0$ (as before, the larger p and δ are, the closer the radial decompositions are). Moreover, both conditions are widely satisfied with $p=1$, e.g. for the class of tempered α-stable processes (see [39] and, for specific examples, Section 3 below).

Theorem 2.3. *Let $\boldsymbol{X}$ be in the domain of non-normal attraction of an α-stable process $\boldsymbol{Z}$.*
(a) *Let $\alpha\in(0,2)\setminus\{1\}$ and Assumptions (C) and (S) hold for some $p\in(0,\infty)\setminus\{\alpha-1\}$, $\delta>0$ and a function G_2 that is slowly varying at 0. Then $\mathcal{W}_q(\boldsymbol{X}^t,\boldsymbol{Z})=\mathcal{O}(G_2(t)^q)$ as $t\downarrow 0$ for any $q\in(0,1]\cap(0,\alpha)\setminus\{\alpha/(p+1),\alpha/(\alpha\delta+1)\}$ with $\mathbb{E}[|\boldsymbol{X}_1|^q]<\infty$.*
(b) *Let $\alpha\in(0,2]$ and define $a(t):=G(1/(2t))/G(1/t)$ for $t>0$. Then for any $q\in(0,1]\cap(0,\alpha)$,*

$$\max\left\{\mathcal{W}_q(\boldsymbol{X}_1^t,\boldsymbol{Z}_1),\mathcal{W}_q(\boldsymbol{X}_1^{2t},\boldsymbol{Z}_1)\right\}\geqslant|1-a(t)^q|\mathbb{E}\big[|\boldsymbol{Z}_1|^q\big]/3,\quad \textit{for all sufficiently small } t>0.$$

Moreover, $|1-a(t)^q|$ cannot be upper bounded by a non-decreasing function ϕ with $\int_0^1\phi(t)t^{-1}\mathrm{d}t<\infty$.

Since G_2 bounds $|G(1/(2t))/G(1/t)-1|\leqslant G_1(1/2)G_2(t)$ for all small $t>0$ and G is not asymptotically equivalent to a positive finite constant, Lemma 7.2 below (which extends [9, Prop., p. 683]) implies G_2 cannot be upper bounded by any non-decreasing function ϕ satisfying $\int_0^1\phi(t)t^{-1}\mathrm{d}t<\infty$. The assumption on the slow variation of G_2 in Theorem 2.3(a) is not essential and may be replaced by assuming that G_2 dominates any (positive) power at zero. However, by Lemma 7.1 below, in most cases of interest such a function G_2 will be slowly varying.

Given a slowly varying function G, the construction of functions G_1 and G_2 satisfying Assumption (S) is not immediately clear. However, in most cases and for a sufficiently regular G, by virtue of Lemma 7.1, Assumption (S) will be satisfied by choosing $G_2(t)\sim t|G'(1/t)|/G(1/t)$ as $t\downarrow 0$ and a slowly varying G_1 (at 0 and ∞) with $G_1(x)\geqslant|\log x|\cdot\sup_{t>0,y\in[x\wedge 1,x\vee 1]}G'(y/t)/G'(1/t)$ (for $a,b\in\mathbb{R}$, we denote $a\wedge b:=\min\{a,b\}$). In such cases, the lower bound in Theorem 2.3 is (by Lemma 7.1) proportional to G_2, i.e. $G_2(t)\sim|1-a(t)|/\log 2$ as $t\downarrow 0$, making the comonotonic coupling rate optimal with respect to the $\mathcal{W}_1$-distance when $\alpha>1$. The following corollary makes this precise and shows that this is the case for a large class of processes in the domain of non-normal attraction.

Corollary 2.4. *Let $\boldsymbol{X}$ be in the domain of non-normal attraction of an α-stable process $\boldsymbol{Z}$.*
(a) *Let $\alpha\in(0,2)\setminus\{1\}$ and Assumption (C) hold for some $\delta>0$ and $p\neq\alpha-1$. Suppose G is*

C^1 with derivative equal to $\widetilde{G}(t)/(c+t)$, where $c \geqslant 0$ and $|\widetilde{G}| \in SV_\infty$ is eventually positive. Further suppose there exists a slowly varying function $\phi(x)$ both at zero and infinity satisfying $\phi(x) \geqslant \sup_{t>0, y\in[x\wedge 1, x\vee 1]} \widetilde{G}(yt)/\widetilde{G}(t)$ for $x>0$. Define $L(t) := |\widetilde{G}(1/t)|/G(1/t)$, then, for any $q \in (0,1] \cap (0,\alpha) \setminus \{\alpha/2, \alpha/(\alpha\delta+1)\}$ with $\mathbb{E}[|\boldsymbol{X}_1|^q] < \infty$, we have $\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{Z}) = \mathcal{O}(L(t)^q)$ as $t \downarrow 0$ and

$$\max\{\mathcal{W}_q(\boldsymbol{X}_1^t, \boldsymbol{Z}_1), \mathcal{W}_q(\boldsymbol{X}_1^{2t}, \boldsymbol{Z}_1)\} \geqslant \frac{q\log 2}{3}\mathbb{E}\big[|\boldsymbol{Z}_1|^q\big] \cdot L(t), \quad \textit{for all sufficiently small } t>0.$$

(b) *Define iteratively the functions $\ell_1(t) = \log(e+t)$ and $\ell_{n+1}(t) = \log(e+\ell_n(t))$ for $t \geqslant 0$ and $n \in \mathbb{N}$. Suppose G is eventually equal to $\ell_n(t)^{q_n} \cdots \ell_m(t)^{q_m}$ where $1 \leqslant n \leqslant m$ in $\mathbb{N}$ and either $q_n, \ldots, q_m \geqslant 0$ with $q_n, q_m > 0$ or $q_n, \ldots, q_m \leqslant 0$ with $q_n, q_m < 0$. Then G satisfies the assumptions of Part (a).*

Remark 2.5. (I) The upper bound in Theorem 2.3(a) is based on the comonotonic coupling in Section 4.2 below. Since the bound is independent of both p and δ, the restriction in $p \neq \alpha - 1$ and $q \notin \{\alpha/(p+1), \alpha/(\alpha\delta+1)\}$ is nonessential. Indeed, if p and δ satisfy Assumption (C), then any $p' \in (0,p)$ and $\delta' \in (0,\delta)$ also satisfy Assumption (C). Moreover, the multiplicative constants in $\mathcal{O}$ can be made explicit and depend on the dimension d only through the characteristics of $\boldsymbol{X}$ and $\boldsymbol{Z}$.
(II) The lower bounds are based on elementary estimates and a universal property of slowly varying functions, see Section 6.1 below for details. When $\alpha = 2$, despite the fact that we do not have an upper bound in Theorem 2.3(a) for this case, the lower bound of Theorem 2.3(b) ensures the nonexistence of a coupling that makes the $\mathcal{W}_q$-distance decay polynomially.
(III) Corollary 2.4 is a consequence of Theorem 2.3 and Lemmas 7.1 & 7.3 below. Furthermore, we stress that the resulting upper and lower bounds may converge slowly and at a rate that is, in some sense, "bounded away from polynomials" even for a very slow function $G = \ell_n$ or $G = 1/\ell_n$, $n \in \mathbb{N}$, see Example 3.2 below. Furthermore, given any $\ell \in \mathrm{SV}_\infty$ with $\ell(1/t) \to 0$ as $t \downarrow 0$ and $\int_1^\infty \ell(t)t^{-1}\mathrm{d}t = \infty$, the functions $G_\pm(t) := \exp(\pm\int_1^t \ell(s)s^{-1}\mathrm{d}s)$ are slowly varying, $G_+(t) \to \infty$, $G_+(t) \to 0$ and the corresponding $G_2(t)$ functions are proportional $\ell(1/t)$ as $t \downarrow 0$. Thus, we may construct processes $\boldsymbol{X}$ such that $\max\{\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{Z}), \mathcal{W}_q(\boldsymbol{X}^{2t}, \boldsymbol{Z})\}$ is asymptotically bounded above and below by multiples of $\ell(1/t)$ as $t \downarrow 0$, see Lemma 7.2 and Example 3.2 below.
(IV) For a given process $\boldsymbol{X}$, we may choose two asymptotically equivalent slowly varying functions G and $\hat{G}$ that have different convergence properties. Indeed, if G' is not asymptotically equivalent to $\hat{G}'$, then the resulting bounds will change (recall that G is only unique up to asymptotic equivalence and that $\boldsymbol{X}_s^t = \boldsymbol{X}_{st}/(t^{1/\alpha}G(1/t))$). For instance, fix $r_1, r_2 \in (0,1)$, denote $\ell(t) = (\log t)^{r_1}$ and let

$$G'(t) := \ell'(t)(1 + (1-\ell'(t)^{r_2})\cos(\ell(t))), \qquad \hat{G}'(t) := \ell'(t)(1 + (1-\ell'(t)^{r_2})\sin(\ell(t))),$$
$$G(t) = \ell(t) + \sin(\ell(t)) + \mathcal{O}((\log t)^{(r_1-1)(r_2+1)+1}), \quad \hat{G}(t) = \ell(t) - \cos(\ell(t)) + \mathcal{O}((\log t)^{(r_1-1)(r_2+1)+1}).$$

Then $tG'(t)$ and $t\hat{G}'(t)$ are slowly varying and $G(t)/\hat{G}(t) \to 1$ as $t \to \infty$ but $\limsup_{t\to\infty} G'(t)/\hat{G}'(t) = \infty$ and $\liminf_{t\to\infty} G'(t)/\hat{G}'(t) = 0$. Optimising the convergence rate within this class appears to be a very difficult task; however, the limitations imposed by Lemma 7.2 would apply to any choice of G. A similar phenomenon was also observed recently in the standard central limit theorem for Lévy processes in [4], where the Kolmogorov distance is shown to satisfy (resp. fail) an integral condition for a non-standard (resp. standard) scaling.
(V) Theorem 2.3 makes full use of Assumption (S), however, a more detailed analysis that does not require G_2 to be slowly varying can be found in our technical result Theorem 5.9 in Section 5 below.
(VI) We note that a lower bound via the Toscani–Fourier distance is plausible but appears suboptimal since the rate has a polynomial factor. Moreover, we believe the slow lower bound in part (b) to hold for $\mathcal{W}_q(\boldsymbol{X}_1^t, \boldsymbol{Z}_1)$ alone (i.e. without taking the maximum value between times t and $2t$). However, this remains a conjecture. ◊

2.3. **Selecting the coupling.** The main idea behind the proof of Theorems 2.1 & 2.3 is a good coupling between $\boldsymbol{X}$ and $\boldsymbol{Z}$. The two couplings we apply in this article, are the thinning coupling and the comonotonic coupling introduced in Sections 4.1 & 4.2. In Theorem 2.3 we solely apply the comonotonic coupling, since this yields clear and concise results. Note that one could apply the thinning coupling to get a similar result in the domain of non-normal attraction. However, since this would require a lengthy argument, and would distort the main story and result, this has been left out of the paper. In comparison, it is easier to use the comonotonic coupling to give bounds for processes in the domain of normal attraction.

Proposition 2.6. *Let $\boldsymbol{Z}$ be α-stable with $\alpha \in (0,2)$ and $\boldsymbol{X}$ be in its domain of normal attraction. Let Assumption (C) (with constant $H \equiv G$ and $Q \equiv 1$) hold for some $p \in (0,\infty) \setminus \{\alpha - 1\}$. Then, for any $q \in (0,1] \cap (0,\alpha)$ with $\mathbb{E}[|\boldsymbol{X}_1|^q] < \infty$, we have, as $t \downarrow 0$,*

$$\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{Z}) = \begin{cases} \mathcal{O}\big(t^{\min\{pq/\alpha, 1-q/\alpha\}}\big), & \alpha \in (0,1), \\ \mathcal{O}\big(t^{q\min\{p/\alpha, 1-1/\alpha\}}\big), & \alpha \in (1,2). \end{cases}$$

Remark 2.7. Proposition 2.6 follows from Theorem 5.9 (see Remark 5.10). The assumptions in Theorem 2.1(a) and Proposition 2.6 are significantly different, making it is necessary to split the upper bounds in two statements. Indeed, as seen in Example 3.4 below, Assumption (C) is slightly stricter than Assumption (T), since we can show that there exist processes for which Assumption (T) is true, where Assumption (C) is no longer valid. In the case where Assumptions (T) and (C) are valid simultaneously with the same parameter $p = 1$, Theorem 2.1 yields an upper bound that is never worse than that of Proposition 2.6. $\diamondsuit$

2.4. **Gaussian domain of attraction.** The domain of attraction to Brownian motion is substantially different as the previously described couplings are inapplicable. Obtaining a coupling between Brownian motion and other Lévy processes that reduced the L^p-distance in uniform norm has been the work of a large body of literature (which we review in Subsection 2.5 below). In this paper, we use a simple independent coupling, which, heuristically, compares the pure-jump component of $\boldsymbol{X}$ with the null process $\boldsymbol{0}$. Let $\langle\cdot,\cdot\rangle$ denote the Euclidean inner product on $\mathbb{R}^{d\times d}$, $\boldsymbol{0}$ denote the zero-vector in $\mathbb{R}^d$ as well as the zero-matrix in $\mathbb{R}^{d\times d}$ and let $\mathbb{R}^d_{\boldsymbol{0}} := \mathbb{R}^d \setminus \{\boldsymbol{0}\}$. Let $\varphi_{\boldsymbol{X}}(\boldsymbol{u}) := \mathbb{E}[e^{i\langle \boldsymbol{u}, \boldsymbol{X}\rangle}]$ for $\boldsymbol{u} \in \mathbb{R}^d$ denote the characteristic function of $\boldsymbol{X}$. Furthermore, let $\psi_{\boldsymbol{S}}$ be the Lévy-Khintchine exponent of $\boldsymbol{S}$, given by $\psi_{\boldsymbol{S}}(\boldsymbol{u}) := t^{-1}\log\varphi_{\boldsymbol{S}_t}(\boldsymbol{u})$ for $\boldsymbol{u} \in \mathbb{R}^d$ and $t > 0$.

Theorem 2.8. *Let $\boldsymbol{\Sigma}$ be a symmetric non-negative definite matrix on $\mathbb{R}^{d\times d}$ and define the process $\boldsymbol{X}^t = ((\boldsymbol{\Sigma}\boldsymbol{B}_{st} + \boldsymbol{S}_{st})/\sqrt{t})_{s\in[0,1]}$ for $t \in (0,1]$ where $(\boldsymbol{B}_t)_{t\in[0,1]}$ is a standard Brownian motion on $\mathbb{R}^d$ independent of the pure-jump Lévy process $\boldsymbol{S}$ with Blumenthal–Getoor index β (defined in (27)).*
(a) *Suppose $\beta \in [0,2)$ and fix any $\beta_* \in (\beta, 2]$ when $\boldsymbol{S}$ is of infinite variation and $\beta_* = 1$ otherwise. Then for any $q > 0$ with $\mathbb{E}[|\boldsymbol{X}_1|^q] < \infty$, we have*

$$\mathcal{W}_q\big(\boldsymbol{X}^t, \boldsymbol{\Sigma}\boldsymbol{B}\big) = \mathcal{O}\big(t^{(q\wedge 1)(\min\{1/q, 1/\beta_*\} - 1/2)}\big), \qquad \text{as } t \downarrow 0.$$

(b) *Pick any $\boldsymbol{u}_* \in \mathbb{R}^d_{\boldsymbol{0}}$ and define $C_* := |\boldsymbol{u}_*|^{-1}|\psi_{\boldsymbol{S}}(\boldsymbol{u}_*)| > 0$. Then for all $q \geqslant 1$, we have*

$$\mathcal{W}_q(\boldsymbol{X}^t_1, \boldsymbol{\Sigma}\boldsymbol{B}_1) \geqslant \frac{C_*}{\sqrt{2}}\sqrt{t} + \mathcal{O}(t^{3/2}), \quad \text{as } t \downarrow 0.$$

(c) *Let λ be the largest eigenvalue of $\boldsymbol{\Sigma}^2$. Suppose there exist $\delta \in [1,2)$ and vectors $(\boldsymbol{u}_r)_{r\in(0,\infty)}$ with $|\boldsymbol{u}_r| = r$ satisfying $c := \inf_{r>1} r^{-\delta}|\psi_{\boldsymbol{S}}(\boldsymbol{u}_r)| > 0$. Then for any $C_* \in (0, ce^{-\lambda/2})$ we have $\mathcal{W}_q(\boldsymbol{X}^t_1, \boldsymbol{\Sigma}\boldsymbol{B}_1) \geqslant (C_*/\sqrt{2})t^{1-\delta/2}$ for all sufficiently small $t > 0$.*

Parts (a) and (c) of Theorem 2.8, with $q = 1$, imply that for processes whose pure jump part is in the domain of attraction of a β-stable process, the upper and lower bounds are essentially proportional to $t^{1/\max\{1,\beta\}-1/2}$ and $t^{1-\max\{1,\beta\}/2}$, respectively. These agree in the finite variation case with rate $\sqrt{t}$ and also as $\beta \to 2$ with an arbitrarily deteriorating convergence rate. As shown in Figure 2 these bounds are not far from each other for fixed β, and the powers of t from the rates are also not far. In the 'limiting case' where $\boldsymbol{S}$ is itself attracted to a Brownian motion and $\beta = 2$, the rescaled process $(\boldsymbol{S}_{st}/\sqrt{t})_{s\in[0,1]}$ is distributionally close to $\ell(t)\boldsymbol{B}$ (see e.g. [26, Thm 2]) for a slowly varying function ℓ satisfying $\lim_{t\downarrow 0}\ell(t) = 0$. It is thus natural to expect that the convergence, in this case, is slow as in Theorem 2.3 above, see Example 6.7.

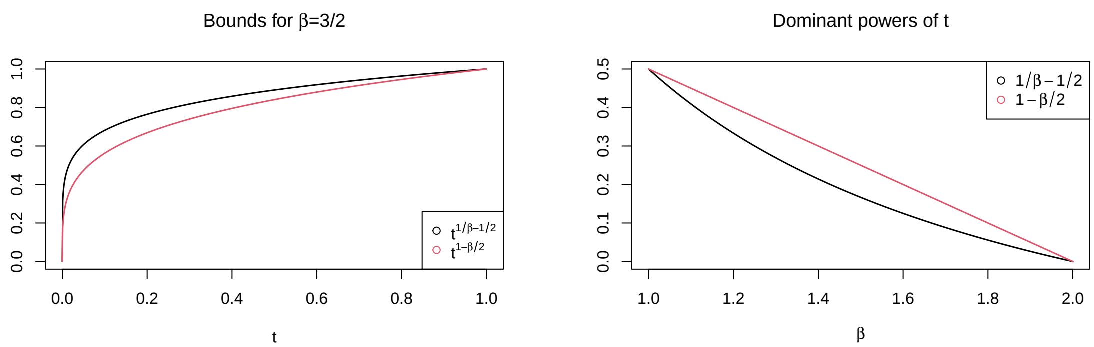


FIGURE 2. In the left picture we see polynomials with the dominant powers from the upper and lower bounds for $\mathcal{W}_1(\boldsymbol{X}_1^t, \boldsymbol{\Sigma}\boldsymbol{B}_1)$ from Theorem 2.8 with $\beta = 3/2$. In the right picture, we see the dominant powers of t in the upper and lower bounds as a function of $\beta \in [1,2]$.

2.5. **Classical bounds are hard to apply!** The couplings and methods used to achieve these bounds are crucial, and they differ significantly from the classical methods used to find rates of convergence. Indeed, if we tried to use standard methods (namely, the Berry–Esseen theorem or [34]) to construct bounds for small-time domain of attraction, the bounds would not converge as $t \downarrow 0$.

The Berry–Essen theorem exploits an increase in the activity of the process to obtain bounds on the distance between a random walk and the limit law. In the small-time regime, the activity is instead decreasing, explaining the unsuitability of this tool in this context (see details below). In fact, the bound would converge to infinity as $t \downarrow 0$ (and in particular does not go to 0) unlike the bounds introduced in this paper. The coupling in [34] couples corresponding components of the Lévy–Itô decompositions for a common small-jump cutoff level. When the time horizon is fixed and the Lévy measure is supported on $[-\varepsilon, \varepsilon]$, the bounds of [34] are asymptotically sharp as $\varepsilon \downarrow 0$. However, for general Lévy measures and as time tends to 0, no time-dependent cutoff level ε_t can be used to obtain convergent bounds. The lack of convergent bounds in the small-time domain of attraction of stable processes is mainly caused by a difference in the jump intensities of the large-jump components (see details below).

We first explain why the Berry–Esseen theorem does not yield suitable bounds even when the limit is Gaussian (see [18, 22]). For the explanation, it is enough to consider the one-dimensional case. Let $(X_t)_{t\in[0,1]}$ be a zero-mean Lévy process on $\mathbb{R}$ with characteristic triplet $(\gamma, \sigma^2, \nu_X)$ (see [40, Def. 8.2]) and finite fourth moment. The variance of X is given by $\mathbb{E}[X_t^2] = (\sigma^2 + \mu_2)t$, where $\mu_2 := \int_{\mathbb{R}\setminus\{0\}} x^2\nu_X(\mathrm{d}x)$. Then, $X_1^t = X_t/\sqrt{(\sigma^2+\mu_2)t}$ is attracted to a standard Gaussian random variable Z as $t \downarrow 0$. Denote by ν_t the Lévy measure of X_1^t, the Berry–Esseen theorem thus implies that there exists some universal constant $C > 0$, such that

$$\mathcal{W}_2(X_1^t, Z) \leqslant C \int_{\mathbb{R}\setminus\{0\}} x^4 \nu_t(\mathrm{d}x) = C \int_{\mathbb{R}\setminus\{0\}} x^4 t \nu_X(\mathrm{d}(\sqrt{(\sigma^2+\mu_2)t}x)) = \frac{C}{t(\sigma^2+\mu_2^2)^2} \int_{\mathbb{R}\setminus\{0\}} x^4 \nu_X(\mathrm{d}x),$$

for all $t \in (0,1]$. As we can see above, this upper bound will tend to ∞ as $t \downarrow 0$ and is therefore not an informative bound in the small-time regime.

For Lévy processes in the domain of attraction of an α-stable law, an application of the bounds in [34] does not yield convergent bounds. The proofs of the bounds in [34] rely on the coupling of small jumps to a Gaussian law. Again, it is enough to consider the one-dimensional case. Let X be symmetric and in the domain of attraction of the symmetric α-stable random variable Z with $\alpha > 1$. Suppose their Lévy measures satisfy $\nu_X(\mathbb{R}\setminus(-x,x)) = x^{-\alpha} + x^{-(\alpha+1)/2}$ and $\nu_Z(\mathbb{R}\setminus(-x,x)) = x^{-\alpha}$ for $x > 0$. In this case we have $X_1^t = X_t/t^{1/\alpha} \xrightarrow{d} Z$ as $t \downarrow 0$.

Let $\eta = (\alpha-1)/2 > 0$ and apply [34, Thm 11] (at time 1 and cutoff ε_t) to obtain:

$$\begin{aligned}\mathcal{W}_1(X_1^t, Z) \leqslant C\varepsilon_t &+ \Big(\big(\tfrac{1}{2-\alpha}\varepsilon_t^{2-\alpha} + \tfrac{1}{3/2-\alpha/2} t^{\eta/\alpha}\varepsilon_t^{3\eta}\big)^{1/2} - \big(\tfrac{1}{2-\alpha}\varepsilon_t^{2-\alpha}\big)^{1/2}\Big)\\ &+ 2\big(\varepsilon_t^{-\alpha} + t^{\eta/\alpha}\varepsilon_t^{-\eta-1}\big)\int_{\varepsilon_t}^\infty \left|\frac{x^{-\alpha}}{\varepsilon_t^{-\alpha}} - \frac{x^{-\alpha}+x^{-(\alpha+1)/2}}{\varepsilon_t^{-\alpha}+t^{\eta/\alpha}\varepsilon_t^{-\eta-1}}\right|\mathrm{d}x\\ &+ 2t^{\eta/\alpha}\varepsilon_t^{-\eta-1}\int_{\varepsilon_t}^\infty x\frac{\alpha x^{-\alpha-1}\mathrm{d}x}{\varepsilon_t^{-\alpha}}, \qquad \text{for all } t>0,\end{aligned}$$

where we used the formula for the L^1-Wasserstein distance in [19, p. 8]. For the first line in the display above to vanish as $t \downarrow 0$, we require $\varepsilon_t = \mathrm{o}(1)$. The term in the middle line of the display above equals

$$\begin{aligned}2\int_{\varepsilon_t}^\infty \left|x^{-\alpha}t^{\eta/\alpha}\varepsilon_t^{-\eta} - x^{-\eta-1}\right|\mathrm{d}x &\geqslant 2\left|t^{\eta/\alpha}\varepsilon_t^{-\eta}\int_{\varepsilon_t}^c x^{-\alpha}\mathrm{d}x - \int_{\varepsilon_t}^c x^{-\eta-1}\mathrm{d}x\right|\\ &= 2\left|\tfrac{1}{\alpha-1}(t^{\eta/\alpha}\varepsilon_t^{-3\eta} - t^{\eta/\alpha}\varepsilon_t^{-\eta}c^{-2\eta}) - \tfrac{2}{\alpha-1}(\varepsilon_t^{-\eta} - c^{-\eta})\right|,\end{aligned}$$

where $c \in (0,\infty]$ is an arbitrary number and the inequality holds for all $t > 0$ for which $\varepsilon_t < c$. For the right-hand side of the display to vanish at $t \downarrow 0$ with $c = \infty$ we must have $\varepsilon_t^{-\eta}(t^{\eta/\alpha}\varepsilon_t^{-2\eta} - 2) = \mathrm{o}(1)$. Then, for $c = 1$, the display above will converge to the constant $4/(\alpha-1)$. In particular, the bound implied by [34, Thm 11] cannot vanish for any choice of ε_t.

3. Examples

In this section, we apply some of the main results from Section 2 on tempered α-stable processes [39, Def. 2.1], that are in the domain of attraction of α-stable processes. We say that a process $(\boldsymbol{X}_t)_{t\in[0,1]}$ is a tempered α-stable process if it has no Gaussian component, and its Lévy measure $\nu_{\boldsymbol{X}}$ has the form

$$\nu_{\boldsymbol{X}}(A) = \int_{\mathbb{S}^{d-1}}\int_0^\infty \mathbb{1}_A(x\boldsymbol{v})\alpha x^{-\alpha-1}q(x,\boldsymbol{v})\mathrm{d}x\sigma(\mathrm{d}\boldsymbol{v}), \qquad \text{for } A \in \mathcal{B}(\mathbb{R}^d_{\mathbf{0}}), \tag{2}$$

where $\mathbb{S}^{d-1}$ is the unit sphere in $\mathbb{R}^d$ and $q(\cdot,\boldsymbol{v}) : (0,\infty)\times\mathbb{S}^{d-1} \mapsto (0,\infty)$ is a completely monotone Borel function (see [39, p. 680]) with $q(\infty,\boldsymbol{v}) = 0$ for all $\boldsymbol{v} \in \mathbb{S}^{d-1}$ and $\mathcal{B}(\mathbb{R}^d_{\mathbf{0}})$ is the Borel σ-algebra on $\mathbb{R}^d\setminus\{0\}$. In Examples 3.1 and 3.2 the process $\boldsymbol{X}$ is a multidimensional tempered α-stable process in the stable domain of attraction. Both of the examples can be easily seen to fulfil Assumption (C) or (T). Example 3.4 shows that Assumption (C) does not imply (T), while Example 3.3 deals with a Gaussian perturbation of a tempered α-stable process.

Example 3.1. Assume that $(\boldsymbol{Z}_t)_{t\in[0,1]}$ is an α-stable process on $\mathbb{R}^d$ and that $(\boldsymbol{X}_t)_{t\in[0,1]}$ is a tempered stable process with Lévy measure as in (2). Assume that $q(x,\boldsymbol{v})=e^{-\lambda(\boldsymbol{v})x}$, for all $x>0$ and $\boldsymbol{v}\in\mathbb{S}^{d-1}$, where $\lambda(\boldsymbol{v})$ is a bounded non-negative function. Thus,

$$|e^{-\lambda(\boldsymbol{v})x}-1|\leqslant(1\wedge|x||\lambda(\boldsymbol{v})|)\leqslant K(1\wedge|x|),\qquad \text{for all } (x,\boldsymbol{v})\in(0,\infty)\times\mathbb{S}^{d-1}.$$

If $\alpha>1$, then Theorem 2.1 with $p=1$ implies that $\mathcal{W}_1(\boldsymbol{X}_t/t^{1/\alpha},\boldsymbol{Z}_1)=\mathcal{O}(t^{1-1/\alpha})$, with lower bound given by $\mathcal{W}_1(\boldsymbol{X}_t/t^{1/\alpha},\boldsymbol{Z}_1)\geqslant Ct^{1-1/\alpha}+\mathcal{O}(t^{2-1/\alpha})$ as $t\downarrow 0$, for some finite constant $C>0$. Thus, the upper and lower bounds have the same rate in this case, yielding a rate-optimal bound. △

Next, we give an example where the function G is non-constant, and see how the rates deteriorate in these cases, as Theorem 2.3 implies. Throughout the paper, we use the notation $f(x)\sim g(x)$ as $x\to a$, if $\lim_{x\to a}f(x)/g(x)=1$.

Example 3.2. Assume that $(\boldsymbol{Z}_t)_{t\in[0,1]}$ is an α-stable process on $\mathbb{R}^d$ with $\alpha\in(1,2)$ and that $\boldsymbol{X}$ is a tempered stable process with Lévy measure as in (2). We assume that $\rho^{\mathsf{c}}_{\boldsymbol{X}}([x,\infty),\boldsymbol{v})=H(x)^\alpha x^{-\alpha}$ for all $x>0$ and $\boldsymbol{v}\in\mathbb{S}^{d-1}$ (see (32)), where H is differentiable and slowly varying, implying $q(x,\boldsymbol{v})\sim H(x)^\alpha$ as $x\downarrow 0$. Then $\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})$ does not depend on $\boldsymbol{v}\in\mathbb{S}^{d-1}$ and its value, denoted $\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x)$, satisfies $\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x)\sim x^{-1/\alpha}H(x^{-1/\alpha})$ as $x\to\infty$ by [6, Cor. 2.3.4]. For any $\boldsymbol{v}\in\mathbb{S}^{d-1}$, by (33), we have

$$G(x)=\int_{\mathbb{S}^{d-1}}H(\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{u}))\sigma(\mathrm{d}\boldsymbol{u})=H(\rho^{\leftarrow}_{\boldsymbol{X}}(x))\sim H(x^{-1/\alpha}),\qquad \text{as } x\to\infty.$$

Theorem 2.3 now yields both the upper and lower bound in terms of G and related functions.

Let ℓ_n be recursively defined as in Lemma 7.3 below: $\ell_1(t)\coloneqq\log(e+t)$ and $\ell_{n+1}(t)=\log(e+\ell_n(t))$ for $t>0$. If either $H(x)=\ell_n(1/x)$ or $H(x)=\ell_n(1/x)^{-1}$ for some $n\in\mathbb{N}$ (i.e. $G(x)\sim\ell_n(x^{1/\alpha})$ or $G(x)\sim\ell_n(x^{1/\alpha})^{-1}$ as $x\to\infty$), then Lemma 7.3 shows that for small $t>0$ we have

$$G_2(t)\coloneqq\prod_{k=1}^n(e+\ell_k(1/t))^{-1}\geqslant|1-G(1/(2t))/G(1/t)|/\log 2.$$

Moreover, by Lemma 7.1, Assumption (S) holds with this G_2. Thus, by Theorem 2.3, there exist constants $0<C_1<C_2$ such that, for all small enough $t>0$, we have

$$\frac{C_1}{\prod_{k=1}^n(e+\ell_k(1/t))}\leqslant\max\{\mathcal{W}_1(\boldsymbol{X}^t_1,\boldsymbol{Z}),\mathcal{W}_1(\boldsymbol{X}^{2t}_1,\boldsymbol{Z})\}\leqslant\frac{C_2}{\prod_{k=1}^n(e+\ell_k(1/t))}.$$

Thus, despite the function ℓ_n being “nearly constant” for large $n\in\mathbb{N}$, the convergence rates of the upper and lower bounds match and are slower than $\log(1/t)^{-1-\varepsilon}$ for any $\varepsilon>0$.

Now consider any $\ell\in\mathrm{SV}_\infty$ with $\ell(t)\downarrow 0$ as $t\to\infty$ and $\int_1^\infty\ell(t)t^{-1}\mathrm{d}t=\infty$. Then $G_\pm(t)\coloneqq\exp(\pm\int_1^t\ell(s)s^{-1}\mathrm{d}s)$ are slowly varying, $G_+(t)\to\infty$, $G_-(t)\to 0$ and $|1-G_\pm(t/2)/G_\pm(t)|\sim\ell(t)\log 2$ as $t\to\infty$ by Lemma 7.1. Thus, by Theorem 2.3, for any $q\in(0,\alpha)\cap(0,1]$, we have $\mathcal{W}_q(\boldsymbol{X}^t,\boldsymbol{Z})\vee\mathcal{W}_q(\boldsymbol{X}^{2t},\boldsymbol{Z})\geqslant C^*\ell(1/t)$ for some $C^*>0$ and all sufficiently small $t>0$. In particular, by taking an appropriate ℓ, e.g., $\ell(t)=1/\ell_n(t)$ where ℓ_n is as in the previous paragraph and $n\in\mathbb{N}$ is large, the convergence in Wasserstein distance may be arbitrarily slow. △

As a last example, we will consider the case of Theorem 2.8, where the pure-jump Lévy process is a tempered α-stable process.

Example 3.3. Let $\boldsymbol{\Sigma}$ be a positive definite matrix on $\mathbb{R}^{d\times d}$ and set $\boldsymbol{X}^t\coloneqq((\boldsymbol{\Sigma}\boldsymbol{B}_{st}+\boldsymbol{S}_{st})/\sqrt{t})_{s\in[0,1]}$ for $t\in(0,1]$ where $(\boldsymbol{B}_t)_{t\in[0,1]}$ is a standard Brownian motion on $\mathbb{R}^d$ independent of the pure-jump tempered α-stable Lévy process $\boldsymbol{S}$. Assume $\alpha\in[1,2)$, that $\boldsymbol{S}$ has zero-mean, and fix any $\beta_*\in(\alpha,2]$. Then, by Theorem 2.8(a), we have the upper bound $\mathcal{W}_1\big(\boldsymbol{X}^t,\boldsymbol{\Sigma}\boldsymbol{B}\big)=\mathcal{O}\big(t^{1/\beta_*-1/2}\big)$ as $t\downarrow 0$. To find the lower bound, we let λ be the largest eigenvalue of $\boldsymbol{\Sigma}^2$, and define $c\coloneqq\inf_{r>1}r^{-\alpha}|\psi_{\boldsymbol{S}}(r\boldsymbol{u})|>0$, for some

$\boldsymbol{u} \in \mathbb{R}^d$ with $|\boldsymbol{u}| = 1$. Then, for any $C_* \in (0, ce^{-\lambda/2})$, Theorem 2.8(c) implies that $\mathcal{W}_1(\boldsymbol{X}_1^t, \boldsymbol{\Sigma B}_1) \geqslant C_* t^{1-\alpha/2}$ for all sufficiently small $t > 0$.

Note that, as α approaches 1, the gap between the lower and upper bound decreases. Indeed, for $\alpha = 1$, we have $\beta_* = 1 + \varepsilon$ for some small $\varepsilon > 0$, so the upper bound is of the rate $t^{1/(1+\varepsilon)-1/2}$, while the lower bound has the rate $\sqrt{t}$, making the quotient of the two bounds proportional to $t^{\varepsilon/(1+\varepsilon)}$. △

Example 3.4. We show in this example, that we can find a process that satisfies Assumption (T) but not Assumption (C). Let $\alpha \in (1, 2)$ and $\alpha' \in (1, \alpha)$. Next, let X be a one-dimensional α-stable process and Y be a α'-stable process that is spectrally negative, with Lévy measures $\nu_X(\mathrm{d}x) = c_1|x|^{-1-\alpha}\mathrm{d}x$ and $\nu_Y(\mathrm{d}x) = c_2 \mathbb{1}_{(-\infty,0)}(x)|x|^{-1-\alpha'}\mathrm{d}x$ for some constants $c_1, c_2 > 0$. We note that $X + Y$ has Lévy measure $\nu_{X+Y}(\mathrm{d}x) = [c_2\mathbb{1}_{(-\infty,0)}(x)|x|^{-1-\alpha'} + c_1|x|^{-1-\alpha}]\mathrm{d}x$, showing that Assumption (T) is indeed fulfilled. We can however note that Assumption (C) cannot be fulfilled, since there doesn't exist the necessary radial decomposition of ν_{X+Y}. △

4. Two couplings of Lévy processes

Let $\boldsymbol{X} = (\boldsymbol{X}_t)_{t\geqslant 0}$ be a Lévy process on $\mathbb{R}^d$ with generating triplet $(\boldsymbol{\gamma_X}, \boldsymbol{\Sigma_X}\boldsymbol{\Sigma_X}^\top, \nu_{\boldsymbol{X}})$ (also called characteristic triplet, see [40, Def. 8.2]) with respect to the cutoff function $\boldsymbol{w} \mapsto \mathbb{1}_{B_{\mathbf{0}}(1)}(\boldsymbol{w})$, where $\boldsymbol{\gamma_X} \in \mathbb{R}^d$, $\boldsymbol{\Sigma_X} \in \mathbb{R}^{d\times d}$ (with transpose $\boldsymbol{\Sigma_X}^\top \in \mathbb{R}^{d\times d}$) and $\boldsymbol{\Sigma_X}\boldsymbol{\Sigma_X}^\top$ a symmetric non-negative definite matrix and $\nu_{\boldsymbol{X}}$ a Lévy measure on $\mathbb{R}^d$. Throughout, we denote by $|\cdot|$ the Euclidean norm of appropriate dimension and let $B_{\mathbf{0}}(r) := \{\boldsymbol{x} \in \mathbb{R}^d : |\boldsymbol{x}| < r\}$ be the open ball in $\mathbb{R}^d$ of radius $r > 0$, centered at the origin $\mathbf{0} \in \mathbb{R}^d$. Fix any $\kappa \in (0, 1]$ and consider the Lévy–Itô decomposition of $\boldsymbol{X}$ given by

$$\boldsymbol{X}_t = \boldsymbol{\gamma}_{\boldsymbol{X},\kappa} t + \boldsymbol{\Sigma_X} \boldsymbol{B}_t^{\boldsymbol{X}} + \boldsymbol{D}_t^{\boldsymbol{X},\kappa} + \boldsymbol{J}_t^{\boldsymbol{X},\kappa}, \quad t \geqslant 0, \tag{3}$$

where $\boldsymbol{\gamma}_{\boldsymbol{X},\kappa} := \boldsymbol{\gamma_X} - \int_{\mathbb{R}^d} \boldsymbol{w}\mathbb{1}_{B_{\mathbf{0}}(1)\setminus B_{\mathbf{0}}(\kappa)}(\boldsymbol{w})\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})$, $\boldsymbol{B}^{\boldsymbol{X}}$ is a standard Brownian motion on $\mathbb{R}^d$, $\boldsymbol{D}^{\boldsymbol{X},\kappa}$ is the small-jump martingale containing all the jumps of $\boldsymbol{X}$ of magnitude less that κ, $\boldsymbol{J}^{\boldsymbol{X},\kappa}$ is the driftless compound Poisson process containing all the jumps of $\boldsymbol{X}$ of magnitude at least κ and all three processes $\boldsymbol{B}^{\boldsymbol{X}}$, $\boldsymbol{D}^{\boldsymbol{X},\kappa}$ and $\boldsymbol{J}^{\boldsymbol{X},\kappa}$ are independent. Moreover, the pure-jump component $\boldsymbol{D}^{\boldsymbol{X},\kappa} + \boldsymbol{J}^{\boldsymbol{X},\kappa}$ of $\boldsymbol{X}$ is a Lévy process with paths of finite variation (i.e. the jumps are summable on any compact time interval) if and only if $\int_{\mathbb{R}^d_{\mathbf{0}}} |\boldsymbol{w}|\mathbb{1}_{B_{\mathbf{0}}(1)}(\boldsymbol{w})\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) < \infty$ [40, Thm 21.9]. In particular, $(\boldsymbol{\gamma_X}, \mathbf{0}, \nu_{\boldsymbol{X}})$ is a characteristic triplet of a Lévy process $\boldsymbol{X}$ without a Gaussian component. Thus, if $\boldsymbol{X}$ has finite variation, then $\boldsymbol{X}$ has *zero natural drift* (i.e. the process equals the sum of its jumps) if and only if $\gamma_{\boldsymbol{X}} = \int_{\mathbb{R}^d_{\mathbf{0}}} \boldsymbol{w}\mathbb{1}_{B_{\mathbf{0}}(1)}(\boldsymbol{w})\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})$.

Similarly, we let $\boldsymbol{Y} = (\boldsymbol{Y}_t)_{t\geqslant 0}$ be a Lévy process on $\mathbb{R}^d$ with characteristic triplet $(\boldsymbol{\gamma_Y}, \boldsymbol{\Sigma_Y}\boldsymbol{\Sigma_Y}^\top, \nu_{\boldsymbol{Y}})$ with respect to the cutoff function $\boldsymbol{w} \mapsto \mathbb{1}_{B_{\mathbf{0}}(1)}(\boldsymbol{w})$ and whose corresponding Lévy–Itô decomposition is given by $\boldsymbol{Y}_t = \boldsymbol{\gamma}_{\boldsymbol{Y},\kappa}t + \boldsymbol{\Sigma_Y}\boldsymbol{B}_t^{\boldsymbol{Y}} + \boldsymbol{D}_t^{\boldsymbol{Y},\kappa} + \boldsymbol{J}_t^{\boldsymbol{Y},\kappa}$, defined as above. The following elementary inequality will be used throughout: for any $q \in (0, 2]$,

$$\mathcal{W}_q(\boldsymbol{X}, \boldsymbol{Y}) \leqslant |\boldsymbol{\gamma}_{\boldsymbol{X},\kappa} - \boldsymbol{\gamma}_{\boldsymbol{Y},\kappa}|^{q\wedge 1} + (2\sqrt{d})^{q\wedge 1}|\boldsymbol{\Sigma_X} - \boldsymbol{\Sigma_Y}|^{q\wedge 1} + \mathcal{W}_q\big(\boldsymbol{D}^{\boldsymbol{X},\kappa}, \boldsymbol{D}^{\boldsymbol{Y},\kappa}\big) + \mathcal{W}_q\big(\boldsymbol{J}^{\boldsymbol{X},\kappa}, \boldsymbol{J}^{\boldsymbol{Y},\kappa}\big), \tag{4}$$

where $|\cdot|$ in the last term denotes the Frobenius norm on $\mathbb{R}^{d\times d}$ (i.e. $|\boldsymbol{\Sigma}|^2 = \sum_{i,j=1}^n \Sigma_{i,j}^2$ for $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$). For completeness, we give a proof of (4) in Appendix C below.

Let $\Xi_{\boldsymbol{X}}$ and $\Xi_{\boldsymbol{Y}}$ be the Poisson random measures on $[0, \infty) \times \mathbb{R}^d_{\mathbf{0}}$ of the jumps of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively, with corresponding compensated measures $\widetilde{\Xi}_{\boldsymbol{X}} = \Xi_{\boldsymbol{X}} - \mathrm{Leb}\otimes\nu_{\boldsymbol{X}}$ and $\widetilde{\Xi}_{\boldsymbol{Y}} = \Xi_{\boldsymbol{Y}} - \mathrm{Leb}\otimes\nu_{\boldsymbol{Y}}$, where Leb denotes the Lebesgue measure on $[0,\infty)$. Since, for every $t \geqslant 0$, we have

$$
\begin{aligned}
\boldsymbol{D}_t^{\boldsymbol{X},\kappa} &= \int_{[0,t]\times\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})\boldsymbol{w}\widetilde{\Xi}_{\boldsymbol{X}}(\mathrm{d}s,\mathrm{d}\boldsymbol{w}), \quad \boldsymbol{J}_t^{\boldsymbol{X},\kappa} = \int_{[0,t]\times\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})\boldsymbol{w}\Xi_{\boldsymbol{X}}(\mathrm{d}s,\mathrm{d}\boldsymbol{w}), \\
\boldsymbol{D}_t^{\boldsymbol{Y},\kappa} &= \int_{[0,t]\times\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})\boldsymbol{w}\widetilde{\Xi}_{\boldsymbol{Y}}(\mathrm{d}s,\mathrm{d}\boldsymbol{w}), \quad \boldsymbol{J}_t^{\boldsymbol{Y},\kappa} = \int_{[0,t]\times\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})\boldsymbol{w}\Xi_{\boldsymbol{Y}}(\mathrm{d}s,\mathrm{d}\boldsymbol{w}),
\end{aligned} \tag{5}
$$

the problem of coupling the jump components of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is reduced to coupling the Poisson random measures $\Xi_{\boldsymbol{X}}$ and $\Xi_{\boldsymbol{Y}}$. Sections 4.1 and 4.2 below each describe such a coupling.

4.1. **Thinning.** Choose any Lévy measure μ on $\mathbb{R}^d_{\boldsymbol{0}}$ that dominates both $\nu_{\boldsymbol{X}}$ and $\nu_{\boldsymbol{Y}}$ with Radon-Nikodym derivatives bounded by 1 μ-a.e., i.e. $f_{\boldsymbol{X}} = \mathrm{d}\nu_{\boldsymbol{X}}/\mathrm{d}\mu \leqslant 1$ and $f_{\boldsymbol{Y}} = \mathrm{d}\nu_{\boldsymbol{Y}}/\mathrm{d}\mu \leqslant 1$ μ-a.e. For instance, a possible choice of μ is $\nu_{\boldsymbol{X}} + \nu_{\boldsymbol{Y}}$. Let $\Xi = \sum_{n\in\mathbb{N}} \delta_{(U_n,\boldsymbol{V}_n)}$ be a Poisson random measure on $(0,1]\times\mathbb{R}^d$, with mean measure $\mathrm{Leb}\otimes\mu$ and the corresponding compensated Poisson random measure $\widetilde{\Xi}(\mathrm{d}s,\mathrm{d}\boldsymbol{w}) = \Xi(\mathrm{d}s,\mathrm{d}\boldsymbol{w}) - \mathrm{d}s\otimes\mu(\mathrm{d}\boldsymbol{w})$. Assume the sequence $(\vartheta_n)_{n\in\mathbb{N}}$ of iid uniform random variables on $[0,1]$ is independent of Ξ. The Marking and Mapping Theorems [31] imply that the following Poisson random measures

$$
\Xi_{\boldsymbol{X}} = \sum_{n\in\mathbb{N}} \mathbb{1}_{\{\vartheta_n \leqslant f_{\boldsymbol{X}}(\boldsymbol{V}_n)\}}\delta_{(U_n,\boldsymbol{V}_n)}, \qquad \text{and} \qquad \Xi_{\boldsymbol{Y}} = \sum_{n\in\mathbb{N}} \mathbb{1}_{\{\vartheta_n \leqslant f_{\boldsymbol{Y}}(\boldsymbol{V}_n)\}}\delta_{(U_n,\boldsymbol{V}_n)}, \tag{6}
$$

have mean measures $\mathrm{Leb}\otimes\nu_{\boldsymbol{X}}$ and $\mathrm{Leb}\otimes\nu_{\boldsymbol{Y}}$, respectively. We couple $\boldsymbol{X}$ and $\boldsymbol{Y}$ by choosing $\boldsymbol{B}^{\boldsymbol{X}} = \boldsymbol{B}^{\boldsymbol{Y}}$ in their Lévy–Itô decompositions and couple their jump parts from (5) via the coupling of the Poisson random measures in (6).

Proposition 4.1. *The coupling* $(\boldsymbol{D}^{\boldsymbol{X},\kappa}, \boldsymbol{D}^{\boldsymbol{Y},\kappa}, \boldsymbol{J}^{\boldsymbol{X},\kappa}, \boldsymbol{J}^{\boldsymbol{Y},\kappa})$ *defined in* (5) *and* (6) *satisfies*

$$
\mathbb{E}\Big[\sup_{t\in[0,1]} \big|\boldsymbol{D}_t^{\boldsymbol{X},\kappa} - \boldsymbol{D}_t^{\boldsymbol{Y},\kappa}\big|^2\Big] \leqslant 4\int_{\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^2|f_{\boldsymbol{X}}(\boldsymbol{w}) - f_{\boldsymbol{Y}}(\boldsymbol{w})|\mu(\mathrm{d}\boldsymbol{w}). \tag{7}
$$

Moreover, if $\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})$ *and* $\nu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{w})\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})$ *have a finite second moment, then*

$$
\mathbb{E}\Big[\sup_{t\in[0,1]} \big|\boldsymbol{D}_t^{\boldsymbol{X},\kappa} + \boldsymbol{J}_t^{\boldsymbol{X},\kappa} - (\boldsymbol{D}_t^{\boldsymbol{Y},\kappa} + \boldsymbol{J}_t^{\boldsymbol{Y},\kappa}) - \boldsymbol{m}_\kappa t\big|^2\Big] \leqslant 4\int_{\mathbb{R}^d_{\boldsymbol{0}}} |\boldsymbol{w}|^2|f_{\boldsymbol{X}}(\boldsymbol{w}) - f_{\boldsymbol{Y}}(\boldsymbol{w})|\mu(\mathrm{d}\boldsymbol{w}) \quad \textit{and}
$$

$$
\mathcal{W}_2(\boldsymbol{X},\boldsymbol{Y}) \leqslant |\boldsymbol{\gamma}_{\boldsymbol{X},\kappa} - \boldsymbol{\gamma}_{\boldsymbol{Y},\kappa} + \boldsymbol{m}_\kappa| + 2d^{1/2}|\boldsymbol{\Sigma}_{\boldsymbol{X}} - \boldsymbol{\Sigma}_{\boldsymbol{Y}}| + 2\bigg(\int_{\mathbb{R}^d_{\boldsymbol{0}}} |\boldsymbol{w}|^2|f_{\boldsymbol{X}}(\boldsymbol{w}) - f_{\boldsymbol{Y}}(\boldsymbol{w})|\mu(\mathrm{d}\boldsymbol{w})\bigg)^{1/2},
$$

where the mean $\boldsymbol{m}_\kappa := \mathbb{E}[\boldsymbol{J}_1^{\boldsymbol{X},\kappa} - \boldsymbol{J}_1^{\boldsymbol{Y},\kappa}] = \int_{\mathbb{R}^d_{\boldsymbol{0}}} \boldsymbol{w}\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})(f_{\boldsymbol{X}}(\boldsymbol{w}) - f_{\boldsymbol{Y}}(\boldsymbol{w}))\mu(\mathrm{d}\boldsymbol{w})$ *is finite.*

Proof. Denote $f^+ := \max\{0,f\}$ for any function mapping into $\mathbb{R}$. Define the Poisson random measures

$$
\Lambda_+ := \sum_{n\in\mathbb{N}} \mathbb{1}_{\{f_{\boldsymbol{Y}}(\boldsymbol{V}_n) < \vartheta_n \leqslant f_{\boldsymbol{X}}(\boldsymbol{V}_n)\}}\delta_{(U_n,\boldsymbol{V}_n)} \qquad \text{and} \qquad \Lambda_- := \sum_{n\in\mathbb{N}} \mathbb{1}_{\{f_{\boldsymbol{X}}(\boldsymbol{V}_n) < \vartheta_n \leqslant f_{\boldsymbol{Y}}(\boldsymbol{V}_n)\}}\delta_{(U_n,\boldsymbol{V}_n)}, \tag{8}
$$

with mean measures $\mathrm{Leb}\otimes(f_{\boldsymbol{X}} - f_{\boldsymbol{Y}})^+\mu$ and $\mathrm{Leb}\otimes(f_{\boldsymbol{Y}} - f_{\boldsymbol{X}})^+\mu$, respectively. Thus $\Xi_{\boldsymbol{X}} - \Xi_{\boldsymbol{Y}} = \Lambda_+ - \Lambda_-$. Note that Λ_+ is independent of Λ_- since they are both thinnings of the same Poisson random measure and have disjoint supports. Let $\widetilde{\Lambda}_+$ and $\widetilde{\Lambda}_-$ denote their respective compensated Poisson random measures and define the Lévy processes $\boldsymbol{D}^\pm = (\boldsymbol{D}_t^\pm)_{t\geqslant 0}$ by $\boldsymbol{D}_t^\pm := \int_{(0,t]\times\mathbb{R}^d_{\boldsymbol{0}}} \boldsymbol{w}\mathbb{1}_{B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})\widetilde{\Lambda}_\pm,(\mathrm{d}s,\mathrm{d}\boldsymbol{w})$, where $\pm \in \{+,-\}$. By construction, $\boldsymbol{D}^+$ and $\boldsymbol{D}^-$ are independent square-integrable martingales, satisfying $\mathbb{E}\big[\boldsymbol{D}_t^+\big] = \mathbb{E}\big[\boldsymbol{D}_t^-\big] = 0$ and $\boldsymbol{D}_t^{\boldsymbol{X},\kappa} - \boldsymbol{D}_t^{\boldsymbol{Y},\kappa} = \boldsymbol{D}_t^+ - \boldsymbol{D}_t^-$ for all $t \in \mathbb{R}_+$. In particular, we have $\mathbb{E}\big[\langle\boldsymbol{D}_t^+, \boldsymbol{D}_t^-\rangle\big] = 0$ and, by Campbell's formula [31, p. 28],

$$
\begin{aligned}
\mathbb{E}\left[|\boldsymbol{D}_t^+|^2\right] &= t\int_{\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^2(f_{\boldsymbol{X}}(\boldsymbol{w}) - f_{\boldsymbol{Y}}(\boldsymbol{w}))^+\mu(\mathrm{d}\boldsymbol{w}), \\
\mathbb{E}\left[|\boldsymbol{D}_t^-|^2\right] &= t\int_{\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^2(f_{\boldsymbol{Y}}(\boldsymbol{w}) - f_{\boldsymbol{X}}(\boldsymbol{w}))^+\mu(\mathrm{d}\boldsymbol{w}).
\end{aligned}
$$

Doob's maximal inequality [29, Prop. 7.16], applied to the submartingale $|\boldsymbol{D}^+ - \boldsymbol{D}^-|$, and the independence of martingales $\boldsymbol{D}^+$ and $\boldsymbol{D}^-$ yield

$$\mathbb{E}\Big[\sup_{t\in[0,1]}\big|\boldsymbol{D}_t^{\boldsymbol{X},\kappa} - \boldsymbol{D}_t^{\boldsymbol{Y},\kappa}\big|^2\Big] = \mathbb{E}\Big[\sup_{t\in[0,1]}\big|\boldsymbol{D}_t^+ - \boldsymbol{D}_t^-\big|^2\Big] \leqslant 4\mathbb{E}\big[\big|\boldsymbol{D}_1^+ - \boldsymbol{D}_1^-\big|^2\big] = 4\mathbb{E}\big[\big|\boldsymbol{D}_1^+\big|^2\big] + 4\mathbb{E}\big[\big|\boldsymbol{D}_1^-\big|^2\big]$$
$$= 4\int_{\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^2|f_{\boldsymbol{X}}(\boldsymbol{w}) - f_{\boldsymbol{Y}}(\boldsymbol{w})|\mu(\mathrm{d}\boldsymbol{w}).$$

Assume, that $\int_{\mathbb{R}^d_{\boldsymbol{0}}} |\boldsymbol{w}|^2\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) < \infty$ and $\int_{\mathbb{R}^d_{\boldsymbol{0}}} |\boldsymbol{w}|^2\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})\nu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{w}) < \infty$, and define the Lévy processes $\boldsymbol{J}^\pm = (\boldsymbol{J}_t^\pm)_{t\geqslant 0}$ by $\boldsymbol{J}_t^\pm := \int_{(0,t]\times\mathbb{R}^d_{\boldsymbol{0}}} \boldsymbol{w}\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})\Lambda_\pm,(\mathrm{d}s,\mathrm{d}\boldsymbol{w})$. By the integrability assumption and construction, $\boldsymbol{J}^+$ and $\boldsymbol{J}^-$ are independent square-integrable processes with $\mathbb{E}[\boldsymbol{J}_t^+ - \boldsymbol{J}_t^-] = t\boldsymbol{m}_\kappa$ and $\boldsymbol{J}_t^{\boldsymbol{X},\kappa} - \boldsymbol{J}_t^{\boldsymbol{Y},\kappa} = \boldsymbol{J}_t^+ - \boldsymbol{J}_t^-$ for all $t \in \mathbb{R}_+$. Thus, Campbell's formula [31, p. 28] yields

$$\mathbb{E}\big[|\boldsymbol{D}_t^+ + \boldsymbol{J}_t^+ - \mathbb{E}[\boldsymbol{J}_t^+]|^2\big] = t\int_{\mathbb{R}^d_{\boldsymbol{0}}} |\boldsymbol{w}|^2(f_{\boldsymbol{X}}(\boldsymbol{w}) - f_{\boldsymbol{Y}}(\boldsymbol{w}))^+\mu(\mathrm{d}\boldsymbol{w}),$$
$$\mathbb{E}\big[|\boldsymbol{D}_t^- + \boldsymbol{J}_t^- - \mathbb{E}[\boldsymbol{J}_t^-]|^2\big] = t\int_{\mathbb{R}^d_{\boldsymbol{0}}} |\boldsymbol{w}|^2(f_{\boldsymbol{Y}}(\boldsymbol{w}) - f_{\boldsymbol{X}}(\boldsymbol{w}))^+\mu(\mathrm{d}\boldsymbol{w}).$$

Next, Doob's maximal inequality applied to the submartingale $|\boldsymbol{D}_t^+ + \boldsymbol{J}_t^+ - (\boldsymbol{D}_t^- + \boldsymbol{J}_t^-) - t\boldsymbol{m}_\kappa|$, and the independence between $\boldsymbol{D}_t^+ + \boldsymbol{J}_t^+ - \mathbb{E}[\boldsymbol{J}_t^+]$ and $\boldsymbol{D}_t^- + \boldsymbol{J}_t^- - \mathbb{E}[\boldsymbol{J}_t^-]$, yield

$$\mathbb{E}\Big[\sup_{t\in[0,1]}\big|\boldsymbol{D}_t^{\boldsymbol{X},\kappa} + \boldsymbol{J}_t^{\boldsymbol{X},\kappa} - \big(\boldsymbol{D}_t^{\boldsymbol{Y},\kappa} + \boldsymbol{J}_t^{\boldsymbol{Y},\kappa}\big) - \boldsymbol{m}_\kappa t\big|^2\Big] = \mathbb{E}\Big[\sup_{t\in[0,1]}\big|\boldsymbol{D}_t^+ + \boldsymbol{J}_t^+ - \big(\boldsymbol{D}_t^- + \boldsymbol{J}_t^-\big) - \boldsymbol{m}_\kappa t\big|^2\Big]$$
$$\leqslant 4\mathbb{E}\big[\big|\boldsymbol{D}_1^+ + \boldsymbol{J}_1^+ - (\boldsymbol{D}_1^- + \boldsymbol{J}_1^-) - \boldsymbol{m}_\kappa\big|^2\big]$$
$$= 4\int_{\mathbb{R}^d_{\boldsymbol{0}}} |\boldsymbol{w}|^2|f_{\boldsymbol{X}}(\boldsymbol{w}) - f_{\boldsymbol{Y}}(\boldsymbol{w})|\mu(\mathrm{d}\boldsymbol{w}). \qquad \square$$

The following bound is required when the big jump components have infinite variance.

Proposition 4.2. *Consider the coupling* $(\boldsymbol{D}^{\boldsymbol{X},\kappa}, \boldsymbol{D}^{\boldsymbol{Y},\kappa}, \boldsymbol{J}^{\boldsymbol{X},\kappa}, \boldsymbol{J}^{\boldsymbol{Y},\kappa})$ *defined in* (5) *and* (6)*. Then*

$$\mathbb{E}\Big[\sup_{t\in[0,1]}\big|\boldsymbol{J}_t^{\boldsymbol{X},\kappa} - \boldsymbol{J}_t^{\boldsymbol{Y},\kappa}\big|^q\Big] \leqslant \int_{\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^q|f_{\boldsymbol{X}}(\boldsymbol{w}) - f_{\boldsymbol{Y}}(\boldsymbol{w})|\mu(\mathrm{d}\boldsymbol{w}), \text{ for any } q \in (0,1]. \tag{9}$$

In particular, the following inequality holds for any $q \in (0,1]$*:*

$$\begin{aligned}\mathcal{W}_q(\boldsymbol{X},\boldsymbol{Y}) \leqslant{}& |\boldsymbol{\gamma}_{\boldsymbol{X},\kappa} - \boldsymbol{\gamma}_{\boldsymbol{Y},\kappa}|^q + \left(4\int_{\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^2|f_{\boldsymbol{X}}(\boldsymbol{w}) - f_{\boldsymbol{Y}}(\boldsymbol{w})|\mu(\mathrm{d}\boldsymbol{w})\right)^{q/2}\\ &+ 2^q d^{q/2}|\boldsymbol{\Sigma}_{\boldsymbol{X}} - \boldsymbol{\Sigma}_{\boldsymbol{Y}}|^q + \int_{\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^q|f_{\boldsymbol{X}}(\boldsymbol{w}) - f_{\boldsymbol{Y}}(\boldsymbol{w})|\mu(\mathrm{d}\boldsymbol{w}).\end{aligned} \tag{10}$$

Note that (9) & (10) hold without assuming $\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})$ and $\nu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{w})\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})$ have a finite q-moment. If this holds, however, the bound is non-trivial because the big jumps in Proposition 4.2 are then controlled by $\int_{\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^q\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) + \int_{\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^q\nu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{w}) < \infty$.

Proof. Recall that $\kappa \in (0,1]$ and let $\kappa' \in (\kappa,\infty)$. Next, for $t \geqslant 0$, we define the processes $\boldsymbol{J}_{t,\kappa'}^{\boldsymbol{X},\kappa} := \int_{[0,t]\times\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{B_{\boldsymbol{0}}(\kappa')\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})\boldsymbol{w}\Xi_{\boldsymbol{X}}(\mathrm{d}s,\mathrm{d}\boldsymbol{w})$ and $\boldsymbol{J}_{t,\kappa'}^{\boldsymbol{Y},\kappa} := \int_{[0,t]\times\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{B_{\boldsymbol{0}}(\kappa')\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})\boldsymbol{w}\Xi_{\boldsymbol{Y}}(\mathrm{d}s,\mathrm{d}\boldsymbol{w})$. Let $\Lambda_\pm$ be as in (8), and define $\boldsymbol{J}_{\cdot,\kappa'}^\pm = (\boldsymbol{J}_{t,\kappa'}^\pm)_{t\geqslant 0}$ as $\boldsymbol{J}_{t,\kappa'}^\pm := \int_{(0,t]\times\mathbb{R}^d_{\boldsymbol{0}}} \boldsymbol{w}\mathbb{1}_{B_{\boldsymbol{0}}(\kappa')\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})\Lambda_\pm(\mathrm{d}s,\mathrm{d}\boldsymbol{w})$ (both being Lévy processes), and note that $\boldsymbol{J}_{t,\kappa'}^+ - \boldsymbol{J}_{t,\kappa'}^- = \boldsymbol{J}_{t,\kappa'}^{\boldsymbol{X},\kappa} - \boldsymbol{J}_{t,\kappa'}^{\boldsymbol{Y},\kappa}$ for $t \in \mathbb{R}_+$ and $\kappa' \in (\kappa,\infty)$. Note that $\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})\mathbb{1}_{B_{\boldsymbol{0}}(\kappa')\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})$ and $\nu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{w})\mathbb{1}_{B_{\boldsymbol{0}}(\kappa')\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})$ have a finite q-moment for all $\kappa' \in (\kappa,\infty)$. Moreover, by the triangle inequality and the fact $(x+y)^q \leqslant x^q + y^q$ for all $x, y \geqslant 0$, we have

$$\sup_{t\in[0,1]}\big|\boldsymbol{J}_{t,\kappa'}^+ - \boldsymbol{J}_{t,\kappa'}^-\big|^q \leqslant \int_{(0,1]\times\mathbb{R}^d_{\boldsymbol{0}}} \mathbb{1}_{B_{\boldsymbol{0}}(\kappa')\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^q(\Lambda_+(\mathrm{d}s,\mathrm{d}\boldsymbol{w}) + \Lambda_-(\mathrm{d}s,\mathrm{d}\boldsymbol{w})). \tag{11}$$

Recall that $\mathrm{Leb}\otimes(f_{\boldsymbol{X}}-f_{\boldsymbol{Y}})^{+}\mu$ and $\mathrm{Leb}\otimes(f_{\boldsymbol{Y}}-f_{\boldsymbol{X}})^{+}\mu$ are the mean measures of Λ_+ and Λ_-, respectively. Thus, by taking expectations in (11) and applying Campbell's formula [31, p. 28], we get

$$\mathbb{E}\bigg[\sup_{t\in[0,1]}\big|\boldsymbol{J}^{\boldsymbol{X},\kappa}_{t,\kappa'}-\boldsymbol{J}^{\boldsymbol{Y},\kappa}_{t,\kappa'}\big|^q\bigg]=\mathbb{E}\bigg[\sup_{t\in[0,1]}\big|\boldsymbol{J}^{+}_{t,\kappa'}-\boldsymbol{J}^{-}_{t,\kappa}\big|^q\bigg]$$
$$\leqslant\int_{\mathbb{R}^d_{\boldsymbol{0}}}\mathbb{1}_{B_{\boldsymbol{0}}(\kappa')\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^q|f_{\boldsymbol{X}}(\boldsymbol{w})-f_{\boldsymbol{Y}}(\boldsymbol{w})|\mu(\mathrm{d}\boldsymbol{w}).$$

Due to the monotone convergence theorem, it follows that, as $\kappa'\to\infty$,

$$\int_{B_{\boldsymbol{0}}(\kappa')\setminus B_{\boldsymbol{0}}(\kappa)}|\boldsymbol{w}|^q|f_{\boldsymbol{X}}(\boldsymbol{w})-f_{\boldsymbol{Y}}(\boldsymbol{w})|\mu(\mathrm{d}\boldsymbol{w})\to\int_{\mathbb{R}^d_{\boldsymbol{0}}}\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^q|f_{\boldsymbol{X}}(\boldsymbol{w})-f_{\boldsymbol{Y}}(\boldsymbol{w})|\mu(\mathrm{d}\boldsymbol{w}).$$

Furthermore, Fatou's lemma together with the above observations imply that

$$\mathbb{E}\bigg[\liminf_{\kappa'\to\infty}\sup_{t\in[0,1]}\big|\boldsymbol{J}^{\boldsymbol{X},\kappa}_{t,\kappa'}-\boldsymbol{J}^{\boldsymbol{Y},\kappa}_{t,\kappa'}\big|^q\bigg]\leqslant\liminf_{\kappa'\to\infty}\mathbb{E}\bigg[\sup_{t\in[0,1]}\big|\boldsymbol{J}^{\boldsymbol{X},\kappa}_{t,\kappa'}-\boldsymbol{J}^{\boldsymbol{Y},\kappa}_{t,\kappa'}\big|^q\bigg]$$
$$\leqslant\int_{\mathbb{R}^d_{\boldsymbol{0}}}\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^q|f_{\boldsymbol{X}}(\boldsymbol{w})-f_{\boldsymbol{Y}}(\boldsymbol{w})|\mu(\mathrm{d}\boldsymbol{w}).$$

We have $\liminf_{\kappa'\to\infty}\sup_{t\in[0,1]}\big|\boldsymbol{J}^{\boldsymbol{X},\kappa}_{t,\kappa'}-\boldsymbol{J}^{\boldsymbol{Y},\kappa}_{t,\kappa'}\big|^q=\sup_{t\in[0,1]}\big|\boldsymbol{J}^{\boldsymbol{X},\kappa}_{t}-\boldsymbol{J}^{\boldsymbol{Y},\kappa}_{t}\big|^q$ a.s., since the largest jump of $\boldsymbol{J}^{\boldsymbol{X},\kappa}$ and $\boldsymbol{J}^{\boldsymbol{Y},\kappa}$ are finite on the time interval $[0,1]$. This implies (9).

Since $\mathcal{W}_q(\boldsymbol{D}^{\boldsymbol{X}},\boldsymbol{D}^{\boldsymbol{Y}})\leqslant\mathcal{W}_2(\boldsymbol{D}^{\boldsymbol{X}},\boldsymbol{D}^{\boldsymbol{Y}})^q$, the inequality in (10) follows from (4), (7) and (9). □

4.2. **Comonotonic coupling.** In this section, we introduce the d-dimensional comonotonic coupling of jumps for any $d\geqslant 1$. We use two ingredients to construct this coupling of the Lévy processes $\boldsymbol{X}$ and $\boldsymbol{Y}$: **(I)** the comonotonic coupling of real-valued random variables ξ and ζ, given by $(\xi,\zeta)=(F^{\leftarrow}_{\xi}(U),F^{\leftarrow}_{\zeta}(U))$, where U is uniform on $(0,1)$ and the functions $F^{\leftarrow}_{\xi}$ and $F^{\leftarrow}_{\zeta}$ are the right inverses of the functions F_ξ and F_ζ; **(II)** LaPage's representation of the Poisson random measures of a Lévy process (see [38, p. 4]).

The comonotonic coupling of the real-valued variables in **(I)** is optimal for the L^p-Wasserstein distance (see [36, Ex. 3.2.14]), $\mathcal{W}_p(\xi,\zeta)^p=\int_0^1|F^{\leftarrow}_{\xi}(u)-F^{\leftarrow}_{\zeta}(u)|^p\mathrm{d}u=\mathbb{E}\big[|F^{\leftarrow}_{\xi}(U)-F^{\leftarrow}_{\zeta}(U)|^p\big]$ for $p\geqslant 1$. The representation in **(II)** decomposes the jumps of a Lévy process into its magnitude (i.e. norm) and angular component. The main idea behind our coupling of the Lévy processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ is to couple their respective Poisson random measures of jumps via a comonotonic coupling of the magnitudes of jumps, while simultaneously aligning their angular components. We now describe this construction.

Recall that the Lévy processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ in $\mathbb{R}^d$ have characteristic triplets $(\boldsymbol{\gamma}_{\boldsymbol{X}},\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{\top},\nu_{\boldsymbol{X}})$ and $(\boldsymbol{\gamma}_{\boldsymbol{Y}},\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{\top},\nu_{\boldsymbol{Y}})$, respectively. Suppose the Lévy measure $\nu_{\boldsymbol{X}}$ (resp. $\nu_{\boldsymbol{Y}}$) of $\boldsymbol{X}$ (resp. $\boldsymbol{Y}$) admits a radial decomposition (see [32, p. 282]), that is, there exists a probability measure $\sigma_{\boldsymbol{X}}$ (resp. $\sigma_{\boldsymbol{Y}}$) on the unit sphere $\mathbb{S}^{d-1}$ (with convention $\mathbb{S}^0\coloneqq\{-1,1\}$) such that:

$$\nu_{\boldsymbol{X}}(B)=\int_{\mathbb{S}^{d-1}}\int_0^\infty\mathbb{1}_B(x\boldsymbol{v})\rho^0_{\boldsymbol{X}}(\mathrm{d}x,\boldsymbol{v})\sigma_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{v}),\ \bigg(\text{resp. }\nu_{\boldsymbol{Y}}(B)=\int_{\mathbb{S}^{d-1}}\int_0^\infty\mathbb{1}_B(x\boldsymbol{v})\rho^0_{\boldsymbol{Y}}(\mathrm{d}x,\boldsymbol{v})\sigma_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{v})\bigg),$$

for any $B\in\mathcal{B}(\mathbb{R}^d\setminus\{\boldsymbol{0}\})$, where $\{\rho^0_{\boldsymbol{X}}(\cdot,\boldsymbol{v})\}_{\boldsymbol{v}\in\mathbb{S}^{d-1}}$ (resp. $\{\rho^0_{\boldsymbol{Y}}(\cdot,\boldsymbol{v})\}_{\boldsymbol{v}\in\mathbb{S}^{d-1}}$) is a measurable family of Lévy measures on $(0,\infty)$. Define the probability measure $\sigma\coloneqq(\sigma_{\boldsymbol{X}}+\sigma_{\boldsymbol{Y}})/2$ on $\mathbb{S}^{d-1}$ and the Radon-Nikodym derivatives $f^\sigma_{\boldsymbol{X}}(\boldsymbol{v})\coloneqq\sigma_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{v})/\sigma(\mathrm{d}\boldsymbol{v})\leqslant 2$ and $f^\sigma_{\boldsymbol{Y}}(\boldsymbol{v})\coloneqq\sigma_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{v})/\sigma(\mathrm{d}\boldsymbol{v})\leqslant 2$ for $\boldsymbol{v}\in\mathbb{S}^{d-1}$. Consider the following radial decompositions of $\nu_{\boldsymbol{X}}$ and $\nu_{\boldsymbol{Y}}$:

$$\nu_{\boldsymbol{X}}(B)=\int_{\mathbb{S}^{d-1}}\int_0^\infty\mathbb{1}_B(x\boldsymbol{v})\rho_{\boldsymbol{X}}(\mathrm{d}x,\boldsymbol{v})\sigma(\mathrm{d}\boldsymbol{v}),\quad\nu_{\boldsymbol{Y}}(B)=\int_{\mathbb{S}^{d-1}}\int_0^\infty\mathbb{1}_B(x\boldsymbol{v})\rho_{\boldsymbol{Y}}(\mathrm{d}x,\boldsymbol{v})\sigma(\mathrm{d}\boldsymbol{v}),\tag{12}$$

for $B\in\mathcal{B}(\mathbb{R}^d\setminus\{\mathbf{0}\})$, where $\rho_{\boldsymbol{X}}(\cdot,\boldsymbol{v})\coloneqq f^\sigma_{\boldsymbol{X}}(\boldsymbol{v})\rho^0_{\boldsymbol{X}}(\cdot,\boldsymbol{v})$ and $\rho_{\boldsymbol{Y}}(\cdot,\boldsymbol{v})\coloneqq f^\sigma_{\boldsymbol{Y}}(\boldsymbol{v})\rho^0_{\boldsymbol{Y}}(\cdot,\boldsymbol{v})$ for $\boldsymbol{v}\in\mathbb{S}^{d-1}$. The advantage of the decomposition in (12), compared to the one in the display above, is that the angular components of jumps are sampled from the same measure σ on $\mathbb{S}^{d-1}$, making it possible to couple the jumps of $\boldsymbol{X}$ and $\boldsymbol{Y}$ by coupling their magnitudes.

For every $\boldsymbol{v}\in\mathbb{S}^{d-1}$, let $u\mapsto\rho^{\leftarrow}_{\boldsymbol{X}}(u,\boldsymbol{v})$ (resp. $u\mapsto\rho^{\leftarrow}_{\boldsymbol{Y}}(u,\boldsymbol{v})$) be the right inverse of $x\mapsto\rho_{\boldsymbol{X}}([x,\infty),\boldsymbol{v})$ (resp. $x\mapsto\rho_{\boldsymbol{Y}}([x,\infty),\boldsymbol{v})$). Let $(U_n)_{n\in\mathbb{N}}$ be a sequence of iid uniform random variables on $[0,1]$, and let $(\Gamma_n)_{n\in\mathbb{N}}$ be a sequence of partial sums of iid standard exponentially distributed random variables that is independent of $(U_n)_{n\in\mathbb{N}}$. Next, independent of $(U_n,\Gamma_n)_{n\in\mathbb{N}}$, we denote by $(\boldsymbol{V}_n)_{n\in\mathbb{N}}$ a sequence of iid random vectors on $\mathbb{S}^{d-1}$ with common distribution σ. Define the Poisson point process Ξ on $[0,1]\times(0,\infty)\times\mathbb{S}^{d-1}$ with measure $\mathrm{Leb}\otimes\mathrm{Leb}\otimes\sigma$ and the compensated Poisson random measure $\widetilde{\Xi}(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v})$ as follows:

$$\Xi\coloneqq\sum_{n\in\mathbb{N}}\delta_{(U_n,\Gamma_n,\boldsymbol{V}_n)},\qquad \widetilde{\Xi}(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v})=\Xi(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v})-\mathrm{d}s\otimes\mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v}). \tag{13}$$

Next, we note that (by Proposition 4.3 below) for any $\varepsilon\in(0,\infty)$ (and even $\varepsilon=\infty$ when $\boldsymbol{X}$ and $\boldsymbol{Y}$ both have jumps of finite variation), the small-jump components of $\boldsymbol{X}$ and $\boldsymbol{Y}$ take the form

$$\boldsymbol{M}^{\boldsymbol{X}}_t\coloneqq\int_{[0,t]\times[\varepsilon,\infty)\times\mathbb{S}^{d-1}}\boldsymbol{v}\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})\widetilde{\Xi}(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v}),\quad \boldsymbol{M}^{\boldsymbol{Y}}_t\coloneqq\int_{[0,t]\times[\varepsilon,\infty)\times\mathbb{S}^{d-1}}\boldsymbol{v}\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v})\widetilde{\Xi}(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v}). \tag{14}$$

The big-jump components of $\boldsymbol{X}$ and $\boldsymbol{Y}$ can similarly be expressed as

$$\boldsymbol{L}^{\boldsymbol{X}}_t\coloneqq\int_{[0,t]\times(0,\varepsilon)\times\mathbb{S}^{d-1}}\boldsymbol{v}\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})\Xi(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v}),\quad \boldsymbol{L}^{\boldsymbol{Y}}_t\coloneqq\int_{[0,t]\times(0,\varepsilon)\times\mathbb{S}^{d-1}}\boldsymbol{v}\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v})\Xi(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v}). \tag{15}$$

Proposition 4.3. *Let Lévy processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ have characteristic triplets $(\boldsymbol{\gamma}_{\boldsymbol{X}},\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{X}}^\top,\nu_{\boldsymbol{X}})$ and $(\boldsymbol{\gamma}_{\boldsymbol{Y}},\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{\Sigma}_{\boldsymbol{Y}}^\top,\nu_{\boldsymbol{Y}})$, respectively. Assume that the Lévy measures of $\nu_{\boldsymbol{X}}$ and $\nu_{\boldsymbol{Y}}$ admit the radial decomposition in (12) and construct the processes $(\boldsymbol{M}^{\boldsymbol{X}},\boldsymbol{M}^{\boldsymbol{Y}},\boldsymbol{L}^{\boldsymbol{X}},\boldsymbol{L}^{\boldsymbol{Y}})$ by (14) and (15), independent of standard Brownian motions $\boldsymbol{B}^{\boldsymbol{X}}$ and $\boldsymbol{B}^{\boldsymbol{Y}}$ on $\mathbb{R}^d$. Then there exists constants $\boldsymbol{\varpi}_{\boldsymbol{X}},\boldsymbol{\varpi}_{\boldsymbol{Y}}\in\mathbb{R}^d$, such that $\boldsymbol{X}_t\overset{d}{=}\boldsymbol{\varpi}_{\boldsymbol{X}}t+\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{B}^{\boldsymbol{X}}_t+\boldsymbol{M}^{\boldsymbol{X}}_t+\boldsymbol{L}^{\boldsymbol{X}}_t$ and $\boldsymbol{Y}_t\overset{d}{=}\boldsymbol{\varpi}_{\boldsymbol{Y}}t+\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{B}^{\boldsymbol{Y}}_t+\boldsymbol{M}^{\boldsymbol{Y}}_t+\boldsymbol{L}^{\boldsymbol{Y}}_t$ for all $t\in[0,1]$. Moreover, this coupling of $\boldsymbol{X}$ and $\boldsymbol{Y}$ satisfies*

$$\mathbb{E}\Big[\sup_{t\in[0,1]}\big|\boldsymbol{M}^{\boldsymbol{X}}_t-\boldsymbol{M}^{\boldsymbol{Y}}_t\big|^2\Big]\leqslant 4\int_{[\varepsilon,\infty)\times\mathbb{S}^{d-1}}(\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v}))^2\mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v}). \tag{16}$$

Furthermore, if $\int_{\mathbb{R}^d_{\mathbf{0}}}|\boldsymbol{w}|^2\mathbb{1}_{\mathbb{R}^d\setminus B_{\mathbf{0}}(1)}(\boldsymbol{w})\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})<\infty$ and $\int_{\mathbb{R}^d_{\mathbf{0}}}|\boldsymbol{w}|^2\mathbb{1}_{\mathbb{R}^d\setminus B_{\mathbf{0}}(1)}(\boldsymbol{w})\nu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{w})<\infty$, then

$$\mathbb{E}\Big[\sup_{t\in[0,1]}\big|\boldsymbol{M}^{\boldsymbol{X}}_t+\boldsymbol{L}^{\boldsymbol{X}}_t-(\boldsymbol{M}^{\boldsymbol{Y}}_t+\boldsymbol{L}^{\boldsymbol{Y}}_t)-\boldsymbol{m}t\big|^2\Big]\leqslant 4\int_{(0,\infty)\times\mathbb{S}^{d-1}}(\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v}))^2\mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v}), \tag{17}$$

were we define $\boldsymbol{m}\coloneqq\mathbb{E}[\boldsymbol{L}^{\boldsymbol{X}}_1-\boldsymbol{L}^{\boldsymbol{Y}}_1]=\int_{(0,\varepsilon)\times\mathbb{S}^{d-1}}\boldsymbol{v}(\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v}))\mathrm{d}s\otimes\sigma(\mathrm{d}\boldsymbol{v})\in\mathbb{R}^d$. In particular,

$$\begin{aligned}\mathcal{W}_2(\boldsymbol{X},\boldsymbol{Y})&\leqslant|\boldsymbol{\varpi}_{\boldsymbol{X}}-\boldsymbol{\varpi}_{\boldsymbol{Y}}+\boldsymbol{m}|+2d^{1/2}|\boldsymbol{\Sigma}_{\boldsymbol{X}}-\boldsymbol{\Sigma}_{\boldsymbol{Y}}|\\&\quad+2\bigg(\int_{(0,\infty)\times\mathbb{S}^{d-1}}(\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v}))^2\mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v})\bigg)^{1/2}.\end{aligned} \tag{18}$$

Coupling the jumps of $\boldsymbol{X}$ and $\boldsymbol{Y}$ via (14) and (15) is based on the idea behind the one-dimensional comonotonic coupling, applied to the magnitudes of the jumps of $\boldsymbol{X}$ and $\boldsymbol{Y}$. Indeed, in the coupling of Proposition 4.3, we align the angular components of the jumps and then couple the magnitudes via the right inverses $\rho^{\leftarrow}_{\boldsymbol{X}}(\cdot,\boldsymbol{v})$ and $\rho^{\leftarrow}_{\boldsymbol{Y}}(\cdot,\boldsymbol{v})$ (of possibly unbounded functions $x\mapsto\rho_{\boldsymbol{X}}([x,\infty),\boldsymbol{v})$ and $x\mapsto\rho_{\boldsymbol{Y}}([x,\infty),\boldsymbol{v})$) evaluated along the sequence $(\Gamma_n)_{n\in\mathbb{N}}$ of partial sums of iid standard exponentially distributed random variables. Note that this construction is analogous to the one-dimensional comonotonic coupling of real random variables described above, but allows for the functions $x\mapsto\rho_{\boldsymbol{X}}([x,\infty),\boldsymbol{v})$ and $x\mapsto\rho_{\boldsymbol{Y}}([x,\infty),\boldsymbol{v})$ to be unbounded.

Proof. We start by showing that there exist $\boldsymbol{\varpi_X}, \boldsymbol{\varpi_Y} \in \mathbb{R}^d$, such that $\boldsymbol{X}_t \overset{d}{=} \boldsymbol{\varpi_X} t + \boldsymbol{\Sigma_X} \boldsymbol{B}_t^{\boldsymbol{X}} + \boldsymbol{M}_t^{\boldsymbol{X}} + \boldsymbol{L}_t^{\boldsymbol{X}}$ and $\boldsymbol{Y}_t \overset{d}{=} \boldsymbol{\varpi_Y} t + \boldsymbol{\Sigma_Y} \boldsymbol{B}_t^{\boldsymbol{Y}} + \boldsymbol{M}_t^{\boldsymbol{Y}} + \boldsymbol{L}_t^{\boldsymbol{Y}}$ for all $t \in [0,1]$. The proof of this fact is essentially given in [38, p. 4], we outline it here for completeness. By the symmetry of the construction, it is sufficient to prove the first equality in law only. Since $\boldsymbol{X}$ is a Lévy process, $\Xi_{\boldsymbol{X}} = \sum_{\{t:\Delta \boldsymbol{X}_t \neq 0\}} \delta_{(t,\Delta \boldsymbol{X}_t)}$ is a Poisson random measure on $[0,1] \times \mathbb{R}^d_{\boldsymbol{0}}$ of the jumps of $\boldsymbol{X}$ with mean measure $\mathrm{Leb} \otimes \nu_{\boldsymbol{X}}$ [40, Thm 19.2]. By (14) and (15), the equality in law $\boldsymbol{X}_t \overset{d}{=} \boldsymbol{\varpi_X} t + \boldsymbol{\Sigma_X} \boldsymbol{B}_t^{\boldsymbol{X}} + \boldsymbol{M}_t^{\boldsymbol{X}} + \boldsymbol{L}_t^{\boldsymbol{X}}$ holds for some $\boldsymbol{\varpi_X} \in \mathbb{R}^d$ if

$$\Xi_{\boldsymbol{X}} \overset{d}{=} \sum_{n=1}^{\infty} \delta_{(U_n, \rho^{\leftarrow}_{\boldsymbol{X}}(\Gamma_n, \boldsymbol{V}_n)\boldsymbol{V}_n)}. \tag{19}$$

To prove this, consider the Poisson random measure Ξ on $[0,1] \times (0,\infty) \times \mathbb{S}^{d-1}$, with mean measure $\mathrm{Leb}\otimes\mathrm{Leb}\otimes\sigma$, defined in (13). Define $h : [0,1]\times(0,\infty)\times\mathbb{S}^{d-1} \to [0,1]\times\mathbb{R}^d$ by $h(t,x,\boldsymbol{v}) \coloneqq (t, \rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})\boldsymbol{v})$. Crucially, by construction, we have $(\mathrm{Leb}\otimes\mathrm{Leb}\otimes\sigma)\circ h^{-1} = \mathrm{Leb}\otimes\nu_{\boldsymbol{X}}$ on $\mathcal{B}(\mathbb{R}^d_{\boldsymbol{0}})$. Thus, by the Mapping Theorem [31, Sec. 2.3], we get $\Xi \circ h^{-1} \overset{d}{=} \Xi_{\boldsymbol{X}}$. Moreover, since $\sum_{n=1}^{\infty} \delta_{(U_n, \rho^{\leftarrow}_{\boldsymbol{X}}(\Gamma_n, \boldsymbol{V}_n)\boldsymbol{V}_n)} = \Xi \circ h^{-1}$ by construction, the equality in law in (19) follows.

Next, we prove that

$$\boldsymbol{M}_t^{\boldsymbol{X}} - \boldsymbol{M}_t^{\boldsymbol{Y}} = \int_{[0,t]\times[\varepsilon,\infty)\times\mathbb{S}^{d-1}} \boldsymbol{v}(\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}) - \rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}))\widetilde{\Xi}(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v})$$

is a square-integrable martingale. Let $\mathcal{F}_t$ to be the σ-field generated by $\Xi((0,s]\times A)$ for $0 \leqslant s \leqslant t$ and $A \in \mathcal{B}([\varepsilon,\infty)\times\mathbb{S}^{d-1})$, then $\boldsymbol{M}^{\boldsymbol{X}} - \boldsymbol{M}^{\boldsymbol{Y}}$ is adapted w.r.t. $(\mathcal{F}_t)_{t\geqslant 0}$ and fulfils the martingale property by virtue of being an integral with respect to a compensated Poisson random measure. Furthermore, by the triangle inequality, $\boldsymbol{M}_t^{\boldsymbol{X}} - \boldsymbol{M}_t^{\boldsymbol{Y}}$ is square integrable since both $\boldsymbol{M}_t^{\boldsymbol{X}}$ and $\boldsymbol{M}_t^{\boldsymbol{Y}}$ are square integrable. Since the process $|\boldsymbol{M}_t^{\boldsymbol{X}} - \boldsymbol{M}_t^{\boldsymbol{Y}}|$ is a submartingale, Doob's maximal inequality [29, Prop. 7.16] and Campbell's formula [31, p. 28] imply

$$\mathbb{E}\Big[\sup_{t\in[0,1]} \big|\boldsymbol{M}_t^{\boldsymbol{X}} - \boldsymbol{M}_t^{\boldsymbol{Y}}\big|^2\Big] \leqslant 4\mathbb{E}\big[\big|\boldsymbol{M}_1^{\boldsymbol{X}} - \boldsymbol{M}_1^{\boldsymbol{X}}\big|^2\big] = 4\int_{[\varepsilon,\infty)\times\mathbb{S}^{d-1}} (\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}) - \rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v}))^2 \mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v}).$$

If $\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(1)}(\boldsymbol{w})\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})$ and $\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(1)}(\boldsymbol{w})\nu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{w})$ have finite second moment, a similar bound can be established for the big-jump components using Doob's maximal inequality and Campbell's formula:

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in[0,1]} \big|\boldsymbol{M}_t^{\boldsymbol{X}} + \boldsymbol{L}_t^{\boldsymbol{X}} - (\boldsymbol{M}_t^{\boldsymbol{Y}} + \boldsymbol{L}_t^{\boldsymbol{Y}}) - \boldsymbol{m}t\big|^2\Big] &\leqslant 4\mathbb{E}\big[\big|\boldsymbol{M}_1^{\boldsymbol{X}} + \boldsymbol{L}_1^{\boldsymbol{X}} - (\boldsymbol{M}_1^{\boldsymbol{Y}} + \boldsymbol{L}_1^{\boldsymbol{Y}}) - \boldsymbol{m}\big|^2\big] \\ &= 4\int_{(0,\infty)\times\mathbb{S}^{d-1}} (\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}) - \rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v}))^2 \mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v}).\end{aligned}$$

Finally, (18) follows directly from (17) and the standard arguments given in Appendix C. □

Proposition 4.4. *Pick $q \in (0,1]$. Assume that the Lévy measures of $\nu_{\boldsymbol{X}}$ and $\nu_{\boldsymbol{Y}}$ admit the radial decomposition in (12) and construct the processes $(\boldsymbol{L}^{\boldsymbol{X}}, \boldsymbol{L}^{\boldsymbol{Y}})$ by (15). For any $\varepsilon \in (0,\infty)$ (we may have $\varepsilon = \infty$ when $\boldsymbol{X}$ and $\boldsymbol{Y}$ are of finite variation), the coupling $(\boldsymbol{L}^{\boldsymbol{X}}, \boldsymbol{L}^{\boldsymbol{Y}})$ satisfies*

$$\mathbb{E}\Big[\sup_{t\in[0,1]} \big|\boldsymbol{L}_t^{\boldsymbol{X}} - \boldsymbol{L}_t^{\boldsymbol{Y}}\big|^q\Big] \leqslant \int_{(0,\varepsilon)\times\mathbb{S}^{d-1}} |\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}) - \rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v})|^q \mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v}). \tag{20}$$

In particular, the following inequality holds

$$\begin{aligned}\mathcal{W}_q(\boldsymbol{X},\boldsymbol{Y}) \leqslant |\boldsymbol{\varpi_X} - \boldsymbol{\varpi_Y}|^q &+ \bigg(4\int_{[\varepsilon,\infty)\times\mathbb{S}^{d-1}} (\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}) - \rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v}))^2 \mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v})\bigg)^{q/2} \\ &+ 2^q d^{q/2}|\boldsymbol{\Sigma_X} - \boldsymbol{\Sigma_Y}|^q + \int_{(0,\varepsilon)\times\mathbb{S}^{d-1}} |\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}) - \rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v})|^q \mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v}).\end{aligned} \tag{21}$$

As was the case for the thinning coupling, we can again note that (20) & (21) hold even without assuming that $\mathbb{1}_{\mathbb{R}^d\setminus B_{\mathbf{0}}(1)}(\boldsymbol{w})\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})$ and $\mathbb{1}_{\mathbb{R}^d\setminus B_{\mathbf{0}}(1)}(\boldsymbol{w})\nu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{w})$ have a finite q-moment. However, under such an assumption, the upper bounds are finite since the integral on the right of (20) is bounded by $\int_{\mathbb{R}^d_{\mathbf{0}}}\mathbb{1}_{\mathbb{R}^d\setminus B_{\mathbf{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^q\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})+\int_{\mathbb{R}^d_{\mathbf{0}}}\mathbb{1}_{\mathbb{R}^d\setminus B_{\mathbf{0}}(\kappa)}(\boldsymbol{w})|\boldsymbol{w}|^q\nu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{w})<\infty$ for some $\kappa>0$.

Proof. For $\kappa\in(0,\varepsilon)$, we denote by $\boldsymbol{L}^{\boldsymbol{X}}_{t,\kappa}$ and $\boldsymbol{L}^{\boldsymbol{Y}}_{t,\kappa}$ the truncated large jumps of $\boldsymbol{X}$ and $\boldsymbol{Y}$, given by

$$\boldsymbol{L}^{\boldsymbol{X}}_{t,\kappa}\coloneqq\int_{[0,t]\times(\kappa,\varepsilon)\times\mathbb{S}^{d-1}}\boldsymbol{v}\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})\Xi(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v}),\ \text{ and }\ \boldsymbol{L}^{\boldsymbol{Y}}_{t,\kappa}\coloneqq\int_{[0,t]\times(\kappa,\varepsilon)\times\mathbb{S}^{d-1}}\boldsymbol{v}\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v})\Xi(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v}).$$

Note that $\boldsymbol{L}^{\boldsymbol{X}}_{t,\kappa}-\boldsymbol{L}^{\boldsymbol{Y}}_{t,\kappa}=\int_{[0,t]\times(\kappa,\varepsilon)\times\mathbb{S}^{d-1}}\boldsymbol{v}(\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v}))\Xi(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v})$, and thus, from the concavity of $x\mapsto x^q$ for $x>0$, it follows that

$$\sup_{t\in[0,1]}\left|\boldsymbol{L}^{\boldsymbol{X}}_{t,\kappa}-\boldsymbol{L}^{\boldsymbol{Y}}_{t,\kappa}\right|^q\leqslant\int_{[0,1]\times(\kappa,\varepsilon)\times\mathbb{S}^{d-1}}|\boldsymbol{v}|^q|\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v})|^q\Xi(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v}).$$

Since $\boldsymbol{v}\in\mathbb{S}^{d-1}$ we have that $|\boldsymbol{v}|^q=1$, and Campbell's theorem [31, p. 28] then implies that

$$\mathbb{E}\left[\int_{[0,1]\times(\kappa,\varepsilon)\times\mathbb{S}^{d-1}}|\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v})|^q\Xi(\mathrm{d}s,\mathrm{d}x,\mathrm{d}\boldsymbol{v})\right]
=\int_{(\kappa,\varepsilon)\times\mathbb{S}^{d-1}}|\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v})|^q\mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v}).$$

Thus, altogether, this implies that

$$\mathbb{E}\left[\sup_{t\in[0,1]}\left|\boldsymbol{L}^{\boldsymbol{X}}_{t,\kappa}-\boldsymbol{L}^{\boldsymbol{Y}}_{t,\kappa}\right|^q\right]\leqslant\int_{(\kappa,\varepsilon)\times\mathbb{S}^{d-1}}|\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v})|^q\mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v}).$$

Due to the monotone convergence theorem, it follows, as $\kappa\downarrow0$, that

$$\int_{(\kappa,\varepsilon)\times\mathbb{S}^{d-1}}|\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v})|^q\mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v})\to\int_{(0,\varepsilon)\times\mathbb{S}^{d-1}}|\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v})|^q\mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v}).$$

Furthermore, Fatou's lemma together with the above observations imply that

$$\begin{aligned}\mathbb{E}\left[\liminf_{\kappa\downarrow0}\sup_{t\in[0,1]}\left|\boldsymbol{L}^{\boldsymbol{X}}_{t,\kappa}-\boldsymbol{L}^{\boldsymbol{Y}}_{t,\kappa}\right|^q\right]&\leqslant\liminf_{\kappa\downarrow0}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\boldsymbol{L}^{\boldsymbol{X}}_{t,\kappa}-\boldsymbol{L}^{\boldsymbol{Y}}_{t,\kappa}\right|^q\right]\\&\leqslant\int_{(0,\varepsilon)\times\mathbb{S}^{d-1}}|\rho^{\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Y}}(x,\boldsymbol{v})|^q\mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v}).\end{aligned}$$

We can now conclude (20), as $\liminf_{\kappa\downarrow0}\sup_{t\in[0,1]}\left|\boldsymbol{L}^{\boldsymbol{X}}_{t,\kappa}-\boldsymbol{L}^{\boldsymbol{Y}}_{t,\kappa}\right|^q=\sup_{t\in[0,1]}\left|\boldsymbol{L}^{\boldsymbol{X}}_{t}-\boldsymbol{L}^{\boldsymbol{Y}}_{t}\right|^q$ a.s., since the largest jumps of $\boldsymbol{L}^{\boldsymbol{X}}_{t,\kappa}$ and $\boldsymbol{L}^{\boldsymbol{Y}}_{t,\kappa}$ are finite on the time interval $[0,1]$.

Inequality (21) then follows from (16), (20) and the elementary arguments in Appendix C. □

5. Upper bounds on the Wasserstein distance in the domain of attraction

The main aim of this section is to prove the upper bounds in Theorems 2.1, 2.3 & 2.8 above. In Section 5.1 we give the characterisation, in terms of their generating triplets, of the Lévy processes in $\mathbb{R}^d$ that are in the stable domain-of-attraction. The proof of the upper bounds in Theorem 2.1, based on the thinning coupling, is given in Section 5.2. The upper bounds in Theorem 2.3 are established in Section 5.3 using the comonotonic coupling. In Section 5.4, we prove the upper bounds of Theorem 2.8 for the Brownian limit. In the proofs, we will rely on the following consequence of Jensen's inequality

$$\mathcal{W}_q(\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}})\leqslant\mathcal{W}_{q'}(\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}})^{\frac{q\wedge1}{q'\wedge1}}\qquad\text{for any }0<q<q'.\tag{22}$$

5.1. **Small-time domain of attraction for Lévy processes.** We start by defining the attractor.

Definition. *For any $\alpha \in (0,2]$, the law of an α-stable Lévy process $\boldsymbol{Z}$ is given by a generating triplet $(\boldsymbol{\gamma_Z}, \boldsymbol{\Sigma_Z}\boldsymbol{\Sigma_Z}^\top, \nu_{\boldsymbol{Z}})$ (for the cutoff function $\boldsymbol{w} \mapsto \mathbb{1}_{B_{\boldsymbol{0}}(1)}(\boldsymbol{w})$) as follows: the Lévy measure equals*

$$\nu_{\boldsymbol{Z}}(A) \coloneqq c_\alpha \int_0^\infty \int_{\mathbb{S}^{d-1}} \mathbb{1}_A(r\boldsymbol{v})\sigma(\mathrm{d}\boldsymbol{v})r^{-\alpha-1}\mathrm{d}r, \quad A \in \mathcal{B}(\mathbb{R}^d_{\boldsymbol{0}}), \tag{23}$$

where σ is a probability measure on $\mathcal{B}(\mathbb{S}^{d-1})$ and $c_\alpha \in [0,\infty)$ an "intensity" parameter, satisfying

- *$\alpha = 2$ [Brownian motion with zero drift]: $\boldsymbol{\Sigma_Z} \neq \boldsymbol{0}$, $\boldsymbol{\gamma_Z} = \boldsymbol{0}$ and $c_\alpha = 0$ (i.e. $\nu_{\boldsymbol{Z}} \equiv 0$);*
- *$\alpha \in (1,2)$ [infinite variation, zero-mean process]: $c_\alpha > 0$, $\boldsymbol{\gamma_Z} = -\int_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(1)} \boldsymbol{x}\nu_{\boldsymbol{Z}}(\mathrm{d}\boldsymbol{x})$ and $\boldsymbol{\Sigma_Z} = \boldsymbol{0}$;*
- *$\alpha = 1$ [Cauchy process]: either $c_\alpha > 0$, with symmetric angular component $\int_{\mathbb{S}^{d-1}} \boldsymbol{v}\sigma(\mathrm{d}\boldsymbol{v}) = \boldsymbol{0}$, or $c_\alpha = 0$ and the process $\boldsymbol{Z}$ is a deterministic nonzero linear drift, i.e. $\boldsymbol{Z}_t = \boldsymbol{\gamma_Z} t$ for all times t;*
- *$\alpha \in (0,1)$ [finite variation and zero natural drift]: $c_\alpha > 0$ and $\boldsymbol{\gamma_Z} = \int_{B_{\boldsymbol{0}}(1)\setminus\{\boldsymbol{0}\}} \boldsymbol{x}\nu_{\boldsymbol{Z}}(\mathrm{d}\boldsymbol{x})$.*

It follows from the definition that an α-stable process $\boldsymbol{Z}$ satisfies the scaling property $(\boldsymbol{Z}_{st})_{s\in[0,1]} \overset{d}{=} (t^{1/\alpha}\boldsymbol{Z}_s)_{s\in[0,1]}$ for $t > 0$. Moreover, for $\alpha \in [1,2)$ (resp. $\alpha \in (0,1)$), a non-deterministic α-stable process $\boldsymbol{Z}$ is of infinite (resp. finite) variation by [40, Thm 21.9], since (23) implies $\int_{B_{\boldsymbol{0}}(1)\setminus\{\boldsymbol{0}\}} |\boldsymbol{x}|\nu_{\boldsymbol{Z}}(\mathrm{d}\boldsymbol{x}) = \infty$ (resp. $\int_{B_{\boldsymbol{0}}(1)\setminus\{\boldsymbol{0}\}} |\boldsymbol{x}|\nu_{\boldsymbol{Z}}(\mathrm{d}\boldsymbol{x}) < \infty$). Note also that in the case of the Cauchy process (stability index $\alpha = 1$), $\boldsymbol{\gamma_Z}$ can be arbitrary if $c_\alpha > 0$ and satisfies $\boldsymbol{\gamma_Z} \in \mathbb{R}^d_{\boldsymbol{0}}$ if $c_\alpha = 0$.

For any $\boldsymbol{a} \in \mathbb{S}^{d-1}$, define $\mathscr{L}_{\boldsymbol{a}}(r) \coloneqq \{\boldsymbol{x} \in \mathbb{R}^d : \langle \boldsymbol{a}, \boldsymbol{x}\rangle \geqslant r\}$ for any $r > 0$. The following known result characterises the Lévy processes in the domain of attraction of an α-stable process defined above. It is a consequence of [29, Thm 15.14] and [26, Thm 2], see Appendix B below for the proof.

Theorem 5.1 (Small-time domains of attraction)**.** *Let $\boldsymbol{X} = (\boldsymbol{X}_t)_{t\in[0,1]}$ and $\boldsymbol{Z} = (\boldsymbol{Z}_t)_{t\in[0,1]}$ be Lévy processes in $\mathbb{R}^d$. Then $(\boldsymbol{X}_{st}/g(t))_{s\in[0,1]} \xrightarrow{d} (\boldsymbol{Z}_s)_{s\in[0,1]}$ as $t \downarrow 0$ in the Skorokhod space for some positive normalising function $g : (0,1] \to (0,\infty)$ if and only if $\boldsymbol{Z}$ is α-stable for some $\alpha \in (0,2]$, the normalising function admits the representation $g(t) = t^{1/\alpha}G(t^{-1})$, where G is a slowly varying function at infinity, and the generating triplets $(\boldsymbol{\gamma_X}, \boldsymbol{\Sigma_X}\boldsymbol{\Sigma_X}^\top, \nu_{\boldsymbol{X}})$ and $(\boldsymbol{\gamma_Z}, \boldsymbol{\Sigma_Z}\boldsymbol{\Sigma_Z}^\top, \nu_{\boldsymbol{Z}})$ (for the cutoff function $\boldsymbol{w} \mapsto \mathbb{1}_{B_{\boldsymbol{0}}(1)}(\boldsymbol{w})$) of $\boldsymbol{X}$ and $\boldsymbol{Z}$, respectively, are related as follows:*

- *if $\alpha = 2$ (attraction to Brownian motion), then*

$$G(t^{-1})^{-2}\left(\boldsymbol{\Sigma_X}\boldsymbol{\Sigma_X}^\top + \int_{B_{\boldsymbol{0}}(g(t))\setminus\{\boldsymbol{0}\}} \boldsymbol{x}\boldsymbol{x}^\top \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})\right) \to \boldsymbol{\Sigma_Z}\boldsymbol{\Sigma_Z}^\top, \quad \text{as } t \downarrow 0; \tag{24}$$

- *if $\alpha \in (1,2)$, we have $\boldsymbol{\Sigma_X} = \boldsymbol{0}$ and*

$$t\nu_{\boldsymbol{X}}(\mathscr{L}_{\boldsymbol{v}}(g(t))) \to \nu_{\boldsymbol{Z}}(\mathscr{L}_{\boldsymbol{v}}(1)), \quad \text{as } t \downarrow 0, \quad \text{for any } \boldsymbol{v} \in \mathbb{S}^{d-1}; \tag{25}$$

- *if $\alpha = 1$ (attraction to Cauchy process), then (25) holds,*

$$G(t^{-1})^{-1}\left(\boldsymbol{\gamma_X} - \int_{B_{\boldsymbol{0}}(1)\setminus B_{\boldsymbol{0}}(g(t))} \boldsymbol{x}\nu(\mathrm{d}\boldsymbol{x})\right) \to \boldsymbol{\gamma_Z}, \quad \text{as } t \downarrow 0, \tag{26}$$

  and, for any $\boldsymbol{v} \in \mathbb{S}^{d-1}$, such that $\langle \boldsymbol{v}, \boldsymbol{X}\rangle$ has finite variation (i.e. $\int_{B_{\boldsymbol{0}}(1)\setminus\{\boldsymbol{0}\}} |\langle \boldsymbol{v}, \boldsymbol{x}\rangle|\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x}) < \infty$) and $\nu_{\boldsymbol{Z}}(\mathscr{L}_{\boldsymbol{v}}(1)) > 0$, the process $\langle \boldsymbol{v}, \boldsymbol{X}\rangle$ has zero natural drift: $\langle \boldsymbol{v}, \boldsymbol{\gamma_X}\rangle = \int_{B_{\boldsymbol{0}}(1)\setminus\{\boldsymbol{0}\}} \langle \boldsymbol{v}, \boldsymbol{x}\rangle \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})$.
- *if $\alpha \in (0,1)$, then (25) holds, $\boldsymbol{X}$ has finite variation (i.e. $\int_{B_{\boldsymbol{0}}(1)\setminus\{\boldsymbol{0}\}} |\boldsymbol{x}|\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x}) < \infty$) and zero natural drift (i.e. $\boldsymbol{\gamma_X} = \int_{B_{\boldsymbol{0}}(1)\setminus\{\boldsymbol{0}\}} \boldsymbol{x}\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})$).*

Moreover, the function g satisfying the weak limit above is asymptotically unique at 0: a positive function $\widetilde{g}$ satisfies $(\boldsymbol{X}_{st}/\widetilde{g}(t))_{s\in[0,1]} \xrightarrow{d} (\boldsymbol{Z}_s)_{s\in[0,1]}$ as $t \downarrow 0$ if and only if $\widetilde{g}(t)/g(t) \to 1$ as $t \downarrow 0$.

Note that in the case $\alpha = 2$ in Theorem 5.1, we may have $\boldsymbol{\Sigma_X} = \boldsymbol{0}$ (see Example 6.7 below), but in this case the function G cannot be asymptotically equal to a positive constant. Moreover, in the case $\alpha \in (1,2)$, the process $\boldsymbol{X}$ does not require centering since its mean is linear in time and thus disappears in the scaling limit. However, in the finite variation case (i.e. when $\alpha \in (0,1)$), the process $\boldsymbol{X}$ must have zero natural drift for the scaling limit to exist.

5.2. **Domain of normal attraction: the thinning coupling.** Let $(\boldsymbol{\gamma}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{X}}^\top, \nu_{\boldsymbol{X}})$ denote the generating triplet [40, Def. 8.2] of $\boldsymbol{X}$ with respect to the cutoff function $\boldsymbol{w} \mapsto \mathbb{1}_{B_{\mathbf{0}}(1)}(\boldsymbol{w})$ on $\boldsymbol{w} \in \mathbb{R}^d$. Define the Blumenthal–Getoor (BG) index β of $\boldsymbol{X}$ by

$$\beta := \inf\{p > 0 : I_p < \infty\} \in [0,2], \qquad I_p := \int_{B_{\mathbf{0}}(1)\setminus\{\mathbf{0}\}} |\boldsymbol{w}|^p \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}). \tag{27}$$

Fix $\beta_+ \in [\beta, 2]$ as follows: $\beta_+ := \beta$ if $I_\beta < \infty$; if $I_\beta = \infty$ and $\beta < 1$, then pick $\beta_+ \in (\beta, 1)$; if $I_\beta = \infty$ and $\beta \geqslant 1$, then $\beta < 2$ and hence choose $\beta_+ \in (\beta, 2)$. In particular, note that $I_{\beta_+} < \infty$ and $\beta_+ > 0$. Furthermore, if $\int_{B_{\mathbf{0}}(1)\setminus\{\mathbf{0}\}} |\boldsymbol{w}|\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) < \infty$ (or, equivalently, if the pure-jump component of the Lévy–Itô decomposition (3) of $\boldsymbol{X}$ is finite variation), we say that $\boldsymbol{X}$ has zero natural drift if $\boldsymbol{\gamma}_{\boldsymbol{X}} = \int_{B_{\mathbf{0}}(1)\setminus\{\mathbf{0}\}} \boldsymbol{w}\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})$ and nonzero natural drift otherwise. Moreover, if $\nu_{\boldsymbol{X}}(B_{\mathbf{0}}(1) \setminus \{\mathbf{0}\}) < \infty$ (or, equivalently, the pure-jump component of $\boldsymbol{X}$ is of finite activity, i.e. a compound Poisson process), then $\beta = \beta_+ = 0$. If $\beta_+ = 0$, throughout the paper we use the convention $1/\beta_+ := \infty$.

The following lemma gives an upper bound on the moments of the supremum of the norm of a general Lévy process. Lemma 5.2 plays an important role in the proofs of Section 5.

Lemma 5.2. *Let $\boldsymbol{X}$ be a Lévy process with generating triplet $(\boldsymbol{\gamma}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{X}}^\top, \nu_{\boldsymbol{X}})$. Recall the Blumenthal–Getoor index β from (27) and the associated quantity $\beta_+ \in [\beta, 2]$. Assume that, for some $p > 0$, we have $\int_{\mathbb{R}^d\setminus B_{\mathbf{0}}(1)} |\boldsymbol{w}|^p \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) < \infty$. Then there exist constants $C_i \in [0,\infty)$, $i = 1,\dots,4$, such that*

$$\mathbb{E}\bigg[\sup_{s\in[0,t]} |\boldsymbol{X}_s|^p\bigg] \leqslant \mathbb{1}_{\{\boldsymbol{\Sigma}_{\boldsymbol{X}}\neq\mathbf{0}\}} C_1 t^{p/2} + C_2 t^p + C_3 t^{\min\{1,p/\beta_+\}}, \qquad \textit{for } t \in [0,1].$$

If $\int_{B_{\mathbf{0}}(1)\setminus\{\mathbf{0}\}} |\boldsymbol{w}|\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) < \infty$ and $\boldsymbol{X}$ has zero natural drift, i.e. $\boldsymbol{\gamma}_{\boldsymbol{X}} = \int_{B_{\mathbf{0}}(1)\setminus\{\mathbf{0}\}} \boldsymbol{w}\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})$, then $C_2 = 0$ in the inequality above.

Note that, by the definition of β_+ above, the pure-jump component of $\boldsymbol{X}$ is a compound Poisson process if and only if $\beta_+ = 0$. In particular, if in addition in this case we have zero natural drift, then the pure-jump component of $\boldsymbol{X}$ is a compound Poisson process. The term $t^{p/2}$ in the bound of Lemma 5.2 is present only if $\boldsymbol{X}$ has a non-trivial Gaussian component.

Lemma 5.2 is a multidimensional generalisation of [23, Lem. 2]. The proof of Lemma 5.2, given in Appendix A below, is likewise a multidimensional generalisation of the arguments in the proof of [23, Lem. 2]. As in [23, Lem. 2], the constants C_i, $i = 1,\dots,4$, can be given explicitly in terms of the characteristic triplet of $\boldsymbol{X}$.

Consider a Lévy process $\boldsymbol{X}$ in $\mathbb{R}^d$ in the domain of normal attraction of the α-stable process $\boldsymbol{Z}$. Thus we may assume that $\boldsymbol{X}^t = (\boldsymbol{X}_{st}/t^{1/\alpha})_{s\in[0,1]}$ converges weakly to $\boldsymbol{Z}$ as $t \downarrow 0$. We will now apply the thinning coupling, described in (5) and (6) of Subsection 4.1 above, to quantify this convergence in terms of the Wasserstein distance under the following assumption.

Assumption (T)**.** *Let the Lévy process $\boldsymbol{X}$ be in the small-time domain of attraction of a stable process $\boldsymbol{Z}$. Assume $\boldsymbol{X}$ has no Gaussian component (i.e. $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \mathbf{0}$) and its Lévy measure has a decomposition $\nu_{\boldsymbol{X}} = \nu_{\boldsymbol{X}}^{\mathrm{c}} + \nu_{\boldsymbol{X}}^{\mathrm{d}}$ satisfying the following: $\nu_{\boldsymbol{X}}^{\mathrm{d}}$ is arbitrary with finite mass $\nu_{\boldsymbol{X}}^{\mathrm{d}}(\mathbb{R}^d_{\mathbf{0}}) < \infty$ and*

$$\nu_{\boldsymbol{X}}^{\mathrm{c}}(\mathrm{d}\boldsymbol{w}) = c^{-1} f_{\boldsymbol{S}}(\boldsymbol{w})\nu_{\boldsymbol{Z}}(\mathrm{d}\boldsymbol{w}) \quad \& \quad |f_{\boldsymbol{S}}(\boldsymbol{w}) - c| \leqslant K_T(1 \wedge |\boldsymbol{w}|^p), \quad \textit{for all} \quad \boldsymbol{w} \in \mathbb{R}^d_{\mathbf{0}},$$

a measurable function $f_{\boldsymbol{S}} : \mathbb{R}^d_{\mathbf{0}} \to [0,1]$ and constants $K_T \in [0,\infty)$, $p \in (0,\infty)$ and $c \in (0,1]$.

Remark 5.3. (a) Condition (25) in Theorem 5.1 suggests that the Lévy measure of the process $\boldsymbol{X}$, which is in the domain of attraction of a stable process $\boldsymbol{Z}$, possesses a decomposition of the type $\nu_{\boldsymbol{X}} = \nu_{\boldsymbol{X}}^{\mathrm{c}} + \nu_{\boldsymbol{X}}^{\mathrm{d}}$. Since Assumption (T) stipulates the regularity of the density $f_{\boldsymbol{S}}$ of $\nu_{\boldsymbol{X}}^{\mathrm{c}}$ with respect to $\nu_{\boldsymbol{Z}}$, it may be interpreted as specifying the rate of convergence in the limit in (25) of Theorem 5.1.
(b) Assumption (T) implies $\int_{\mathbb{R}^d\setminus B_{\mathbf{0}}(1)} |\boldsymbol{w}|^q \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \leqslant (1+K_T c^{-1}) \int_{\mathbb{R}^d\setminus B_{\mathbf{0}}(1)} |\boldsymbol{w}|^q \nu_{\boldsymbol{Z}}(\mathrm{d}\boldsymbol{w}) < \infty$ for all $q \in (0,\alpha)$. Hence, by [40, Thm 25.3], the component $\boldsymbol{S}$ of $\boldsymbol{X}$ with Lévy measure $\nu_{\boldsymbol{X}}^{\mathrm{c}}$ has as many moments as the limit $\boldsymbol{Z}$. Note that this is not a restriction on $\boldsymbol{X}$ since $\nu_{\boldsymbol{X}}^{\mathrm{d}}$ may contain all the mass of $\nu_{\boldsymbol{X}}$ outside of some neighborhood of $\mathbf{0}$. ♢

Remark 5.4. Under Assumption (T), we may decompose the process $\boldsymbol{X}$ as the sum $\boldsymbol{S}+\boldsymbol{R}$ of independent Lévy processes $\boldsymbol{S}$ and $\boldsymbol{R}$ with generating triplets $(\boldsymbol{\gamma_S}, \mathbf{0}, \nu_{\boldsymbol{X}}^{\mathrm{c}})$ and $(\boldsymbol{\gamma_R}, \mathbf{0}, \nu_{\boldsymbol{X}}^{\mathrm{d}})$, respectively, such that, when $\alpha \in (0,1)$, both processes have zero natural drift (note that for $\alpha \in (0,1)$, Assumption (T) and Theorem 5.1 imply that $\boldsymbol{X}$ has zero natural drift), and when $\alpha \in (1,2)$ then $\boldsymbol{R}$ has zero natural drift. For $t \in (0,1]$, let $\boldsymbol{S}^t = (\boldsymbol{S}_{st}/t^{1/\alpha})_{s\in[0,1]}$ and $\boldsymbol{R}^t = (\boldsymbol{R}_{st}/t^{1/\alpha})_{s\in[0,1]}$ and note that $\boldsymbol{X}^t$ has the same law as $\boldsymbol{S}^t + \boldsymbol{R}^t$. We couple $\boldsymbol{S}^t$ and $\boldsymbol{Z}$ via the coupling $(\boldsymbol{D}^{\boldsymbol{S}^t,\kappa}, \boldsymbol{D}^{\boldsymbol{Z},\kappa}, \boldsymbol{J}^{\boldsymbol{S}^t,\kappa}, \boldsymbol{J}^{\boldsymbol{Z},\kappa})$ given in (5) and (6) of Subsection 4.1. ♢

Theorem 5.5. *Let $\alpha \in (0,2)\setminus\{1\}$ and Assumption (T) hold for some $p>0$. Then, for any $q \in (0,1]$ with $\int_{\mathbb{R}^d\setminus B_{\mathbf{0}}(1)} |\boldsymbol{w}|^q \nu_{\boldsymbol{X}}^{\mathrm{d}}(\mathrm{d}\boldsymbol{w}) < \infty$, we have $\mathbb{E}\big[\sup_{s\in[0,1]} |\boldsymbol{R}_{st}/t^{1/\alpha}|^q\big] = \mathcal{O}\big(t^{1-q/\alpha}\big)$ as $t \downarrow 0$. Moreover, for any $q \in (0,\alpha)\cap(0,1]$, we let $\kappa(t) \coloneqq t^r$ for $t\in(0,1]$ and some $r \geqslant -1/\alpha$. Then, as $t\downarrow 0$, we have*

$$\mathcal{W}_q\big(\boldsymbol{J}^{\boldsymbol{S}^t,\kappa(t)}, \boldsymbol{J}^{\boldsymbol{Z},\kappa(t)}\big) = \begin{cases} \mathcal{O}\big(t^{1-q/\alpha}\big(1+\log(1/t)\mathbb{1}_{\{p+q=\alpha,\, r\neq -1/\alpha\}}\big)\big), & \text{for } p+q \geqslant \alpha, \\ \mathcal{O}\big(t^{p/\alpha + r(p+q-\alpha)}\big), & \text{for } p+q<\alpha. \end{cases} \tag{28}$$

$$\mathcal{W}_2\big(\boldsymbol{D}^{\boldsymbol{S}^t,\kappa(t)}, \boldsymbol{D}^{\boldsymbol{Z},\kappa(t)}\big) = \mathcal{O}\big(t^{p/\alpha + r(p-\alpha+2)}\big) \tag{29}$$

$$|\boldsymbol{\gamma}_{\boldsymbol{S}^t,\kappa(t)} - \boldsymbol{\gamma}_{\boldsymbol{Z},\kappa(t)}| = \begin{cases} \mathcal{O}\big(t^{1-1/\alpha}(1+\mathbb{1}_{\{p+1=\alpha,\, r\neq -1/\alpha\}}\log(1/t))\big), & p \geqslant \alpha-1>0, \\ \mathcal{O}(t^{p/\alpha + r(p-\alpha+1)}), & p<\alpha-1 \text{ or } \alpha\in(0,1). \end{cases} \tag{30}$$

Remark 5.6. By (4) and (22), we have

$$\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{Z}) \leqslant \mathbb{E}\Big[\sup_{s\in[0,1]} |\boldsymbol{R}_s^t|^q\Big] + \mathcal{W}_q\big(\boldsymbol{J}^{\boldsymbol{S}^t,\kappa(t)}, \boldsymbol{J}^{\boldsymbol{Z},\kappa(t)}\big) + \mathcal{W}_2\big(\boldsymbol{D}^{\boldsymbol{S}^t,\kappa(t)}, \boldsymbol{D}^{\boldsymbol{Z},\kappa(t)}\big)^q + |\boldsymbol{\gamma}_{\boldsymbol{S}^t,\kappa(t)} - \boldsymbol{\gamma}_{\boldsymbol{Z},\kappa(t)}|^q.$$

A careful case-by-case analysis reveals that the upper bound implied by Theorem 5.5 on the distance above (which decreases as fast as the slowest of the terms on the right) decreases the fastest when r is chosen as follows (recall $\alpha \in (0,2)\setminus\{1\}$):

$$r = \begin{cases} 0, & \alpha>1, \\ \dfrac{p}{\alpha(\alpha-p)}, & \alpha<1,\ \alpha>p+q, \\ \dfrac{\alpha-q(p+1)}{\alpha q(p+1-\alpha)}, & \alpha<1,\ \alpha\leqslant p+q. \end{cases}$$

Moreover, in that case, we have

$$\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{Z}) = \begin{cases} \mathcal{O}\big(t^{\min\{\alpha/q-1,\, p,\, \alpha-1\}q/\alpha}\big(1+|\log t|\mathbb{1}_{\{p+q=\alpha\}} + |\log t|^q\mathbb{1}_{\{p+1=\alpha\}}\big)\big), & \alpha>1, \\ \mathcal{O}\big(t^{\min\{1-q/\alpha,\, pq/(\alpha(\alpha-p))\}}\big), & \alpha<1,\ \alpha>p+q, \\ \mathcal{O}\big(t^{1-q/\alpha}\big(1+|\log t|\mathbb{1}_{\{p+q=\alpha\}}\big)\big), & \alpha<1,\ \alpha\leqslant p+q. \end{cases}$$

Since the above bounds are not easily interpretable because of the multiple cases depending on the parameters (α,p,q), we decided to only present the case $p=1$ in Theorem 2.1 above. In particular, this removes the possibility of a logarithmic term appearing in the upper bound. ♢

Proof of Theorem 5.5. The bound on $\boldsymbol{R}^t$ follows directly from Lemma 5.2 with $\beta_+ = 0$ and the construction of $\boldsymbol{R}^t$.

We now consider the process $\boldsymbol{S}^t$. Define the measure

$$\mu(A) := c^{-1}\nu_{\boldsymbol{Z}}(A) = \frac{c_\alpha}{c}\int_{\mathbb{S}^{d-1}}\int_0^\infty \mathbb{1}_A(x\boldsymbol{v})\frac{\mathrm{d}x}{x^{\alpha+1}}\sigma(\mathrm{d}\boldsymbol{v}), \qquad \text{for } A \in \mathcal{B}(\mathbb{R}^d_{\boldsymbol{0}}),$$

where c (resp. c_α) is in Assumption (T) (resp. in (23) of the definition of $\boldsymbol{Z}$). The Radon–Nikodym derivative $f_{\boldsymbol{S}^t}(\boldsymbol{w}) := \nu_{\boldsymbol{S}^t}(\mathrm{d}\boldsymbol{w})/\mu(\mathrm{d}\boldsymbol{w})$ equals $f_{\boldsymbol{S}}(t^{1/\alpha}\boldsymbol{w})$ on the support of μ, since Lévy–Khintchine exponent satisfies $t\psi_{\boldsymbol{S}}(\boldsymbol{u}/t^{1/\alpha}) = \psi_{\boldsymbol{S}^t}(\boldsymbol{u})$ and hence

$$\nu_{\boldsymbol{S}^t}(\mathrm{d}\boldsymbol{w}) = t^{1+d/\alpha}\nu^{\mathsf{c}}_{\boldsymbol{X}}(\mathrm{d}(t^{1/\alpha}\boldsymbol{w})) = t^{1+d/\alpha}f_{\boldsymbol{S}}(t^{1/\alpha}\boldsymbol{w})\mu(\mathrm{d}(t^{1/\alpha}\boldsymbol{w})) = f_{\boldsymbol{S}}(t^{1/\alpha}\boldsymbol{w})\mu(\mathrm{d}\boldsymbol{w}).$$

First we bound the large-jump component $\boldsymbol{J}^{\boldsymbol{S}^t,\kappa(t)} - \boldsymbol{J}^{\boldsymbol{Z},\kappa(t)}$: inequality (9) of Proposition 4.2 yields

$$\mathbb{E}\Big[\sup_{s\in[0,1]}|\boldsymbol{J}^{\boldsymbol{S}^t,\kappa(t)}_s - \boldsymbol{J}^{\boldsymbol{Z},\kappa(t)}_s|^q\Big] \leqslant \frac{c_\alpha}{c}\int_{\mathbb{S}^{d-1}}\int_{\kappa(t)}^\infty |f_{\boldsymbol{S}}(t^{1/\alpha}x\boldsymbol{v}) - c|x^{q-\alpha-1}\mathrm{d}x\sigma(\mathrm{d}\boldsymbol{v}).$$

Recall that $f(x) \lesssim g(x)$ as $x \downarrow 0$ means that there exists some $c_0, x_0 > 0$ such that $f(x) \leqslant c_0 g(x)$ for all $x \leqslant x_0$. Using Assumption (T), as $t \downarrow 0$, we obtain

$$\mathbb{E}\Big[\sup_{s\in[0,1]}\big|\boldsymbol{J}^{\boldsymbol{S}^t,\kappa(t)}_s - \boldsymbol{J}^{\boldsymbol{Z},\kappa(t)}_s\big|^q\Big] \lesssim \int_{t^r}^{t^{-1/\alpha}} t^{p/\alpha}x^{p+q-\alpha-1}\mathrm{d}x + \int_{t^{-1/\alpha}}^\infty x^{q-\alpha-1}\mathrm{d}x,$$

where $\int_{t^{-1/\alpha}}^\infty x^{q-\alpha-1}\mathrm{d}x = \mathcal{O}(t^{1-q/\alpha})$. Next, as $t \downarrow 0$, we note that

$$\int_{t^r}^{t^{-1/\alpha}} t^{p/\alpha}x^{p+q-\alpha-1}\mathrm{d}x = \begin{cases}\mathcal{O}\big(t^{1-q/\alpha}\big), & \text{for } p+q>\alpha,\\ \mathcal{O}\big(t^{1-q/\alpha}\big(1+\log(1/t)\mathbb{1}_{\{r\neq -1/\alpha\}}\big)\big), & \text{for } p+q=\alpha,\\ \mathcal{O}\big(t^{p/\alpha+r(p+q-\alpha)}\big), & \text{for } p+q<\alpha.\end{cases}$$

Thus, since $r \geqslant -1/\alpha$, altogether we have, as $t \downarrow 0$,

$$\mathbb{E}\Big[\sup_{s\in[0,1]}\big|\boldsymbol{J}^{\boldsymbol{S}^t,\kappa(t)}_s - \boldsymbol{J}^{\boldsymbol{Z},\kappa(t)}_s\big|^q\Big] = \begin{cases}\mathcal{O}\big(t^{1-q/\alpha}\big(1+\log(1/t)\mathbb{1}_{\{p+q=\alpha,\, r\neq -1/\alpha\}}\big)\big), & \text{for } p+q\geqslant\alpha,\\ \mathcal{O}\big(t^{p/\alpha+r(p+q-\alpha)}\big), & \text{for } p+q<\alpha.\end{cases}$$

Next, we find the rate for the small-jump component $\boldsymbol{D}^{\boldsymbol{S}^t,\kappa(t)} - \boldsymbol{D}^{\boldsymbol{Z},\kappa(t)}$. Assumption (T) and (7) of Proposition 4.1 imply that, as $t \downarrow 0$,

$$\begin{aligned}\mathbb{E}\Big[\sup_{s\in[0,1]}\big|\boldsymbol{D}^{\boldsymbol{S}^t,\kappa(t)}_s - \boldsymbol{D}^{\boldsymbol{Z},\kappa(t)}_s\big|^2\Big] &\leqslant 4\frac{c_\alpha}{c}\int_{\mathbb{S}^{d-1}}\int_0^{\kappa(t)} |f_{\boldsymbol{S}}(t^{1/\alpha}x\boldsymbol{v}) - c|x^{1-\alpha}\mathrm{d}x\sigma(\mathrm{d}\boldsymbol{v})\\ &\leqslant 4K_T\frac{c_\alpha}{c}\int_0^{\kappa(t)} t^{p/\alpha}x^{p-\alpha+1}\mathrm{d}x = \mathcal{O}\big(t^{p/\alpha+r(p-\alpha+2)}\big).\end{aligned}$$

Next, we control the difference $|\boldsymbol{\gamma}_{\boldsymbol{S}^t,\kappa(t)} - \boldsymbol{\gamma}_{\boldsymbol{Z},\kappa(t)}|$ of the drift terms. First, consider the infinite variation case $\alpha \in (1,2)$. Since $\boldsymbol{Z}$ has zero mean, representation (3) implies

$$\begin{aligned}\boldsymbol{\gamma}_{\boldsymbol{S}^t,\kappa(t)} &= t^{1-1/\alpha}\mathbb{E}[\boldsymbol{S}_1] - \int_{\mathbb{R}^d}\boldsymbol{w}\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa(t))}(\boldsymbol{w})f_{\boldsymbol{S}}(t^{1/\alpha}\boldsymbol{w})\mu(\mathrm{d}\boldsymbol{w}), \qquad \text{and}\\ \boldsymbol{\gamma}_{\boldsymbol{Z},\kappa(t)} &= -\int_{\mathbb{R}^d}\boldsymbol{w}\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa(t))}(\boldsymbol{w})c\mu(\mathrm{d}\boldsymbol{w}).\end{aligned}$$

Thus, we obtain

$$\boldsymbol{\gamma}_{\boldsymbol{S}^t,\kappa(t)} - \boldsymbol{\gamma}_{\boldsymbol{Z},\kappa(t)} = t^{1-1/\alpha}\mathbb{E}[\boldsymbol{S}_1] - \int_{\mathbb{R}^d}\boldsymbol{w}\mathbb{1}_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(\kappa(t))}(\boldsymbol{w})(f_{\boldsymbol{S}}(t^{1/\alpha}\boldsymbol{w}) - c)\mu(\mathrm{d}\boldsymbol{w}). \tag{31}$$

By Assumption (T), the integral in the display satisfies

$$\begin{aligned}&\left|\int_{\mathbb{R}^d}\boldsymbol{w}\mathbb{1}_{\mathbb{R}^d\setminus B_{\mathbf{0}}(\kappa(t))}(\boldsymbol{w})(f_{\boldsymbol{S}}(t^{1/\alpha}\boldsymbol{w})-c)\mu(\mathrm{d}\boldsymbol{w})\right|\\&\leqslant\int_{\mathbb{R}^d}|\boldsymbol{w}|\mathbb{1}_{\mathbb{R}^d\setminus B_{\mathbf{0}}(\kappa(t))}(\boldsymbol{w})|f_{\boldsymbol{S}}(t^{1/\alpha}\boldsymbol{w})-c|\mu(\mathrm{d}\boldsymbol{w})=\frac{c_\alpha}{c}\int_{\mathbb{S}^{d-1}}\int_{\kappa(t)}^\infty x^{-\alpha}|f_{\boldsymbol{S}}(t^{1/\alpha}x\boldsymbol{v})-c|\mathrm{d}x\,\sigma(\mathrm{d}\boldsymbol{v})\\&\leqslant K_T\frac{c_\alpha}{c}\left(t^{p/\alpha}\int_{\kappa(t)}^{t^{-1/\alpha}}x^{p-\alpha}\mathrm{d}x+\int_{t^{-1/\alpha}}^\infty x^{-\alpha}\mathrm{d}x\right)=\begin{cases}\mathcal{O}(t^{1-1/\alpha}), & p>\alpha-1,\\ \mathcal{O}(t^{1-1/\alpha}(1+\log(1/t)\mathbb{1}_{\{r\neq-1/\alpha\}})), & p=\alpha-1,\\ \mathcal{O}(t^{p/\alpha+r(p-\alpha+1)}), & p<\alpha-1,\end{cases}\end{aligned}$$

where we used the fact that $r\geqslant-1/\alpha$. By (31), we obtain

$$|\boldsymbol{\gamma}_{\boldsymbol{S}^t,\kappa(t)}-\boldsymbol{\gamma}_{\boldsymbol{Z},\kappa(t)}|=\begin{cases}\mathcal{O}\big(t^{1-1/\alpha}(1+\mathbb{1}_{\{p+1=\alpha,\,r\neq-1/\alpha\}}\log(1/t))\big), & p\geqslant\alpha-1,\\ \mathcal{O}(t^{p/\alpha+r(p-\alpha+1)}), & p<\alpha-1.\end{cases}$$

In the finite variation case $\alpha\in(0,1)$, recall that $\boldsymbol{S}$ and $\boldsymbol{Z}$ have zero natural drift, so that

$$\boldsymbol{\gamma}_{\boldsymbol{S}^t,\kappa(t)}=\int_{\mathbb{R}^d}\boldsymbol{w}\mathbb{1}_{B_{\mathbf{0}}(\kappa(t))}(\boldsymbol{w})f_{\boldsymbol{S}}(t^{1/\alpha}\boldsymbol{w})\mu(\mathrm{d}\boldsymbol{w}),\quad \boldsymbol{\gamma}_{\boldsymbol{Z},\kappa(t)}=\int_{\mathbb{R}^d}\boldsymbol{w}\mathbb{1}_{B_{\mathbf{0}}(\kappa(t))}(\boldsymbol{w})c\mu(\mathrm{d}\boldsymbol{w}).$$

Thus, we have, by Assumption (T),

$$\begin{aligned}|\boldsymbol{\gamma}_{\boldsymbol{S}^t,\kappa(t)}-\boldsymbol{\gamma}_{\boldsymbol{Z},\kappa(t)}|&\leqslant\int_{\mathbb{R}^d}|\boldsymbol{w}|\mathbb{1}_{B_{\mathbf{0}}(\kappa(t))}(\boldsymbol{w})|f_{\boldsymbol{S}}(t^{1/\alpha}\boldsymbol{w})-c|\mu(\mathrm{d}\boldsymbol{w})\\&\leqslant K_T\frac{c_\alpha}{c}\int_0^{\kappa(t)}t^{p/\alpha}x^{p-\alpha}\mathrm{d}x=\mathcal{O}(t^{p/\alpha+r(p-\alpha+1)}).\end{aligned}$$ □

5.3. **Domain of non-normal attraction: the comonotonic coupling.** Let $\boldsymbol{Z}$ be an α-stable process on $\mathbb{R}^d$ for some $\alpha\in(0,2)$, defined as in Section 5.1, with "intensity" parameter c_α, probability measure σ on $\mathcal{B}(\mathbb{S}^{d-1})$ and Lévy measure $\nu_{\boldsymbol{Z}}$ in (23). Define the measure $\rho_{\boldsymbol{Z}}(\mathrm{d}x,\boldsymbol{v}):=c_\alpha x^{-\alpha-1}\mathrm{d}x$ on $\mathcal{B}((0,\infty))$ and note that the right inverse of its tail $x\mapsto\rho_{\boldsymbol{Z}}([x,\infty),\boldsymbol{v})$ is given by $\rho_{\boldsymbol{Z}}^{\leftarrow}(x,\boldsymbol{v})=(c_\alpha/\alpha)^{1/\alpha}x^{-1/\alpha}$ for all $x>0$ and $\boldsymbol{v}\in\mathbb{S}^{d-1}$. The comonotonic coupling of $\boldsymbol{Z}$ and a Lévy process $\boldsymbol{X}$ requires the following assumption on the generating triplet $(\boldsymbol{\gamma}_{\boldsymbol{X}},\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{X}}^\top,\nu_{\boldsymbol{X}})$ of $\boldsymbol{X}$.

Assumption (C)**.** *$\boldsymbol{X}$ has no Gaussian component $\boldsymbol{\Sigma}_{\boldsymbol{X}}=\mathbf{0}$ and $\nu_{\boldsymbol{X}}=\nu_{\boldsymbol{X}}^{\mathsf{c}}+\nu_{\boldsymbol{X}}^{\mathsf{d}}$, where the measure $\nu_{\boldsymbol{X}}^{\mathsf{d}}$ is arbitrary with finite mass $\nu_{\boldsymbol{X}}^{\mathsf{d}}(\mathbb{R}^d_{\mathbf{0}})<\infty$ and the Lévy measure $\nu_{\boldsymbol{X}}^{\mathsf{c}}$ can be expressed as*

$$\nu_{\boldsymbol{X}}^{\mathsf{c}}(B)=\int_{\mathbb{S}^{d-1}}\int_0^\infty\mathbb{1}_B(x\boldsymbol{v})\rho_{\boldsymbol{X}}^{\mathsf{c}}(\mathrm{d}x,\boldsymbol{v})\sigma(\mathrm{d}\boldsymbol{v})\quad\&\quad\rho_{\boldsymbol{X}}^{\mathsf{c}}([x,\infty),\boldsymbol{v})=\frac{c_\alpha}{\alpha}(1+h(x,\boldsymbol{v}))H(x)^\alpha x^{-\alpha}\tag{32}$$

for all $B\in\mathcal{B}(\mathbb{R}^d_{\mathbf{0}})$, $x>0$, $\boldsymbol{v}\in\mathbb{S}^{d-1}$ and some monotonic function $H:(0,\infty)\to(0,\infty)$, slowly varying at 0, and a measurable $h:(0,\infty)\times\mathbb{S}^{d-1}\to[-1,\infty)$. Assume that the functions H, h and

$$G(x):=\int_{\mathbb{S}^{d-1}}H(\rho_{\boldsymbol{X}}^{\mathsf{c}\leftarrow}(x,\boldsymbol{u}))\sigma(\mathrm{d}\boldsymbol{u}),\qquad x>0,\tag{33}$$

where $\rho_{\boldsymbol{X}}^{\mathsf{c}\leftarrow}(\cdot,\boldsymbol{v})$ is the right inverse of $x\mapsto\rho_{\boldsymbol{X}}^{\mathsf{c}}([x,\infty),\boldsymbol{v})$, satisfy

$$|h(x,\boldsymbol{v})|\leqslant K_h(1\wedge x^p)\quad\&\quad|H(\rho_{\boldsymbol{X}}^{\mathsf{c}\leftarrow}(x,\boldsymbol{v}))/G(x)-1|\leqslant K_Q(1\wedge x^{-\delta})\quad\text{for all } x>0,\boldsymbol{v}\in\mathbb{S}^{d-1}\tag{34}$$

and some constants $p,\delta>0$ and $K_h,K_Q\geqslant0$.

Under Assumption (C), G is monotonic and slowly varying at ∞. In fact, for any $\boldsymbol{v}\in\mathbb{S}^{d-1}$, we have $G(x)\sim H(\rho_{\boldsymbol{X}}^{\mathsf{c}\leftarrow}(x,\boldsymbol{v}))$ as $x\to\infty$ by virtue of (34), which is slowly varying by [6, Prop. 1.5.7(ii)] (since $\rho_{\boldsymbol{X}}^{\mathsf{c}\leftarrow}([x,\infty),\boldsymbol{v})$ is regularly varying and H is slowly varying). Note also that H may be either non-increasing or non-decreasing.

Remark 5.7. Condition (25) in Theorem 5.1 states that the Lévy measure of the process $\boldsymbol{X}$ in the domain of attraction of a stable process $\boldsymbol{Z}$ behaves as the Lévy measure of $\boldsymbol{Z}$ in every half-space of the form $\mathscr{L}_{\boldsymbol{v}}(c)$ for every $\boldsymbol{v} \in \mathbb{S}^{d-1}$ and small $c > 0$. Assumptions (S) and (C) may thus be interpreted as a refinement of this condition, requiring the Lévy measure of $\boldsymbol{X}$ to satisfy an analogue of (25) but on every ray directed by $\boldsymbol{v} \in \mathbb{S}^{d-1}$ and quantifying how fast such a limit holds. In particular, under Assumptions (S) and (C) and for G defined in (33), we let $g(t) := t^{1/\alpha}G(1/t)$ with $t \in (0,1]$. For such a g it follows that $(\boldsymbol{X}_{st}/g(t))_{s\in[0,1]} \xrightarrow{d} (\boldsymbol{Z}_s)_{s\in[0,1]}$, i.e. $\boldsymbol{X}$ is in the small-time domain of non-normal attraction of $\boldsymbol{Z}$ (by Theorem 5.1 above). ◊

Remark 5.8. We decompose the process $\boldsymbol{X}$ as the sum $\boldsymbol{S}+\boldsymbol{R}$ of the independent Lévy processes $\boldsymbol{S}$ and $\boldsymbol{R}$ with generating triplets $(\boldsymbol{\gamma_S}, \boldsymbol{0}, \nu^{\mathsf{c}}_{\boldsymbol{X}})$ and $(\boldsymbol{\gamma_R}, \boldsymbol{0}, \nu^{\mathsf{d}}_{\boldsymbol{X}})$, respectively, such that, when $\alpha \in (0,1)$, both processes have zero natural drift, and when $\alpha \in (1,2)$ then $\boldsymbol{R}$ has zero natural drift. Let the processes $(\boldsymbol{M}^{\boldsymbol{S}^t}, \boldsymbol{M}^{\boldsymbol{Z}}, \boldsymbol{L}^{\boldsymbol{S}^t}, \boldsymbol{L}^{\boldsymbol{Z}})$ be coupled as in (14) and (15) from Subsection 4.2, where $\boldsymbol{S}^t = (\boldsymbol{S}_{st}/g(t))_{s\in[0,1]}$ and $\boldsymbol{R}^t = (\boldsymbol{R}_{st}/g(t))_{s\in[0,1]}$ for $t \in (0,1]$. Note that $\boldsymbol{X}^t$ has the same law as $\boldsymbol{S}^t + \boldsymbol{R}^t$ for $t \in (0,1]$ and that, under Assumption (C), $\boldsymbol{S}$ has a finite q-moment for every $q \in (0,\alpha)$ by (32). ◊

Theorem 5.9. *Let $\alpha \in (0,2)\setminus\{1\}$, $\boldsymbol{X}$ and $\boldsymbol{Z}$ be as above and Assumptions (S) & (C) hold. Then, for every $q \in (0,1]$ with $\int_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(1)} |\boldsymbol{w}|^q \nu^{\mathsf{d}}_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) < \infty$ we have $\mathbb{E}\big[\sup_{s\in[0,1]} |\boldsymbol{R}^t_s|^q\big] = \mathcal{O}\big(t^{1-q/\alpha}G(t)^{-q}\big)$ as $t \downarrow 0$ and, if $p \neq \alpha - 1$ and $q \in (0,\alpha)\cap(0,1]$ satisfy $q \notin \{\alpha/(p+1), \alpha/(\alpha\delta+1)\}$, we have, as $t \downarrow 0$,*

$$\mathcal{W}_2\big(\boldsymbol{M}^{\boldsymbol{S}^t}, \boldsymbol{M}^{\boldsymbol{Z}}\big) = \mathcal{O}\big(G_2(t) + (1 + G(1/t)^p)t^{\min\{p/\alpha,\delta\}}\big), \tag{35}$$

$$\mathcal{W}_q\big(\boldsymbol{L}^{\boldsymbol{S}^t}, \boldsymbol{L}^{\boldsymbol{Z}}\big) = \mathcal{O}\big(G_2(t)^q + (1 + G(1/t)^{pq})(1 + G_1(x))^{q(1+p)}t^{\min\{pq/\alpha,q\delta,1-q/\alpha\}}\big), \quad \textit{and} \tag{36}$$

$$|\boldsymbol{\varpi}_{\boldsymbol{S}^t} - \boldsymbol{\varpi}_{\boldsymbol{Z}}| = \begin{cases} \mathcal{O}\big(G_2(t) + t^{1-1/\alpha}G(1/t)^{-1} + (1+G_1(t))t^{\min\{1-1/\alpha,\delta\}} \\ \qquad + G(1/t)^p(1+G_1(t))^{1+p}t^{\min\{1-1/\alpha,p/\alpha\}}\big), & \alpha \in (1,2), \\ \mathcal{O}\big(G_2(t) + t^{p/\alpha}G(1/t)^p + t^\delta\big), & \alpha \in (0,1). \end{cases} \tag{37}$$

Remark 5.10. If $H \equiv 1$, then G is constant and hence Theorem 5.9 is also applicable to the domain of normal attraction: set $G_1 \equiv 1$, $G_2 \equiv 0$ and δ arbitrarily large, then for $p \neq \alpha - 1$ and $q \in (0,\alpha)\cap(0,1]$ with $\int_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(1)} |\boldsymbol{w}|^q \nu^{\mathsf{d}}_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) < \infty$ and $q \neq \alpha/(p+1)$, we have, as $t \downarrow 0$,

$$\mathcal{W}_2\big(\boldsymbol{M}^{\boldsymbol{S}^t}, \boldsymbol{M}^{\boldsymbol{Z}}\big) = \mathcal{O}\big(t^{p/\alpha}\big), \quad \mathcal{W}_q\big(\boldsymbol{L}^{\boldsymbol{S}^t}, \boldsymbol{L}^{\boldsymbol{Z}}\big) = \mathcal{O}\big(t^{\min\{pq/\alpha,1-q/\alpha\}}\big), \quad \text{and}$$

$$|\boldsymbol{\varpi}_{\boldsymbol{S}^t} - \boldsymbol{\varpi}_{\boldsymbol{Z}}| = \begin{cases} \mathcal{O}\big(t^{\min\{1-1/\alpha,p/\alpha\}}\big), & \alpha \in (1,2), \\ \mathcal{O}(t^{p/\alpha}), & \alpha \in (0,1). \end{cases}$$

Note that, when $q(p+1) > \alpha$, these rates match the ones in Theorem 5.5, established under more general conditions using the thinning coupling. ◊

Lemma 5.11. *Under Assumption (C), there exists a function $\widetilde{h} : (0,\infty)\times\mathbb{S}^{d-1} \to (-1,\infty)$ and a constant $K_{\widetilde{h}} \geqslant 0$, such that, for all $x > 0$ and $\boldsymbol{v} \in \mathbb{S}^{d-1}$,*

$$\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}) = (c_\alpha/c)^{1/\alpha}\, x^{-1/\alpha}G(x)(1+\widetilde{h}(x,\boldsymbol{v})) \quad \textit{and} \quad |\widetilde{h}(x,\boldsymbol{v})| \leqslant K_{\widetilde{h}}\big(1 \wedge (x^{-p/\alpha}G(x)^p + x^{-\delta})\big). \tag{38}$$

Proof. Note that $\rho^{\mathsf{c}}_{\boldsymbol{X}}([\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}),\infty),\boldsymbol{v}) = x$ for all $\boldsymbol{v} \in \mathbb{S}^{d-1}$ and $x > 0$. Hence, for all $\boldsymbol{v} \in \mathbb{S}^{d-1}$ and $x > 0$, $x = \frac{c_\alpha}{\alpha}(1 + h(\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}),\boldsymbol{v}))H(\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}))^\alpha \rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})^{-\alpha}$, implying that

$$\begin{aligned} \rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}) &= \left(\frac{c_\alpha}{\alpha}\right)^{1/\alpha} x^{-1/\alpha}H(\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}))(1 + h(\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}))^{1/\alpha} \\ &= \left(\frac{c_\alpha}{\alpha}\right)^{1/\alpha} x^{-1/\alpha}G(x)\frac{H(\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}))}{G(x)}\big(1 + h(\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}))\big)^{1/\alpha}. \end{aligned}$$

Thus, the first part of (38) holds if $\widetilde{h}(x,\boldsymbol{v}) := (H(\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}))/G(x))(1+h(\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}))^{1/\alpha}-1 \in (-1,\infty)$.

Suppose now that (34) in Assumption (C) holds for some $p,\delta>0$. Since h is bounded by K_h and by (34), we obtain $\widetilde{h}(x,\boldsymbol{v}) \leqslant (K_Q+1)(1+K_h)^{1/\alpha}-1$ for all $x>0$ and $\boldsymbol{v}\in\mathbb{S}^{d-1}$. Moreover, the elementary inequality $|(1+x)^r-1|\leqslant|x|$ for any $r\in[0,1]$ and $x\geqslant-1$ and the triangle inequality yield $|(1+y)(1+x)^r-1|\leqslant|x|+|y|(1+x)^r$, implies for all $x>0$ and $\boldsymbol{v}\in\mathbb{S}^{d-1}$, that

$$\begin{aligned}
|\widetilde{h}(x,\boldsymbol{v})| &\leqslant |h(\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}))| + \left|\frac{H(\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}))}{G(x)}-1\right|(1+K_h)^{1/\alpha}\\
&\leqslant K_h\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v})^p + K_Q(1+K_h)^{1/\alpha}x^{-\delta}\\
&\leqslant K_h(c_\alpha/\alpha)^{p/\alpha}x^{-p/\alpha}G(x)^p\left(\frac{H(\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}))}{G(x)}\right)^p(1+h(\rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x,\boldsymbol{v}))^{p/\alpha} + K_Q(1+K_h)^{1/\alpha}x^{-\delta}\\
&\leqslant K_h(1+K_h)^{p/\alpha}(1+K_Q)^p(c_\alpha/\alpha)^{p/\alpha}x^{-p/\alpha}G(x)^p + K_Q(1+K_h)^{1/\alpha}x^{-\delta}.
\end{aligned}$$

Choosing $K_{\widetilde{h}} := \max\{(c_\alpha/\alpha)^{p/\alpha}K_h(1+K_h)^{p/\alpha}(1+K_Q)^p, (K_Q+1)(1+K_h)^{1/\alpha}, 1\}$, concludes the last part of (38). □

Lemma 5.12. *Let $\alpha\in(0,2)\setminus\{1\}$ and $q\in(0,\alpha)\cap(0,1]$ satisfy $\int_{\mathbb{R}^d\setminus B_{\mathbf{0}}(1)}|\boldsymbol{w}|^q\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})<\infty$, where $q\neq\alpha/(p+1)$ and $q\neq\alpha/(\alpha\delta+1)$. Then, under Assumptions (S) & (C) we have, as $t\downarrow 0$,*

$$\mathcal{W}_q\big(\boldsymbol{L}^{\boldsymbol{S}^t},\boldsymbol{L}^{\boldsymbol{Z}}\big) = \mathcal{O}\big(G_2(t)^q + (1+G_1(t))^q t^{\min\{1-q/\alpha,q\delta\}} + G(1/t)^{pq}(1+G_1(t))^{q(1+p)}t^{\min\{1-q/\alpha,pq/\alpha\}}\big).$$

Proof. By Lemma 5.11, for all $x>0$, $t\in(0,1]$ and $\boldsymbol{v}\in\mathbb{S}^{d-1}$, it holds that

$$\rho^{\leftarrow}_{\boldsymbol{S}^t}(x,\boldsymbol{v}) = \rho^{\mathsf{c}\leftarrow}_{\boldsymbol{X}}(x/t,\boldsymbol{v})/g(t) = \left(\frac{c_\alpha}{\alpha}\right)^{1/\alpha}x^{-1/\alpha}\frac{G(x/t)}{G(1/t)}\big(1+\widetilde{h}(x/t,\boldsymbol{v})\big). \tag{39}$$

Hence, Proposition 4.4 now implies that

$$\begin{aligned}
\mathcal{W}_q\big(\boldsymbol{L}^{\boldsymbol{S}^t},\boldsymbol{L}^{\boldsymbol{Z}}\big) &\leqslant \int_{\mathbb{S}^{d-1}}\int_0^\varepsilon|\rho^{\leftarrow}_{\boldsymbol{S}^t}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Z}}(x,\boldsymbol{v})|^q\mathrm{d}x\sigma(\mathrm{d}\boldsymbol{v})\\
&= \left(\frac{c_\alpha}{\alpha}\right)^{q/\alpha}\int_{\mathbb{S}^{d-1}}\int_0^\varepsilon\left|\frac{G(x/t)}{G(1/t)}\big(1+\widetilde{h}(x/t,\boldsymbol{v})\big)-1\right|^q x^{-q/\alpha}\mathrm{d}x\sigma(\mathrm{d}\boldsymbol{v}) =: I(t).
\end{aligned}$$

To bound $I(t)$, we use the triangle inequality and the fact that $x\mapsto x^q$ is concave, to obtain

$$\left(\frac{\alpha}{c_\alpha}\right)^{q/\alpha}I(t) \leqslant \int_0^\varepsilon x^{-q/\alpha}\left|\frac{G(x/t)}{G(1/t)}-1\right|^q\mathrm{d}x + \int_{\mathbb{S}^{d-1}}\int_0^\varepsilon x^{-q/\alpha}\left|\frac{G(x/t)}{G(1/t)}\widetilde{h}(x/t,\boldsymbol{v})\right|^q\mathrm{d}x\,\sigma(\mathrm{d}\boldsymbol{v}).$$

We consider each integral on its own. Assumption (S) implies that the first integral $I_1(t)$ in the display above is bounded by $G_2(t)^q\int_0^\varepsilon x^{-q/\alpha}G_1(x)^q\mathrm{d}x<\infty$. Next, we bound the second integral $I_2(t)$ in the display above. Assumption (S) and (34) yield, as $t\downarrow 0$,

$$\begin{aligned}
I_2(t) &\lesssim \int_{\mathbb{S}^{d-1}}\int_0^\varepsilon(1+G_1(x)G_2(t))^q x^{-q/\alpha}|\widetilde{h}(x/t,\boldsymbol{v})|^q\mathrm{d}x\,\sigma(\mathrm{d}\boldsymbol{v})\\
&\lesssim \int_0^t(1+G_1(x))^q x^{-q/\alpha}\mathrm{d}x + t^{q\delta}\int_t^\varepsilon(1+G_1(x))^q x^{-q/\alpha-q\delta}\mathrm{d}x\\
&\qquad + t^{pq/\alpha}G(1/t)^{pq}\int_t^\varepsilon(1+G_1(x))^q x^{-q/\alpha-pq/\alpha}\frac{G(x/t)^{pq}}{G(1/t)^{pq}}\mathrm{d}x\\
&\lesssim (1+G_1(t))^q t^{\min\{1-q/\alpha,q\delta\}} + t^{pq/\alpha}G(1/t)^{pq}\int_t^\varepsilon(1+G_1(x))^{q(1+p)}x^{-q/\alpha-pq/\alpha}\mathrm{d}x\\
&= \mathcal{O}\big((1+G_1(t))^q t^{\min\{1-q/\alpha,q\delta\}} + G(1/t)^{pq}(1+G_1(t))^{q(1+p)}t^{\min\{1-q/\alpha,pq/\alpha\}}\big).
\end{aligned}$$ □

Lemma 5.13. *Let $\alpha\in(0,2)\setminus\{1\}$. Under Assumptions (S) & (C) we have*

$$\mathcal{W}_2\big(\boldsymbol{M}^{\boldsymbol{S}^t},\boldsymbol{M}^{\boldsymbol{Z}}\big) = \mathcal{O}\big(G_2(t)+t^{p/\alpha}G(1/t)^p+t^\delta\big), \qquad \text{as } t\downarrow 0. \tag{40}$$

Proof. Proposition 4.3 together with (39) shows that

$$\begin{aligned}\mathcal{W}_2\big(\boldsymbol{M}^{\boldsymbol{S}^t},\boldsymbol{M}^{\boldsymbol{Z}}\big)^2 &\leqslant 4\int_{[\varepsilon,\infty)\times\mathbb{S}^{d-1}}(\rho^{\leftarrow}_{\boldsymbol{S}^t}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Z}}(x,\boldsymbol{v}))^2\mathrm{d}x\otimes\sigma(\mathrm{d}\boldsymbol{v})\\ &=4\bigg(\frac{c_\alpha}{\alpha}\bigg)^{2/\alpha}\int_{\mathbb{S}^{d-1}}\int_\varepsilon^\infty\bigg(\frac{G(x/t)}{G(1/t)}\big(1+\widetilde{h}(x/t,\boldsymbol{v})\big)-1\bigg)^2x^{-2/\alpha}\mathrm{d}x\sigma(\mathrm{d}\boldsymbol{v})=:I(t).\end{aligned}$$

To bound $I(t)$, we use the elementary inequality $(x+y)^2\leqslant 2(x^2+y^2)$, which implies,

$$\frac{1}{8}\bigg(\frac{\alpha}{c_\alpha}\bigg)^{2/\alpha}I(t)\leqslant\int_\varepsilon^\infty x^{-2/\alpha}\bigg(\frac{G(x/t)}{G(1/t)}-1\bigg)^2\mathrm{d}x+\int_{\mathbb{S}^{d-1}}\int_\varepsilon^\infty x^{-2/\alpha}\bigg(\frac{G(x/t)}{G(1/t)}\widetilde{h}(x/t,\boldsymbol{v})\bigg)^2\mathrm{d}x\,\sigma(\mathrm{d}\boldsymbol{v}).$$

By Assumption (S), the first integral $I_1(t)$ above is bounded by $G_2(t)^2\int_\varepsilon^\infty x^{-2/\alpha}G_1(x)^2\mathrm{d}x<\infty$. Assumption (S) and (38) imply, as $t\downarrow 0$,

$$\begin{aligned}I_2(t)&\lesssim t^{2p/\alpha}G(1/t)^{2p}\int_\varepsilon^\infty(1+G_1(x))^2\frac{G(x/t)^{2p}}{G(1/t)^{2p}}x^{-2/\alpha-2p/\alpha}\mathrm{d}x+t^{2\delta}\int_\varepsilon^\infty(1+G_1(x))^2x^{-2/\alpha-2\delta}\mathrm{d}x\\ &\lesssim t^{2p/\alpha}G(1/t)^{2p}\int_\varepsilon^\infty(1+G_1(x))^{2(1+p)}x^{-2/\alpha-2p/\alpha}\mathrm{d}x+t^{2\delta}=\mathcal{O}(t^{2p/\alpha}G(1/t)^{2p}+t^{2\delta}).\end{aligned}$$ □

In the following lemma, we find at what rate the drifts converge.

Lemma 5.14. (a) Let $\alpha\in(1,2)$, and $p,\delta>0$ where $p\neq\alpha-1$ and $\delta\neq(\alpha-1)/\alpha$. Then, under Assumptions (S) & (C), we have, as $t\downarrow 0$,

$$\begin{aligned}|\varpi_{\boldsymbol{S}^t}-\varpi_{\boldsymbol{Z}}|=\mathcal{O}\big(&G_2(t)+t^{1-1/\alpha}G(1/t)^{-1}+(1+G_1(t))t^{\min\{1-1/\alpha,\delta\}}\\ &+G(1/t)^p(1+G_1(t))^{1+p}t^{\min\{1-1/\alpha,p/\alpha\}}\big).\end{aligned}$$

(b) Let $\alpha\in(0,1)$. Then, under Assumptions (S) & (C) we have, as $t\downarrow 0$,

$$|\varpi_{\boldsymbol{S}^t}-\varpi_{\boldsymbol{Z}}|=\mathcal{O}\big(G_2(t)+t^{p/\alpha}G(1/t)^p+t^\delta\big).$$

Proof. First, assume that $\alpha\in(1,2)$. The proof in this setting follows the steps of the proof of Lemma 5.12. Note that

$$\begin{aligned}\varpi_{\boldsymbol{S}^t}&=t^{1-1/\alpha}G(1/t)^{-1}(\varpi_{\boldsymbol{S}}-\mathbb{E}[\boldsymbol{S}_1])-\int_{\mathbb{S}^{d-1}}\int_0^\varepsilon\rho^{\leftarrow}_{\boldsymbol{S}^t}(x,\boldsymbol{v})\mathrm{d}x\sigma(\mathrm{d}\boldsymbol{v}),\\ \varpi_{\boldsymbol{Z}}&=-\int_{\mathbb{S}^{d-1}}\int_0^\varepsilon\rho^{\leftarrow}_{\boldsymbol{Z}}(x,\boldsymbol{v})\mathrm{d}x\sigma(\mathrm{d}\boldsymbol{v}),\end{aligned}$$

and since $\boldsymbol{S}$ has a finite first moment, it follows that

$$\varpi_{\boldsymbol{S}^t}-\varpi_{\boldsymbol{Z}}=t^{1-1/\alpha}G(1/t)^{-1}(\varpi_{\boldsymbol{S}}+\mathbb{E}[\boldsymbol{S}_1])-\int_{\mathbb{S}^{d-1}}\int_0^\varepsilon\rho^{\leftarrow}_{\boldsymbol{S}^t}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Z}}(x,\boldsymbol{v})\mathrm{d}x\sigma(\mathrm{d}\boldsymbol{v}).$$

Recall from (39), that $\rho^{\leftarrow}_{\boldsymbol{S}^t}(x,\boldsymbol{v})=(c_\alpha/\alpha)^{1/\alpha}x^{-1/\alpha}\big(1+\widetilde{h}(x/t,\boldsymbol{v})\big)G(x/t)/G(1/t)$, which implies that

$$\begin{aligned}\bigg|\int_{\mathbb{S}^{d-1}}\int_0^\varepsilon\rho^{\leftarrow}_{\boldsymbol{S}^t}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Z}}(x,\boldsymbol{v})\mathrm{d}x\sigma(\mathrm{d}\boldsymbol{v})\bigg|&\leqslant\int_{\mathbb{S}^{d-1}}\int_0^\varepsilon|\rho^{\leftarrow}_{\boldsymbol{S}^t}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Z}}(x,\boldsymbol{v})|\mathrm{d}x\sigma(\mathrm{d}\boldsymbol{v})\\ =\bigg(\frac{c_\alpha}{\alpha}\bigg)^{1/\alpha}\int_{\mathbb{S}^{d-1}}\int_0^\varepsilon\bigg|\frac{G(x/t)}{G(1/t)}\big(1+\widetilde{h}(x/t,\boldsymbol{v})\big)-1\bigg|x^{-1/\alpha}\mathrm{d}x\sigma(\mathrm{d}\boldsymbol{v})&=:I(t).\end{aligned}$$

The triangle inequality now implies, that

$$\bigg(\frac{\alpha}{c_\alpha}\bigg)^{1/\alpha}I(t)\leqslant\int_0^\varepsilon x^{-1/\alpha}\bigg|\frac{G(x/t)}{G(1/t)}-1\bigg|\mathrm{d}x+\int_{\mathbb{S}^{d-1}}\int_0^\varepsilon x^{-1/\alpha}\bigg|\frac{G(x/t)}{G(1/t)}\widetilde{h}(x/t,\boldsymbol{v})\bigg|\mathrm{d}x\,\sigma(\mathrm{d}\boldsymbol{v}).$$

The two terms in the upper bound are denoted by $I_1(t)$ and $I_2(t)$. Following the calculations made in the proof of Lemma 5.12, we see by Assumption (S) and (38), that $I_1(t)$ in the display above is bounded by $G_2(t)\int_0^\varepsilon x^{-1/\alpha}G_1(x)\mathrm{d}x<\infty$, and, as $t\downarrow 0$,

$$I_2(t)=\mathcal{O}\big((1+G_1(t))t^{\min\{1-1/\alpha,\delta\}}+G(1/t)^p(1+G_1(t))^{1+p}t^{\min\{1-1/\alpha,p/\alpha\}}\big).$$

Assume $\alpha\in(0,1)$. Recall $\varpi_{\boldsymbol{S}^t}=\int_{\mathbb{S}^{d-1}}\int_\varepsilon^\infty \rho^{\leftarrow}_{\boldsymbol{S}^t}(x,\boldsymbol{v})\mathrm{d}x\,\sigma(\mathrm{d}\boldsymbol{v})$ and $\varpi_{\boldsymbol{Z}}=\int_{\mathbb{S}^{d-1}}\int_\varepsilon^\infty \rho^{\leftarrow}_{\boldsymbol{Z}}(x,\boldsymbol{v})\mathrm{d}x\,\sigma(\mathrm{d}\boldsymbol{v})$ and hence, by Lemma 5.11 and (39),

$$\begin{aligned}|\varpi_{\boldsymbol{S}^t}-\varpi_{\boldsymbol{Z}}|&\leqslant\int_{\mathbb{S}^{d-1}}\int_\varepsilon^\infty|\rho^{\leftarrow}_{\boldsymbol{S}^t}(x,\boldsymbol{v})-\rho^{\leftarrow}_{\boldsymbol{Z}}(x,\boldsymbol{v})|\mathrm{d}x\,\sigma(\mathrm{d}\boldsymbol{v})\\&=\left(\frac{c_\alpha}{\alpha}\right)^{1/\alpha}\int_{\mathbb{S}^{d-1}}\int_\varepsilon^\infty\left|\frac{G(x/t)}{G(1/t)}\big(1+\widetilde{h}(x/t,\boldsymbol{v})\big)-1\right|x^{-1/\alpha}\mathrm{d}x\,\sigma(\mathrm{d}\boldsymbol{v})=:I(t).\end{aligned}$$

Bounding $I(t)$ using the triangle inequality, yields

$$\left(\frac{\alpha}{c_\alpha}\right)^{1/\alpha}I(t)\leqslant\int_\varepsilon^\infty x^{-1/\alpha}\left|\frac{G(x/t)}{G(1/t)}-1\right|\mathrm{d}x+\int_{\mathbb{S}^{d-1}}\int_\varepsilon^\infty x^{-1/\alpha}\left|\frac{G(x/t)}{G(1/t)}\widetilde{h}(x/t,\boldsymbol{v})\right|\mathrm{d}x\,\sigma(\mathrm{d}\boldsymbol{v}),$$

where the upper bound is denoted $I_1(t)+I_2(t)$. Assumption (S) and (34) with Lemma 5.11, imply that $I_1(t)\leqslant G_2(t)\int_\varepsilon^\infty x^{-1/\alpha}G_1(x)\mathrm{d}x<\infty$ and

$$\begin{aligned}I_2(t)&\lesssim t^{p/\alpha}G(1/t)^p\int_\varepsilon^\infty(1+G_1(x))\frac{G(x/t)^p}{G(1/t)^p}x^{-1/\alpha-p/\alpha}\mathrm{d}x+t^\delta\int_\varepsilon^\infty(1+G_1(x))x^{-1/\alpha-\delta}\mathrm{d}x\\&=\mathcal{O}\big(t^{p/\alpha}G(1/t)^p+t^\delta\big),\qquad\text{as } t\downarrow 0.\end{aligned}$$

□

Proof of Theorem 5.9. The bound on $\boldsymbol{R}^t$ follows directly from its definition and Lemma 5.2 with $\beta_+=0$. The bounds on the big-jump components, the small-jump components and the drifts, follow directly from Lemmas 5.12, 5.13 & 5.14, respectively. □

5.4. **Brownian limits: upper bounds.** In this subsection, we construct upper bounds on the distance between a Lévy process with nonzero Gaussian component and its attracting Brownian motion. Recall that $(\boldsymbol{\gamma}_{\boldsymbol{X}},\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{X}}^\top,\nu_{\boldsymbol{X}})$ denotes the characteristic triplet [40, Def. 8.2] of $\boldsymbol{X}$ with respect to the cutoff function $\boldsymbol{w}\mapsto\mathbb{1}_{B_{\boldsymbol{0}}(1)}(\boldsymbol{w})$ on $\boldsymbol{w}\in\mathbb{R}^d$ and β_+ is given in terms of the BG index defined in (27).

Proposition 5.15. *Let $\boldsymbol{X}$ be a Lévy process on $\mathbb{R}^d$ with the characteristic triplet $(\boldsymbol{\gamma}_{\boldsymbol{X}},\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{X}}^\top,\nu_{\boldsymbol{X}})$. Let $\boldsymbol{X}^t=(\boldsymbol{X}_{st}/\sqrt{t})_{s\in[0,1]}$ for $t\in(0,1]$ and assume $\int_{\mathbb{R}^d\setminus B_{\boldsymbol{0}}(1)}|\boldsymbol{w}|^q\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})<\infty$ for some $q\in(0,2]$. Let $\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{B}$ be the Gaussian component of $\boldsymbol{X}$ in its Lévy–Itô decomposition (3) and define $\boldsymbol{S}:=\boldsymbol{X}-\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{B}$.*
(a) *If $\boldsymbol{S}$ is of infinite variation or has finite variation with infinite activity and zero natural drift, then*

$$\mathcal{W}_q\big(\boldsymbol{X}^t,\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{B}\big)=\mathcal{O}\big(t^{(q\wedge1)(\min\{1/q,1/\beta_+\}-1/2)}\big),\qquad\textit{as } t\downarrow 0.$$

(b) *If $\boldsymbol{S}$ has finite variation and nonzero natural drift, then*

$$\mathcal{W}_q\big(\boldsymbol{X}^t,\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{B}\big)=\mathcal{O}\big(t^{(q\wedge1)\min\{1/q-1/2,1/2\}}\big),\qquad\textit{as } t\downarrow 0.$$

Note that, if $\boldsymbol{S}$ has infinite activity we have $\beta_+>0$ and if the BG index $\beta<2$, then $1/\beta_+>1/2$. Hence, Proposition 5.15 provides bounds on the rate of convergence in the appropriate L^q-Wasserstein distance for the weak limit in Theorem 5.1 (case $\alpha=2$ and G asymptotically constant). In the case $\beta=2$, it is well-known that $\boldsymbol{X}_1^t$ converges weakly to the Gaussian law of $\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{B}_1$ (see e.g. [5, Prop. I.2(i)]), but the convergence of $\mathcal{W}_q\big(\boldsymbol{X}^t,\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{B}\big)$ could be arbitrarily slow, see Example 6.7 below. It is thus not surprising that Proposition 5.15 gives no information about the rate of convergence. Note also that Proposition 5.15 covers the case $\boldsymbol{\Sigma}_{\boldsymbol{X}}=\boldsymbol{0}$. Moreover, the bound on the L^q-Wasserstein distance when the BG index is less than one is sharper if the natural drift is zero, than if it is not.

Proof of Proposition 5.15. Fix $t \in (0,1]$, let $\boldsymbol{B}$ be the Brownian motion in the Lévy–Itô decomposition (3) of the Lévy process $\boldsymbol{X}$ and recall $\boldsymbol{X}^t = (\boldsymbol{X}_{st}/\sqrt{t})_{s\in[0,1]}$. Since the Brownian motion $\boldsymbol{B}$ satisfies the identity in law $(t^{-1/2}\boldsymbol{B}_{st})_{s\in[0,1]} \overset{d}{=} (\boldsymbol{B}_s)_{s\in[0,1]}$ by self-similarity (Lévy's characterisation theorem), there exists a coupling $(\boldsymbol{X}^t, \boldsymbol{B}')$, such that $\boldsymbol{B}' = (t^{-1/2}\boldsymbol{B}_{st})_{s\in[0,1]}$ and $\boldsymbol{B}' \overset{d}{=} \boldsymbol{B}$. Recalling $\boldsymbol{S} = \boldsymbol{X} - \boldsymbol{\Sigma_X B}$, we obtain

$$(41)\qquad \mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{\Sigma_X B})^{q\vee 1} \leqslant \mathbb{E}\Big[\sup_{s\in[0,1]} |\boldsymbol{\Sigma_X B}_{st}/\sqrt{t} + \boldsymbol{S}_{st}/\sqrt{t} - \boldsymbol{\Sigma_X B}'_s|^q\Big] = t^{-q/2}\mathbb{E}\Big[\sup_{s\in[0,t]} |\boldsymbol{S}_s|^q\Big].$$

Note that the characteristic triplet $(\boldsymbol{\gamma_Y}, \boldsymbol{0}, \nu_{\boldsymbol{S}})$ of $\boldsymbol{S}$ is given by $\boldsymbol{\gamma_Y} = \boldsymbol{\gamma_X}$ and $\nu_{\boldsymbol{S}} = \nu_{\boldsymbol{X}}$. In particular, the BG index of $\boldsymbol{S}$ equals that of $\boldsymbol{X}$.

Part (a). Assume $\boldsymbol{S}$ is of infinite variation. Since, $\boldsymbol{S}$ has no Gaussian component, by [40, Thm 21.9] we have $\int_{\mathbb{R}^d_{\boldsymbol{0}}} |\boldsymbol{w}|\mathbb{1}_{B_{\boldsymbol{0}}(1)}(\boldsymbol{w})\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) = \infty$, implying that the BG index of $\boldsymbol{S}$ satisfies $\beta \geqslant 1$. Hence the associated quantity $\beta_+ \in [\beta, 2]$ satisfies: $\min\{1/q, 1/\beta_+\} - 1/2 \leqslant 1/\beta_+ - 1/2 \leqslant 1/2$. Thus, by Lemma 5.2, we have $t^{-q/2}\mathbb{E}\Big[\sup_{s\in[0,t]} |\boldsymbol{S}_s|^q\Big] \leqslant C_2 t^{q/2} + C_3 t^{q(\min\{1/q,1/\beta_+\}-1/2)}$, implying that $\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{\Sigma_X B})^{q\vee 1} \leqslant 2\max\{C_2, C_3\} t^{q(\min\{1/\beta_+,1/q\}-1/2)}$ for $t \in (0,1]$.

If $\boldsymbol{S}$ has finite variation and zero natural drift, then the bound in Lemma 5.2 with $C_3 = 0$ yields $\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{\Sigma_X B})^{q\vee 1} \leqslant C_3 t^{q(\min\{1/q,1/\beta_+\}-1/2)}$. Noting that $q/(q\vee 1) = q \wedge 1$ implies Part (a).

Part (b). Since $\boldsymbol{S}$ has finite variation, by definition (27) and [40, Thm 21.9], we have $\beta \in [0,1]$ and $I_1 < \infty$, thus implying $\beta_+ \in [0,1]$. By Lemma 5.2 applied to $\boldsymbol{S}$, for $t \in [0,1]$, we find

$$t^{-q/2}\mathbb{E}\Big[\sup_{s\in[0,t]} |\boldsymbol{S}_s|^q\Big] \leqslant C_2 t^{q/2} + C_3 t^{q(\min\{1/q,1/\beta_+\}-1/2)}, \quad t \in (0,1].$$

Since $1 \leqslant 1/\beta_+$, we have $1/2 \leqslant 1/\beta_+ - 1/2$ and $\min\{1/q - 1/2, 1/2\} \leqslant \min\{1/q, 1/\beta_+\} - 1/2$. Thus, for any $\beta_+ \in [0,1]$, by (41) we get $\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{\Sigma_X B})^{q\vee 1} \leqslant 2\max\{C_2, C_3\} t^{q\min\{1/q-1/2,1/2\}}$. As in Part (a), note that $q/(q\vee 1) = q\wedge 1$, implying the claim in Part (b). □

6. Lower bounds on the Wasserstein distance in the domain of attraction

In this section we prove the lower bounds from Theorems 2.1, 2.3 & 2.8. We first cover the domain of non-normal attraction and then turn to the domain of normal attraction.

6.1. Domain of non-normal attraction.

The lower bound on the rate of decay of $\mathcal{W}_q(\boldsymbol{X}^t, \boldsymbol{Z})$ as $t \downarrow 0$ is much greater than polynomial when the scaling function $g(t) = t^{1/\alpha}G(1/t)$ is such that G, which is slowly varying at 0, does not convergent to a positive constant (i.e. the process $\boldsymbol{X}$ is in the domain of non-normal attraction). To show this, we start with the following result, which can be viewed as an extension of [9, Thm 1] from random walks to multidimensional Lévy processes, stated for the L^q-Wasserstein distance. (We remark here that an extension for the Prokhorov distance, used in [9], is also possible in this context.) Our proof below was inspired by that of [9, Thm 1]. Our main tool in the proof of the lower bound in Theorem 2.3 is the following.

Proposition 6.1. *Let $\boldsymbol{X}$ be in the domain of non-normal attraction of an α-stable process $\boldsymbol{Z}$ with $\alpha \in (0,2]$ and define $a(t) \coloneqq G(1/(2t))/G(1/t)$ for $t > 0$. Then, for all $t > 0$ and $q \in (0,1] \cap (0,\alpha)$,*

$$2^{1-q/\alpha}\mathcal{W}_q(\boldsymbol{X}^t_1, \boldsymbol{Z}_1) + a(t)^q \mathcal{W}_q(\boldsymbol{X}^{2t}_1, \boldsymbol{Z}_1) \geqslant |1 - a(t)^q| \mathbb{E}\big[|\boldsymbol{Z}_1|^q\big].$$

In Lemma 6.2 we state some well-known facts used in the proof of Proposition 6.1.

Lemma 6.2. (a) *Let $\boldsymbol{\xi}$ be a random vector in L^q, i.e. $\mathbb{E}[|\boldsymbol{\xi}|^q] < \infty$, for some $q \in (0,1]$. Then,*

$$\mathcal{W}_q(\boldsymbol{\xi}, a\boldsymbol{\xi}) \geqslant |1 - a^q|\mathbb{E}\big[|\boldsymbol{\xi}|^q\big] \quad \textit{for any constant } a \in (0,\infty).$$

(b) *Assume that the random vectors* $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2$ *are in* L^q*, for some* $q \in (0,1]$*, and that* $(\boldsymbol{\xi}_1, \boldsymbol{\zeta}_1)$ *and* $(\boldsymbol{\xi}_2, \boldsymbol{\zeta}_2)$ *are independent. Then the following inequality holds:*

$$\mathcal{W}_q(\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2, \boldsymbol{\zeta}_1 + \boldsymbol{\zeta}_2) \leqslant \mathcal{W}_q(\boldsymbol{\xi}_1, \boldsymbol{\zeta}_1) + \mathcal{W}_q(\boldsymbol{\xi}_2, \boldsymbol{\zeta}_2).$$

Proof. (a) By the subadditivity of $t \mapsto t^q$ on $\mathbb{R}_+$, we have $|\boldsymbol{x}|^q \leqslant (|\boldsymbol{y}|+|\boldsymbol{x}-\boldsymbol{y}|)^q \leqslant |\boldsymbol{y}|^q + |\boldsymbol{x}-\boldsymbol{y}|^q$ for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$. A similar inequality holds by reversing the roles of $\boldsymbol{x}$ and $\boldsymbol{y}$, implying $|\boldsymbol{x}-\boldsymbol{y}|^q \geqslant ||\boldsymbol{x}|^q - |\boldsymbol{y}|^q|$. Hence, we have

$$\mathcal{W}_q(\boldsymbol{\xi}, a\boldsymbol{\xi}) = \inf_{(\boldsymbol{\xi},\boldsymbol{\zeta}),\, \boldsymbol{\zeta} \overset{d}{=} a\boldsymbol{\xi}} \mathbb{E}\big[|\boldsymbol{\xi} - \boldsymbol{\zeta}|^q\big] \geqslant \inf_{(\boldsymbol{\xi},\boldsymbol{\zeta}),\, \boldsymbol{\zeta} \overset{d}{=} a\boldsymbol{\xi}} \big|\mathbb{E}\big[|\boldsymbol{\xi}|^q\big] - \mathbb{E}\big[|\boldsymbol{\zeta}|^q\big]\big| = |1 - a^q|\mathbb{E}\big[|\boldsymbol{\xi}|^q\big].$$

(b) By [41, Thm 4.1] and [35, Main Thm] there exist minimal couplings $(\boldsymbol{\xi}_1, \boldsymbol{\zeta}_1)$ and $(\boldsymbol{\xi}_1, \boldsymbol{\zeta}_2)$, satisfying $\mathbb{E}[|\boldsymbol{\xi}_1 - \boldsymbol{\zeta}_1|^q] = \mathcal{W}_q(\boldsymbol{\xi}_1, \boldsymbol{\zeta}_1)$ and $\mathbb{E}[|\boldsymbol{\xi}_2 - \boldsymbol{\zeta}_2|^q] = \mathcal{W}_q(\boldsymbol{\xi}_2, \boldsymbol{\zeta}_2)$. The product of these two probability spaces yields a coupling of all four vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2$, such that $(\boldsymbol{\xi}_1, \boldsymbol{\zeta}_1)$ and $(\boldsymbol{\xi}_2, \boldsymbol{\zeta}_2)$ are independent. Thus,

$$\begin{aligned}\mathcal{W}_q(\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2, \boldsymbol{\zeta}_1 + \boldsymbol{\zeta}_2) &\leqslant \mathbb{E}[|\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2 - \boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2|^q] \\ &\leqslant \mathbb{E}[|\boldsymbol{\xi}_1 - \boldsymbol{\zeta}_1|^q] + \mathbb{E}[|\boldsymbol{\xi}_2 - \boldsymbol{\zeta}_2|^q] = \mathcal{W}_q(\boldsymbol{\xi}_1, \boldsymbol{\zeta}_1) + \mathcal{W}_q(\boldsymbol{\xi}_2, \boldsymbol{\zeta}_2).\end{aligned}$$
□

Proof of Proposition 6.1. Recall $\boldsymbol{X}_1^t = \boldsymbol{X}_t/g(t)$, and note that $\boldsymbol{X}_{2t} = \boldsymbol{X}_{2t} - \boldsymbol{X}_t + \boldsymbol{X}_t$, where $\boldsymbol{X}_{2t} - \boldsymbol{X}_t$ and $\boldsymbol{X}_t$ are independent and equal in distribution. Furthermore let $\boldsymbol{Z}^{(1)}, \boldsymbol{Z}^{(2)}$ be independent copies of $\boldsymbol{Z}$. Recall that $g(t) = t^{1/\alpha}G(1/t)$ and note that $\boldsymbol{Z}_1 \overset{d}{=} 2^{-1/\alpha}\boldsymbol{Z}_1^{(1)} + 2^{-1/\alpha}\boldsymbol{Z}_1^{(2)}$. This together with Lemma 6.2(b) implies that

$$\begin{aligned}\mathcal{W}_q\left(\frac{G(1/(2t))}{G(1/t)}\boldsymbol{X}_1^{2t}, \boldsymbol{Z}_1\right) &= \mathcal{W}_q\left(\frac{\boldsymbol{X}_{2t} - \boldsymbol{X}_t}{(2t)^{1/\alpha}G(1/t)} + \frac{\boldsymbol{X}_t}{(2t)^{1/\alpha}G(1/t)}, \frac{\boldsymbol{Z}_1^{(1)}}{2^{1/\alpha}} + \frac{\boldsymbol{Z}_1^{(2)}}{2^{1/\alpha}}\right) \\ &\leqslant 2\mathcal{W}_q\left(\frac{\boldsymbol{X}_t}{(2t)^{1/\alpha}G(1/t)}, \frac{\boldsymbol{Z}_1^{(1)}}{2^{1/\alpha}}\right) = 2^{1-q/\alpha}\mathcal{W}_q\left(\boldsymbol{X}_1^t, \boldsymbol{Z}_1\right).\end{aligned}$$

The scaling property for the $\mathcal{W}_q$-distance implies that $\mathcal{W}_q(a(t)\boldsymbol{X}_1^{2t}, a(t)\boldsymbol{Z}_1) = a(t)^q\mathcal{W}_q(\boldsymbol{X}_1^{2t}, \boldsymbol{Z}_1)$. Putting everything together and applying the triangle inequality with Lemma 6.2(a), yields

$$\begin{aligned}2^{1-q/\alpha}\mathcal{W}_q(\boldsymbol{X}_1^t, \boldsymbol{Z}_1) + a(t)^q\mathcal{W}_q(\boldsymbol{X}_1^{2t}, \boldsymbol{Z}_1) &\geqslant \mathcal{W}_q\left(a(t)\boldsymbol{X}_1^{2t}, \boldsymbol{Z}_1\right) + \mathcal{W}_q\left(a(t)\boldsymbol{X}_1^{2t}, a(t)\boldsymbol{Z}_1\right) \\ &\geqslant \mathcal{W}_q\left(\boldsymbol{Z}_1, a(t)\boldsymbol{Z}_1\right) \geqslant |1 - a(t)^q|\mathbb{E}\big[|\boldsymbol{Z}_1|^q\big].\end{aligned}$$
□

6.2. **Domain of normal attraction and the Toscani-Fourier lower bounds.** We begin with the following technical result, used in the proofs of Theorems 2.1 & 2.8. Given two d-dimensional random vectors $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ with characteristic functions $\varphi_{\boldsymbol{\xi}}(\boldsymbol{u}) := \mathbb{E}[\exp(i\langle \boldsymbol{u}, \boldsymbol{\xi}\rangle)]$ and $\varphi_{\boldsymbol{\zeta}}(\boldsymbol{u}) := \mathbb{E}[\exp(i\langle \boldsymbol{u}, \boldsymbol{\zeta}\rangle)]$, respectively, and define the Toscani–Fourier distance (see [3, Eq. (1)]) as

$$T_s(\boldsymbol{\xi}, \boldsymbol{\zeta}) := \sup_{\boldsymbol{u} \in \mathbb{R}^d_{\boldsymbol{0}}} \frac{|\varphi_{\boldsymbol{\xi}}(\boldsymbol{u}) - \varphi_{\boldsymbol{\zeta}}(\boldsymbol{u})|}{|\boldsymbol{u}|^s}, \qquad \text{for } s > 0.$$

The following lemma is an extension of [34, Prop. 2] to the multivariate case and to L^q-Wasserstein distances for $q \in (0,1]$, and the proof is inspired by the proof in the one-dimensional case. For completeness, we give a simple proof below.

Lemma 6.3. *For any random vectors* $\boldsymbol{\xi}, \boldsymbol{\zeta}$ *and* $q \in (0,1]$*, we have* $\mathcal{W}_q(\boldsymbol{\xi}, \boldsymbol{\zeta}) \geqslant 2^{q-1}T_q(\boldsymbol{\xi}, \boldsymbol{\zeta})$.

Proof. Fix $q \in (0,1]$. Since the map $\psi : x \mapsto e^{ix}$, $x \in \mathbb{R}$, satisfies $|1 - \psi(x)| \leqslant 2\min\{|x/2|, 1\} \leqslant 2|x/2|^q$ for $x \in \mathbb{R}$, we have $|\psi(x) - \psi(y)| \leqslant 2^{1-q}|x - y|^q$ for any $x, y \in \mathbb{R}$. Hence, for any $\boldsymbol{u} \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$,

$$\mathbb{E}\big[2^{1-q}|\boldsymbol{\xi} - \boldsymbol{\zeta}|^q\big] \geqslant \frac{\mathbb{E}\big[2^{1-q}|\langle \boldsymbol{u}, \boldsymbol{\xi}\rangle - \langle \boldsymbol{u}, \boldsymbol{\zeta}\rangle|^q\big]}{|\boldsymbol{u}|^q} \geqslant \frac{\mathbb{E}\big[|\psi(\langle \boldsymbol{u}, \boldsymbol{\zeta}\rangle) - \psi(\langle \boldsymbol{u}, \boldsymbol{\zeta}\rangle)|\big]}{|\boldsymbol{u}|^q} \geqslant \frac{|\varphi_{\boldsymbol{\xi}}(\boldsymbol{u}) - \varphi_{\boldsymbol{\zeta}}(\boldsymbol{u})|}{|\boldsymbol{u}|^q}.$$

Since $\boldsymbol{u} \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$ is arbitrary, the result follows. □

6.2.1. *Heavy-tailed domain of normal attraction.* Let $(\boldsymbol{X}_t)_{t\geqslant 0}$ be a Lévy process on $\mathbb{R}^d$ in the domain of attraction of an α-stable process $\boldsymbol{Z}$, such that $\boldsymbol{X}_1^t = \boldsymbol{X}_t/t^{1/\alpha} \xrightarrow{d} \boldsymbol{Z}_1$ as $t \downarrow 0$.

Lemma 6.4. *Let $\boldsymbol{X}$ be a Lévy process that differs in law from the α-stable process $\boldsymbol{Z}$, $\alpha \in (0,2]$. Let $\psi_{\boldsymbol{X}}$ and $\psi_{\boldsymbol{Z}}$ denote their Lévy–Khintchine exponents. Pick any $q \in (0,1] \cap (0,\alpha)$ and $\boldsymbol{u}_* \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$ for which $C_* := 2^{q-1}|\boldsymbol{u}_*|^{-q}|\psi_{\boldsymbol{X}_1}(\boldsymbol{u}_*) - \psi_{\boldsymbol{Z}_1}(\boldsymbol{u}_*)| > 0$. Then, we have*

$$\mathcal{W}_q(\boldsymbol{X}_1^t, \boldsymbol{Z}_1) \geqslant C_* t^{1-q/\alpha} + \mathcal{O}(t^{2-q/\alpha}), \qquad \text{as } t \downarrow 0.$$

Proof. First, Lemma 6.3 implies that

$$\mathcal{W}_q(\boldsymbol{X}_1^t, \boldsymbol{Z}_1) \geqslant 2^{q-1} T_q(\boldsymbol{X}_1^t, \boldsymbol{Z}_1) \geqslant \frac{2^{q-1}}{|\boldsymbol{u}|^q} |\varphi_{\boldsymbol{X}_1^t}(\boldsymbol{u}) - \varphi_{\boldsymbol{Z}_1}(\boldsymbol{u})|, \qquad \text{for any } \boldsymbol{u} \in \mathbb{R}^d_{\boldsymbol{0}},$$

where $\varphi_{\boldsymbol{\xi}}$ denotes the characteristic function of the random vector $\boldsymbol{\xi}$. Second, set $\boldsymbol{u} = t^{1/\alpha}\boldsymbol{u}_*$ with $\boldsymbol{u}_* \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$ as in the statement of the lemma and note that

$$\varphi_{\boldsymbol{X}_1^t}(\boldsymbol{u}) = \mathbb{E}\big[\exp(i\langle \boldsymbol{X}_1^t, t^{1/\alpha}\boldsymbol{u}_*\rangle)\big] = \mathbb{E}\big[\exp(i\langle \boldsymbol{X}_t/t^{1/\alpha}, t^{1/\alpha}\boldsymbol{u}_*\rangle)\big] = e^{t\psi_{\boldsymbol{X}}(\boldsymbol{u}_*)}.$$

Similarly, since $\boldsymbol{Z}_1 \stackrel{d}{=} \boldsymbol{Z}_t/t^{1/\alpha}$, we have $\varphi_{\boldsymbol{Z}_1}(\boldsymbol{u}) = \exp(t\psi_{\boldsymbol{Z}}(\boldsymbol{u}_*))$ and hence

$$\mathcal{W}_q(\boldsymbol{X}_1^t, \boldsymbol{Z}_1) \geqslant \frac{2^{q-1}}{|\boldsymbol{u}_*|^q t^{q/\alpha}} \big|e^{t\psi_{\boldsymbol{X}}(\boldsymbol{u}_*)} - e^{t\psi_{\boldsymbol{Z}}(\boldsymbol{u}_*)}\big|.$$

Since for any $z \in \mathbb{C}$ we have $e^z = 1 + z + \mathcal{O}(z^2)$ as $|z| \to 0$, it follows that $|e^{at} - e^{bt}| = |at - bt + \mathcal{O}(t^2)| = |a-b|t + \mathcal{O}(t^2)$ for $a = \psi_{\boldsymbol{X}}(\boldsymbol{u}_*)$ and $b = \psi_{\boldsymbol{Z}}(\boldsymbol{u}_*)$. The result then follows. □

6.2.2. *Brownian domain of normal attraction.* The domain of normal attraction to a Brownian motion consists of the class of Lévy processes with a nontrivial Brownian component (see e.g. [26] and [5, Prop. I.2(i)]). To construct a lower bound on the distance between the Lévy process and its Brownian limit require the following lower estimates.

Lemma 6.5. *Let $\boldsymbol{Y}$ be a nonzero pure-jump Lévy process on $\mathbb{R}^d$, let $\psi_{\boldsymbol{Y}}(\boldsymbol{u})$ denote its Lévy-Khintchine exponent and $\nu_{\boldsymbol{Y}}$ its Lévy measure.*
(a) *If $\boldsymbol{Y}$ has finite variation and nonzero drift with direction $\boldsymbol{u}_* \in \mathbb{S}^{d-1}$. Then $|\psi_{\boldsymbol{Y}}(r\boldsymbol{u}_*)| \geqslant cr$ for some $c > 0$ and all sufficiently large $r > 0$.*
(b) *Suppose there exist a locally finite measure ρ on $(0,\infty)$ and a probability measure σ on $\mathbb{S}^{d-1}$ with*

$$\nu_{\boldsymbol{Y}}(A) \geqslant \int_{\mathbb{S}^{d-1}} \int_{(0,\infty)} \mathbb{1}_A(r\boldsymbol{v})\rho(\mathrm{d}r)\sigma(\mathrm{d}\boldsymbol{v}), \qquad A \in \mathcal{B}(\mathbb{R}^d_{\boldsymbol{0}}).$$

Define $\upsilon(r) := \int_{(0,r)} r^2\rho(\mathrm{d}r)$ for $r > 0$, and, given any $c \in (0,1)$, let $\boldsymbol{u}_* \in \mathbb{S}^{d-1}$ be such that the set $C_{c,\boldsymbol{u}_*} := \{\boldsymbol{v} \in \mathbb{S}^{d-1} : |\langle \boldsymbol{u}_*, \boldsymbol{v}\rangle| \geqslant c\}$ has positive σ-measure $m := \sigma(C_{c,\boldsymbol{u}_*}) > 0$. Then we have

$$|\psi_{\boldsymbol{Y}}(r\boldsymbol{u}_*)| \geqslant \frac{c^2 m}{3} r^2 \upsilon(r^{-1}) \qquad \text{for all } r > 0.$$

In particular, if $c_\delta := \inf_{r\in(0,1)} r^{\delta-2}\upsilon(r) > 0$ for some $\delta \in (0,2)$, then $|\psi_{\boldsymbol{Y}}(r\boldsymbol{u}_*)| \geqslant (c_\delta c^2 m/3) r^\delta$, $r > 1$.

Proof. Let $\Re z$ and $\Im z$ denote the real and imaginary parts of $z \in \mathbb{C}$, respectively.
(a) Since $\boldsymbol{Y}$ has finite variation, it is clear from the Lévy-Khintchine formula without compensator that $|\psi_{\boldsymbol{Y}}(r\boldsymbol{u}_*)| \geqslant |\Im\psi_{\boldsymbol{Y}}(r\boldsymbol{u}_*)| \geqslant cr$ for some $c > 0$ and all sufficiently large $r > 0$. Indeed, this follows from [5, Prop. 2(ii)] applied to the finite variation Lévy process $\langle \boldsymbol{u}_*, \boldsymbol{Y}\rangle$.

(b) Note at first, that $1-\Re e^{ix} = 1-\cos(x) \geqslant \frac{1}{3}x^2\mathbb{1}_{\{|x|<1\}}$ for all $x\in\mathbb{R}$. Thus, the Lévy-Khintchine formula applied to $\Re\psi$ yields

$$\begin{aligned}3|\psi_{\boldsymbol{Y}}(r\boldsymbol{u}_*)| \geqslant 3|\Re\psi(r\boldsymbol{u}_*)| &\geqslant \int_{\mathbb{R}^d_{\boldsymbol{0}}} |\langle r\boldsymbol{u}_*,\boldsymbol{w}\rangle|^2\mathbb{1}_{\{|\langle r\boldsymbol{u}_*,\boldsymbol{w}\rangle|<1\}}\nu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{w})\\ &\geqslant \int_{\mathbb{R}^d_{\boldsymbol{0}}} |\langle r\boldsymbol{u}_*,\boldsymbol{w}\rangle|^2\mathbb{1}_{\{r|\boldsymbol{w}|<1\}}\nu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{w}) \geqslant \int_0^{1/r}\int_{C_{c,\boldsymbol{u}_*}} |\langle r\boldsymbol{u}_*,s\boldsymbol{v}\rangle|^2\sigma(\mathrm{d}\boldsymbol{v})\rho(\mathrm{d}s)\\ &\geqslant \int_0^{1/r}\int_{C_{c,\boldsymbol{u}_*}} c^2r^2s^2\sigma(\mathrm{d}\boldsymbol{v})\rho(\mathrm{d}s) \geqslant c^2m\int_0^{1/r} r^2s^2\rho(\mathrm{d}s) = c^2mr^2\upsilon(r^{-1}).\end{aligned}$$

This proves the first claim in Part (b). The second claim follows from the additional assumption. □

Lemma 6.6. *Let $\boldsymbol{X}$ be a Lévy process on $\mathbb{R}^d$ with the characteristic triplet $(\boldsymbol{\gamma_X},\boldsymbol{\Sigma_X}\boldsymbol{\Sigma_X}^\top,\nu_{\boldsymbol{X}})$. Let $\boldsymbol{X}^t = (\boldsymbol{X}_{st}/\sqrt{t})_{s\in[0,1]}$ for $t\in(0,1]$. Moreover, let $\boldsymbol{\Sigma_X B}$ be the Gaussian component of $\boldsymbol{X}$ in its Lévy–Itô decomposition (3) and define $\boldsymbol{S} \coloneqq \boldsymbol{X}-\boldsymbol{\Sigma_X B}$ with Lévy–Khintchine exponent $\psi_{\boldsymbol{S}}$.*
(a) *Pick any $\boldsymbol{u}_*\in\mathbb{R}^d_{\boldsymbol{0}}$ and define $C_* \coloneqq |\boldsymbol{u}_*|^{-1}|\psi_{\boldsymbol{S}_1}(\boldsymbol{u}_*)|>0$. Then, we have*

$$\mathcal{W}_1(\boldsymbol{X}^t_1,\boldsymbol{\Sigma_X B}_1) \geqslant C_*\sqrt{t}+\mathcal{O}(t^{3/2}),\quad \text{as } t\downarrow 0.$$

(b) *Let λ be the largest eigenvalue of $\boldsymbol{\Sigma_X}\boldsymbol{\Sigma_X}^\top$. Suppose there exist $\delta\in[1,2)$ and vectors $(\boldsymbol{u}_r)_{r\in(0,\infty)}$ on $\mathbb{R}^d_{\boldsymbol{0}}$ satisfying $|\boldsymbol{u}_r|=r$ and $c\coloneqq\inf_{r>1}r^{-\delta}|\psi_{\boldsymbol{S}}(\boldsymbol{u}_r)|>0$. Then for any $C_*\in(0,ce^{-\lambda/2})$ we have*

$$\mathcal{W}_1(\boldsymbol{X}^t_1,\boldsymbol{\Sigma_X B}_1) \geqslant C_*t^{1-\delta/2}\qquad \text{for all sufficiently small } t>0.$$

Proof. Since $\boldsymbol{B}$ and $\boldsymbol{S}$ are independent, we have $\varphi_{\boldsymbol{X}^t_1}(\boldsymbol{u}) = \varphi_{\boldsymbol{\Sigma_X B}_1}(\boldsymbol{u})\varphi_{\boldsymbol{S}_t}(\boldsymbol{u}/\sqrt{t})$. Hence, Lemma 6.3 gives, for any $\boldsymbol{u}\in\mathbb{R}^d_{\boldsymbol{0}}$,

$$\mathcal{W}_1(\boldsymbol{X}^t_1,\boldsymbol{\Sigma_X B}_1) \geqslant T_1(\boldsymbol{X}^t_1,\boldsymbol{\Sigma_X B}_1) \geqslant \frac{1}{|\boldsymbol{u}|}|\varphi_{\boldsymbol{\Sigma_X B}_1}(\boldsymbol{u})|\cdot|\varphi_{\boldsymbol{S}_t}(\boldsymbol{u}/\sqrt{t})-1|, \tag{42}$$

(a) Let $\boldsymbol{u}_*\in\mathbb{R}^d$ be as in the statement. The result then follows from Lemma 6.4.

(b) The proof follows as in that of Lemma 6.4. The main idea is to use the fact that, if $a_t\to0$ as $t\downarrow0$ and $\sup_{t>0}|b_t|<\infty$, then

$$|e^{a_t+b_t}-e^{b_t}| = |e^{a_t}-1|\cdot|e^{b_t}| \geqslant |a_t|\cdot\inf_{s>0}|e^{b_s}|+\mathcal{O}(|a_t|^2),\quad \text{as } t\downarrow 0.$$

Set $b_t = -(t/2)\boldsymbol{u}_{t^{-1/2}}^\top\boldsymbol{\Sigma}^2_{\boldsymbol{X}}\boldsymbol{u}_{t^{-1/2}}$ and $a_t = t\psi_{\boldsymbol{S}}(\boldsymbol{u}_{t^{-1/2}})$ for $t>0$. Note that $b_t=\mathcal{O}(1)$ as $t\downarrow0$ and $|e^{b_t}|\geqslant e^{-\lambda/2}$. Since $\boldsymbol{S}$ does not have a Brownian component, we have $\lim_{|\boldsymbol{u}|\to\infty}|\boldsymbol{u}|^{-2}\cdot|\psi_{\boldsymbol{S}}(\boldsymbol{u})|=0$ (see, e.g. [40, Lem. 43.11]) and thus $a_t\to0$ as $t\downarrow0$. Moreover, $|a_t|\geqslant ct^{1-\delta/2}$ for $t<1$ by assumption. Thus, applying (42) with $\boldsymbol{u}=\sqrt{t}\boldsymbol{u}_{t^{-1/2}}$ yields Part (b). □

Example 6.7. Consider an example inspired by [26, Ex. 4.2.1]. Let S be a real-valued martingale Lévy process with Lévy measure $\nu(\mathrm{d}y) = py^{-3}\log^{-1-p}(y)\mathbb{1}_{(0,1)}(y)\mathrm{d}y$ for some $p>0$ and all $y\in\mathbb{R}$. Let $B=(B_s)_{s\in[0,1]}$ be a standard Brownian motion independent of S and let $\sigma^2>0$. Define $X^t\coloneqq(S_{st}/\sqrt{t}+\sigma B_{st}/\sqrt{t})_{s\in[0,1]}$ for $t>0$. Choose g to satisfy $g(t)^2\log(1/g(t))^p\sim t$ as $t\downarrow0$. The function g is regularly varying at 0 with index 1/2, and therefore $g(t)\sim\sqrt{t/\log(1/t)^p}$ as $t\downarrow0$ by [6, Thm 1.5.12]. Note that $S_t/g(t)$ converges in law to a standard normal distribution by [26, Thm 2(i)]. Hence, the Lévy–Khintchine exponent ψ_S of S satisfies $t\psi_S(u/g(t))\to-u^2/2$ as $t\downarrow0$ for any $u\in\mathbb{R}$. In particular, by taking $t=g^{-1}(\sqrt{s})\sim s\log(1/s)^p$, which tends to 0 as $s\downarrow0$, we obtain $s\log(1/s)^p\psi_S(1/\sqrt{s})\to-1/2$ as $s\downarrow0$. Then, a slight modification of the argument in the proof of Lemma 6.6(b) shows that $\liminf_{t\downarrow0}\log(1/t)^p\mathcal{W}_1(X^t_1,\sigma B_1)>0$. △

7. Proofs of Section 2

In this section we give the proofs of the results stated in Section 2.

Proof of Theorem 2.1. Part (a). Recall from Assumption (T) and Remark 5.4, that we may decompose $\boldsymbol{X}$ as the sum $\boldsymbol{S}+\boldsymbol{R}$ of independent Lévy processes $\boldsymbol{S}$ and $\boldsymbol{R}$ with generating triplets $(\boldsymbol{\gamma_S},\boldsymbol{0},\nu^{\mathrm{c}}_{\boldsymbol{X}})$ and $(\boldsymbol{\gamma_R},\boldsymbol{0},\nu^{\mathrm{d}}_{\boldsymbol{X}})$, respectively. For $t\in(0,1]$, denote $\boldsymbol{S}^t=(\boldsymbol{S}_{st}/t^{1/\alpha})_{s\in[0,1]}$ and $\boldsymbol{R}^t=(\boldsymbol{R}_{st}/t^{1/\alpha})_{s\in[0,1]}$. Let $\kappa(t)\coloneqq t^r$ for $t\in(0,1]$ and some $r\geqslant -1/\alpha$. Assume the processes $(\boldsymbol{D}^{\boldsymbol{S}^t,\kappa(t)},\boldsymbol{D}^{\boldsymbol{Z},\kappa(t)},\boldsymbol{J}^{\boldsymbol{S}^t,\kappa(t)},\boldsymbol{J}^{\boldsymbol{Z},\kappa(t)})$ are coupled as in Subsection 4.1 (i.e. (5) and (6)).

Note that $\mathcal{W}_q(\boldsymbol{X}^t,\boldsymbol{Z})\leqslant\mathcal{W}_q(\boldsymbol{S}^t,\boldsymbol{Z})+\mathbb{E}[\sup_{t\in[0,1]}|\boldsymbol{R}^t_s|^q]$ by the triangle inequality and the definition of $\mathcal{W}_q$. Theorem 5.5 (with $p=1$ and $q\in(0,1]\cap(0,\alpha)$) yields a bound on $\mathcal{W}_q(\boldsymbol{X}^t,\boldsymbol{Z})$ via (4) as follows:

$$\mathbb{E}\bigg[\sup_{t\in[0,1]}\big|\boldsymbol{R}_{st}/t^{1/\alpha}\big|^q\bigg]=\mathcal{O}\big(t^{1-q/\alpha}\big),$$

$$\mathcal{W}_q\big(\boldsymbol{J}^{\boldsymbol{S}^t,\kappa(t)},\boldsymbol{J}^{\boldsymbol{Z},\kappa(t)}\big)=\begin{cases}\mathcal{O}\big(t^{1/\alpha+r(q+1-\alpha)}\big), & q<\alpha-1,\\ \mathcal{O}\big(t^{1-q/\alpha}(1+\log(1/t)\mathbb{1}_{\{q=\alpha-1,\,r\neq-1/\alpha\}})\big), & q\geqslant\alpha-1,\end{cases}\quad\text{by (28)},$$

$$\mathcal{W}_q\big(\boldsymbol{D}^{\boldsymbol{S}^t,\kappa(t)},\boldsymbol{D}^{\boldsymbol{Z},\kappa(t)}\big)\leqslant\mathcal{W}_2\big(\boldsymbol{D}^{\boldsymbol{S}^t,\kappa(t)},\boldsymbol{D}^{\boldsymbol{Z},\kappa(t)}\big)^q=\mathcal{O}\big(t^{q/\alpha+rq(3-\alpha)}\big),\quad\text{by (22) and (29)},$$

$$|\boldsymbol{\gamma}_{\boldsymbol{S}^t,\kappa(t)}-\boldsymbol{\gamma}_{\boldsymbol{Z},\kappa(t)}|^q=\begin{cases}\mathcal{O}\big(t^{q/\alpha+rq(2-\alpha)}\big), & \alpha\in(0,1),\\ \mathcal{O}\big(t^{q(1-1/\alpha)}\big), & \alpha\in(1,2),\end{cases}\quad\text{by (30)}.$$

Part (a) can now be deduced by optimising r as follows. If $\alpha<1$ then $q>0>\alpha-1$ and taking r sufficiently large makes all terms become $\mathcal{O}(t^{1-q/\alpha})$. If $\alpha>1$ and $q<\alpha-1$, then the bounds are $\mathcal{O}(t^{1-q/\alpha})$, $\mathcal{O}(t^{1/\alpha+r(q+1-\alpha)})$, $\mathcal{O}(t^{q/\alpha+rq(3-\alpha)})$ and $\mathcal{O}(t^{q(1-1/\alpha)})$. Since $q\leqslant1$ and $1-1/\alpha\leqslant1/\alpha$, all these bounds can be made $\mathcal{O}(t^{q(1-1/\alpha)})$ by picking $r=0$. If $\alpha>1$ and $q\geqslant\alpha-1$, then the bounds are $\mathcal{O}(t^{1-q/\alpha})$, $\mathcal{O}(t^{1-q/\alpha}(1+\log(1/t)\mathbb{1}_{\{q=\alpha-1,\,r\neq-1/\alpha\}}))$, $\mathcal{O}(t^{q/\alpha+rq(3-\alpha)})$ and $\mathcal{O}(t^{q(1-1/\alpha)})$ and, as before, these can all be made $\mathcal{O}(t^{q(1-1/\alpha)})$ by picking $r=0$. Note here that the logarithmic term never arises in the dominant term, as it would require $1=q=\alpha-1$, but $\alpha<2$.

Part (b). The claim follows from a direct application of Lemma 6.4. □

Proof of Theorem 2.8. Proposition 5.15 gives Part (a). Parts (b) and (c) follow from Lemma 6.6. □

In preparation for the proofs of Theorem 2.3 and Corollary 2.4, we establish Lemmas 7.1, 7.2 and 7.3 about slowly varying functions.

Lemma 7.1. *Let ℓ be C^1 and slowly varying such that its derivative equals $t\mapsto\widetilde{\ell}(t)/(c+t)$ for some $c\geqslant0$, where $|\widetilde{\ell}|$ is positive and slowly varying at infinity. Then, for each $x>0$, we have $(\ell(t)-\ell(xt))/\widetilde{\ell}(t)\to-\log x$ as $t\to\infty$.*

For $x>0$, define $L_x(t)\coloneqq1-\ell(xt)/\ell(t)$ for all $t>0$. The function $|L_x|$ is asymptotically equivalent to $|\widetilde{\ell}(t)\log x|/\ell(t)\sim|L_x(t)|$ as $t\to\infty$ and, if $x\neq1$, slowly varying at infinity. Moreover, for $x>0$, the function $\Sigma(x)\coloneqq\sup_{t>0,y\in[x\wedge1,x\vee1]}\widetilde{\ell}(yt)/\widetilde{\ell}(t)$ satisfies

$$\big|L_x(t)\big|=\bigg|\frac{\ell(xt)}{\ell(t)}-1\bigg|\leqslant\frac{|\widetilde{\ell}(t)|}{\ell(t)}\cdot\Sigma(x)|\log x|\qquad\textit{for all } t,x>0.$$

Proof. Since $|\widetilde{\ell}|$ is positive and $\widetilde{\ell}$ is continuous, $\widetilde{\ell}$ is either eventually positive or negative and either $\widetilde{\ell}$ or $-\widetilde{\ell}$ is slowly varying at infinity, respectively. By [6, Thm 1.2.1], we have $\sup_{y\in[a,b]}|\widetilde{\ell}(yt)/\widetilde{\ell}(t)-1|\to0$ as $t\to\infty$ for any $0<a<b<\infty$. Thus for all sufficiently large $t>0$ we have

$$\frac{\widetilde{\ell}(ty)}{\widetilde{\ell}(t)}\frac{1}{c/t+y}\leqslant\big(1+\sup_{z\in[x\wedge1,x\vee1]}|\widetilde{\ell}(zt)/\widetilde{\ell}(t)-1|\big)\frac{1}{y}\leqslant\frac{2}{y}\qquad\text{for all } y\in[x\wedge1,x\vee1].$$

The dominated convergence theorem now yields

$$\frac{\ell(t)-\ell(xt)}{\widetilde{\ell}(t)}=\frac{1}{\widetilde{\ell}(t)}\int_{xt}^{t}\widetilde{\ell}(y)\frac{\mathrm{d}y}{c+y}=\int_{x}^{1}\frac{\widetilde{\ell}(ty)}{\widetilde{\ell}(t)}\frac{\mathrm{d}y}{c/t+y}\to\int_{x}^{1}\frac{\mathrm{d}y}{y}=-\log x,\quad \text{as } t\to\infty,$$

which establishes the first claim. Since $L_x(t)=\big((\ell(t)-\ell(xt))/\widetilde{\ell}(t)\big)\big(\widetilde{\ell}(t)/\ell(t)\big)$, the function $|L_x|$ is positive on a neighbourhood of infinity and asymptotically equivalent to $|\widetilde{\ell}(t)\log x|/\ell(t)$ by the limit in the previous display. Moreover, since $x>0$ in the limit was arbitrary, for any $\lambda>0$ we have

$$\frac{L_x(\lambda t)}{L_x(t)}=\frac{\ell(t)}{\ell(\lambda t)}\cdot\frac{\widetilde{\ell}(\lambda t)^{-1}(\ell(\lambda t)-\ell(\lambda tx))}{\widetilde{\ell}(t)^{-1}(\ell(t)-\ell(xt))}\cdot\frac{\widetilde{\ell}(\lambda t)}{\widetilde{\ell}(t)}\to 1,\quad \text{as } t\to\infty,$$

implying that $|L_x|$ is slowly varying at infinity.

To establish the non-asymptotic inequality in the lemma, fix $x>0$ and note that

$$|\ell(xt)-\ell(t)|\leqslant\left|\int_t^{xt}\widetilde{\ell}(y)\frac{\mathrm{d}y}{c+y}\right|\leqslant\left|\int_t^{xt}\widetilde{\ell}(y)\frac{\mathrm{d}y}{y}\right|\leqslant\int_{[x\wedge 1,x\vee 1]}\big|\widetilde{\ell}(yt)\big|\frac{\mathrm{d}y}{y}\leqslant|\widetilde{\ell}(t)|\cdot\Sigma(x)|\log x|.\qquad\square$$

Note that the assumption in Lemma 7.1 requires ℓ to be eventually strictly monotone. Moreover, if ℓ satisfies the conditions of Lemma 7.1, then so does ℓ^q for any $q>0$.

Lemma 7.2. (a) *Let ℓ be slowly varying at infinity. Suppose that, for some $\lambda\in(0,\infty)\setminus\{1\}$ and non-increasing function ϕ_λ, we have $\phi_\lambda(t)\geqslant|1-\ell(\lambda t)/\ell(t)|$ for all $t\geqslant 1$ and $\int_1^\infty\phi_\lambda(t)t^{-1}\mathrm{d}t<\infty$. Then ℓ has a positive finite limit at infinity.*
(b) *Let ϕ be slowly varying at infinity with $\phi(t)\to 0$ as $t\to\infty$ and $\int_1^\infty\phi(t)t^{-1}\mathrm{d}t=\infty$. Then the functions $\ell_\pm(t):=\exp(\pm\int_1^t\phi(s)s^{-1}\mathrm{d}s)$ are slowly varying at infinity, $\ell_+(t)\to\infty$, $\ell_-(t)\to 0$ and $|1-\ell_\pm(\lambda t)/\ell_\pm(t)|\sim|\log\lambda|\phi(t)$ as $t\to\infty$ for any $\lambda>0$.*

Note that the smallest non-increasing function φ_λ satisfying $\phi_\lambda(t)\geqslant|1-\ell(\lambda t)/\ell(t)|$ for all $t\geqslant 1$ is given by $\varphi_\lambda(t):=\sup_{s\geqslant t}|1-\ell(\lambda s)/\ell(s)|$.

Proof of Lemma 7.2. Part (a). First assume $\lambda>1$. Define $U_\lambda(t):=\sup_{x\in[1,\lambda]}|1-\ell(xt)/\ell(t)|$ and note that $U_\lambda(t)\to 0$ as $t\to\infty$ by the uniform convergence theorem [6, Thm 1.2.1]. As $\int_1^\infty\phi_\lambda(t)t^{-1}\mathrm{d}t<\infty$, we also have $\phi_\lambda(t)\to 0$ as $t\to 0$, making $\eta:=\inf\{t\geqslant 1:\max\{\phi_\lambda(s),U_\lambda(s)\}<1/2 \text{ for all } s\geqslant t\}$ finite. Since $1+U_\lambda(t)\geqslant\ell(xt)/\ell(t)\geqslant 1-U_\lambda(t)$ and $U_\lambda(t)\leqslant 1/2$ for all $t>\eta$ and $x\in[1,\lambda]$, we obtain

$$U_\lambda(t)\geqslant\log(1+U_\lambda(t))\geqslant\log\left(\frac{\ell(xt)}{\ell(t)}\right)\geqslant\log(1-U_\lambda(t))\geqslant-2U_\lambda(t),\ \text{implying}\ \left|\log\left(\frac{\ell(xt)}{\ell(t)}\right)\right|\leqslant 2U_\lambda(t)$$

for all $t>\eta$ and $x\in[1,\lambda]$. Similarly, for $t>\eta$ we have $|\log(\ell(\lambda t)/\ell(t))|\leqslant 2\phi_\lambda(t)$. For any $T>t\geqslant\eta'$ set $n:=\lfloor\log(T/t)/\log\lambda\rfloor$, implying $T/(\lambda^n t)\in[1,\lambda)$. By the monotonicity of ϕ_λ we obtain

$$\begin{aligned}|\log\ell(T)-\log\ell(t)|&\leqslant\left|\log\left(\frac{\ell(T)}{\ell(\lambda^n t)}\right)\right|+\sum_{k=1}^n\left|\log\left(\frac{\ell(\lambda^k t)}{\ell(\lambda^{k-1}t)}\right)\right|\\&\leqslant 2U_\lambda(\lambda^n t)+\sum_{k=1}^n 2\phi_\lambda\big(\lambda^{k-1}t\big)\\&\leqslant 2U_\lambda(\lambda^n t)+\sum_{k=1}^n\frac{2}{\log\lambda}\int_{\lambda^{k-2}t}^{\lambda^{k-1}t}\phi_\lambda(s)\frac{\mathrm{d}s}{s}\\&\leqslant 2U_\lambda(\lambda^n t)+\frac{2}{\log\lambda}\int_{\lambda^{-1}t}^{\infty}\phi_\lambda(s)\frac{\mathrm{d}s}{s}\overset{t\to\infty}{\longrightarrow}0\qquad(\text{uniformly in } T\in[t,\infty)).\end{aligned}\tag{43}$$

If we had $\limsup_{t\to\infty}\log\ell(t)>\liminf_{t\to\infty}\log\ell(t)$, there would exist an increasing sequence $(t_k)_{k\in\mathbb{N}}$ and $\epsilon>0$ such that $t_k\to\infty$ and $|\log\ell(t_{k+1})-\log\ell(t_k)|>\epsilon$ for all $k\in\mathbb{N}$, contradicting (43). Hence the limit $\lim_{t\to\infty}\log\ell(t)$ exists. By taking the limit as $T\to\infty$ on the left-hand side of (43) for any fixed t, it follows that $\lim_{t\to\infty}|\log\ell(t)|\neq\infty$. Thus ℓ has a finite and positive limit. The case $\lambda<1$ can be established in a similar way.

Part (b). The statement follows from a direct application of Lemma 7.1. □

Lemma 7.3. *Define iteratively the functions* $\ell_1(t)=\log(e+t)$ *and* $\ell_{n+1}(t)=\log(e+\ell_n(t))$ *for* $t\geqslant 0$ *and* $n\in\mathbb{N}$. *Then, the following statements hold.*
(a) *For any* $x>0$ *and* $c\geqslant 0$, *we have* $(c+\ell_n(xt))/(c+\ell_n(t))\leqslant 1+\mathbb{1}_{\{x>1\}}\log x$.
(b) *We have* $(e+t)\ell_n'(t)=\prod_{i=1}^{n-1}(e+\ell_k(t))^{-1}$.
(c) *Suppose* $\ell(t)=\ell_n(t)^{q_n}\cdots\ell_m(t)^{q_m}$ *for some* $1\leqslant n\leqslant m$ *in* $\mathbb{N}$ *and either* $q_n,\ldots,q_m\geqslant 0$ *with* $q_n,q_m>0$ *or* $q_n,\ldots,q_m\leqslant 0$ *with* $q_n,q_m<0$. *Set* $\widetilde{\ell}(t):=(e+t)\ell'(t)$, *then we have*

$$\Sigma(x):=\sup_{t>0,\,y\in[x\wedge 1,x\vee 1]}\frac{\widetilde{\ell}(yt)}{\widetilde{\ell}(t)}\leqslant\begin{cases}(1+\log x)^{\sum_{j=n}^m q_j^+}, & x\geqslant 1,\\ (1+|\log x|)^{m+\sum_{j=n}^m q_j^-}, & x<1.\end{cases}$$

Proof. For $x<1$ we have $\ell_n(xt)\leqslant\ell_n(t)$. For $x>1$, we have

$$\ell_n(xt)-\ell_n(t)=\int_t^{xt}\frac{1}{\prod_{k=1}^{n-1}(e+\ell_k(s))}\frac{\mathrm{d}s}{e+s}\leqslant\frac{1}{\prod_{k=1}^{n-1}(e+\ell_k(t))}\int_t^{xt}\frac{\mathrm{d}s}{s}=\frac{\log x}{\prod_{k=1}^{n-1}(e+\ell_k(t))}.$$

In particular, we may add $c\geqslant 0$ to $\ell_n(xt)$ and $\ell_n(t)$ and divide by $c+\ell_n(t)$ to obtain

$$\frac{c+\ell_n(xt)}{c+\ell_n(t)}\leqslant 1+\frac{\log x}{(c+\ell_n(t))\prod_{k=1}^{n-1}(e+\ell_k(t))}\leqslant 1+\log x,\quad n\in\mathbb{N},\,x>1,\,t>0,$$

implying Part (a). Part (b) is obvious, so we need only establish Part (c).

It is simple to show that for $a_1,a_2,b_1,b_2>0$, the fraction $(a_1+a_2)/(b_1+b_2)$ lies between a_1/b_1 and a_2/b_2. An inductive argument implies that for any $a_1,\ldots,a_k,b_1,\ldots,b_k>0$, we have

$$\min_{j\in\{1,\ldots,k\}}\frac{a_j}{b_j}\leqslant\frac{\sum_{j=1}^k a_j}{\sum_{j=1}^k b_j}\leqslant\max_{j\in\{1,\ldots,k\}}\frac{a_j}{b_j}.$$

Thus, by virtue of Parts (a) and (b) and denoting $\widetilde{\ell}_j(t):=(e+t)\ell_j'(t)$, we have

$$\begin{aligned}\Sigma(x)&=\sup_{t>0,\,y\in[x\wedge 1,x\vee 1]}\frac{\ell(yt)\sum_{j=n}^m|q_j|\widetilde{\ell}_j(yt)/\ell_j(yt)}{\ell(t)\sum_{j=n}^m|q_j|\widetilde{\ell}_j(t)/\ell_j(t)}\leqslant\sup_{t>0,\,y\in[x\wedge 1,x\vee 1]}\max_{j\in\{n,\ldots,m\},\,q_j\neq 0}\frac{\ell(yt)\widetilde{\ell}_j(yt)/\ell_j(yt)}{\ell(t)\widetilde{\ell}_j(t)/\ell_j(t)}\\&=\sup_{t>0,\,y\in[x\wedge 1,x\vee 1]}\max_{j\in\{n,\ldots,m\},\,q_j\neq 0}\prod_{i=1}^{j-1}\frac{e+\ell_i(t)}{e+\ell_i(yt)}\cdot\prod_{i=n}^m\left(\frac{\ell_i(yt)}{\ell_i(t)}\right)^{q_i-\mathbb{1}_{\{i=j\}}}\\&\leqslant\begin{cases}(1+\log x)^{\sum_{j=n}^m q_j^+}, & x\geqslant 1,\\ (1+|\log x|)^{m+\sum_{j=n}^m q_j^-}, & x<1.\end{cases}\end{aligned}$$

□

Proof of Theorem 2.3. Part (a). Recall from Remark 5.8, that we can decompose $\boldsymbol{X}$ as the sum $\boldsymbol{S}+\boldsymbol{R}$ of the independent processes $\boldsymbol{S}$ and $\boldsymbol{R}$ with generating triplets $(\boldsymbol{\gamma_S},\boldsymbol{0},\nu_{\boldsymbol{X}}^{\mathrm{c}})$ and $(\boldsymbol{\gamma_R},\boldsymbol{0},\nu_{\boldsymbol{X}}^{\mathrm{d}})$, respectively. For $t\in[0,1]$, let $\boldsymbol{S}^t=(\boldsymbol{S}_{st}/g(t))_{s\in[0,1]}$ and $\boldsymbol{R}^t=(\boldsymbol{R}_{st}/g(t))_{s\in[0,1]}$. Assume that $(\boldsymbol{M}^{\boldsymbol{S}^t},\boldsymbol{M}^{\boldsymbol{Z}},\boldsymbol{L}^{\boldsymbol{S}^t},\boldsymbol{L}^{\boldsymbol{Z}})$ is coupled as in (14) and (15).

Note that $\mathcal{W}_q(\boldsymbol{X}^t,\boldsymbol{Z})\leqslant\mathcal{W}_q(\boldsymbol{S}^t,\boldsymbol{Z})+\mathbb{E}[\sup_{t\in[0,1]}|\boldsymbol{R}_s^t|^q]$ by the triangle inequality. Next, we apply (4) with $p=1$ to $\mathcal{W}_q(\boldsymbol{S}^t,\boldsymbol{Z})$ and use Theorem 5.9 and Potter's bounds [6, Thm 1.5.6] (applied to the slowly varying function G_2) to show that each resulting term is $\mathcal{O}(G_2(t)^q)$:

- $\mathcal{W}_q(\boldsymbol{M}^{\boldsymbol{S}^t},\boldsymbol{M}^{\boldsymbol{Z}})\leqslant\mathcal{W}_2(\boldsymbol{M}^{\boldsymbol{S}^t},\boldsymbol{M}^{\boldsymbol{Z}})^q$ by (22), and (35) then yields the bound;
- $\mathcal{W}_q(\boldsymbol{L}^{\boldsymbol{S}^t},\boldsymbol{L}^{\boldsymbol{Z}})$ is bounded by (36);
- $|\boldsymbol{\varpi}_{\boldsymbol{S}^t}-\boldsymbol{\varpi}_{\boldsymbol{Z}}|^q$ is bounded by (37).

Similarly, by Potter's bounds and Theorem 5.9, $\mathbb{E}[\sup_{t\in[0,1]}|\boldsymbol{R}_s^t|^q]=\mathcal{O}(t^{1-q/\alpha}G(1/t)^{-q})=\mathcal{O}(G_2(t)^q)$.
Part (b). Recall $a(t)=G(1/(2t))/G(1/t)\to 1$ as $t\to 0$. Directly from Proposition 6.1, we see that

$$3\max\{\mathcal{W}_q(\boldsymbol{X}_1^t,\boldsymbol{Z}_1),\mathcal{W}_q(\boldsymbol{X}_1^{2t},\boldsymbol{Z}_1)\}\geqslant|1-a(t)^q|\mathbb{E}[|\boldsymbol{Z}_1|^q]\quad\text{for all sufficiently small } t>0,$$

yielding the first claim of part (b). The second claim follows from Lemma 7.2 since G is assumed not to have a positive finite limit. □

Proof of Corollary 2.4. Part (b) follows from Lemma 7.3 above. Given Theorem 2.3, it suffices to show that the assumptions in Corollary 2.4(a) imply those of Theorem 2.3 and that the upper and lower bounds have the desired form. These facts follow from Lemmas 7.1 & 7.3. Indeed, for instance, the function G_1 in Assumption (S) is given by the upper bound on $x\mapsto\Sigma(x)|\log x|$ given in Lemma 7.3, where Σ is as in Lemma 7.1. □

8. Concluding remarks

Over small time horizons, a Lévy process may be attracted to an α-stable process with heavy tails (i.e. $\alpha\in(0,2)$) or Brownian motion (i.e. $\alpha=2$). In this paper, we established upper and lower bounds on the rate of convergence in L^q-Wasserstein distance in both regimes, as listed below.

- For $\alpha\in(0,2]\setminus\{1\}$ and processes in the domain of non-normal attraction, the Wasserstein distance is bounded above and below by slowly varying functions (see Theorem 2.3), both of which are slower than any power of logarithm greater than 1.
- For $\alpha\in(0,2)\setminus\{1\}$ and processes in the domain of normal attraction, we establish upper and lower bounds that are polynomial in t (see Theorem 2.1). The established bounds are often rate optimal in L^q-Wasserstein distance for $q<\alpha<1$ or $q=1<\alpha$ and proportional to $t^{1-q/\alpha}$.
- For $\alpha=2$ (i.e. Brownian limit) and processes in the domain of normal attraction, the established upper and lower bounds are also polynomial in t (see Theorem 2.8). In this case, the bounds are rate optimal when the Blumenthal–Getoor index $\beta\leqslant 1$ and otherwise there is a polynomial gap. This suggests at least one of the bounds is not sharp. Establishing sharper bounds in this special case is nontrivial (as classical tools such as the Berry–Esseen theorem fail to provide converging bounds) and is therefore left for future work.

The process $\boldsymbol{R}$ in Assumption (T) (resp. (C)) in the thinning (resp. comonotonic) coupling is assumed to have finitely many jumps on compact time intervals. Our results can be extended to the case where this process has infinitely many jumps on compact time intervals and a Blumenthal–Getoor index $\beta<\alpha$. For such an extension, the moments of $\sup_{s\in[0,1]}|\boldsymbol{R}_s^t|$, as a function of t, can be controlled via Lemma 5.2 and would result in worse convergence rates as $\beta\uparrow\alpha$. We chose not to include this simple extension as the convergence rates would be much harder to express in terms of all the model parameters, resulting in a less concise presentation of our results.

The tools developed in Section 4 could be used for the omitted case $\alpha=1$ to establish upper and lower bounds on the Wasserstein distance in the domains of normal and non-normal attraction. However, as multiple cases would arise, requiring careful treatment of the emerging slowly varying functions, we leave such extension for future work.

The upper bounds in the heavy-tailed case $\alpha\in(0,2)\setminus\{1\}$ are based on two distinct couplings introduced in Section 4: the comonotonic and thinning couplings. We mention briefly that it is possible to combine both couplings. Consider two Lévy processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ with Lévy measures $f_{\boldsymbol{X}}\mathrm{d}\mu$ and $f_{\boldsymbol{Y}}\mathrm{d}\mu$ where $0\leqslant f_{\boldsymbol{X}}\leqslant 1$ and $0\leqslant f_{\boldsymbol{Y}}\leqslant 1$ are measurable and μ is a Lévy measure. It is then possible to first apply the thinning coupling to synchronise the jumps arising from the Lévy measure $g\mathrm{d}\mu$, where $g:=\min\{f_{\boldsymbol{X}},f_{\boldsymbol{Y}}\}$, and then apply the comonotonic coupling to the remaining jumps of

$\boldsymbol{X}$ and $\boldsymbol{Y}$ with corresponding Lévy measures $(f_{\boldsymbol{X}} - g)\mathrm{d}\mu$ and $(f_{\boldsymbol{Y}} - g)\mathrm{d}\mu$. It appears, however, that this combined coupling does not yield improved rates of convergence to heavy-tailed stable limits as each coupling already attains optimal rates of convergence in most cases. We expect such a combined coupling to reduce the L^p-distance between coupled processes by a constant factor.

When $\alpha \in (0, 2)$, the α-stable limit has heavy tails and its q-moment is finite if and only if $q < \alpha$, making it impossible to obtain general converging bounds in the L^2-Wasserstein distance, which play a key role in various applications including Multilevel Monte Carlo. However, substituting the standard Euclidean metric on $\mathbb{R}^d$ with an equivalent *bounded* metric would remove this obstruction. We expect our couplings to perform well and have a fast converging L^2-Wasserstein distance under the bounded metric on $\mathbb{R}^d$. Such an extension of our results is also left for future work.

The present work focused on the small-time stable domain of attraction where the small jumps of the process dominate the activity. Finally we remark that it is natural to expect that the couplings developed in this paper could also typically achieve asymptotically optimal convergence rate in the Wasserstein distance in the scaling limits of the long-time stable domain of attraction. This is because in the long-time horizon regime the activity in the limit is dominated by the large jumps of the Lévy process, which are also efficiently coupled under the couplings of Section 4 above.

Acknowledgements

JGC and AM were supported by EPSRC grant EP/V009478/1 and The Alan Turing Institute under the EPSRC grant EP/N510129/1; AM was also supported by EPSRC grant EP/W006227/1; DKB was funded by the CDT in Mathematics and Statistics at The University of Warwick and is supported by AUFF NOVA grant AUFF-E-2022-9-39. The authors would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support during the INI satellite programme *Heavy tails in machine learning*, hosted by The Alan Turing Institute, London, and the INI programme *Stochastic systems for anomalous diffusion* hosted at INI in Cambridge, where work on this paper was undertaken. This work was supported by EPSRC grant EP/R014604/1.

References

[1] S. Asmussen and J. Ivanovs. Discretization error for a two-sided reflected Lévy process. *Queueing Syst.*, 89(1-2):199–212, 2018.

[2] S. Asmussen and J. Rosiński. Approximations of small jumps of Lévy processes with a view towards simulation. *J. Appl. Probab.*, 38(2):482–493, 2001.

[3] G. Auricchio, A. Codegoni, S. Gualandi, G. Toscani, and M. Veneroni. The equivalence of Fourier-based and Wasserstein metrics on imaging problems. *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.*, 31(3):627–649, 2020.

[4] D. Bang, J. González Cázares, and A. Mijatović. A Gaussian approximation theorem for Lévy processes. *Statist. Probab. Lett.*, 178:Paper No. 109187, 4, 2021.

[5] J. Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.

[6] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989.

[7] K. Bisewski and J. Ivanovs. Zooming-in on a Lévy process: failure to observe threshold exceedance over a dense grid. *Electron. J. Probab.*, 25:Paper No. 113, 33, 2020.

[8] J. Blanchet and K. Murthy. Quantifying distributional model risk via optimal transport. *Math. Oper. Res.*, 44(2):565–600, 2019.

[9] C. Börgers and C. Greengard. Slow convergence in generalized central limit theorems. *C. R. Math. Acad. Sci. Paris*, 356(6):679–685, 2018.

[10] M. Broadie, P. Glasserman, and S. Kou. A continuity correction for discrete barrier options. *Math. Finance*, 7(4):325–349, 1997.

[11] M. Broadie, P. Glasserman, and S. G. Kou. Connecting discrete and continuous path-dependent options. *Finance Stoch.*, 3(1):55–82, 1999.

[12] M. E. Caballero, J. C. Pardo, and J. L. Pérez. On Lamperti stable processes. *Probab. Math. Statist.*, 30(1):1–28, 2010.

[13] S. Cohen and J. Rosiński. Gaussian approximation of multivariate Lévy processes with applications to simulation of tempered stable processes. *Bernoulli*, 13(1):195–210, 2007.

[14] E. H. A. Dia and D. Lamberton. Connecting discrete and continuous lookback or hindsight options in exponential Lévy models. *Adv. in Appl. Probab.*, 43(4):1136–1165, 2011.

[15] E. H. A. Dia and D. Lamberton. Continuity correction for barrier options in jump-diffusion models. *SIAM J. Financial Math.*, 2(1):866–900, 2011.

[16] C. Duval, T. Jalal, and E. Mariucci. Nonparametric density estimation for the small jumps of lévy processes. 2024.

[17] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. For insurance and finance.

[18] V. Fomichov, J. González Cázares, and J. Ivanovs. Implementable coupling of Lévy process and Brownian motion. *Stochastic Process. Appl.*, 142:407–431, 2021.

[19] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *INTERNAT. STATIST. REV.*, pages 419–435, 2002.

[20] B. V. Gnedenko and A. N. Kolmogorov. *Limit distributions for sums of independent random variables.* Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills., Ont., revised edition, 1968. Translated from the Russian, annotated, and revised by K. L. Chung, With appendices by J. L. Doob and P. L. Hsu.

[21] J. González Cázares, D. Kramer-Bang, and A. Mijatović. Presentation on "asymptotically optimal Wasserstein couplings for the small-time stable domain of attraction". YouTube presentation on the YouTube channel Prob-AM, 2024.

[22] J. González Cázares and A. Mijatović. Simulation of the drawdown and its duration in Lévy models via stick-breaking Gaussian approximation. *Finance Stoch.*, 26(4):671–732, 2022.

[23] J. I. González Cázares, A. Mijatović, and G. Uribe Bravo. Geometrically convergent simulation of the extrema of lévy processes. *Mathematics of Operations Research*, 47(2):1141–1168, 2022.

[24] F. Götze. On the rate of convergence in the multivariate CLT. *Ann. Probab.*, 19(2):724–739, 1991.

[25] P. Hall. Two-sided bounds on the rate of convergence to a stable law. *Z. Wahrsch. Verw. Gebiete*, 57(3):349–364, 1981.

[26] J. Ivanovs. Zooming in on a Lévy process at its supremum. *Ann. Appl. Probab.*, 28(2):912–940, 2018.

[27] J. Ivanovs and J. D. Thøstesen. Discretization of the Lamperti representation of a positive self-similar Markov process. *Stochastic Process. Appl.*, 137:200–221, 2021.

[28] O. Johnson and R. Samworth. Central limit theorem and convergence to stable laws in Mallows distance. *Bernoulli*, 11(5):829–845, 2005.

[29] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[30] W. S. Kendall, M. B. Majka, and A. Mijatović. Optimal Markovian coupling for finite activity Lévy processes. *Bernoulli*, 30(4):2821–2845, 2024.

[31] J. F. C. Kingman. *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1993. Oxford Science Publications.

[32] R. LePage. Multidimensional infinitely divisible variables and processes. II. In *Probability in Banach spaces, III (Medford, Mass., 1980)*, volume 860 of *Lecture Notes in Math.*, pages 279–284. Springer, Berlin-New York, 1981.

[33] S. Manou-Abi. Rate of convergence to alpha stable law using Zolotarev distance. *Journal of Statistics: Advances in Theory and Applications*, 18:166–177, 2017.

[34] E. Mariucci and M. Reiß. Wasserstein and total variation distance between marginals of Lévy processes. *Electron. J. Stat.*, 12(2):2482–2514, 2018.

[35] P. Pegon, F. Santambrogio, and D. Piazzoli. Full characterization of optimal transport plans for concave costs. *Discrete Contin. Dyn. Syst.*, 35(12):6113–6132, 2015.

[36] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.

[37] E. Rio. Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):802–817, 2009.

[38] J. Rosiński. Series representations of Lévy processes from the perspective of point processes. In *Lévy processes*, pages 401–415. Birkhäuser Boston, Boston, MA, 2001.

[39] J. Rosiński. Tempering stable processes. *Stochastic Process. Appl.*, 117(6):677–707, 2007.

[40] K.-i. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2013. Translated from the 1990 Japanese original, Revised edition of the 1999 English translation.

[41] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.

Appendix A. Proof of Lemma 5.2

Given any $\kappa \in (0,1]$ consider the Lévy–Itô decomposition $\boldsymbol{X}_t = \boldsymbol{\gamma}_{\boldsymbol{X}}^{(\kappa)} t + \boldsymbol{\Sigma}_{\boldsymbol{X}} \boldsymbol{B}_t + \boldsymbol{D}_t^{(\kappa)} + \boldsymbol{J}_t^{(\kappa)}$ given in (3), where $\boldsymbol{B}$ is a standard Brownian motion, $\boldsymbol{D}^{(\kappa)}$ is the pure-jump martingale containing all the jumps of $\boldsymbol{X}$ of magnitude less than κ and $\boldsymbol{J}^{(\kappa)}$ is the driftless compound Poisson process containing all the jumps of $\boldsymbol{X}$ of magnitude at least κ. Since $|\boldsymbol{\Sigma}_{\boldsymbol{X}} \boldsymbol{B}_s| \leqslant |\boldsymbol{\Sigma}_{\boldsymbol{X}}| \cdot |\boldsymbol{B}_s|$, we have

$$\sup_{s\in[0,t]} |\boldsymbol{X}_s| \leqslant |\boldsymbol{\gamma}_{\boldsymbol{X}}^{(\kappa)}| t + |\boldsymbol{\Sigma}_{\boldsymbol{X}}| \sup_{s\in[0,t]} |\boldsymbol{B}_s| + \sup_{s\in[0,t]} |\boldsymbol{D}_s^{(\kappa)}| + \sup_{s\in[0,t]} |\boldsymbol{J}_s^{(\kappa)}|, \qquad \text{for } t \in [0,1]. \tag{44}$$

By the elementary bound $(\sum_{i=1}^n x_i)^p \leqslant n^{(p-1)^+} \sum_{i=1}^n x_i^p$, $p > 0$, $(p-1)^+ = \max\{p-1, 0\}$ and $x_i \geqslant 0$, $i \in \{1, \ldots, n\}$, we only need to bound the p-th moment of each summand on the right-hand side of the display above. Recall that β_+ is the quantity associated to the BG index of $\boldsymbol{X}$ defined in (27).

Case $\beta_+ > 0$. Define $\kappa := t^{1/\beta_+}$. To bound the drift term $|\boldsymbol{\gamma}_{\boldsymbol{X}}^{(\kappa)}| t$, first assume $\beta_+ \geqslant 1$ and note that

$$\begin{aligned}
|\boldsymbol{\gamma}_{\boldsymbol{X}}^{(\kappa)}| = \left| \boldsymbol{\gamma}_{\boldsymbol{X}} - \int_{B_{\boldsymbol{0}}(1)\setminus B_{\boldsymbol{0}}(\kappa)} \boldsymbol{w} \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \right| &\leqslant |\boldsymbol{\gamma}_{\boldsymbol{X}}| + \int_{B_{\boldsymbol{0}}(1)\setminus B_{\boldsymbol{0}}(\kappa)} |\boldsymbol{w}| \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \\
&\leqslant |\boldsymbol{\gamma}_{\boldsymbol{X}}| + \int_{B_{\boldsymbol{0}}(1)\setminus B_{\boldsymbol{0}}(\kappa)} \kappa^{1-\beta_+} |\boldsymbol{w}|^{\beta_+} \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \leqslant |\boldsymbol{\gamma}_{\boldsymbol{X}}| + \kappa^{1-\beta_+} I_{\beta_+}.
\end{aligned}$$

Thus, $(|\boldsymbol{\gamma}_{\boldsymbol{X}}^{(\kappa)}|t)^p$ is bounded by a constant multiple of $t^p + t^{p/\beta_+}$. If $\beta_+ \in (0,1)$ and the natural drift of $\boldsymbol{X}$ is zero (i.e. $\boldsymbol{\gamma}_{\boldsymbol{X}} = \int_{B_{\mathbf{0}}(1)\setminus\{\mathbf{0}\}} \boldsymbol{w}\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})$), then $|\boldsymbol{\gamma}_{\boldsymbol{X}}^{(\kappa)}|$ is bounded (and convergent) as $t \downarrow 0$,

$$|\boldsymbol{\gamma}_{\boldsymbol{X}}^{(\kappa)}| = \left| \int_{B_{\mathbf{0}}(\kappa)\setminus\{\mathbf{0}\}} \boldsymbol{w}\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \right| \leqslant \int_{B_{\mathbf{0}}(\kappa)\setminus\{\mathbf{0}\}} |\boldsymbol{w}|\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \\ \leqslant \int_{B_{\mathbf{0}}(\kappa)\setminus\{\mathbf{0}\}} \kappa^{1-\beta_+}|\boldsymbol{w}|^{\beta_+}\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \leqslant \kappa^{1-\beta_+} I_{\beta_+},$$

making $(|\boldsymbol{\gamma}_{\boldsymbol{X}}^{(\kappa)}|t)^p$ bounded by a multiple of t^{p/β_+}. Hence, in this case, we may take $C_2 = 0$ in Lemma 5.2 (it will become clear from the remainder of the proof that none of the other summands on the right-hand side of the inequality in (44) will produce a term of order t^p).

The p-th moment of the Brownian term is easily bounded by a constant multiple of $t^{p/2}$ since we have $\sup_{s\in[0,t]} |\boldsymbol{B}_s| \overset{d}{=} t^{1/2} \sup_{s\in[0,1]} |\boldsymbol{B}_s|$. If $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ is a zero matrix, then $|\boldsymbol{\Sigma}_{\boldsymbol{X}}| = 0$ and hence $C_1 = 0$.

Next, we bound the big-jump term $\boldsymbol{J}^{(\kappa)}$. Let $A_\kappa := \mathbb{R}^d \setminus B_{\mathbf{0}}(\kappa)$ and recall that $\boldsymbol{J}_t = \sum_{k=1}^{N_t} \boldsymbol{R}_k$ for some Poisson random variable N_t with mean $t\nu_{\boldsymbol{X}}(A_\kappa)$ and iid random vectors $(\boldsymbol{R}_n)_{n\in\mathbb{N}}$ independent of N_t with law $\nu_{\boldsymbol{X}}(\cdot \cap A_\kappa)/\nu_{\boldsymbol{X}}(A_\kappa)$. Recall, from the formula for the moments of a Poisson random variable, that $\mathbb{E}[N_t^k] = \sum_{j=1}^k \left\{ {k \atop j} \right\} (t\nu_{\boldsymbol{X}}(A_\kappa))^j$, where $\left\{ {k \atop j} \right\}$ denotes the Stirling number of the second kind. Note that the triangle inequality of the Euclidean norm $|\cdot|$ implies that

$$|\boldsymbol{J}_s^{(\kappa)}|^p \leqslant \left(\sum_{k=1}^{N_s} |\boldsymbol{R}_k| \right)^p \leqslant \left(\sum_{k=1}^{N_t} |\boldsymbol{R}_k| \right)^p \leqslant N_t^{(p-1)^+} \sum_{k=1}^{N_t} |\boldsymbol{R}_k|^p, \qquad \text{for every } s \in [0,t].$$

Denote $\lceil p \rceil := \inf\{n \in \mathbb{N} : n \geqslant p\}$, and note that since $\boldsymbol{R}_k$, $k \in \mathbb{N}$, are iid and independent of N_t and $1 \leqslant (p-1)^+ + 1 \leqslant \lceil p \rceil$, we find

$$\mathbb{E}\left[\sup_{s\in[0,t]} |\boldsymbol{J}_s^{(\kappa)}|^p \right] \leqslant \mathbb{E}\big[|\boldsymbol{R}_1|^p\big]\mathbb{E}\Big[N_t^{\lceil p \rceil}\Big] = \int_{A_\kappa} |\boldsymbol{w}|^p \frac{\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})}{\nu_{\boldsymbol{X}}(A_\kappa)} \cdot \sum_{k=1}^{\lceil p \rceil} \left\{ {\lceil p \rceil \atop k} \right\} (t\nu_{\boldsymbol{X}}(A_\kappa))^k \\ = t \int_{A_\kappa} |\boldsymbol{w}|^p \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \cdot \sum_{k=1}^{\lceil p \rceil} \left\{ {\lceil p \rceil \atop k} \right\} (t\nu_{\boldsymbol{X}}(A_\kappa))^{k-1}. \tag{45}$$

Note that $\nu_{\boldsymbol{X}}(A_\kappa) \leqslant \nu_{\boldsymbol{X}}(A_1) + \int_{B_{\mathbf{0}}(1)\setminus B_{\mathbf{0}}(\kappa)} \kappa^{-\beta_+}|\boldsymbol{w}|^{\beta_+}\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \leqslant \nu_{\boldsymbol{X}}(A_1) + \kappa^{-\beta_+} I_{\beta_+}$ and hence $t\nu_{\boldsymbol{X}}(A_\kappa)$ is bounded in $t \in [0,1]$ (recall that $\kappa = t^{1/\beta_+}$), making the sum in the display also bounded. Denote $I_p' = \int_{A_1} |\boldsymbol{w}|^p \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})$, which we assumed finite, and hence

$$\int_{A_\kappa} |\boldsymbol{w}|^p \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \leqslant I_p' + \int_{B_{\mathbf{0}}(1)\setminus B_{\mathbf{0}}(\kappa)} |\boldsymbol{w}|^p \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \\ \leqslant I_p' + \int_{B_{\mathbf{0}}(1)\setminus B_{\mathbf{0}}(\kappa)} \kappa^{-(\beta_+-p)^+}|\boldsymbol{w}|^{\max\{\beta_+,p\}}\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \leqslant I_p' + \kappa^{-(\beta_+-p)^+} I_{\max\{\beta_+,p\}}.$$

Thus, there is a finite constant $C > 0$ such that

$$\mathbb{E}\left[\sup_{s\in[0,t]} |\boldsymbol{J}_s^{(\kappa)}|^p \right] \leqslant Ct(I_p' + \kappa^{-(\beta_+-p)^+} I_{\max\{\beta_+,p\}}) = C(I_p' t + I_{\max\{\beta_+,p\}} t^{\min\{1,p/\beta_+\}}).$$

In the case where $\beta_+ > 0$, it remains to bound the small-jump term $\boldsymbol{D}^{(\kappa)}$. In this case, we show that the p-th moment is bounded by a multiple of t^{p/β_+}. We may assume without loss of generality that $p > 1$, since the other cases would follow by Jensen's inequality since $\mathbb{E}[|\boldsymbol{\xi}|^q] \leqslant \mathbb{E}[|\boldsymbol{\xi}|^p]^{q/p}$ for any $q \leqslant p$. Since $|\boldsymbol{D}_t^{(\kappa)}|$ is a submartingale, Doob's maximal inequality and the elementary inequality $|x|^p \leqslant (p/e)^p e^{|x|}$ imply

$$\mathbb{E}\left[\sup_{s\in[0,t]} |\boldsymbol{D}_s^{(\kappa)}|^p \right] \leqslant \left(\frac{p}{p-1} \right)^p \mathbb{E}\big[|\boldsymbol{D}_t^{(\kappa)}|^p\big] = \left(\frac{\kappa p}{p-1} \right)^p \mathbb{E}\big[|\kappa^{-1}\boldsymbol{D}_t^{(\kappa)}|^p\big] \leqslant \left(\frac{\kappa p^2/e}{p-1} \right)^p \mathbb{E}\Big[e^{\kappa^{-1}|\boldsymbol{D}_t^{(\kappa)}|}\Big],$$

for $t \in [0,1]$. Thus, to complete the proof it suffices to show that the expectation on the right is bounded as $t \downarrow 0$. Let $\{\boldsymbol{e}_i\}_{i=1}^{2^d}$ be the vertices of the hypercube centered at the origin with sides parallel to the axes and side length 2 (e.g., the vectors $(1,1,\dots,1)$ and $(-1,-1,\dots,-1)$ are opposite vertices of this hypercube). Note that $e^{|\boldsymbol{w}|} \leqslant e^{|\boldsymbol{w}|_1} \leqslant \sum_{i=1}^{2^d} e^{\langle \boldsymbol{e}_i, \boldsymbol{w}\rangle}$ where $|(s_1,\dots,s_d)|_1 := \sum_{i=1}^d |s_i|$ denotes the ℓ^1-norm in $\mathbb{R}^d$. Hence, it suffices to show that $\mathbb{E}[\exp(\langle \kappa^{-1}\boldsymbol{e}_i, \boldsymbol{D}_t^{(\kappa)}\rangle)]$ is bounded as $t \downarrow 0$ for each $i \in \{1,\dots,2^d\}$. The Lévy–Khintchine formula, the elementary inequality $e^x - 1 - x \leqslant e^c x^2$ for $x \in [-c,c]$, $c>0$, and the Cauchy-Schwarz inequality $|\langle \kappa^{-1}\boldsymbol{e}_i, \boldsymbol{w}\rangle| \leqslant |\boldsymbol{e}_i| = \sqrt{d}$ for all $\boldsymbol{w} \in B_{\boldsymbol{0}}(\kappa)$ and $i \in \{1,\dots,2^d\}$ yield

$$\begin{aligned}\log \mathbb{E}\big[e^{\langle \kappa^{-1}\boldsymbol{e}_i, \boldsymbol{D}_t^{(\kappa)}\rangle}\big] &= t\int_{B_{\boldsymbol{0}}(\kappa)\setminus\{\boldsymbol{0}\}} \big(e^{\langle \kappa^{-1}\boldsymbol{e}_i, \boldsymbol{w}\rangle} - 1 - \langle \kappa^{-1}\boldsymbol{e}_i, \boldsymbol{w}\rangle\big)\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})\\ &\leqslant t\int_{B_{\boldsymbol{0}}(\kappa)\setminus\{\boldsymbol{0}\}} e^{\sqrt{d}}\kappa^{-2}\langle \boldsymbol{e}_i, \boldsymbol{w}\rangle^2\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \leqslant t\int_{B_{\boldsymbol{0}}(\kappa)\setminus\{\boldsymbol{0}\}} de^{\sqrt{d}}\kappa^{-2}|\boldsymbol{w}|^2\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})\\ &\leqslant t\int_{B_{\boldsymbol{0}}(\kappa)\setminus\{\boldsymbol{0}\}} de^{\sqrt{d}}\kappa^{-\beta_+}|\boldsymbol{w}|^{\beta_+}\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w}) \leqslant de^{\sqrt{d}} I_{\beta_+},\end{aligned}$$

completing the proof in the case $\beta_+ > 0$.

Case $\beta_+ = 0$. Note in this case, that the pure-jump component of $\boldsymbol{X}$ is compound Poisson. Thus, as in (44), we have that $\sup_{s\in[0,t]}|\boldsymbol{X}_s| \leqslant |\boldsymbol{\gamma}_{\boldsymbol{X}}|t + |\boldsymbol{\Sigma}_{\boldsymbol{X}}|\sup_{s\in[0,t]}|\boldsymbol{B}_s| + \sup_{s\in[0,t]}|\widetilde{\boldsymbol{J}}_s|$ for all $t\in[0,1]$, where $\widetilde{\boldsymbol{J}}_t = \boldsymbol{X}_t - \boldsymbol{\gamma}_{\boldsymbol{X}}t - \boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{B}_t$ for all $t\in[0,1]$. The bound on the p-moment of the Brownian term follows exactly as in the case of $\beta_+>0$ and is a constant multiple of $t^{p/2}$. From the term $(|\boldsymbol{\gamma}_{\boldsymbol{X}}|t)^p$, we get a multiple of t^p. Note that $\widetilde{\boldsymbol{J}}_s$ is a compound Poisson process with finitely many jumps on $\mathbb{R}^d\setminus\{\boldsymbol{0}\}$, with $\beta=0$, and hence $\widetilde{\boldsymbol{J}}_t = \sum_{n=1}^{N_t}\widetilde{\boldsymbol{R}}_k$ for some Poisson random variable N_t with mean $t\nu_{\boldsymbol{X}}(\mathbb{R}^d\setminus\{\boldsymbol{0}\})$ and iid random vectors $(\widetilde{\boldsymbol{R}}_n)_{n\in\mathbb{N}}$ independent of N_t with law $\nu_{\boldsymbol{X}}(\cdot\cap(\mathbb{R}^d\setminus\{\boldsymbol{0}\}))/\nu_{\boldsymbol{X}}(\mathbb{R}^d\setminus\{\boldsymbol{0}\})$. For the term $\mathbb{E}[\sup_{s\in[0,t]}|\widetilde{\boldsymbol{J}}_s|^p]$, we now use the same proof as in the case of $\beta_+>0$, until (45). Hence, we see that

$$\mathbb{E}\Big[\sup_{s\in[0,t]}|\widetilde{\boldsymbol{J}}_s|^p\Big] \leqslant d^{p/2}t\int_{\mathbb{R}^d\setminus\{\boldsymbol{0}\}}|\boldsymbol{w}|^p\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})\cdot\sum_{k=1}^{\lceil p\rceil}\left\{\begin{matrix}\lceil p\rceil\\ k\end{matrix}\right\}(t\nu_{\boldsymbol{X}}(\mathbb{R}^d\setminus\{\boldsymbol{0}\}))^{k-1}.$$

Since $\widetilde{\boldsymbol{J}}$ has finite activity, it follows that $\int_{\mathbb{R}^d\setminus\{\boldsymbol{0}\}}|\boldsymbol{w}|^p\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{w})<\infty$. Moreover, since the sum in the display above is bounded in $t\in[0,1]$, we get that $\mathbb{E}[\sup_{s\in[0,t]}|\widetilde{\boldsymbol{J}}_s|^p]$ is bounded by a multiple of t, concluding the proof of Lemma 5.2.

Appendix B. Small-time domains of attraction - proof of Theorem 5.1

The proof is essentially a consequence of [29, Thm 15.14] and [26, Thm 2]. Recall that $B_{\boldsymbol{a}}(r) = \{\boldsymbol{x}\in\mathbb{R}^d : |\boldsymbol{x}-\boldsymbol{a}|<r\}$ denotes the open ball in $\mathbb{R}^d$ with center $\boldsymbol{a}\in\mathbb{R}^d$ and radius $r>0$, by $\mathbb{S}^{d-1}$ the unit sphere in $\mathbb{R}^d$ and define $\mathscr{L}_{\boldsymbol{a}}(r) := \{\boldsymbol{x}\in\mathbb{R}^d : \langle \boldsymbol{a},\boldsymbol{x}\rangle \geqslant r\}$.

Since $\boldsymbol{X}$ and $\boldsymbol{Z}$ are Lévy processes, the stated weak convergence is equivalent to $\langle\boldsymbol{\lambda},\boldsymbol{X}_t\rangle/g(t)\xrightarrow{d}\langle\boldsymbol{\lambda},\boldsymbol{Z}_1\rangle$ as $t\downarrow 0$ for any $\boldsymbol{\lambda}\in\mathbb{R}^d$ by [29, Cor. 15.7]. By [26, Thm 2], it follows that, for some $\alpha\in(0,2]$, $g(t)=t^{1/\alpha}G(t^{-1})$ for $t>0$ where G is a slowly varying function at infinity and, moreover, $\langle\boldsymbol{\lambda},\boldsymbol{Z}_1\rangle$ is α-stable for all $\boldsymbol{\lambda}\in\mathbb{R}^d$. Thus, $\boldsymbol{Z}$ is itself α-stable. We then have the following cases.

If $\alpha=2$, then, by [26, Thm 2(i)], the weak convergence in the direction $\boldsymbol{v}\in\mathbb{S}^{d-1}$ is equivalent to

$$G(t^{-1})^{-2}\bigg(|\boldsymbol{v}^\top\boldsymbol{\Sigma}_{\boldsymbol{X}}|^2 + \int_{B_{\boldsymbol{0}}(g(t))\setminus\{\boldsymbol{0}\}}|\langle\boldsymbol{v},\boldsymbol{x}\rangle|^2\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})\bigg)\to|\boldsymbol{v}^\top\boldsymbol{\Sigma}_{\boldsymbol{Z}}|^2,\quad \text{as } t\downarrow 0,$$

so the weak convergence in $\mathbb{R}^d$ is equivalent to (24), completing the proof in this case.

If $\alpha\in(1,2)$, the weak convergence in the direction $\boldsymbol{v}\in\mathbb{S}^{d-1}$ is equivalent to (25) by [26, Thm 2(iii)], completing the proof in this case.

If $\alpha \in (0,1)$, the weak convergence in the direction $\boldsymbol{v} \in \mathbb{S}^{d-1}$ is equivalent to (25) and $\langle \boldsymbol{v}, \boldsymbol{X}\rangle$ having zero natural drift by [26, Thm 2(iii)]. Since the latter condition is required for all $\boldsymbol{v} \in \mathbb{R}^d$, it is equivalent to $\boldsymbol{X}$ having zero natural drift $\boldsymbol{\gamma_X} = \int_{B_{\boldsymbol{0}}(1)\setminus\{\boldsymbol{0}\}} \boldsymbol{x}\nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})$, completing the proof in this case.

If $\alpha = 1$, the weak convergence in the direction $\boldsymbol{v} \in \mathbb{S}^{d-1}$ may be different depending on the behaviour of the limiting process in this direction. If $\nu_{\boldsymbol{Z}}(\mathscr{L}_{\boldsymbol{v}}(1)) = 0$ then $\langle \boldsymbol{v}, \boldsymbol{Z}\rangle$ is a linear drift and the weak convergence, by [26, Thm 2(ii)], is equivalent to the following two limits as $t \downarrow 0$:

$$\frac{t}{g(t)}\bigg(\langle \boldsymbol{v}, \boldsymbol{\gamma_X}\rangle - \int_{B_{\boldsymbol{0}}(1)\setminus B_{\boldsymbol{0}}(g(t))} \langle \boldsymbol{v}, \boldsymbol{x}\rangle \nu(\mathrm{d}\boldsymbol{x})\bigg) \to \langle \boldsymbol{v}, \boldsymbol{\gamma_Z}\rangle, \quad \text{and}$$
$$g(t)\nu_{\boldsymbol{X}}(\mathscr{L}_{\boldsymbol{v}}(g(t)))\bigg(\langle \boldsymbol{v}, \boldsymbol{\gamma_X}\rangle - \int_{B_{\boldsymbol{0}}(1)\setminus B_{\boldsymbol{0}}(g(t))} \langle \boldsymbol{v}, \boldsymbol{x}\rangle \nu(\mathrm{d}\boldsymbol{x})\bigg)^{-1} \to 0, \quad \text{whenever } \langle \boldsymbol{v}, \boldsymbol{\gamma_Z}\rangle \neq 0,$$

(where we recall that $t/g(t) = 1/G(t^{-1})$). By the first limit, the second limit is equivalent to $t\nu_{\boldsymbol{X}}(\mathscr{L}_{\boldsymbol{v}}(g(t))) \to 0 = \nu_{\boldsymbol{Z}}(\mathscr{L}_{\boldsymbol{v}}(1))$. If instead $\nu_{\boldsymbol{Z}}(\mathscr{L}_{\boldsymbol{v}}(1)) > 0$, then, by [26, Thm 2(ii)], the weak limit is equivalent to the following. The process $\langle \boldsymbol{v}, \boldsymbol{X}\rangle$ has zero natural drift whenever it has finite variation and $\langle \boldsymbol{v}, \boldsymbol{Z}\rangle$ and the following two limits hold as $t \downarrow 0$:

$$\frac{\nu_{\boldsymbol{Z}}(\mathscr{L}_{\boldsymbol{v}}(1))}{g(t)\nu_{\boldsymbol{X}}(\mathscr{L}_{\boldsymbol{v}}(g(t)))}\bigg(\langle \boldsymbol{v}, \boldsymbol{\gamma_X}\rangle - \int_{B_{\boldsymbol{0}}(1)\setminus B_{\boldsymbol{0}}(g(t))} \langle \boldsymbol{v}, \boldsymbol{x}\rangle \nu(\mathrm{d}\boldsymbol{x})\bigg) \to \langle \boldsymbol{v}, \boldsymbol{\gamma_Z}\rangle, \quad \text{and}$$
$$t\nu_{\boldsymbol{X}}(\mathscr{L}_{\boldsymbol{v}}(g(t))) \to \nu_{\boldsymbol{Z}}(\mathscr{L}_{\boldsymbol{v}}(1)).$$

By the second limit, the first limit can be rewritten as the first limit in the display above. Thus, in either case, the conditions are equivalent to those stated in Theorem 5.1 in the direction $\boldsymbol{v}$. Since the directional limits are equivalent to the corresponding limits in $\mathbb{R}^d$, the result follows. □

Appendix C. Proof of the inequality in (4)

Recall that the two Lévy processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ in $\mathbb{R}^d$ have the Lévy–Itô decompositions $\boldsymbol{X}_t = \boldsymbol{\gamma}_{\boldsymbol{X},\kappa}t + \boldsymbol{\Sigma_X}\boldsymbol{B}_t^{\boldsymbol{X}} + \boldsymbol{D}_t^{\boldsymbol{X},\kappa} + \boldsymbol{J}_t^{\boldsymbol{X},\kappa}$ and $\boldsymbol{Y}_t = \boldsymbol{\gamma}_{\boldsymbol{Y},\kappa}t + \boldsymbol{\Sigma_Y}\boldsymbol{B}_t^{\boldsymbol{Y}} + \boldsymbol{D}_t^{\boldsymbol{Y},\kappa} + \boldsymbol{J}_t^{\boldsymbol{Y},\kappa}$, see Section 4. Recall that we chose coupling $\boldsymbol{B}^{\boldsymbol{X}} = \boldsymbol{B}^{\boldsymbol{Y}}$, implying $|(\boldsymbol{\Sigma_X} - \boldsymbol{\Sigma_Y})\boldsymbol{B}_t^{\boldsymbol{X}}| \leqslant |\boldsymbol{\Sigma_X}\boldsymbol{B}_t^{\boldsymbol{X}} - \boldsymbol{\Sigma_Y}\boldsymbol{B}_t^{\boldsymbol{Y}}| \leqslant |\boldsymbol{\Sigma_X} - \boldsymbol{\Sigma_Y}| \cdot |\boldsymbol{B}_t^{\boldsymbol{X}}|$ (where $|\boldsymbol{\Sigma_X} - \boldsymbol{\Sigma_Y}|$ is the Frobenius norm of the matrix $\boldsymbol{\Sigma_X} - \boldsymbol{\Sigma_Y}$). Applying Doob's maximal inequality, we obtain

$$\begin{aligned} \mathcal{W}_q\big(\boldsymbol{\Sigma_X}\boldsymbol{B}^{\boldsymbol{X}}, \boldsymbol{\Sigma_Y}\boldsymbol{B}^{\boldsymbol{Y}}\big) &\leqslant \mathcal{W}_2\big(\boldsymbol{\Sigma_X}\boldsymbol{B}^{\boldsymbol{X}}, \boldsymbol{\Sigma_Y}\boldsymbol{B}^{\boldsymbol{Y}}\big)^{q\wedge 1} \\ &\leqslant \mathbb{E}\bigg[\sup_{t\in[0,1]} |\boldsymbol{\Sigma_X}\boldsymbol{B}_t^{\boldsymbol{X}} - \boldsymbol{\Sigma_Y}\boldsymbol{B}_t^{\boldsymbol{Y}}|^2\bigg]^{(q\wedge 1)/2} \leqslant (2\sqrt{d}|\boldsymbol{\Sigma_X} - \boldsymbol{\Sigma_Y}|)^{q\wedge 1}. \end{aligned} \tag{46}$$

Applying the triangle inequality, we obtain

$$\begin{aligned} &\sup_{t\in[0,1]} |\boldsymbol{\gamma}_{\boldsymbol{X},\kappa}t + \boldsymbol{\Sigma_X}\boldsymbol{B}_t^{\boldsymbol{X}} + \boldsymbol{D}_t^{\boldsymbol{X},\kappa} + \boldsymbol{J}_t^{\boldsymbol{X},\kappa} - \boldsymbol{\gamma}_{\boldsymbol{Y},\kappa}t - \boldsymbol{\Sigma_Y}\boldsymbol{B}_t^{\boldsymbol{Y}} - \boldsymbol{D}_t^{\boldsymbol{Y},\kappa} - \boldsymbol{J}_t^{\boldsymbol{Y},\kappa}| \\ &\quad\leqslant |\boldsymbol{\gamma}_{\boldsymbol{X},\kappa} - \boldsymbol{\gamma}_{\boldsymbol{Y},\kappa}| + \sup_{t\in[0,1]} |\boldsymbol{\Sigma_X}\boldsymbol{B}_t^{\boldsymbol{X}} - \boldsymbol{\Sigma_Y}\boldsymbol{B}_t^{\boldsymbol{Y}}| + \sup_{t\in[0,1]} |\boldsymbol{D}_t^{\boldsymbol{X},\kappa} - \boldsymbol{D}_t^{\boldsymbol{Y},\kappa}| + \sup_{t\in[0,1]} |\boldsymbol{J}_t^{\boldsymbol{X},\kappa} - \boldsymbol{J}_t^{\boldsymbol{Y},\kappa}|. \end{aligned}$$

For $q \in (0,1]$ (resp. $q \in (1,2]$) inequality (4) follows by subadditivity $(a+b)^q \leqslant a^q + b^q$ for $a, b \geqslant 0$ (resp. Minkowski's inequality).