

**Manuscript version: Author's Accepted Manuscript**

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

**Persistent WRAP URL:**

<http://wrap.warwick.ac.uk/158645>

**How to cite:**

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

© 2021 Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International <http://creativecommons.org/licenses/by-nc-nd/4.0/>.



**Publisher's statement:**

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk).

# IMPLEMENTABLE COUPLING OF LÉVY PROCESS AND BROWNIAN MOTION

VLADIMIR FOMICHOV, JORGE GONZÁLEZ CÁZARES AND JEVGENIJS IVANOVŠ

ABSTRACT. We provide a simple algorithm for construction of Brownian paths approximating those of a Lévy process on a finite time interval. It requires knowledge of the Lévy process trajectory on a chosen regular grid and the law of its endpoint, or the ability to simulate from that. This algorithm is based on reordering of Brownian increments, and it can be applied in a recursive manner. We establish an upper bound on the mean squared maximal distance between the paths and determine a suitable mesh size in various asymptotic regimes. The analysis proceeds by reduction to the comonotonic coupling of increments. Applications to model risk and multilevel Monte Carlo are discussed in detail, and numerical examples are provided.

## 1. INTRODUCTION

Let  $X = (X(t), t \geq 0)$  be a Lévy process with  $X(1)$  having zero mean and unit variance. We aim to construct a standard Brownian motion  $W$  on the same probability space (or its extension) such that the mean squared maximal distance

$$\mathbb{E} \sup_{t \in [0,1]} |X(t) - W(t)|^2 \tag{1}$$

is small. This, in particular, provides an upper bound on the respective Wasserstein distance between the laws of the given Lévy process and the standard Brownian motion on the time interval  $[0, 1]$ , assuming the above expectation can be computed with some guaranteed accuracy. Such bounds are needed in a number of applications including model risk and distributionally robust optimization. Our focus, however, is on simple explicit constructions allowing to efficiently generate Brownian trajectories, which are paramount, for example, in the multilevel Monte Carlo method discussed below. It should be mentioned that our problem is different from so-called Markov couplings of Lévy processes [6, 25], where the two processes of interest have the same transition probabilities, but different initial distributions.

### 1.1. Applications.

*Model Risk and Distributionally Robust Optimization.* Every model is only an approximation of reality and thus it is important to understand the impact of model misspecification (going beyond model parameters) on the quantities of interest. A popular approach is to find the worst-case values for a family of plausible models in some neighbourhood of a chosen baseline model, which can be viewed as a systematic stress test [7]. The Wasserstein distance is a natural choice to define such an ambiguity ball of models in various settings, and it often leads to simple worst-case expressions [4]. Furthermore, this approach has close links to distributionally robust optimization, where allowing for some freedom in the model helps to mitigate the optimizer's curse [12], which is a well-known phenomenon in stochastic programming.

A critical task in these procedures is to estimate the radius of the ambiguity ball, and in the case of Lévy-driven models our coupling provides the necessary tool. For example, in the context of ruin theory one may use a diffusion approximation as a baseline model, while insisting that a certain compound Poisson process (corresponding to the classical Cramér–Lundberg risk model) belongs to the ambiguity ball. In [4] it has been suggested to simulate coupled paths (using the Brownian

---

*Key words and phrases.* Approximation; distributional model risk; Lévy processes; multilevel Monte Carlo; Wasserstein distance.

embedding in [23]) and to choose the radius according to the square-root of the empirical counterpart of (1). In this setting an upper bound on the Wasserstein distance between the processes would be sufficient, given it is not overly conservative. Note, however, that the implementable coupling approach can be based on the available historical claim sizes (jumps) without the knowledge of the exact underlying distribution.

*Multilevel Monte Carlo.* Suppose we want to estimate the value of the expectation  $\mathbb{E}g(X)$  for some appropriate function  $g$  of the path of  $X$  using Monte Carlo simulation. Sampling  $g(X)$  exactly is not feasible in general. One standard way around this problem is to fix a small truncation level  $\varepsilon_n > 0$  and to consider an approximation  $X_n$  of  $X$  obtained by replacing the martingale of jumps in  $[-\varepsilon_n, \varepsilon_n]$  by an appropriately scaled Brownian motion [1]. Apart from estimating  $\mathbb{E}g(X_n)$  one also needs to control the bias  $\mathbb{E}[g(X_n) - g(X)]$ . A significant improvement in the computational complexity can be often obtained by the multilevel Monte Carlo method [16], see also [13, 17] for the Lévy process setting. This, however, requires the ability to sample jointly  $g(X_n)$  and  $g(X_{n+1})$  in a way that the level variance  $\text{Var}[g(X_{n+1}) - g(X_n)]$  is as small as possible, which must be done without significantly increasing the cost of drawing such a joint sample compared with the single level sample.

It is clear from the construction of approximations that the only difficult part is to couple the martingale of jumps in  $[-\varepsilon_n, -\varepsilon_{n+1}) \cup (\varepsilon_{n+1}, \varepsilon_n]$ , where  $\varepsilon_{n+1} < \varepsilon_n$ , with an appropriately scaled Brownian motion, since the other parts can be reused. Note that such a Brownian motion must be constructed and it is not sufficient to simply claim its existence. Assuming  $g$  is sufficiently regular, it typically suffices to control the mean squared maximal distance between the paths as in (1). Our coupling provides the necessary tool, and the corresponding asymptotic analysis is presented in §5. In particular, we show that a substantial gain in computational complexity can be achieved for processes  $X$  of unbounded variation on compacts over the algorithms available in the existing literature (see §5.3 and Figure 2 below for a comparison with [9, 10]).

**1.2. The construction.** Here we describe in words our construction of the coupled Brownian motion  $W(t)$  on the time interval  $[0, 1]$ , postponing its precise definition and related notation to §2. This construction depends on an integer  $k \geq 1$ , and it can be viewed in a number of alternative ways. Firstly, we construct a skeleton of a standard Brownian motion  $W'(i/k)$  for  $i = 0, \dots, k$  such that  $W'(1)$  is (nearly) comonotonically coupled with  $X(1)$ , but is otherwise independent. The law of  $X(1)$  may not be explicit, in which case we may use either fast Fourier inversion [20] or an empirical counterpart that only requires the ability to simulate  $X$  at time 1 (see Appendix B). Secondly, we reorder the  $k$  increments  $\Delta_i^k W' = W'(i/k) - W'((i-1)/k)$  so that they match the ordering of the increments  $\Delta_i^k X$  of the process  $X$ ; the ties in the latter are resolved randomly. The accumulated reordered increments form the skeleton of another Brownian motion  $W$  over the grid  $0, 1/k, \dots, 1$ , and we supplement this skeleton with independently sampled Brownian bridges connecting these points, see Figure 1 for an illustration.

*Remark 1.1.* The reordering of the increments matches their ranks, so a random draw of any pair of increments is equivalent to a draw from the comonotonically coupled empirical measures based on the sampled increments. In other words, our reordering procedure exploits the convergence of empirical measures to their true distributions and the simple form of the comonotonic coupling between discrete distributions with uniform weights.

It is clear that the quality of our coupling strongly depends on the choice of  $k$ . For  $k = 1$  only the end-points are coupled, but the Brownian bridge is independent of  $X$ . The case of a large  $k$  can be understood using the results in [18], where we employed the idea of reordering independently sampled normal increments to recover the Brownian component of  $X$  up to a random linear drift. More precisely, letting  $W_X$  be the Brownian component of  $X$  normalized to have unit variance, the Brownian bridge  $W(t) - tW(1)$  converges in probability in supremum norm to the bridge  $W_X(t) - tW_X(1)$

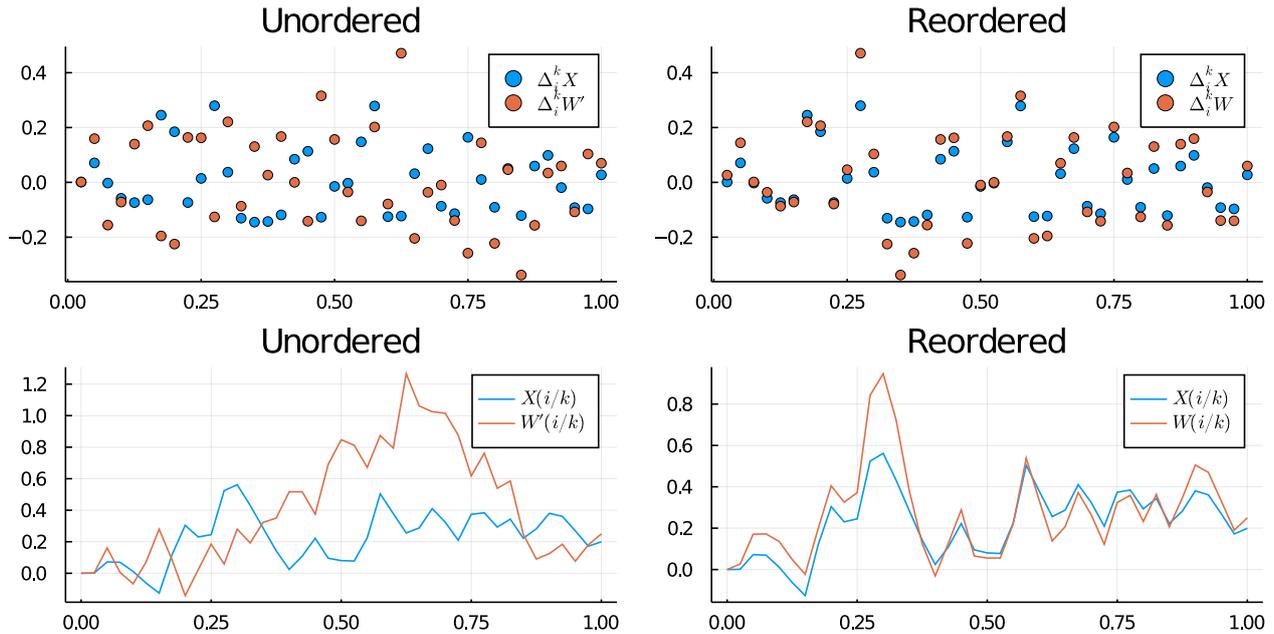


FIGURE 1. Increments (top) and skeletons (bottom) of a gamma process  $X$  and Brownian motions  $W'$ ,  $W$  over the uniform grid  $\{0, 1/k, \dots, 1\}$  for  $k = 40$ .

as  $k \rightarrow \infty$ . In the case when  $X$  has no Brownian component (under a further minor condition) the weak limit of the pair of processes  $(W(t) - tW(1), X(t))$  has independent components. In such a case  $k = 1$  and  $k = \infty$  are analogous in the sense that they both lead to an independent Brownian bridge. Therefore, one needs to choose a moderately large  $k$ .

Our main result, Corollary 3.3, provides some asymptotic theory suggesting an adequate choice of  $k$ . We further investigate it in §4 for various limiting regimes, apply it to multilevel Monte Carlo method in §5, and then also illustrate various choices numerically in §6.

*Remark 1.2.* Note that the coupled skeleton of  $W$  is supplemented with independent Brownian bridges. Alternatively, one may repeat the partitioning and reordering procedure for each interval  $[(i-1)/k, i/k]$  hoping to construct bridges which better approximate the evolution of  $X$  within these intervals. This leads to a hierarchical construction with possibly different  $k$  at each level. In §6 we examine this idea numerically.

**1.3. Literature and related methods.** Coupling of random variables and processes is a classical area of probability theory, see monographs [36, 32] and references therein. The famous result of Skorokhod in its modern interpretation [22, Thm. 15.17] states that given a sequence of Lévy processes  $X_n$  with  $X_n(1) \xrightarrow{d} W(1)$  it is possible to construct  $X_n$  and a standard Brownian motion  $W$  on a common probability space so that  $\sup_{t \in [0,1]} |X_n(t) - W(t)| \xrightarrow{\mathbb{P}} 0$ . This existence result relies on Skorokhod's representation theorem [22, Thm. 4.30], and it does not yield a feasible algorithm for constructing such couplings.

Another related classical result is Strassen's random walk approximation by a Brownian motion [34] underlying the functional LIL and based on Skorokhod's embedding theorem, see also [22, Ch. 14]. It yields an explicit coupling, implementation of which requires sampling Brownian paths up to a certain stopping time. It is, however, not obvious how to do this both accurately and efficiently. An embedding of a given Lévy process into a Brownian motion by means of a random time change was used in [27] to investigate its  $p$ -variation. Furthermore, a construction of a compensated Poisson process  $X$  from a Brownian motion  $W$  was suggested in [23] and then used to study

crossings of empirical processes. The analogous construction is also suggested for a compound Poisson process with negative jumps, but the resultant process in general exhibits positive dependence between inter-arrival times and subsequent jumps, and so it is not compound Poisson.

Yet another coupling of random walks with exponential moments, often called the Hungarian embedding or the KMT (Komlós–Major–Tusnády) coupling, was proposed in [24]; it was later extended by Zaitsev to the multidimensional case in [37]. Although, in a certain sense, these couplings are optimal, the fact that they are based on conditional distributions makes them hardly suitable for numerical computations. The same critique applies to the coupling of solutions of two SDEs with respect to a Lévy process and a Brownian motion constructed in [15] and of a Lévy process with small jumps and a Brownian motion introduced in [9], both of which are based on the KMT coupling and its multidimensional extension by Zaitsev.

A somewhat related problem consists in estimating the Wasserstein distance between two infinitely divisible distributions, or equivalently between the marginal distributions of two Lévy processes. The optimal, or near-optimal, bounds in the general case were recently obtained in [26] relying on the corresponding bound in [31] on the error in the central limit theorem, see also Lemma 2.1 below.

## 2. DEFINITIONS AND FUNDAMENTAL PROPERTIES

We start by setting up the notation and stating a result about the quality of comonotonic coupling of the marginals. Next, we introduce an auxiliary coupling method based on comonotonic coupling of  $k$  increments, show that the respective Brownian bridges are closely related, and investigate associate discretization errors. These notions will be used in §3 to show (under mild conditions) that our reordering coupling is not worse than the above auxiliary coupling. Finally, the quality of the latter is analysed to arrive at our main result in Corollary 3.3. Unlike in §3, where we work with a convergent sequence  $X_n$  of Lévy processes, here we deal with a single process  $X$  which greatly simplifies presentation.

Throughout the paper, unless explicitly stated otherwise, we assume that

$$\mathbb{E}X(1) = 0, \quad \mathbb{E}X^2(1) = 1, \quad \mathbb{E}X^4(1) < \infty. \quad (2)$$

The Lévy measure of  $X$  will be denoted by  $\Pi$ , and we also define its fourth moment and the tail function:

$$\mu_4 := \int_{\mathbb{R}} x^4 \Pi(dx), \quad \bar{\Pi}(x) := \Pi(\mathbb{R} \setminus [-x, x]), \quad x > 0.$$

Recall that the marginal distribution functions are continuous unless  $X$  is a drifted compound Poisson process, and so this fundamental class of processes should be treated with care.

**2.1. Wasserstein distance between the marginal laws.** Consider two random variables  $\zeta_1$  and  $\zeta_2$  with distribution functions  $F_1$  and  $F_2$ , respectively. Recall that every coupling  $(\zeta_1, \zeta_2)$  can be obtained by setting  $\zeta_2 = h(\zeta_1, U)$  for an appropriate measurable function  $h$  and a standard uniform  $U$  independent of  $\zeta_1$ . That is, every joint law with marginals  $F_1$  and  $F_2$  can be retrieved in this way. Comonotonic coupling is usually defined by  $(F_1^{-1}(U), F_2^{-1}(U))$ , where  $F_i^{-1}(u) = \inf\{x | F_i(x) \geq u\}$  denotes the left inverse of  $F_i$ . Alternatively, it can be obtained by taking

$$h(x, u) = F_2^{-1}(\mathbb{P}(\zeta_1 < x) + u\mathbb{P}(\zeta_1 = x)) \quad (3)$$

in the specification of  $\zeta_2$ , see [33]. For a continuous  $F_1$  there is no need to further randomize using  $U$ , and we see that  $\zeta_2 = F_2^{-1}(F_1(\zeta_1))$  is a monotone transform of  $\zeta_1$ . The Wasserstein distance  $\mathcal{W}_2$  between the laws of  $\zeta_1$  and  $\zeta_2$  (with some abuse of notation) is defined as

$$\mathcal{W}_2(\zeta_1, \zeta_2) := \inf_{(\zeta'_1, \zeta'_2)} (\mathbb{E}|\zeta'_1 - \zeta'_2|^2)^{1/2},$$

where the infimum is taken over all possible couplings of  $\zeta_1$  and  $\zeta_2$ . It is a standard fact that the infimum is achieved by the comonotonic coupling [30, Ex. 3.2.14].

The following lemma provides an upper bound for the Wasserstein distance between the marginals of a Lévy process and a Brownian motion in terms of  $\mu_4$ , see also [15, 26].

**Lemma 2.1.** *There exists a constant  $C > 0$  such that for any Lévy process  $X$  satisfying condition (2) and a standard Brownian motion  $B$  we have*

$$\forall t \geq 0 : \quad \mathcal{W}_2^2(X(t), B(t)) \leq C\mu_4.$$

*Proof.* Obviously, we can assume that  $t > 0$ . For arbitrary  $n \geq 1$  we have

$$X(t) = \sum_{i=1}^n [X(it/n) - X((i-1)t/n)],$$

where the summands are i.i.d., and so applying [31, Thm. 4.1] yields

$$\mathcal{W}_2^2(X(t), B(t)) = t\mathcal{W}_2^2(X(t)/\sqrt{t}, B(1)) \leq tCn\mathbb{E}(|X(t/n)|/\sqrt{t})^4 = C(n/t)\mathbb{E}X^4(t/n).$$

By [14, Thm. 1.1], the upper bound tends to the upper bound of the stated inequality as  $n \rightarrow \infty$ .  $\square$

**2.2. Two coupling methods.** Generally speaking, coupling marginal distributions of two processes does not allow to construct a coupling of the whole processes. Nevertheless, we may comonotonically couple the increments of a Brownian motion and  $X$  over a grid. This auxiliary method is closely related to the proposed coupling based on reordering of increments, and it will be important for the asymptotic analysis of the latter. Note that the auxiliary coupling is much more challenging to implement efficiently in a common scenario when the distribution function of the increments of  $X$  is not readily available. Here we provide precise definitions of the two methods, both of which depend on

integer  $k \geq 1$  specifying the number of increments used.

We will need standard uniforms  $U, U_1, \dots, U_k$  and also  $k+1$  Brownian bridges, which are mutually independent and also independent of  $X$ .

*Reordering of increments.* Let us now give a precise definition of the process  $W$  described in words in §1.2. Start by taking a standard Brownian motion  $W'$  with  $W'(1) = h(X(1), U)$  for an appropriate measurable function  $h$  and such that  $(W'(t) - tW'(1), t \in [0, 1])$  is a standard Brownian bridge independent of everything else. Recall that this construction allows for an arbitrary coupling of the end-points  $X(1)$  and  $W'(1)$ . Importantly, the processes  $W'$  and  $X$  are independent given  $X(1)$ .

Let  $\pi$  be an a.s. unique random permutation on  $\{1, \dots, k\}$  such that for all  $i \neq j$ :

$$\Delta_{\pi(i)}^k W' < \Delta_{\pi(j)}^k W' \quad \text{iff} \quad \Delta_i^k X < \Delta_j^k X \quad \text{or} \quad \Delta_i^k X = \Delta_j^k X, \quad U_i < U_j. \quad (4)$$

That is, the ties in  $\Delta_i^k X$ , only possible when  $X$  is a compound Poisson process with drift, are resolved uniformly at random. Define a stochastic process  $W$  by setting  $W(0) := 0$  and

$$W(t) := W\left(\frac{i-1}{k}\right) + W'\left(\frac{\pi(i)-1}{k} + t - \frac{i-1}{k}\right) - W'\left(\frac{\pi(i)-1}{k}\right), \quad \frac{i-1}{k} < t \leq \frac{i}{k}, \quad i = 1, \dots, k. \quad (5)$$

Thus defined  $W$  is indeed a standard Brownian motion, see Lemma 2.2 below.

*Comonotonic coupling of increments.* Let  $h_k$  be the function defined in (3), where  $\zeta_1 = X(1/k)$  and  $\zeta_2$  is mean zero normal with variance  $1/k$ . By taking  $\xi_i = h_k(\Delta_i^k X, U_i)$  we produce Brownian increments comonotonically coupled with the increments of  $X$ . Let also  $\beta_1, \dots, \beta_k$  be independent Brownian bridges on the time interval  $[0, 1/k]$  that are independent of everything else. Define a stochastic process  $\widehat{W}$  by setting  $\widehat{W}(0) := 0$  and

$$\widehat{W}(t) := \widehat{W}\left(\frac{i-1}{k}\right) + \beta_i\left(t - \frac{i-1}{k}\right) + k\left(t - \frac{i-1}{k}\right)\xi_i, \quad \frac{i-1}{k} < t \leq \frac{i}{k}, \quad i = 1, \dots, k. \quad (6)$$

In words,  $\widehat{W}$  is a stochastic process whose increments over the given grid are comonotonically coupled with the increments of  $X$  and whose “interval bridges” are independent of  $X$ .

**Lemma 2.2.** *The processes  $W$  and  $\widehat{W}$  are standard Brownian motions and their increments have the same ordering:*

$$\Delta_i^k W < \Delta_j^k W \quad \text{iff} \quad \Delta_i^k \widehat{W} < \Delta_j^k \widehat{W}$$

with probability 1.

*Proof.* The same ordering of increments follows from (4) and the fact that  $h_k(x_1, u_1) < h_k(x_2, u_2)$  when either  $x_1 < x_2$  or  $x_1 = x_2$  and  $u_1 < u_2$  with  $\mathbb{P}(X(1/k) = x_1) > 0$ ; here  $u_i \in (0, 1)$ . It is essential that the same sequence of  $U_i$  was used to resolve the ties in the first method and to provide extra randomness in the second method.

It is easy to see that  $\widehat{W}$  is a standard Brownian motion, since the increments  $\xi_i$  are independent and have the required distribution. Checking that  $W$  is a Brownian motion is more complicated, and we need to take care of both reordering and coupling the end-points. Recall that the process  $X$ , Brownian bridge  $W'(t) - tW'(1)$ , and uniforms  $U, U_1, \dots, U_k$  are mutually independent. Conditionally on  $X(1)$ , the pairs  $(\Delta_i^k X, U_i)$  are exchangeable, and thus  $W(t) - tW(1)$  is still a standard Brownian bridge. The latter can be seen as obtained from  $W'(t) - tW'(1)$  by an independent uniform permutation of increments. Note that  $W(1) = W'(1) = h(X(1), U)$  to conclude.  $\square$

*Remark 2.3.* If  $X$  is not a compound Poisson process with drift, then there is no need in the uniforms  $U_1, \dots, U_k$  in the above two constructions. In such a case there are no ties in the first method, the increments  $\xi_i$  are monotone transforms of  $\Delta_i^k X$  in the second method, and the same ordering of the increments of  $W$  and  $\widehat{W}$  is automatic.

Finally, let us point out that we have constructed a trivariate process  $(X, W, \widehat{W})$ , and not just two couplings  $(X, W)$  and  $(X, \widehat{W})$ . This is crucial for the proofs, where the quality of the reordering coupling is related to the quality of the auxiliary coupling, see Theorem 3.1 below.

**2.3. Proximity of the two Brownian bridges.** The following lemma is an extension of Lemma 7 from [18], which concerned convergence in probability, to convergence in mean.

**Lemma 2.4.** *Let  $Z_1, \dots, Z_k$  and  $Z'_1, \dots, Z'_k$  be two independent sets of i.i.d. standard normal random variables, and let  $Z_{(1)}, \dots, Z_{(k)}$  and  $Z'_{(1)}, \dots, Z'_{(k)}$  be their order statistics. Then*

$$\lim_{k \rightarrow \infty} \frac{1}{\log \log k} \mathbb{E} \sum_{i=1}^k ((Z_{(i)} - \bar{Z}_k) - (Z'_{(i)} - \bar{Z}'_k))^2 = 2,$$

where  $\bar{Z}_k = (Z_1 + \dots + Z_k)/k$  and  $\bar{Z}'_k$  denote the respective arithmetic means.

*Proof.* Let  $G_k$  and  $G'_k$  be the empirical distribution functions of  $Z_1, \dots, Z_k$  and  $Z'_1, \dots, Z'_k$ , respectively. We have

$$\frac{1}{\log \log k} \sum_{i=1}^k (Z_{(i)} - Z'_{(i)})^2 = \frac{k}{\log \log k} \int_0^1 (G_k^{-1}(x) - G'_k{}^{-1}(x))^2 dx = \frac{k \mathcal{W}_2^2(G_k, G'_k)}{\log \log k},$$

where the expectation of the last term converges to 2 by [2, Rem. 2.4]. Also, note that

$$\sum_{i=1}^k ((Z_{(i)} - \bar{Z}_k) - (Z'_{(i)} - \bar{Z}'_k))^2 = \sum_{i=1}^k (Z_{(i)} - Z'_{(i)})^2 - k(\bar{Z}_k - \bar{Z}'_k)^2,$$

where the difference  $\bar{Z}_k - \bar{Z}'_k$  has distribution  $\mathcal{N}(0, 2/k)$ .  $\square$

We are now ready to upper bound the mean squared maximal distance between the bridges corresponding to  $W$  and  $\widehat{W}$  over the grid with mesh size  $1/k$ . Note that both processes depend on  $k$  as well.

**Lemma 2.5.** *For the Brownian motions  $W$  and  $\widehat{W}$  defined in (5) and (6) using  $k \geq 1$  increments it holds that*

$$\mathbb{E} \max_{1 \leq i \leq k} \left| [W(\frac{i}{k}) - \frac{i}{k}W(1)] - [\widehat{W}(\frac{i}{k}) - \frac{i}{k}\widehat{W}(1)] \right|^2 = O(\log \log k/k), \quad k \rightarrow \infty,$$

uniformly for all processes  $X$ .

*Proof.* Observe that the bridge  $W(t) - tW(1)$  is obtained by sampling the bridge  $W'(t) - tW'(1)$  independently and then reordering its increments (together with in-between evolutions) according to the increments of the bridge  $\widehat{W}(t) - t\widehat{W}(1)$ , see Lemma 2.2 in particular. Here we rely on the obvious fact that adding a linear trend has no influence on the ordering of increments. Denoting the increments by

$$\widehat{\eta}_i = \Delta_i^k \widehat{W} - \Delta_i^k t \cdot \widehat{W}(1), \quad \eta_i = \Delta_i^k W - \Delta_i^k t \cdot W(1),$$

observe, using the notation of Lemma 2.4, that the vectors  $(\sqrt{k}\widehat{\eta}_i)$  and  $(\sqrt{k}\eta_i)$  jointly have the law of  $(Z_i - \overline{Z}_k)$  and  $(Z'_i - \overline{Z}'_k)$  with the latter reordered according to the former. Since the summation order is arbitrary, we find using Lemma 2.4 that

$$\mathbb{E} \sum_{i=1}^k (\sqrt{k}\widehat{\eta}_i - \sqrt{k}\eta_i)^2 = \mathbb{E} \sum_{i=1}^k ((Z_{(i)} - \overline{Z}_k) - (Z'_{(i)} - \overline{Z}'_k))^2 = O(\log \log k). \quad (7)$$

Note that this quantity does not depend on the underlying process  $X$ .

Furthermore, the pairs  $(\widehat{\eta}_i, \eta_i)$  are exchangeable and

$$(\widehat{\eta}_1 - \eta_1) + \dots + (\widehat{\eta}_k - \eta_k) = (\widehat{\eta}_1 + \dots + \widehat{\eta}_k) - (\eta_1 + \dots + \eta_k) = 0.$$

We are going to use Garsia's inequality (see [8, p. 26] and references therein), which states that for any  $p \geq 1$  and any  $x_1, \dots, x_k \in \mathbb{R}$  such that  $x_1 + \dots + x_k = 0$  we have

$$\frac{1}{k!} \sum_{\sigma \in S_k} \max_{1 \leq i \leq k} |x_{\sigma(1)} + \dots + x_{\sigma(i)}|^p \leq C_p \left( \sum_{i=1}^k x_i^2 \right)^{p/2},$$

where the sum is taken over all permutations  $\sigma$  of order  $k$  and the constant  $C_p > 0$  depends only on  $p$ . We apply this inequality to  $p = 2$  and  $x_i = \widehat{\eta}_i - \eta_i$ , and take expectations of both sides. Owing to the exchangeability, the expectations of all maximums are equal. Therefore, we obtain

$$\mathbb{E} \max_{1 \leq i \leq k} |(\widehat{\eta}_1 + \dots + \widehat{\eta}_i) - (\eta_1 + \dots + \eta_i)|^2 \leq C_2 \mathbb{E} \sum_{i=1}^k (\widehat{\eta}_i - \eta_i)^2,$$

which is  $O(\log \log k/k)$  according to (7).  $\square$

**2.4. Discretization error.** Finally, we need appropriate bounds on the mean-squared-maximal discretization error of the involved processes. For an integer  $k \geq 1$  let  $h^{[k]}(t) = h(\lfloor kt \rfloor / k)$  be a discretization of the function  $h$  over the grid  $\{0, 1/k, 2/k, \dots, 1\}$ . It is a basic fact that a standard Brownian motion  $B$  satisfies

$$\mathbb{E} \sup_{t \in [0,1]} (B(t) - B^{[k]}(t))^2 = O(\log k/k), \quad k \rightarrow \infty, \quad (8)$$

see [28, Lem. 4.4]. This also can be seen using extreme value theory and the scaling property of  $B$ . We are now ready to extend the result of Lemma 2.5 to the whole interval  $[0, 1]$ .

**Proposition 2.6.** *For the Brownian motions  $W$  and  $\widehat{W}$  defined in (5) and (6) using  $k \geq 1$  increments it holds that*

$$\mathbb{E} \sup_{t \in [0,1]} \left| [W(t) - tW(1)] - [\widehat{W}(t) - t\widehat{W}(1)] \right|^2 = O(\log k/k)$$

as  $k \rightarrow \infty$ , uniformly for all processes  $X$ .

*Proof.* Observe the following upper bound:

$$\begin{aligned} \sup_{t \in [0,1]} \left| [W(t) - tW(1)] - [\widehat{W}(t) - t\widehat{W}(1)] \right| &\leq \max_{1 \leq i \leq k} \left| [W(\frac{i}{k}) - \frac{i}{k}W(1)] - [\widehat{W}(\frac{i}{k}) - \frac{i}{k}\widehat{W}(1)] \right| \\ &+ \sup_{t \in [0,1]} |W(t) - W^{[k]}(t)| + \frac{1}{k}|W(1)| + \sup_{t \in [0,1]} |\widehat{W}(t) - \widehat{W}^{[k]}(t)| + \frac{1}{k}|\widehat{W}(1)|, \end{aligned}$$

where the second line is an upper bound on the discretization of the two bridges. It is sufficient to bound the second moments of all the terms, which is achieved by Lemma 2.5 and (8).  $\square$

With respect to the discretization of the process  $X$  we note that  $\sup_{t \in [0,1]} |X(t) - X^{[k]}(t)|$  converges to the size of the largest jump a.s., which has been the motivation for studying the integrated discretization error in [21]. Hence the mean-squared-maximal error can be small only when  $X$  is close to a Brownian motion in some sense. We need the following bound in terms of  $\mu_4$ , see also Theorem 3.2 below, even though various alternative bounds may be easier to establish. The quantity  $\mu_4$  is normally small when  $X$  is close to a Brownian motion, see Lemma 3.5 for a precise result.

**Proposition 2.7.** *There is a constant  $C > 0$  such that for any Lévy process  $X$  satisfying (2) and for any  $k \geq 2$  the following bound holds:*

$$\mathbb{E} \sup_{t \in [0,1]} (X(t) - X^{[k]}(t))^2 \leq C(k\mu_4 + \log k/k).$$

The proof is rather lengthy so we postpone it to Appendix A.

### 3. ASYMPTOTIC QUALITY OF COUPLING METHODS

**3.1. Main results.** Given a sequence of Lévy processes  $X_n$  and integers  $k_n \geq 1$ , let  $W_n$  and  $\widehat{W}_n$  be the Brownian motions constructed using, respectively, the reordering method and comonotonic coupling of increments with the common discretization parameter  $k_n$ . Note that the coupling of  $X(1)$  and  $W(1)$  is arbitrary. We start by showing that the two couplings have asymptotically equivalent quality given that the respective coupling of the end-points is sufficiently good.

**Theorem 3.1.** *Consider a sequence of Lévy processes  $X_n$  and a sequence of integers  $k_n \rightarrow \infty$ . If  $\varepsilon_n \downarrow 0$  is such that  $\varepsilon_n k_n / \log k_n$  is bounded away from 0 and*

$$\mathbb{E}|X_n(1) - W_n(1)|^2 = O(\varepsilon_n), \quad \mathbb{E} \sup_{t \in [0,1]} |X_n(t) - \widehat{W}_n(t)|^2 = O(\varepsilon_n),$$

then also

$$\mathbb{E} \sup_{t \in [0,1]} |X_n(t) - W_n(t)|^2 = O(\varepsilon_n).$$

*This result is also true when  $W$  and  $\widehat{W}$  are swapped.*

*Proof.* Start with an obvious bound

$$\mathbb{E} \sup_{t \in [0,1]} |X_n(t) - W_n(t)|^2 \leq 2\mathbb{E} \sup_{t \in [0,1]} |X_n(t) - \widehat{W}_n(t)|^2 + 2\mathbb{E} \sup_{t \in [0,1]} |W_n(t) - \widehat{W}_n(t)|^2.$$

In view of the assumptions, it is only required to consider the second term

$$\begin{aligned} \mathbb{E} \sup_{t \in [0,1]} |W_n(t) - \widehat{W}_n(t)|^2 &\leq 2\mathbb{E} \sup_{t \in [0,1]} |[W_n(t) - tW_n(1)] - [\widehat{W}_n(t) - t\widehat{W}_n(1)]|^2 \\ &+ 2\mathbb{E}|W_n(1) - \widehat{W}_n(1)|^2. \end{aligned}$$

According to Proposition 2.6 we find that the first term on the right hand side is  $O(\log k_n/k_n) = O(\varepsilon_n)$ . Finally,

$$\mathbb{E}|W_n(1) - \widehat{W}_n(1)|^2 \leq 2\mathbb{E}|X_n(1) - W_n(1)|^2 + 2\mathbb{E}|X_n(1) - \widehat{W}_n(1)|^2 = O(\varepsilon_n),$$

by assumption. The final statement is proven analogously.  $\square$

Observe that, for comonotonically coupled end-points  $W_n(1)$  and  $X_n(1)$ , the assumption on their mean squared distance is automatically satisfied whenever  $\mathcal{W}_2(W_n(1), X_n(1)) \rightarrow 0$ . This follows from the optimality of the comonotonic coupling. Next we state an upper bound on the quality of the auxiliary method based on comonotonic coupling of increments.

**Theorem 3.2.** *Let  $X_n$  be a sequence of Lévy processes satisfying (2) with Lévy measures  $\Pi_n$  such that*

$$\mu_{4,n} := \int_{\mathbb{R}} x^4 \Pi_n(dx) \rightarrow 0.$$

*Then for any sequence  $k_n \rightarrow \infty$  we have*

$$\mathbb{E} \sup_{t \in [0,1]} |X_n(t) - \widehat{W}_n(t)|^2 = O(k_n \mu_{4,n} + \log k_n/k_n).$$

*Proof.* An upper bound on the maximal distance  $\sup_{t \in [0,1]} |X_n(t) - \widehat{W}_n(t)|$  is given by

$$\max_{1 \leq i \leq k_n} |X_n(i/k_n) - \widehat{W}_n(i/k_n)| + \sup_{t \in [0,1]} |\widehat{W}_n(t) - \widehat{W}_n^{[k_n]}(t)| + \sup_{t \in [0,1]} |X_n(t) - X_n^{[k_n]}(t)|. \quad (9)$$

By construction,  $(X(i/k_n) - \widehat{W}(i/k_n), 1 \leq i \leq k_n)$  is a zero-mean random walk, and so applying Doob's maximal inequality [22, Prop. 7.16] we obtain

$$\mathbb{E} \max_{1 \leq i \leq k_n} |X(i/k_n) - \widehat{W}(i/k_n)|^2 \leq 4\mathbb{E}|X(1) - \widehat{W}(1)|^2 = 4k_n \mathbb{E}|X(1/k_n) - \widehat{W}(1/k_n)|^2. \quad (10)$$

But the latter expectation is  $\mathcal{W}_2^2(X(1/k_n), B(1/k_n))$  again by construction. According to Lemma 2.1 we see that the upper bound in (10) is  $O(k_n \mu_{4,n})$ . The other two terms in (9) are  $O(\log k_n/k_n)$  and  $O(k_n \mu_{4,n} + \log k_n/k_n)$  according to (8) and Proposition 2.7, respectively.  $\square$

The following is our main result.

**Corollary 3.3.** *Consider a sequence of Lévy processes  $X_n$  satisfying (2) and  $\mu_{4,n} \rightarrow 0$ . Then for any  $k_n \rightarrow \infty$  such that  $k_n \mu_{4,n} \rightarrow 0$  we have*

$$\mathbb{E} \sup_{t \in [0,1]} |X_n(t) - W_n(t)|^2 = O(k_n \mu_{4,n} + \log k_n/k_n),$$

*provided  $W_n(1)$  is chosen so that  $\mathbb{E}(X_n(1) - W_n(1))^2$  is of the same order.*

*In particular, taking  $k_n \sim \sqrt{|\log \mu_{4,n}|/\mu_{4,n}}$  and coupling  $X_n(1)$  and  $W_n(1)$  comonotonically yields*

$$\mathbb{E} \sup_{t \in [0,1]} |X_n(t) - W_n(t)|^2 = O(\log k_n/k_n) = O(\sqrt{|\mu_{4,n}| \log \mu_{4,n}}).$$

*Proof.* Theorems 3.1 and 3.2 yield the first part. The second part follows from Lemma 2.1 implying  $\mathbb{E}|X_n(1) - W_n(1)|^2 = O(\mu_{4,n})$ .  $\square$

**Remark 3.4.** Note that any upper bound  $\bar{\mu}_{4,n} \rightarrow 0$  on  $\mu_{4,n}$  can be taken instead in Corollary 3.3. In that case it is sufficient to choose  $k_n \sim c(|\log \bar{\mu}_{4,n}|/\bar{\mu}_{4,n})^{1/2}$  for any  $c > 0$  to get the bound  $O((\bar{\mu}_{4,n} |\log \bar{\mu}_{4,n}|)^{1/2})$  on the mean squared maximal distance.

It is generally not necessary to couple  $X_n(1)$  and  $W_n(1)$  comonotonically to obtain the rate in Corollary 3.3. This is important, since the distribution function of  $X(1)$  is rarely explicit, and we may produce a near-comonotonic coupling by sampling independent copies of  $X(1)$ , see Lemma B.1 below.

Finally, we note that a slightly better rate can be produced when restricting comparison of the paths to the grid points:

$$\mathbb{E} \max_{1 \leq i \leq k_n} |X_n(i/k_n) - W_n(i/k_n)|^2 = O(k_n \mu_{4,n} + \log \log k_n/k_n).$$

This stems from a slightly better bound in Lemma 2.5 as compared to Proposition 2.6.

**3.2. On the fourth moment.** Here we provide a general condition implying that  $\mu_{4,n} \rightarrow 0$ .

**Lemma 3.5.** *Let  $X_n$  be a sequence of Lévy processes converging to a Brownian motion and having Lévy measures  $\Pi_n$ . Then  $\mu_{4,n} \rightarrow 0$  if and only if*

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \int_{|x| > M} x^4 \Pi_n(dx) = 0. \quad (11)$$

*Proof.* Fix  $h > 0$ . According to [22, Thm. 15.14], the convergence  $X_n(1) \xrightarrow{d} W(1)$  implies

$$\limsup_{n \rightarrow \infty} \int_{|x| \leq h} x^2 \Pi_n(dx) \leq 1, \quad \bar{\Pi}_n(h) := \Pi_n(\mathbb{R} \setminus [-h, h]) \rightarrow 0.$$

Define the measures  $\Pi'_n(dx) := \Pi_n(dx) \mathbf{1}_{\{|x| > h\}} + (1 - \bar{\Pi}_n(h)) \delta_0(dx)$ , which converge weakly to  $\delta_0$ , the point mass at 0, and are probability measures for all sufficiently large  $n$ . Thus

$$\int_{|x| > h} x^4 \Pi_n(dx) = \int_{\mathbb{R}} x^4 \Pi'_n(dx) \rightarrow 0,$$

if and only if (11) holds, which is the respective uniform integrability condition, see [22, Lem. 4.12]. It is left to note that  $\limsup_{n \rightarrow \infty} \int_{|x| \leq h} x^4 \Pi_n(dx) \leq h^2$ , and to recall that  $h > 0$  was arbitrary.  $\square$

Some standard conditions implying (11) can be also provided. For example, it is sufficient to assume that  $\int_{|x| > 1} |x|^{4+\delta} \Pi_n(dx)$  is bounded for some small  $\delta > 0$ . Alternatively, one can assume a bound on the tails:  $\bar{\Pi}_n(x) \leq \bar{\Pi}(x)$  for  $x$  sufficiently large, where the Lévy measure  $\Pi$  satisfies  $\int_{|x| > 1} x^4 \Pi(dx) < \infty$ .

#### 4. LIMITING REGIMES

Here we explore three limiting regimes: the classical scaling regime, perturbation of a Brownian motion by an independent Lévy process, and a small-jump Brownian approximation. We write  $W_n$  for the Brownian motion constructed from  $X_n$  according to our increment reordering coupling, assuming that the end-points are coupled comonotonically. In the latter two regimes it is more natural to index the sequence of processes by  $\varepsilon \downarrow 0$  instead of  $n \rightarrow \infty$ .

**4.1. Zooming out.** For a Lévy process  $X$  satisfying condition (2), define a sequence of Lévy processes  $X_n(t) = X(nt)/\sqrt{n}$ ,  $n \geq 1$ . Then the associated Lévy measures  $\Pi_n$  satisfy

$$\mu_{4,n} = \int_{\mathbb{R}} x^4 \Pi_n(dx) = \frac{1}{n} \int_{\mathbb{R}} x^4 \Pi(dx) \rightarrow 0, \quad n \rightarrow \infty.$$

Thus, if  $X_n(1)$  and  $W_n(1)$  are comonotonically coupled and  $k_n \sim c\sqrt{n \log n}$ , for some  $c > 0$ , then

$$\mathbb{E} \sup_{t \in [0,1]} |X_n(t) - W_n(t)|^2 = O(\sqrt{\log n/n}), \quad n \rightarrow \infty.$$

The coupling in [23] for a compensated Poisson process and a Brownian motion admits a similar rate. More precisely, there is an a.s. upper bound of order  $(\log \log n \log^2 n/n)^{1/2}$  on the squared maximal distance. The same a.s. rate can be deduced for the coupling between a time-changed random walk  $S_{[tn]}/\sqrt{n}$  with a Brownian motion as in [35], given that  $\mathbb{E}S_1^4 < \infty$ .

**4.2. Perturbed Brownian motion.** Consider a standard Brownian motion  $B$  perturbed by an independent Lévy process  $Y$  satisfying condition (2):

$$X_\varepsilon(t) = \sqrt{1 - \varepsilon^2} B(t) + \varepsilon Y(t), \quad 0 \leq t \leq 1,$$

with  $\varepsilon \downarrow 0$ . Letting  $\Pi$  stand for the Lévy measure of  $Y$ , we obtain  $\mu_{4,\varepsilon} = \varepsilon^4 \int_{\mathbb{R}} x^4 \Pi(dx) \rightarrow 0$ . According to Corollary 3.3, we choose  $k_\varepsilon \sim \sqrt{|\log \varepsilon|}/\varepsilon^2$  to get

$$\mathbb{E} \sup_{t \in [0,1]} |X_\varepsilon(t) - W_\varepsilon(t)|^2 = O(\varepsilon^2 \sqrt{|\log \varepsilon|}), \quad \varepsilon \rightarrow 0.$$

Note that  $\mathbb{E} \sup_{t \in [0,1]} |X_\varepsilon(t) - B(t)|^2 = O(\varepsilon^2)$ , which is smaller than the bound in the display by a logarithmic factor. To obtain this rate, we would need a simulatable  $W_\varepsilon$  that is sufficiently close to  $B$ . If the path of  $X_\varepsilon$  is already given, then this can be done by taking a sufficiently large  $k_\varepsilon$ . Indeed, in this particular case, according to [18], for fixed  $\varepsilon$  the Brownian bridge  $W_\varepsilon(t) - tW_\varepsilon(1)$  converges in probability in supremum norm to the bridge  $B(t) - tB(1)$  as  $k_\varepsilon \rightarrow \infty$ . Letting  $W'_\varepsilon(t) = B(t) - t(B(1) - W_\varepsilon(1))$  be the corresponding limiting process we find (after some straightforward manipulations) that

$$\mathbb{E} \sup_{t \in [0,1]} |X_\varepsilon(t) - W'_\varepsilon(t)|^2 \leq O(\varepsilon^2) + 2\mathbb{E}|X_\varepsilon(1) - W_\varepsilon(1)|^2 = O(\varepsilon^2).$$

Hence, the rate  $O(\varepsilon^2)$  can be obtained by taking a sufficiently large  $k_\varepsilon$ , increasing the cost. We stress that increasing  $k_\varepsilon$  does not always lead to an improvement. Indeed, if  $X_\varepsilon$  has no Brownian part, then infinite  $k_\varepsilon$  results in an independent bridge, as in the case of  $k_\varepsilon = 1$ , see [18, Prop. 3] for details.

**4.3. Small-jump Gaussian approximation.** A widely used idea in simulation of Lévy processes is to approximate the small jump martingale by an appropriately scaled Brownian motion. For every cutoff level  $\varepsilon \in (0, 1)$  we let  $M_\varepsilon(t)$  be the martingale containing the compensated jumps of  $X$  in  $[-\varepsilon, \varepsilon]$ . We denote its variance at time 1 by

$$\sigma_\varepsilon^2 := \int_{[-\varepsilon, \varepsilon]} x^2 \Pi(dx).$$

According to [1] the process  $X_\varepsilon(t) = M_\varepsilon(t)/\sigma_\varepsilon$  weakly converges to  $W(t)$  as  $\varepsilon \downarrow 0$  under a minor regularity condition, such as  $\sigma_\varepsilon \neq 0$  and  $\varepsilon/\sigma_\varepsilon \rightarrow 0$ , which we assume in the following. Below, we investigate the quality of our coupling in this limiting regime.

Observe that

$$\mu_{4,\varepsilon} = \sigma_\varepsilon^{-4} \int_{[-\varepsilon, \varepsilon]} x^4 \Pi(dx) \leq \sigma_\varepsilon^{-4} \varepsilon^2 \int_{[-\varepsilon, \varepsilon]} x^2 \Pi(dx) = \sigma_\varepsilon^{-2} \varepsilon^2.$$

Corollary 3.3 (see also Remark 3.4), readily gives a bound on the mean squared maximal error:

$$\mathbb{E} \sup_{t \in [0,1]} |X_\varepsilon(t) - W_\varepsilon(t)|^2 = O(\sigma_\varepsilon^{-1} \varepsilon \sqrt{|\log(\sigma_\varepsilon^{-1} \varepsilon)|}),$$

where  $|\log(\sigma_\varepsilon^{-1} \varepsilon)| = O(|\log \varepsilon|)$ , achievable by choosing  $k_\varepsilon \sim c \sigma_\varepsilon \varepsilon^{-1} \sqrt{|\log \varepsilon|}$  for some  $c > 0$ . In words, the quality of our coupling is directly linked to the condition  $\varepsilon/\sigma_\varepsilon \rightarrow 0$  and its rate of convergence.

Letting  $\beta$  be the Blumenthal–Gettoor index of  $X$ , i.e.

$$\beta = \inf \left\{ p \geq 0 : \int_{[-1,1]} |x|^p \Pi(dx) < \infty \right\} \in [0, 2], \quad (12)$$

we find that  $\sigma_\varepsilon^2 \leq \varepsilon^{2-\beta_+} \int_{[-\varepsilon, \varepsilon]} x^{\beta_+} \Pi(dx) = O(\varepsilon^{2-\beta_+})$  for any  $\beta_+ > \beta$ . This readily yields an upper bound on the distance of scaled processes (as arising in applications):

$$\mathbb{E} \sup_{t \in [0,1]} |M_\varepsilon(t) - \sigma_\varepsilon W_\varepsilon(t)|^2 = O(\varepsilon^{2-\beta_+/2}), \quad \text{with } k_\varepsilon \sim \varepsilon^{-\beta/2}.$$

More can be said under an additional lower bound assumption on the jump activity of  $X$ . For instance, if  $\bar{\Pi}(\varepsilon) = \Pi(\mathbb{R} \setminus [-\varepsilon, \varepsilon])$  is regularly varying at 0, then the corresponding index of regular variation must be  $-\beta$ , and by standard theory [3, Sec. 1.5 and 1.6] we find that  $\sigma_\varepsilon^2$  is regularly varying with index  $2 - \beta$ . Thus,  $\varepsilon/\sigma_\varepsilon$  is regularly varying with index  $\beta/2$  and we get an upper bound

$$\mathbb{E} \sup_{t \in [0,1]} |X_\varepsilon(t) - W_\varepsilon(t)|^2 = O(\varepsilon^{\beta-/2}), \quad \beta_- < \beta.$$

In this case  $k_\varepsilon$ , as prescribed by Remark 3.4, is regularly varying with index  $-\beta/2$ .

## 5. APPLICATION TO THE MULTILEVEL MONTE CARLO METHOD

Let us now return to the multilevel Monte Carlo method for the computation of  $\mathbb{E}g(X)$  with  $\mathbb{E}g^2(X) < \infty$  as discussed in the introduction. Recall that  $X_n$  is an approximation of  $X$  obtained by replacing the martingale of jumps in  $[-\varepsilon_n, \varepsilon_n]$  by a scaled Brownian motion with the same variance

$$\sigma_n^2 := \int_{|x| \leq \varepsilon_n} x^2 \Pi(dx).$$

We choose a geometric sequence of truncation levels, say  $\varepsilon_n = 2^{-n}$ , and assume that the Lévy measure  $\Pi$  of  $X$  has the Blumenthal–Gettoor index  $\beta \in [0, 2]$ , defined in (12). We will assume that  $g$  is Lipschitz in supremum norm with a constant  $L > 0$ .

**5.1. Coupling and level variance.** The crux of the method is to construct a pair  $(X_n, X_{n+1})$  of successive approximations of  $X$  so that the variance  $\text{Var}[g(X_{n+1}) - g(X_n)]$  is small. Consider the decomposition

$$X_{n+1} = M'_n + R'_n,$$

where  $M'_n$  is the martingale of jumps in  $[-\varepsilon_n, -\varepsilon_{n+1}) \cup (\varepsilon_{n+1}, \varepsilon_n]$  and  $R'_n$  is the remainder, an independent process consisting of a Brownian motion and a drifted compound Poisson process with jumps exceeding  $\varepsilon_n$ . We form  $X_n$  as an independent sum of  $R'_n$  and  $\sigma'_n W'_n$ , where  $\sigma_n'^2 = \sigma_n^2 - \sigma_{n+1}^2$  and  $W'_n$  is a certain standard Brownian motion. Note that the constructed  $X_n$  indeed has the desired law. Since  $g$  is Lipschitz in the supremum norm, we have the upper bound

$$\text{Var}[g(X_{n+1}) - g(X_n)] \leq L^2 \cdot \mathbb{E} \sup_{t \in [0,1]} |M'_n(t) - \sigma'_n W'_n(t)|^2. \quad (13)$$

*Remark 5.1.* In the literature (see, e.g. [9, 10]), the process  $W'_n$  is typically an independent Brownian motion. Note that this still provides a coupling of  $X_n$  and  $X_{n+1}$  via the same  $R'_n$ . In this case, by Doob's martingale inequality, the bound in (13) is strictly of order  $O(\sigma_n'^2)$ , which is further upper bounded by  $O(\varepsilon_n^{2-\beta_+})$  for any  $\beta_+ > \beta$ .

Here we propose to construct  $W'_n$  from  $X'_n = M'_n/\sigma'_n$  according to the algorithm with increment reordering presented in this paper with  $k'_n$  as prescribed in Corollary 3.3. We assume that  $\sigma'_n/\varepsilon_n \rightarrow \infty$ , making the Brownian approximation of  $X'_n$  valid in the limiting sense. Note that  $X'_n$  is a compound Poisson process, which we may easily evaluate on any chosen grid. Furthermore, we assume that the comonotonic coupling of the end-points can be implemented with sufficient accuracy, and return to this issue later in §5.4. Finally, as in §4.3, we find by Corollary 3.3 that

$$\mathbb{E} \sup_{t \in [0,1]} |M'_n(t) - \sigma'_n W'_n(t)|^2 = O(\sigma'_n \varepsilon_n \sqrt{|\log \varepsilon_n|}) = O(\varepsilon_n^{2-\beta_+/2}), \quad \beta_+ > \beta,$$

which implies the same upper bound on the level variance in (13). In particular, we have improved the variance by a factor  $\varepsilon_n/\sigma'_n \rightarrow 0$  up to a log term. Furthermore, our chosen number of increments satisfies  $k'_n \sim \sigma'_n \varepsilon_n^{-1} \sqrt{|\log \varepsilon_n|} = O(\varepsilon_n^{-\beta_+/2})$ .

**5.2. Computational complexity.** First, we consider the expected cost of sampling a pair of processes  $(X_n, X_{n+1})$ , where sampling does not include specification of  $k'_n$  independent Brownian bridges for each process. The expected number of jumps of  $X_{n+1}$  is  $\bar{\Pi}(\varepsilon_{n+1}) = O(\varepsilon_n^{-\beta_+})$ , since we have assumed that  $\varepsilon_{n+1} = \varepsilon_n/2$ . This is also the expected cost of drawing all jumps and their times, and calculation of  $k'_n$  increments of  $M'_n$  as well as drawing the Brownian increments, since  $k'_n = O(\varepsilon_n^{-\beta_+/2})$  is smaller. The reordering procedure incurs the cost  $O(k'_n \log k'_n)$ , which is again of smaller order. Thus the cost associated to the level  $n$  is of order  $O(\varepsilon_n^{-\beta_+})$ . In other words, simulating a pair of coupled processes has nearly the same cost as simply drawing a sample of the marginal  $X_{n+1}(1)$ . In addition, we assume that  $g(X_n)$  can be constructed from the above described skeleton at a comparable or lower cost. This is true for a number of functions  $g$ , including the functions  $g(X) = \int_0^1 X(t)dt$  and  $g(X) = \sup_{t \in [0,1]} X(t)$ . In §5.3 below we consider evaluation of  $g$  via further path discretization.

Secondly, to control the bias, we use [9, Cor. 6.2], which implies that

$$|\mathbb{E}g(X) - \mathbb{E}g(X_n)| = O(\varepsilon_n |\log \varepsilon_n|).$$

Note that any coupling can be used to produce a weak bound here, as it does not need to be implementable. Attempting to control the bias with a square root of the second moment and using a bound from §4.3 leads to a slightly worse estimate  $O(\varepsilon_n^{1-\beta_+/4})$ .

Finally, we recall that  $\varepsilon_n = 2^{-n}$  and apply [16, Thm. 2.1] (with  $\alpha' < 1$ ,  $\beta' < 2 - \beta/2$  and  $\gamma' > \beta$  arbitrarily close to their boundary values) to find an MLMC algorithm with mean squared error smaller than  $\delta^2$  and computational complexity  $\mathcal{C}_\delta$  satisfying

$$\mathbb{E}\mathcal{C}_\delta = \begin{cases} O(\delta^{-2}), & \beta < 4/3, \\ O(\delta^{-\frac{3}{2}\beta_+}), & \beta \geq 4/3. \end{cases}$$

**5.3. On discretization of paths.** Even though the structure of  $X_n$  allows to simulate  $g(X_n)$  exactly for a number of functions  $g$ , we assume that a discretization is used to approximate  $g(X_n)$ . That is, we discretize the Brownian component of  $X_n$  on a fine grid including the times  $\{0, 1/m_n, \dots, 1\}$  and the above sampled points of  $X_n$ . The error (both bias and level variance) of the resulting approximation  $g(Y_n)$  may be controlled using an analogue of (8) below by virtue of the Lipschitz continuity of  $g$ :

$$\mathbb{E}|g(Y_n) - g(X_n)|^2 \leq L^2 \cdot \mathbb{E} \sup_{t \in [0,1]} |Y_n(t) - X_n(t)|^2 = O(\log m_n/m_n).$$

The orders of the bias, level variance and cost corresponding to the use of the approximations  $Y_n$  to estimate  $X$  are thus

$$O(\varepsilon_n |\log \varepsilon_n| + \sqrt{\log m_n/m_n}), \quad O(\varepsilon_n^{2-\beta_+/2} + \log m_n/m_n), \quad O(\varepsilon_n^{-\beta_+} + m_n),$$

respectively. Taking  $m_n$  much greater than  $\varepsilon_n^{-2}$  increases the cost without reducing the order of the bias or level variance. Similarly, taking  $m_n$  much smaller than  $\varepsilon_n^{-\beta}$  is also suboptimal, since the bias grows without reducing the order of the level variance and cost.

Letting  $m_n \sim \varepsilon_n^{-p}$  for some  $p \in [\beta, 2]$  we may again apply [16, Thm. 2.1] (but with  $\alpha' < p/2$ ,  $\beta' < (2 - \beta/2) \wedge p$  and  $\gamma' > \beta \vee p$  arbitrarily close to their boundary values). This leads to the minimization of  $(\beta \vee p - (2 - \beta/2) \wedge p)/p$  achieved by  $p = \beta$ , and the expected complexity bound

$$\mathbb{E}\mathcal{C}_\delta = O(\delta^{-(5-4/\beta)\vee 2-\varepsilon}), \quad \varepsilon > 0,$$

for the required precision  $\delta \downarrow 0$ . A more detailed analysis shows that the upper bound for  $\beta < 4/3$  can be improved to  $O(\delta^{-2} |\log \delta|^3)$ . Since  $m_n$  is of the same order as the expected number of jumps, it is, in fact, not strictly required to further discretize the path. The randomness in jump times and their number will affect the small order term and not the power.

Finally, we stress that the standard way of taking an independent Brownian motion  $W'_n$ , as discussed in Remark 5.1, results in the upper bounds  $\mathbb{E}\mathcal{C}_\delta = O(\delta^{-(6-4/\beta)\vee 2-\varepsilon})$  and  $\mathbb{E}\mathcal{C}_\delta = O(\delta^{-(2\beta)\vee 2-\varepsilon})$

for any  $\varepsilon > 0$  when using  $g(Y_n)$  and  $g(X_n)$ , respectively. Indeed, this follows from an analogous analysis and another application of [16, Thm. 2.1] (see also [9, Cor. 1.2]). Thus our coupling leads to an improved computational complexity in the case when  $\beta > 1$ , and this improvement is substantial for  $\beta$  away from 1. Graphical comparison of the respective powers is presented in Figure 2, where we also include the case when  $g(X_n)$  can be sampled exactly. For further comparison, it is noted that an ordinary MC results in  $\mathbb{E}\mathcal{C}_\delta = O(\delta^{-4-\varepsilon})$  when paths are discretized and  $\mathbb{E}\mathcal{C}_\delta = O(\delta^{-2-\beta-\varepsilon})$  when exact simulation of  $g(X_n)$  is possible.

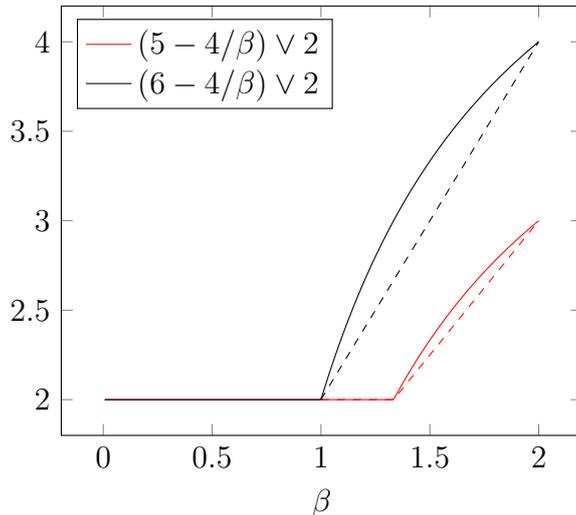


FIGURE 2. Asymptotic upper bounds on  $\log \mathbb{E}\mathcal{C}_\delta / |\log \delta|$  in MLMC as precision  $\delta \downarrow 0$ . The coupling proposed in this paper is in red; solid (resp. dashed) lines correspond to the use of  $g(Y_n)$  (resp.  $g(X_n)$ ).

**5.4. Near-comonotonic coupling.** Ideally we want to sample  $X'_n(1)$  and  $W'_n(1)$  comonotonically since this coupling minimizes the  $L^2$ -distance. Sampling from a coupling that does not increase the order of the  $L^2$ -distance is a rudimentary fundamental problem. One way is described in Lemma B.1 and it consists of drawing a large number of independent copies of both random variables for each required sample. This, however, comes at the expense of significantly increasing the computational cost. Simple analysis based on Corollary 3.3 and Lemma B.1 shows that the order of  $\varepsilon_n^{-\beta}$  of final values needs to be generated for each sample, unless one is willing to reuse these values while controlling the induced dependence.

Another way is to numerically evaluate the distribution function of  $X'_n(1)$  including identification of the associated atoms if such exist. In this regard, we observe that the Brownian component of  $X_{n+1}$  could have been pushed into  $M'_n$  instead of  $R'_n$  when constructing the coupled pairs  $(X_n, X_{n+1})$ . This change makes our problem more similar to the one studied in §4.2 and affects neither the convergence rates nor the bounds on the MLMC computational complexity. It results in a smoothing effect, since now  $X'_n$  has a Brownian component and the distribution of  $X'_n(1)$  is continuous. Fast Fourier inversion allows for computationally cheap evaluation of the distribution function with good numerical performance [11, 20]. It is, nevertheless, hard to incorporate the resulting numerical errors in our MLMC complexity analysis.

## 6. NUMERICS

**6.1. Coupling implementation.** For the numerical experiments below we consider a Lévy measure

$$\Pi^0(dx) = (0.4|x|^{-\alpha-1}\mathbf{1}_{\{x \in (-\varepsilon_1, -\varepsilon_2)\}} + 0.6x^{-\alpha-1}\mathbf{1}_{\{x \in (\varepsilon_2, \varepsilon_1)\}}) dx, \quad \alpha = 1.5,$$

with some truncation levels  $0 < \varepsilon_2 < \varepsilon_1$ . Note that it corresponds to a truncation of the Lévy measure of an 1.5-stable process with skewness 0.2. The process  $X$  is the respective drifted compound Poisson process with zero mean, rescaled so that  $\mathbb{E}X^2(1) = 1$ . The comonotonic coupling of the end-points  $W(1)$  and  $X(1)$  is implemented using a close approximation of the distribution function of the latter obtained from 30000 independent realizations; it is fixed within every numerical experiment. A few smoothing options are possible here but they have no visible influence on the results below. The root-mean-squared-maximal distance, the square root of (1), will be computed over the grid of mesh  $2^{-12} \approx 0.00025$  using 1000 independent samples of pairs of paths.

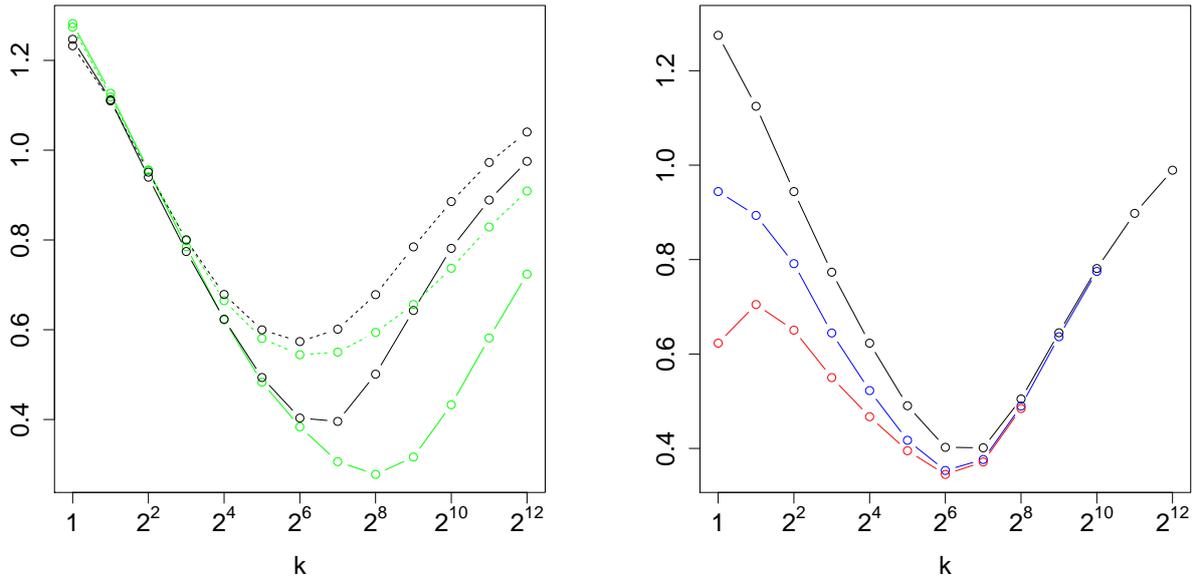


FIGURE 3. Root-mean-squared-maximal distances between  $X$  and  $W$  as a function of  $k$ . Left: processes  $X$  corresponding to  $\varepsilon_1 \in \{0.1, 1\}$  (solid, dashed) and  $\varepsilon_2 \in \{0.01, 0.03\}$  (green and black). Right: second-level reordering with  $\varepsilon_1 = 0.1$ ,  $\varepsilon_2 = 0.03$  and  $k_2 \in \{1, 4, 16\}$  (black, blue, red).

In our first experiment we investigate the quality of our coupling for various values of the parameter  $k$ , which is the number of incremental processes to be permuted, see §1.2. We take  $\varepsilon_1 \in \{0.1, 1\}$  (solid, dashed) and  $\varepsilon_2 \in \{0.01, 0.03\}$  (green and black), resulting in four different processes  $X$ . We apply our coupling for  $k = 2^i$ ,  $i = 0, \dots, 12$ , and plot in Figure 3 (left) the estimated root-mean-squared-maximal distance. Observe that the optimal  $k$  (among powers of 2) is larger for processes better approximating the Brownian motion, that is, when  $\varepsilon_1, \varepsilon_2$  are smaller. As explained in the Introduction, the case of a very large  $k$  need not be good, and should eventually result in the same error as  $k = 1$ .

In our second experiment we take  $\varepsilon_1 = 0.1, \varepsilon_2 = 0.03$  (solid black) and apply the reordering idea on two levels, see Remark 1.2. Firstly, we use  $k = 2^i$  as above, and then reorder  $k_2 \in \{1, 4, 16\}$  (black, blue, red) increments in each of the  $k$  pieces. Note that  $k_2 = 1$  corresponds to the first experiment. The corresponding root-mean-squared-maximal distances are presented in Figure 3 (right). In this case the second-level reordering is beneficial, and the optimal  $k$  is similar in the three considered scenarios. The smallest value is 0.40 for the standard coupling with  $k = 2^7$  ( $k = 2^6$  gives almost the same result) and 0.34 for the two-level procedure with  $k = 2^6, k_2 = 16$ , which is about 15% less. In applications, one may first use our standard coupling and find a good  $k$  and then try second-level reordering for a few  $k_2$ .

Next, we illustrate the coupling for the process  $X$  (still corresponding to  $\varepsilon_1 = 0.1, \varepsilon_2 = 0.03$ ) using the above found optimal  $k = 2^7$ . Figure 4 (top left) presents a histogram of maximal absolute distances resulted from 1000 independent trials, which indeed has root-mean-squared value of 0.40.

Furthermore, we plot the pairs of paths corresponding to the 0.05, 0.50, 0.95 quantiles (top-right to bottom-right). Importantly, the root-mean-squared error at the end-point,  $\sqrt{\mathbb{E}(W(1) - X(1))^2}$ , is estimated to be 0.008, which is negligible when compared to the error for the sample paths.

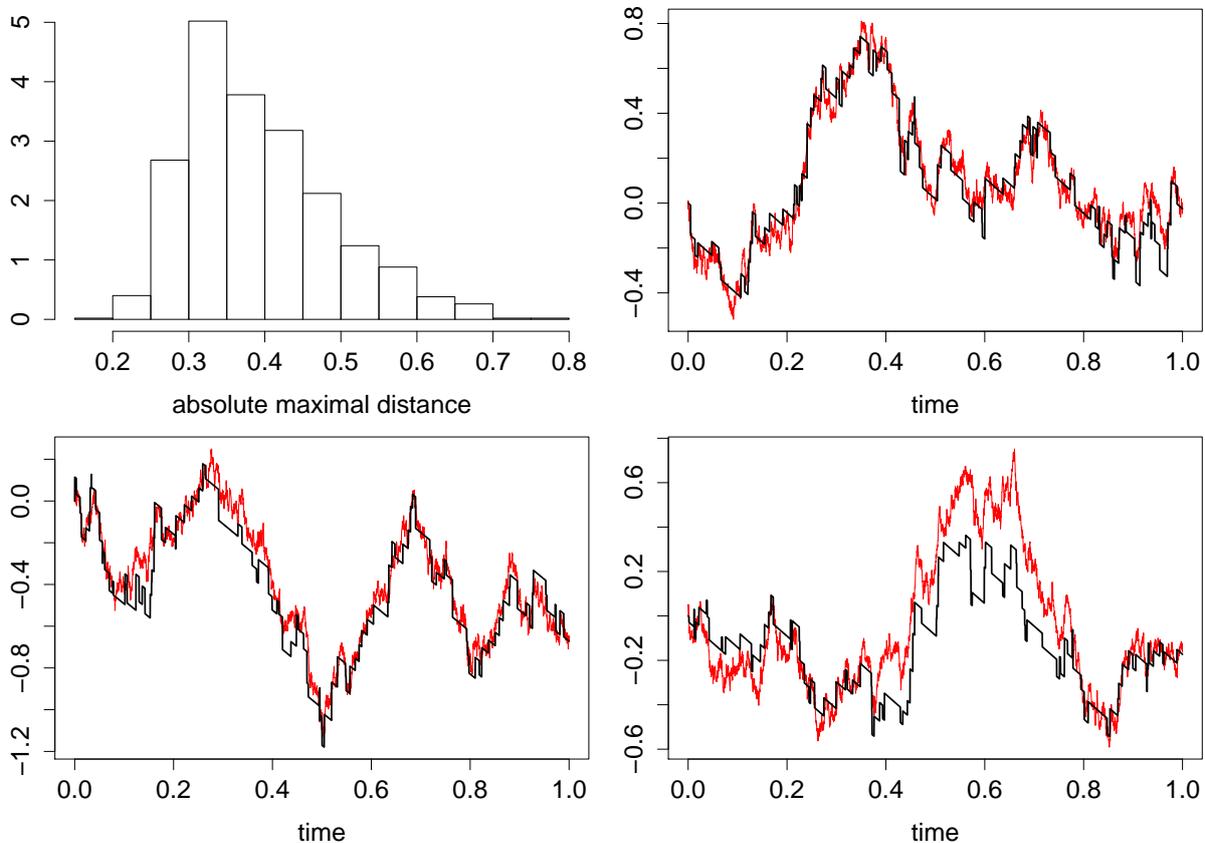


FIGURE 4. Illustration of coupling using  $k = 2^7$ : histogram of maximal absolute distances and three pairs of paths corresponding to 0.05, 0.50, 0.95 quantiles.

Finally, we provide an illustration of the limit result in Corollary 3.3 in the setting similar to §5. Take a geometric sequence of truncation levels  $\varepsilon_{1,n} = 2^{-n}$  and  $\varepsilon_{2,n} = 2^{-n-1}$ , and let  $X_n$  correspond to a rescaled martingale of jumps in  $(-2^{-n}, -2^{-n-1}) \cup (2^{-n-1}, 2^{-n})$  as above. We estimate the root-mean-squared-maximal distance for our coupling using different choices of  $k_n$  (powers of 2) and denote the smallest such distance by  $d_n^*$  and the respective  $k_n$  by  $k_n^*$ . These are plotted in Figure 5 and also compared to their theoretical counterparts in Corollary 3.3. We find a rather good prediction of both the error and the adequate number of increments, which suggests that our theoretical upper bound is rather tight.

In conclusion, we would like to stress that implementation of the coupling presented in this paper is rather straightforward, but, nevertheless, some issues may arise. The increments of a compound Poisson process will normally exhibit ties which must be resolved randomly. Furthermore, for a drifted process such ties may not be detected (numerical rounding) leading to strange non-Brownian trajectories. A simple solution of this numerical problem is to add a negligible Brownian component to the process.

**6.2. Multilevel Monte Carlo.** To test the performance of our coupling algorithm within the context of the multilevel Monte Carlo analysis of §5, we will consider a tempered stable Lévy process with zero mean and Lévy measure

$$\Pi(dx) = 0.05|x|^{-\alpha-1}e^{-|x|}dx, \quad \alpha \in \{1.2, 1.5\},$$

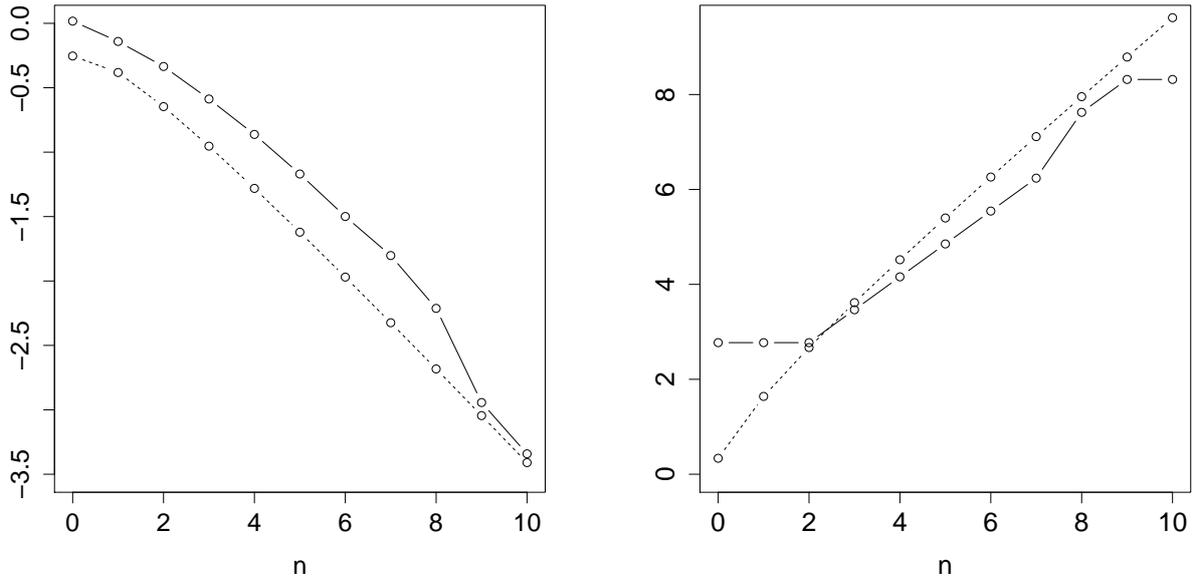


FIGURE 5. The optimal root-mean-squared-maximal distance  $d_n^*$ , the respective number  $k_n^*$ , and their theoretical counterparts for  $\varepsilon_{1,n} = 2^{-n}$  and  $\varepsilon_{2,n} = 2^{-n-1}$ . Left:  $\log d_n^*$  (solid) and  $\log(\mu_{4,n} |\log \mu_{4,n}|) / 4$  (dashed). Right:  $\log k_n^*$  (solid) and  $\log(|\log \mu_{4,n}| / \mu_{4,n}) / 2$  (dashed).

and the truncation levels  $\varepsilon_n = 2^{-n/4}$ . To couple the endpoints  $X'_n(1)$  and  $W'_n(1)$ , we will take the samples we produced for the estimation and match the pairs by rank order, as discussed in §5.4. To test the empirical performance of the multilevel Monte Carlo estimation, it suffices to verify that the bias and level variance converge at the predicted rate or faster. Testing the accuracy of our estimation would require access to the value of  $\mathbb{E}g(X)$ , which is rarely available analytically. To obtain a numerical approximation, we consider the function  $g(X) = \sup_{t \in [0,1]} X(t)$  and use the methodology proposed in [19], which is more limited but very efficient for this example.

Following the analysis of §5, we chose  $k'_n \sim \varepsilon_n^{-\alpha/2} \sqrt{|\log \varepsilon_n|}$  and  $m_n \sim \varepsilon_n^{-p}$ . In multilevel Monte Carlo, the number of samples per level depends on the desired accuracy of the estimator, the level variance and simulation cost, they decrease exponentially in  $n$  and are typically given adaptively as the simulations are run. For simplicity, we drew  $10^{6-\lfloor n/15 \rfloor}$  samples of level  $n$  for each  $n$ . Figure 6 shows the convergence rate of the bias and level variance of the Monte Carlo estimate in §5 as a function of the truncation levels  $\varepsilon_n$  with the top (resp. bottom) figures corresponding to the case  $\alpha = 1.2$  (resp.  $\alpha = 1.5$ ). The dashed lines indicate the predicted rate of convergence in both cases (see details in §5). The bias and level variance both satisfy the stated bounds, with a particularly good agreement between the the level variance and its bound.

#### APPENDIX A. BOUNDS FOR DISCRETIZED PROCESSES

Proposition 2.7 provides an upper bound for a similar discretization error for a Lévy process. Its proof is based on the following three auxiliary results.

**Lemma A.1.** *Let  $\varepsilon > 0$  and  $X$  be a Lévy process without a Brownian component with  $\mathbb{E}X(1) = 0$ ,  $\mathbb{E}X^2(1) \leq 1$  and the corresponding Lévy measure  $\Pi$  supported on  $[-\varepsilon, \varepsilon]$ . Then for any  $t, x, u > 0$  we have*

$$\mathbb{P}\left(\sup_{s \in [0,t]} X(s) \geq x\right) \leq \exp(-ux + u^2 e^{u\varepsilon} t).$$

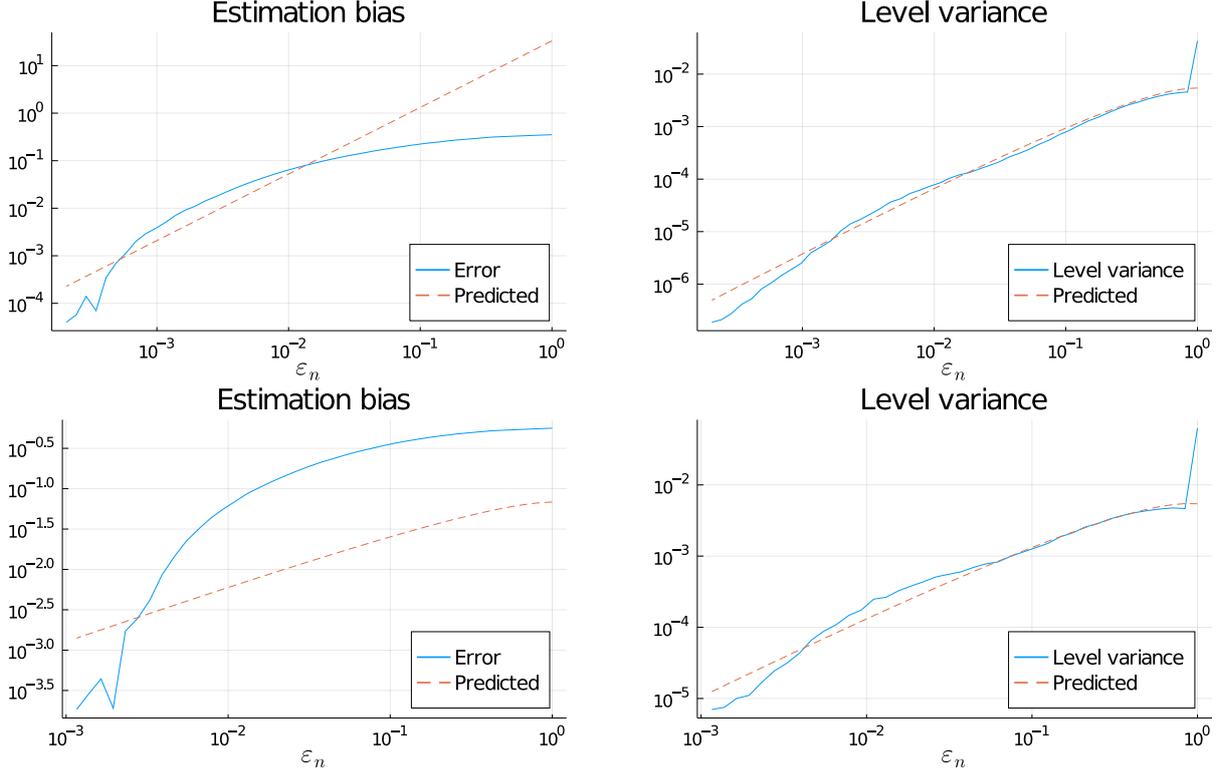


FIGURE 6. The figures show the bias and level variance of the multilevel Monte Carlo method discussed in §5. Top:  $\alpha = 1.2$ , bottom:  $\alpha = 1.5$ , left: decay of the bias  $\mathbb{E}(g(X) - g(Y_n))$  (solid) and its bound  $\varepsilon_n^{\alpha/2} \sqrt{|\log \varepsilon_n|}$  (dashed), right: decay of the level variance  $\text{Var}(g(Y_{n+1}) - g(Y_n))$  (solid) and its bound  $\max\{\varepsilon_n^{2-\alpha/2}, \varepsilon_n^\alpha |\log \varepsilon_n|\}$  (dashed).

*Proof.* The inequality  $e^z - 1 - z \leq z^2 e^{|z|}$  yields

$$\Psi(u) := \int_{[-\varepsilon, \varepsilon]} (e^{uz} - 1 - uz) \Pi(dz) \leq u^2 e^{u\varepsilon} \int_{[-\varepsilon, \varepsilon]} z^2 \Pi(dz) \leq u^2 e^{u\varepsilon}.$$

Since  $\exp(uX(t) - t\Psi(u))$  is a martingale and  $\Psi(u)$  is non-negative,  $\exp(uX(t))$  is a submartingale and Doob's maximal inequality [22, Prop. 7.15] yields

$$\mathbb{P}\left(\sup_{s \in [0, t]} X(s) \geq x\right) = \mathbb{P}\left(e^{u \sup_{s \in [0, t]} X(s)} \geq e^{ux}\right) \leq e^{-ux} \mathbb{E} e^{uX(t)} = e^{-ux + t\Psi(u)}$$

implying the stated bound.  $\square$

**Lemma A.2.** *There is a constant  $C > 0$  such that*

$$\mathbb{E} \sup_{t \in [0, 1]} (X(t) - X^{[k]}(t))^2 \leq C \log k / k$$

for all  $k \geq 2$  and every Lévy process  $X$  satisfying the conditions of Lemma A.1 with  $\varepsilon \leq (k \log k)^{-1/2}$ .

*Proof.* Let  $E_1, E_2, \dots$  be i.i.d. standard exponential random variables. For any  $u, x > 0$  we have

$$\mathbb{P}(E_1/u + ue^{u\varepsilon}/k \geq x) = \exp(-ux + u^2 e^{u\varepsilon}/k) \wedge 1,$$

which is also an upper bound on  $\mathbb{P}(\sup_{s \in [0, 1/k]} X(s) \geq x)$  by Lemma A.1. Thus,  $\sup_{s \in [0, 1/k]} X(s)$  and similarly  $-\inf_{s \in [0, 1/k]} X(s)$  are stochastically bounded by  $E_1/u + ue^{u\varepsilon}/k$ .

Let  $(\bar{G}_i, \underline{G}_i)$  be independent copies of  $(\sup_{s \in [0, 1/k]} X(s), -\inf_{s \in [0, 1/k]} X(s))$ . Then

$$\mathbb{E} \sup_{t \in [0, 1]} (X(t) - X^{[k]}(t))^2 = \mathbb{E} \max_{1 \leq i \leq k} (\bar{G}_i \vee \underline{G}_i)^2 \leq \mathbb{E} \max_{1 \leq i \leq k} (\bar{G}_i^2 + \underline{G}_i^2) \leq \mathbb{E} \max_{1 \leq i \leq k} \bar{G}_i^2 + \mathbb{E} \max_{1 \leq i \leq k} \underline{G}_i^2.$$

But each of the latter is upper bounded by

$$\mathbb{E} \max_{1 \leq i \leq k} (E_i/u + ue^{u\varepsilon}/k)^2 \leq \frac{2}{u^2} \mathbb{E} \max_{1 \leq i \leq k} E_i^2 + \frac{2u^2}{k^2} e^{2u\varepsilon}.$$

It is a basic fact that

$$\max_{1 \leq i \leq k} E_i - \log k \xrightarrow{d} G, \quad k \rightarrow \infty,$$

where  $G$  has the standard Gumbel distribution. From [29] we also have

$$\mathbb{E} \left( \max_{1 \leq i \leq k} E_i - \log k \right) \rightarrow \mathbb{E}G, \quad \mathbb{E} \left( \max_{1 \leq i \leq k} E_i - \log k \right)^2 \rightarrow \mathbb{E}G^2,$$

implying  $\mathbb{E} \max_{1 \leq i \leq k} E_i^2 = \log^2 k + O(\log k)$ . Hence we may choose a constant  $C'$  such that

$$\mathbb{E} \sup_{t \in [0,1]} (X(t) - X^{[k]}(t))^2 \leq C' \frac{2}{u^2} \log^2 k + \frac{2u^2}{k^2} e^{2u\varepsilon}$$

for all  $k \geq 2$  and all  $X$  satisfying our assumptions. Choose  $u = \sqrt{k \log k}$  to get the result.  $\square$

**Lemma A.3.** *There is a universal constant  $C > 0$  such that*

$$\mathbb{E} \sup_{t \in [0,1]} (X(t) - X^{[k]}(t))^2 \leq C(\varepsilon^{-2} \mu_4 / \sqrt{k} + \sqrt{\mu_4}),$$

for any  $k \geq 1$ , any  $\varepsilon > 0$  and any compensated compound Poisson process  $X$  with jumps of size at least  $\varepsilon$ .

*Proof.* Applying Doob's martingale inequality [22, Prop. 7.16], we obtain

$$\mathbb{E} \sup_{t \in [0,1]} (X(t) - X^{[k]}(t))^4 \leq k \mathbb{E} \sup_{t \in [0,1/k]} X^4(t) \leq (4/3)^4 \cdot k \mathbb{E} X^4(1/k).$$

Using the moment formula in terms of cumulants and the equality  $\mathbb{E}X(1) = 0$ , we obtain

$$\mathbb{E} X^4(t) = 3t^2 \left( \int_{|x| \geq \varepsilon} x^2 \Pi(dx) \right)^2 + t \int_{|x| \geq \varepsilon} x^4 \Pi(dx),$$

and noting that

$$\int_{|x| \geq \varepsilon} x^2 \Pi(dx) \leq \varepsilon^{-2} \int_{|x| \geq \varepsilon} x^4 \Pi(dx) = \varepsilon^{-2} \mu_4,$$

we get

$$k \mathbb{E} X^4(1/k) \leq 3\varepsilon^{-4} \mu_4^2 / k + \mu_4.$$

By Jensen's inequality,

$$\mathbb{E} \sup_{t \in [0,1]} (X(t) - X^{[k]}(t))^2 \leq 16/9 \cdot (3\varepsilon^{-4} \mu_4^2 / k + \mu_4)^{1/2},$$

and the result follows.  $\square$

*Proof of Proposition 2.7.* Let  $\varepsilon = (k \log k)^{-1/2}$  and consider the Lévy-Itô decomposition  $X(t) = \sigma B(t) + Y(t) + Z(t)$ , where  $B$  is a standard Brownian motion,  $Y$  is a compensated compound Poisson process with jumps of size exceeding  $\varepsilon$ , and  $Z$  is a pure-jump martingale with jumps of size at most  $\varepsilon$ . Since  $\sigma^2 \leq \mathbb{E}X^2(1) \leq 1$ , we have

$$\sup_{t \in [0,1]} |X(t) - X^{[k]}(t)| \leq \sup_{t \in [0,1]} |B(t) - B^{[k]}(t)| + \sup_{t \in [0,1]} |Y(t) - Y^{[k]}(t)| + \sup_{t \in [0,1]} |Z(t) - Z^{[k]}(t)|.$$

Bounding each term in the squared mean sense by using relation (8), Lemma A.2 and Lemma A.3, for all  $k \geq 2$  we get the bound

$$C(\sqrt{k} \log k \cdot \mu_4 + \sqrt{\mu_4} + \log k/k).$$

It is left to note that  $\log k < \sqrt{k}$  and  $\mu_4/(k\mu_4 + \log k/k)^2 \leq 1/(2 \log k) \leq 1/(2 \log 2)$ .  $\square$

We note that Proposition 2.7 holds under less restrictive assumptions. It is sufficient to assume instead of (2) that  $\mathbb{E}X^2(1) \leq 1$ ,  $\mathbb{E}X^4(1) < \infty$ , since the additional drift introduces an error of order  $[\mathbb{E}X(1)/k]^2$ , which does not change the bound.

#### APPENDIX B. NEAR-COMONOTONIC COUPLING VIA REPEATED SIMULATIONS

**Lemma B.1.** *Let  $\xi, \xi_1, \xi_2, \dots$  and  $\zeta, \zeta_1, \zeta_2, \dots$  be two independent i.i.d. sequences with laws  $\mu_\xi$  and  $\mu_\zeta$ , respectively, with finite fourth moment. For any  $n \geq 2$  let  $\xi_{(1)} \leq \dots \leq \xi_{(n)}$  and  $\zeta_{(1)} \leq \dots \leq \zeta_{(n)}$  be the ranked values of  $\xi_1, \dots, \xi_n$  and  $\zeta_1, \dots, \zeta_n$ , respectively. Then for any independent uniform random variable  $U$  on  $\{1, \dots, n\}$ , we have  $\xi_{(U)} \sim \mu_\xi$ ,  $\zeta_{(U)} \sim \mu_\zeta$  and*

$$\mathbb{E}(\xi_{(U)} - \zeta_{(U)})^2 \leq 3\mathcal{W}_2^2(\mu_\xi, \mu_\zeta) + \frac{c \log n}{\sqrt{n}} (\sqrt{\mathbb{E}\xi^4} + \sqrt{\mathbb{E}\zeta^4}),$$

where  $c > 0$  is a universal constant independent of  $\mu_\xi, \mu_\zeta$  and  $n$ .

*Proof.* The independence and exchangeability imply that  $\xi_{(U)} \sim \mu_\xi$  and  $\zeta_{(U)} \sim \mu_\zeta$ . Define the empirical distribution functions  $\mu_\xi^{(n)} := n^{-1} \sum_{k=1}^n \delta_{\xi_k}$  and  $\mu_\zeta^{(n)} := n^{-1} \sum_{k=1}^n \delta_{\zeta_k}$ . By [5, Cor. 7.18] (and the discussion preceding it), there exists a universal constant  $c > 0$  such that then

$$\mathbb{E}\mathcal{W}_2^2(\mu_\xi^{(n)}, \mu_\xi) \leq \frac{c \log n}{\sqrt{n}} \sqrt{\mathbb{E}\xi^4}, \quad \text{and} \quad \mathbb{E}\mathcal{W}_2^2(\mu_\zeta^{(n)}, \mu_\zeta) \leq \frac{c \log n}{\sqrt{n}} \sqrt{\mathbb{E}\zeta^4}, \quad \forall n \geq 2. \quad (14)$$

Note that

$$\begin{aligned} \mathbb{E}(\xi_{(U)} - \zeta_{(U)})^2 &= \frac{1}{n} \sum_{k=1}^n \mathbb{E}(\xi_{(k)} - \zeta_{(k)})^2 = \mathbb{E}\mathcal{W}_2^2(\mu_\xi^{(n)}, \mu_\zeta^{(n)}) \\ &\leq 3\mathcal{W}_2^2(\mu_\xi, \mu_\zeta) + 3\mathbb{E}\mathcal{W}_2^2(\mu_\xi^{(n)}, \mu_\xi) + 3\mathbb{E}\mathcal{W}_2^2(\mu_\zeta^{(n)}, \mu_\zeta). \end{aligned}$$

Thus, an application of (14) gives the claim.  $\square$

#### ACKNOWLEDGMENTS

VF and JI gratefully acknowledge financial support of Sapere Aude Starting Grant 8049-00021B ‘‘Distributional Robustness in Assessment of Extreme Risk’’. JGC is grateful for the support of The Alan Turing Institute under EPSRC grant EP/N510129/1 and CoNaCyT scholarship 2018-000009-01EXTF-00624 CVU699336.

#### REFERENCES

- [1] 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.
- [2] P. Berthet and J. C. Fort. Exact rate of convergence of the expected  $W_2$  distance between the empirical and true Gaussian distribution. *Electron. J. Probab.*, 25(12):1–16, 2020.
- [3] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*. Cambridge University Press, Cambridge, 1989.
- [4] J. Blanchet and K. Murthy. Quantifying distributional model risk via optimal transport. *Math. Oper. Res.*, 44(2):565–600, 2019.
- [5] S. Bobkov and M. Ledoux. One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.*, 261(1259):v+126, 2019.
- [6] B. Böttcher, R. L. Schilling, and J. Wang. Constructions of coupling processes for Lévy processes. *Stochastic Process. Appl.*, 121(6):1201–1216, 2011.
- [7] T. Breuer and I. Csizsár. Systematic stress tests with entropic plausibility constraints. *J. Bank. Finance*, 37(5):1552–1559, 2013.
- [8] S. Chobanyan, S. Levental, and H. Salehi. A distribution maximum inequality for rearrangements of summands. *Bulletin of the Georgian National Academy of Sciences*, 5(3):25–30, 2011.
- [9] S. Dereich. Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *Ann. Appl. Probab.*, 21(1):283–311, 2011.

- [10] S. Dereich and F. Heidenreich. A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stochastic Process. Appl.*, 121:1565–1587, 2011.
- [11] P. Embrechts and M. Frei. Panjer recursion versus FFT for compound distributions. *Math. Methods Oper. Res.*, 69(3):497–508, 2009.
- [12] P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Math. Program.*, 171(1):115–166, 2018.
- [13] A. Ferreiro-Castilla, A. Kyprianou, R. Scheichl, and G. Suryanarayana. Multilevel Monte Carlo simulation for Lévy processes based on the Wiener–Hopf factorisation. *Stochastic Process. Appl.*, 124(2):985–1010, 2014.
- [14] J. E. Figueroa-López. Small-time moment asymptotics for Lévy processes. *Stat. Probab. Lett.*, 78(18):3355–3365, 2008.
- [15] N. Fournier. Simulation and approximation of Lévy-driven stochastic differential equations. *ESAIM: PS*, 15:233–248, 2011.
- [16] M. B. Giles. Multilevel Monte Carlo methods. *Acta Numer.*, 24:259–328, 2015.
- [17] M. B. Giles and Y. Xia. Multilevel Monte Carlo for exponential Lévy models. *Finance Stoch.*, 21(4):995–1026, 2017.
- [18] J. González Cázares and J. Ivanovs. Recovering Brownian and jump parts from high-frequency observations of a Lévy process. *Bernoulli*, 2021. (in print) arxiv:2003.05363.
- [19] J. I. González Cázares and A. Mijatović. Monte carlo algorithm for the extrema of tempered stable processes. 2021.
- [20] P. Hughett. Error bounds for numerical inversion of a probability characteristic function. *SIAM J. Numer. Anal.*, 35(4):1368–1392, 1998.
- [21] J. Jacod, A. Jakubowski, and J. Mémin. On asymptotic errors in discretization of processes. *Ann. Probab.*, 31(2):592–608, 2003.
- [22] O. Kallenberg. *Foundations of modern probability*. Springer-Verlag, New York, second edition, 2002.
- [23] D. Khoshnevisan. An embedding of compensated compound Poisson processes with applications to local times. *Ann. Probab.*, 21(1):340–361, 1993.
- [24] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent rv’s, and the sample df. I. *Z. Wahrsch. verw. Geb.*, 32:111–131, 1975.
- [25] M. B. Majka. Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes. *Stochastic Process. Appl.*, 127(12):4083–4125, 2017.
- [26] E. Mariucci and M. Reiß. Wasserstein and total variation distance between marginals of Lévy processes. *Electron. J. Stat.*, 12:2482–2514, 2018.
- [27] I. Monroe. On the  $\gamma$ -variation of processes with stationary independent increments. *Ann. Math. Stat.*, 43(4):1213–1220, 1972.
- [28] R. Petterson. Approximations for stochastic differential equations with reflecting convex boundaries. *Stochastic Process. Appl.*, 59:295–308, 1995.
- [29] J. Pickands III. Moment convergence of sample extremes. *Ann. Math. Stat.*, 39(3):881–889, 1968.
- [30] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I*. Springer-Verlag, New York, 1998.
- [31] E. Rio. Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):802–817, 2009.
- [32] E. Rio. *Asymptotic theory of weakly dependent random processes*. Springer-Verlag, Berlin, Heidelberg, 2017.
- [33] L. Rüschendorf. On the distributional transform, Sklar’s theorem, and the empirical copula process. *J. Stat. Planning and Inference*, 139(11):3921–3927, 2009.
- [34] V. Strassen. An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. verw. Geb.*, 3(3):211–226, 1964.
- [35] V. Strassen. Almost sure behavior of sums of independent random variables and martingales. *Proc. 5th Berkeley Sympos. Math. Statist. Probab.*, 5.2A:315–343, 1967.
- [36] H. Thorisson. *Coupling, stationarity, and regeneration*. Springer-Verlag, New York, 2000.
- [37] A. Y. Zaitsev. Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. *ESAIM: PS*, 2:41–108, 1998.