Strict Kantorovich contractions for Markov chains and Euler schemes with general noise

Citation for published version:
Huang, L-J, Majka, MB & Wang, J 2022, 'Strict Kantorovich contractions for Markov chains and Euler schemes with general noise', Stochastic Processes and their Applications, vol. 151, pp. 307-341. https://doi.org/10.1016/j.spa.2022.06.011

Digital Object Identifier (DOI):
10.1016/j.spa.2022.06.011

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Peer reviewed version

Published In:
Stochastic Processes and their Applications

Publisher Rights Statement:
© 2022 Elsevier B.V.

General rights
Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
STRICT KANTOROVICH CONTRACTIONS FOR MARKOV CHAINS AND EULER SCHEMES WITH GENERAL NOISE

LU-JING HUANG  MATEUSZ B. MAJKA  JIAN WANG

Abstract. We study contractions of Markov chains on general metric spaces with respect to some carefully designed distance-like functions, which are comparable to the total variation and the standard $L^p$-Wasserstein distances for $p \geq 1$. We present explicit lower bounds of the corresponding contraction rates. By employing the refined basic coupling and the coupling by reflection, the results are applied to Markov chains whose transitions include additive stochastic noises that are not necessarily isotropic. This can be useful in the study of Euler schemes for SDEs driven by Lévy noises. In particular, motivated by recent works on the use of heavy tailed processes in Markov Chain Monte Carlo, we show that chains driven by the $\alpha$-stable noise can have better contraction rates than corresponding chains driven by the Gaussian noise, due to the heavy tails of the $\alpha$-stable distribution.

Keywords: Markov chain, strict Kantorovich contracitivity, total variation, Wasserstein distance, refined basic coupling, coupling by reflection

MSC 2020: 60J05, 60J22, 65C30, 65C40.

1. Introduction

Let $(S,d)$ be a separable metric space. For any probability measures $\mu$ and $\nu$ on $S$, the Kantorovich distance ($L^1$-Wasserstein distance) with respect to the metric $d$ is defined by

$W_d(\mu,\nu) = \inf_{\xi \sim \mu, \eta \sim \nu} \mathbb{E}[d(\xi,\eta)],$

where the infimum is taken over all pairs of random variables $(\xi,\eta)$ defined on a common probability space $(\Omega,\mathcal{F})$ such that $\xi$ (resp. $\eta$) is distributed as $\mu$ (resp. $\nu$). A Markov chain with the transition kernel $p(x,dz)$ on $(S,d)$ is called strictly Kantorovich contractive with respect to $d$, if there exists a constant $c \in (0,1)$ such that

$W_d(p(x,\cdot),p(y,\cdot)) \leq (1-c)d(x,y), \quad x, y \in S.$

Strict Kantorovich contractivity goes back to the famous work [9] by Dobrushin, and (1.2) is also known as the “Dobrushin uniqueness condition”. It is closely related to the ergodicity of Markov processes, particle systems or other random dynamical systems, see e.g. [3, 8] and the references therein. In particular, strict Kantorovich contractivity provides an estimate for the spectral gap, cf. [3]. Ollivier in [28] introduced the concept of the “Ricci curvature” of Markov chains on metric spaces, according to which a Markov chain has a “positive Ricci curvature” if (1.2) holds true. Lévy-Gromov-like Gaussian concentration theorem and log-Sobolev inequalities are established in [28] under positive Ricci curvature. See [29] for further extensions in this direction. Strict Kantorovich contractivity also plays an important role in the Monte Carlo method. For example, it has been shown in [19] that the constant $c$ in (1.2) provides upper bounds for the biases of empirical means when we use the Markov chain to simulate a given target distribution. Moreover, strict Kantorovich contractivity is crucial in the study of the perturbation theory for...
Markov chains, and for obtaining quantitative bounds of the biases for Markov Chain Monte Carlo (MCMC) algorithms [30] [18] [32] [25] [31].

However, verifying (1.2) is usually highly non-trivial in applications. In particular, requiring strict Kantorovich contractivity with respect to the underlying distance $d$ of the state space can be too restrictive. Instead, a natural approach is to modify the original distance involved in the strict Kantorovich contractivity condition (1.2). For this purpose, we need to take into consideration the following two issues. One is to find a good Markov coupling $\langle (X, Y), P_{x,y} \rangle$ for the transition kernels $p(x, dz)$ and $p(y, dz)$, and the second is to design a suitable distance-like function $\rho$ such that

$$E[\rho(X, Y)] \leq (1 - c_* \rho(x, y), \quad x, y \in S$$

holds for some $c_* \in (0, 1)$. Here, $\rho(x, y)$ is called a distance-like function on $S \times S$, if $\rho(x, y) = 0$ if and only if $x = y$, and $\rho(x, y) = \rho(y, x) > 0$ for all $x \neq y \in S$. Note that (1.3) naturally implies bounds in a related Kantorovich metric or semi-metric, cf. [13] Lemma 2.1. This approach has been widely used in the research on the ergodicity of SDEs driven by Brownian motions or pure jump Lévy processes (see [11] [12] [22] [34] [24] [23]), and it has proven to be very useful in the study of MCMC (see [16] [12] [13] [23]).

The present paper is strongly motivated by [13], where two different kinds of distance-like functions $\rho$ for Markov chains on general metric state space were provided, and some quantitative bounds on contraction rates for (1.3) were also obtained. Here, on the one hand, we focus on the construction of other distance functions such that (1.3) is satisfied; on the other hand, we present explicit Markov couplings for general Markov chains on $\mathbb{R}^d$, whose transitions have the following form

$$x \mapsto x + hb(x) + g(h)\xi, \quad x \in \mathbb{R}^d$$

with $h, g(h) > 0, b : \mathbb{R}^d \to \mathbb{R}^d$ and $\xi$ being a “noise” random variable, with an arbitrary probability distribution $\mu$. Compared to [13], the improvements of our paper are as follows:

- In [13], chains given by (1.4) were studied only with the Gaussian noise $\xi$. We cover a much larger class of (not necessarily isotropic) additive stochastic noises; including the $\alpha$-stable law with $\alpha \in (1, 2)$. We construct the corresponding Markov couplings such that (1.3) holds with appropriately designed distance-like functions $\rho(x, y)$. For any isotropic noise $\xi$ our coupling is a natural extension of the coupling from [13] (which combines the maximal coupling with the reflection coupling). However, for non-isotropic noise $\xi$ we require a new approach that adapts the so-called refined basic coupling (originally introduced for Lévy-driven SDEs in [22]) to the setting of Markov chains, see the discussion in Section 3.1 for details.

- The designed distance-like functions $\rho(x, y)$ in (1.3) and the associated lower bound estimates for the constant $c_*$ are more straightforward than in [13], which allows for a comparison of contraction rates for chains such as (1.4) driven by different types of noise. In particular, we show how contraction rates corresponding to the $\alpha$-stable noise can be larger than those corresponding to the Gaussian noise (cf. Remark 3.11). This is relevant in the context of applications of chains (1.4) in MCMC methods, see the discussion in the remaining part of the introduction.

- Besides contractions in terms of the total variation or the standard $L^1$-Wasserstein distance as in [13], we study contractions in terms of the $L^p$-Wasserstein distance (with $p > 1$) as well. Note that the $L^p$-Wasserstein distance $W^p_d$, by analogy to (1.1), is defined by $W^p_d(\mu, \nu) = (\inf_{\xi \sim \mu - \nu} E[|\xi|]^{1/p})^{1/p}$, and obtaining contractions with respect to $W^p_d$ requires using functions $\rho$ in (1.3) that are convex, rather than concave, at infinity.

We would like to point out that some of the contraction rates from [13] for chains (1.4) with the Gaussian noise have been recently improved in [7], and related coupling methods have been also extended in [10] to study functional autoregressive processes (again, only with the Gaussian noise). However, in the present paper, as explained above, we focus on extending the results from [13] in different directions.
Markov chains (1.4) with Gaussian noise $\xi$ have been extensively applied in MCMC methods to construct approximate samples from high-dimensional probability distributions $\pi(dx) \propto e^{-U(x)}dx$ for a potential $U: \mathbb{R}^d \to \mathbb{R}$. It is well-known that for a sufficiently regular $U$, such $\pi$ can be obtained as the unique stationary distribution of the Langevin SDE

$$dX_t = -\nabla U(X_t)dt + \sqrt{2}dW_t,$$

where $(W_t)_{t \geq 0}$ is the standard Brownian motion in $\mathbb{R}^d$. If under some assumptions on $U$ we can prove that $(X_t)_{t \geq 0}$ is ergodic, then we can use its discretisation (1.4), with $\xi$ being a multivariate standard normal random variable, $q(h) = (2h)^{1/2}$ and $b(x) = -\nabla U(x)$, as a basis for an algorithm that produces approximate samples from $\pi$.

Recently there has been some interest in MCMC methods utilizing (1.4) with a heavy-tailed noise [27, 33, 35], based on the intuition that such chains may explore the state space better than their Gaussian counterparts, and hence the corresponding approximate sampling algorithms may exhibit faster convergence. Obviously, the analysis of such chains creates numerous new technical challenges. First, for a given measure $\pi(dx) \propto e^{-U(x)}dx$ and a given noise $(Z_t)_{t \geq 0}$, it may be difficult to construct a correct drift $b: \mathbb{R}^d \to \mathbb{R}^d$ such that the SDE

$$dX_t = b(X_t)dt + dZ_t$$

has the stationary distribution $\pi$. Second, even if such $b$ can be constructed, it may be difficult to evaluate it numerically for the purpose of the simulation of the corresponding sampling algorithm. For example, even for the case of a symmetric $\alpha$-stable noise $(Z_t)_{t \geq 0}$, the corresponding drift $b$ is rather complicated and it is given by

$$b(x) = -C_{d,2-\alpha}e^{V(x)}\int_{\mathbb{R}^d} \frac{e^{-V(y)}\nabla V(y)}{|x-y|^{d-(2-\alpha)}} dy$$

when $d > 2 - \alpha$, see [33, 17]. However, as pointed out in [33], it is possible to use algorithms with a heavy-tailed noise in which the correct drift is replaced by a suitable approximation. In particular, in the $\alpha$-stable case described above, the drift $b$ defined by (1.5) can be approximated via a finite series representation (see [33, (9)]) which is easier to evaluate numerically, and whose zero-th term is $-\nabla U$. With such an approximation, we obviously do not sample from the correct target measure $\pi$, but rather from its perturbed version. Still, numerical experiments in [33] demonstrate that this perturbation can be close enough to the target, so that the heavy-tailed chain (1.4) with an incorrect drift, can outperform a corresponding Gaussian chain with a correct drift in certain scenarios. This can be true for instance for measures with double-well potentials, where a chain with a Gaussian noise can become stuck in one well, while a corresponding chain with a heavy-tailed noise manages to leave that well due to its larger jumps (see [33] Figure 2).

The framework presented in our paper is well-suited for studying chains (1.4) with drifts $b(x) = -\nabla U(x)$ for potentials $U$ that are strongly convex outside of a compact set (such as double-well potentials). For such drifts, the discussion in Remark 3.11 demonstrates that chains with an $\alpha$-stable noise can have better contraction rates than corresponding chains with a Gaussian noise. Obviously, in practice these two types of noise require using different drifts and hence Remark 3.11 does not offer a full understanding of the corresponding MCMC problem described in the previous paragraph. To achieve such understanding, we would need a precise analysis of the difference between the exact drift used for the $\alpha$-stable noise and its approximation that we choose to implement numerically (and a quantification of the error between the desired target measure $\pi$ and its perturbation due to the approximated drift). This seems to be a challenging problem that falls beyond the scope of the present paper. However, we believe that any such analysis to be obtained in the future, will require estimates for corresponding contraction rates as presented below, and hence our results constitute an important first step in that direction. Moreover, our results for $L^p$-Wasserstein distances with $p \geq 2$ can be also applied to study Multi-level Monte Carlo (MLMC) methods based on chains (1.4), as explained in Remark 3.17.

The rest of the paper is arranged as follows. In Section 2 we present general results for strict Kantorovich contractions of Markov chains on a separable metric space. In particular, the
contractions in terms of the total variation, the $L^1$-Wasserstein distance and the $L^p$-Wasserstein distance with $p > 1$ are studied. In Section 3, we apply the results from Section 2 to Markov chains whose transitions involve additive stochastic noises. In particular, three different Markov couplings (via the reflection coupling and the refined basic coupling as well as their variants) are constructed in order to illustrate the practicality of the results in Section 2. Section 4 is devoted to proofs of the contractivity results presented in Section 2 and those of some key lemmas required to obtain results for the Euler scheme in Section 3.

2. Strict Kantorovich contractions for Markov chains

Let $p(x, dz)$ be a Markov transition kernel on a separable metric space $(S, d)$. Assume that $((X, Y), P_{x,y})$ is a Markov coupling of $p(x, \cdot)$ and $p(y, \cdot)$ with $x, y \in S$. That is, there are random variables $X, Y : \Omega \to S$ defined on a common measurable space $(\Omega, \mathcal{F})$ and a probability kernel $(x, y, A) \mapsto P_{x,y}(A)$ from $S \times S \times \mathcal{B}(S)$ to $[0, 1]$ such that

$$X \sim p(x, \cdot) \text{ and } Y \sim p(y, \cdot) \text{ under } P_{x,y}.$$ 

Throughout this paper, we always assume that for all $x \in S$,

$$P_{x,x}(d(X, Y) = 0) = P_{x,x}(X = Y) = 1.$$ 

We denote by $E_{x,y}$ the expectation with respect to $P_{x,y}$.

The aim of this section is, for a given Markov coupling $((X, Y), P_{x,y})$, to construct various distance-like functions $\rho(x, y)$ such that the Markov coupling $((X, Y), P_{x,y})$ is strictly contractive in the sense that for any $x, y \in S$, $\rho(x, y)$ holds with some $c_\rho \in (0, 1)$.

To present the distance-like function $\rho(x, y)$ in an explicit way, we need the following notation. For any fixed $x, y \in S$ and $l \geq 0$, define

$$\pi(x, y) = P_{x,y}(d(X, Y) = 0), \quad \beta(x, y) = E_{x,y}[d(X, Y) - d(x, y)],$$

$$\alpha_l(x, y) = \frac{1}{2} E_{x,y}[(d(X, Y) - d(x, y))^2 1_{\{d(X,Y)<d(x,y)+l\}}].$$

(2.1)

For brevity, we denote $\alpha(x, y) = \alpha_0(x, y)$. Roughly speaking, $\pi(x, y)$ indicates the probability that the Markov coupling $((X, Y), P_{x,y})$ will succeed after one step; $\beta(x, y)$ is the drift of the Markov coupling, while $\alpha_l(x, y)$ reflects the fluctuations that mainly decrease the distance of the Markov coupling. Note that the functions $\pi(x, y)$ and $\beta(x, y)$ have been used in $\mathbb{I}_3$ (2.12) and (2.10) before, while $\alpha_l(x, y)$ here is a little different from (indeed is larger than) the corresponding functions in $\mathbb{I}_3$ (2.13) and (2.22).

Furthermore, for any $r > 0$ and $l \geq 0$, set

$$\pi(r) = \inf_{d(x,y) = r} \pi(x, y), \quad \beta(r) = \sup_{d(x,y) = r} \beta(x, y), \quad \alpha_l(r) = \inf_{d(x,y) = r} \alpha_l(x, y).$$

(2.2)

For brevity we write $\alpha(r) = \alpha_0(r)$.

In the remaining part of this section, we will present four results on establishing $\rho(x, y)$ for several different distance-like functions $\rho$. In the first two instances (Theorems 2.1 and 2.3), $\rho$ will be comparable with the total variation metric, in the third instance (Theorem 2.5) with the $L^1$-Wasserstein distance and in the final instance (Theorem 2.7) with the $L^p$-Wasserstein distance for $p > 1$. Hence the first three results can be considered as analogues of the results from $\mathbb{I}_3$ Section 2, albeit with more straightforward formulas for $\rho$ and the corresponding contractivity constant $c_\rho$. This will be crucial in our analysis of contractions for Euler schemes in Section 3, especially in the comparison between schemes with different noises, which would be more cumbersome if attempted directly based on the results from $\mathbb{I}_3$. The fourth result has no counterpart in $\mathbb{I}_3$. The proofs are postponed to Section 4.
2.1. Contraction in terms of metrics comparable to the total variation. In order to consider the contraction of Markov chains in terms of the total variation, we require a positive probability that the Markov coupling \((X, Y), \mathbb{P}_{x,y}\) will succeed after one step when \(d(x, y)\) is small; see (a1) in Assumption (A).

We first make the following assumption.

**Assumption (A)** There exist positive constants \(c_0\) and \(r_0 \leq r_1\) such that

\[
\begin{align*}
(a1) \ & \ \inf_{r \in (0, r_0]} \pi(r) > 0; \\
(a2) \ & \ \inf_{r \in (r_0, r_1]} a(r) > 0 \ \text{and} \ \sup_{r \in (0, r_1]} \bar{\beta}(r) < \infty; \\
(a3) \ & \ \bar{\beta}(r) \leq -c_0 \rho \ \text{for all} \ r \in (r_1, \infty).
\end{align*}
\]

Assumption (A)(a3) is a dissipative condition on the drift term for large distances. Such conditions have been used to establish the exponential ergodicity of diffusions or SDEs with jumps; see [11, 12, 22, 34, 24, 23, 21].

Let

\[
\rho(x, y) = a 1_{\{d(x, y) > 0\}} + f_0(d(x, y)),
\]

where \(f_0(r) = 1 - e^{-cr} + ce^{-cr_1}r\), and the constants \(a, c > 0\) are chosen such that

\[
c \geq \sup_{r \in (r_0, r_1]} \frac{4\bar{\beta}(r)_+}{\alpha(r)} + 1, \quad a \geq 2c(1 + e^{-cr_1}) \sup_{r \in (0, r_0]} \frac{\bar{\beta}(r)_+}{\pi(r)} + 1
\]

with \(\bar{\beta}(r)_+ = \bar{\beta}(r) \vee 0\). By Assumption (A), the function \(\rho\) is well defined, and it is easy to check that

\[
f_0'(r) = ce^{-cr} + ce^{-cr_1} > 0, \quad f_0''(r) = -c^2e^{-cr} < 0, \quad f_0^{(3)}(r) = c^3e^{-cr} > 0.
\]

It also holds that there is a constant \(\bar{c} \geq 1\) such that for all \(x, y \in S\),

\[
\bar{c}^{-1} \left( 1_{\{d(x, y) > 0\}} + d(x, y) \right) \leq \rho(x, y) \leq \bar{c} \left( 1_{\{d(x, y) > 0\}} + d(x, y) \right).
\]

Hence the contractivity inequality [13] for such \(\rho\) implies upper bounds on the total variation distance (although we do not get a true contraction in the total variation, since the right hand side of (2.5) cannot be controlled from above by \(1_{\{d(x, y) > 0\}}\)).

**Theorem 2.1.** Suppose that Assumption (A) is satisfied. Let \(\rho(x, y)\) be defined by (2.3). Then for all \(x, y \in S\),

\[
E_{x,y}[\rho(X, Y)] \leq (1 - c_*) \rho(x, y),
\]

where

\[
c_* := \min \left\{ \frac{a \inf_{r \in (0, r_0]} \pi(r)}{2(a + 1 + cr_0e^{-cr_1})}, \frac{c_0}{2(a + 1 + cr_1e^{-cr_1})}, \frac{c_0}{1 + (1 + a)r_1^{-1}e^{-cr_1}} \right\}.
\]

**Remark 2.2.** In many applications, the constant \(r_1\) in Assumption (A)(a3) is uniquely determined as a parameter of the model under consideration. For instance, if we consider a Markov chain corresponding to a discretised SDE with the drift \(b(x) = -\nabla U(x)\) (cf. the discussion in the introduction), then \(r_1\) is the radius of the compact set, outside of which \(U\) is strongly convex, see Lemma 3.1 for more details on how Assumption (A) can be verified for such chains. However, for many models, condition (a1) in Assumption (A) can hold with any \(r_0 \in (0, r_1]\), as long as the transition probability of the Markov chain has an unbounded support. Hence, in such cases one can easily simplify the statement of Theorem 2.1 by taking \(r_0 = r_1\) and disregarding \(\alpha(r)\). Then the bounds for the constants \(c\) and \(a\) specified in (2.4) become simplified and we can take

\[
c := 1, \quad a \geq 2(1 + e^{-r_1}) \sup_{r \in (0, r_1]} \frac{\bar{\beta}(r)_+}{\pi(r)} + 1,
\]

whereas the contractivity inequality \(E_{x,y}[\rho(X, Y)] \leq (1 - c_*) \rho(x, y)\) in the statement of Theorem 2.1 holds with

\[
c_* = \min \left\{ \frac{a \inf_{r \in (0, r_1]} \pi(r)}{2(a + 1 + r_1e^{-r_1})}, \frac{c_0}{1 + (1 + a)r_1^{-1}e^{-r_1}} \right\}.
\]
We will use this simplified version of Theorem 2.1 in Section 3 in our analysis of contraction rates for Euler schemes with different types of noise.

In the following, we consider a variation of Theorem 2.1 in which we will replace (a2) and (a3) in Assumption (A) by the following two conditions respectively:

(a2*) For any \( r > r_0 \), \( \inf_{s \in (r_0, r]} \alpha(s) > 0 \) and \( \sup_{s \in (0, r]} \beta(s) < \infty \);

(a3*) There exist a measurable function \( V : S \to [0, \infty) \), and constants \( C_0 \in (0, \infty) \) and \( \lambda \in (0, 1) \) such that

1. for all \( x \in S \),
   \[
   \int_S V(z) p(x, dz) \leq (1 - \lambda)V(x) + C_0;
   \]

2. \( \lim_{r \to \infty} \inf_{d(x,y)=r} [V(x) + V(y)] = \infty \) and there is a constant \( c_0 \geq 0 \) such that \( \lim_{r \to \infty} \sup_{d(x,y)=r} e^{-r \beta(x,y)} \geq 0 \), where \( \beta(x,y) = \beta(x, y) \lor 0 \).

Assumption (a3*)(i) is a standard Lyapunov condition for the exponential ergodicity of the Markov chain with transition kernel \( p(x, dz) \); see [25]. Assumption (a3*)(ii) is a technical condition, which in particular is satisfied when the function \( r \mapsto \sup_{d(x,y)=r} V(x) + V(y) \) grows at most exponentially. Note that (a3*) implies substantially weaker than the analogous assumption [13] (A4)(b)], which requires the function \( r \mapsto \sup_{d(x,y)=r} \frac{\beta(x,y)}{V(x) + V(y)} \) to converge to zero when \( r \to \infty \).

For any fixed \( K > 0 \), we define
\[
(2.7) \quad r_1 = \sup \left\{ d(x, y) : x, y \in S \text{ with } \frac{\beta(x, y)e^{-c_0 d(x, y)}}{V(x) + V(y)} \geq K \text{ or } V(x) + V(y) \leq 4C_0 \lambda \right\}.
\]

In particular, by (ii) in condition (a3*), \( r_1 < \infty \). Under conditions (a1), (a2*) and (a3*), we are concerned with the following distance-like function
\[
(2.8) \quad \rho(x, y) = a \mathbb{I}_{\{d(x, y) > 0\}} + f_1(d(x, y)) + \epsilon (V(x) + V(y)) \mathbb{I}_{\{d(x, y) > 0\}},
\]
where \( f_1(r) = 1 - e^{-cr} \).

(2.9) \quad a = 2 \left( c \sup_{r \in (0, r_1]} \sup_{r \in (0, r_1]} \beta(r) \right) + 2eC_0 \left[ \inf_{r \in (0, r_1]} \frac{\pi(r)}{\pi(r)} \right]^{-1}, \quad \epsilon = \frac{1}{8C_0} e^{-c_0} \inf_{r \in (0, r_1]} \frac{\beta(r)}{\alpha(r)} \sup_{r \in (0, r_1]} \frac{\alpha(r)}{\alpha(r)}
\]

and
\[
(2.10) \quad c = 2 \sup_{r \in (0, r_1]} \frac{2\beta(r)}{\alpha(r)} + A + c_0 + 1, \quad A = \frac{16KC_0 e^{-c}}{\lambda \inf_{r \in (0, r_1]} \alpha(r)}.
\]

Consequently, there is a constant \( \bar{c} \geq 1 \) such that for all \( x, y \in S \),
\[
\bar{c}^{-1} \mathbb{I}_{\{d(x, y) > 0\}} (1 + V(x) + V(y)) \leq \rho(x, y) \leq \bar{c} \mathbb{I}_{\{d(x, y) > 0\}} (1 + V(x) + V(y)).
\]

Thus, using the distance-like function \( \rho(x, y) \) defined by (2.8), we can consider the convergence to equilibrium in terms of the weighted total variation metric; see [13] for related discussions on this topic.

**Theorem 2.3.** Suppose that (a1) in Assumption (A) and (a2*) as well as (a3*) hold. Let \( \rho(x, y) \) be defined by (2.8). Then for all \( x, y \in S \),
\[
\mathbb{E}_{x, y}[\rho(X, Y)] \leq (1 - c_*) \rho(x, y),
\]
where
\[
c_* := \min \left\{ \lambda, \frac{a \inf_{r \in (0, r_1]} \pi(r)}{2(a + 1)} , \frac{c_0 e^{-c} \inf_{r \in (0, r_1]} \alpha(r)}{4(a + 1)} , \frac{2C_0 \eta}{\lambda (a + 1)} , \frac{2}{\eta} \right\}
\]

with
\[
\eta := \frac{\lambda}{16C_0} e^{-c} \left( \inf_{r \in (0, r_1]} \alpha(r) \right) \left[ \sup_{r \in (0, r_1]} \frac{2\beta(r)}{\alpha(r)} + c_0 + 1 \right].
\]
Remark 2.4. Similarly to Remark 2.2 that follows Theorem 2.1, we would like to point out that in many applications \( r_0 \) in condition (a1) in Assumption (A) can be chosen arbitrarily. Hence, if we choose \( r_0 = r_1 \), where \( r_1 \) is given by (2.7), we can simplify the statement of Theorem 2.3. Then we have
\[
\epsilon = \frac{1}{8c_0} 2 e^{-cr_1}, \quad c = A + c_0 + 1, \quad A = \frac{16KC_0 e^{\epsilon r_1}}{\lambda},
\]
whereas \( a \) is given by the same formula as in (2.9), with \( r_0 = r_1 \). Hence the constant \( c_* \) in the statement of Theorem 2.3 becomes
\[
c_* := \min \left\{ \lambda, \frac{a \inf_{r \in (0, r_1]} \frac{\Psi(r)}{2(a+1)}}{\lambda (a+1)}, \frac{2C_0 \eta}{\lambda (a+1)}, \frac{\eta}{2c} \right\}
\]
with \( \eta := \frac{1}{16c_0} e^{-cr_1}(c_0 + 1) \).

2.2. Contraction in terms of metrics comparable to the Wasserstein distance. In this subsection, we will adopt a slightly different approach, that allows us to study contractions in terms of the Wasserstein-type distance, whose associated distance function \( \rho(x, y) \) satisfies \( \lim_{d(x, y) \to 0} \rho(x, y) = 0 \). In particular, this approach covers even couplings with zero probability of being successful. Indeed, when a Markov coupling \( ((X, Y), \mathbb{P}_{x,y}) \) starting from \( (x, y) \) can not succeed after one step, one may not expect the contraction of \( ((X, Y), \mathbb{P}_{x,y}) \) in terms of the total variation distance.

Note that the first part of this subsection is similar in spirit to [13 Subsection 2.2]. However, in the second part we will expand our approach to cover contractions in the \( L^p \)-Wasserstein distances for \( p > 1 \), which have not been considered in [13].

Recall that \( \alpha_l(x, y) \) and \( \alpha_l(r) \) are defined by (2.11) and (2.2), respectively. We suppose that the following assumption holds:

Assumption (B): There are constants \( l_0, r_1, c_0 > 0 \) and a nonnegative \( C([0, \infty)) \cap C^2([0, \infty)) \) function \( \Psi \), which satisfies that \( \Psi(0) = 0, \Psi' > 0, \Psi'' \leq 0 \) and \( \Psi'' \) is non-increasing on \( (0, \infty) \), such that
\[
\begin{align*}
(b1) \quad & \inf_{r \in (0, r_1]} \frac{\Psi(r)}{r} > 0, \\
(b2) \quad & \sup_{r \in (0, r_1]} \frac{2\beta(r)}{\Psi(r + l_0) \Psi_0(r)} + 1 \Psi(l_0) \leq \log 2; \\
(b3) \quad & \beta(r) \leq -c_0 r \text{ for all } r \in (r_1, \infty).
\end{align*}
\]

Note that we use \( \alpha_l(r) \) with \( l > 0 \) in Assumption (B) rather than \( \alpha_l(r) := \alpha_{l_0}(r) \). Indeed, by the definition of \( \alpha_l(r) \), we have \( \alpha_l(r) \leq r^2/2 \) and so (b1) is not satisfied. That is, using \( \alpha_l(r) \) with \( l > 0 \) is crucial in this subsection. Define
\[
f_3(r) := \begin{cases} 
\int_0^r e^{-c\Psi(s)} \, ds, & 0 \leq r \leq r_1 + l_0, \\
\int_0^r e^{-c\Psi(s)} \, ds + \frac{f_3(r_1 + l_0)}{2} \int_{r_1 + l_0}^r \left[ 1 + \exp \left( \frac{2\beta(r_1 + l_0)}{f_3(r_1 + l_0)} (s - (r_1 + l_0)) \right) \right] \, ds, & r > r_1 + l_0,
\end{cases}
\]
where
\[
c := \sup_{r \in (0, r_1]} \frac{2\beta(r_1 + l_0)}{\Psi(r + l_0) \alpha_0(r)} + 1.
\]
In particular, there is a constant \( \bar{c} \geq 1 \) such that for all \( r \geq 0 \),
\[
\bar{c}^{-1} r \leq f_3(r) \leq c r.
\]

Theorem 2.5. Suppose that Assumption (B) is satisfied. Let \( \rho(x, y) = f_3(d(x, y)) \), where \( f_3 \) is defined by (2.11). Then for all \( x, y \in S \),
\[
E_{x,y}[\rho(X, Y)] \leq (1 - c_*) \rho(x, y),
\]
where
\[
c_* := \min \left\{ \frac{\Psi'(r_1 + l_0) e^{-c\Psi(r_1 + l_0)} \inf_{r \in (0, r_1]} \frac{\alpha_l(r)}{r}}{2 c_0 e^{-c(r_1 + l_0)}}, \frac{1}{\max \left\{ 1, \frac{\int_0^{r_1 + l_0} e^{-c\Psi(s)} \, ds}{r_1 + l_0} \right\}^{-1}} \right\}.
\]
Remark 2.6. Take $\Psi(r) = r$, then (b2) in Assumption (B) becomes
\begin{align*}
(b2^*) \left[ \sup_{r \in (0, r_1]} \frac{2\overline{\beta}(r)+1}{\overline{\omega}_0(r)} \right] l_0 \leq \log 2.
\end{align*}
With this special choice, (b1) and (b2*) are equivalent to \cite{[13]} (B1) and (B2) in Subsection 2.2.
Moreover, in this case $f_3$ defined by (2.11) becomes
\begin{align*}
f_3(r) = \begin{cases} 
\frac{1}{e}(1-e^{-c_2(r_1+l_0)}) + \frac{1}{e}e^{-c_2(r_1+l_0)}(r-(r_1+l_0)), & 0 \leq r \leq r_1 + l_0, \\
\frac{1}{e}(1-e^{-c_2(r_1+l_0)}) + \frac{1}{e}e^{-c_2(r_1+l_0)}(r-(r_1+l_0)) + \frac{1}{4}e^{-c_2(r_1+l_0)}(1-e^{-2c_2(r-(r_1+l_0))}), & r > r_1 + l_0.
\end{cases}
\end{align*}

In the following, we will consider the contraction in terms of the $L^p$-Wasserstein distance with $p > 1$. For this, we assume that there is a constant $l > 0$ such that $r_1 \geq l + 1$ and the Markov coupling $((X, Y), \mathbb{P}_{x,y})$ satisfies $d(X, Y) \leq d(x, y) + l$ for all $x, y \in S$, where $r_1$ is the constant given in Assumption (B). An example of such coupling will be used in Section 3 see (3.10) and (3.11) therein. Let $\Psi$ be the function given in Assumption (B), and let $l_0$, $r_1$ and $c_0$ be the constants given therein. For any $p > 1$, we now define
\begin{align*}
f_4(r) = \begin{cases} 
f_0^r e^{-c\Psi(s)} ds, & 0 \leq r \leq r_1 + l_0, \\
f_4(r_1 + l_0) + A (r^p - (r_1 + l_0)^p) + B \left(1 - e^{-c(r-(r_1+l_0))}\right), & r > r_1 + l_0,
\end{cases}
\end{align*}
where
\begin{align*}
c := \sup_{r \in (0, r_1]} \frac{2\overline{\beta}(r)}{\overline{\Psi}'(r_1+l_0)\overline{\omega}_0(r)} + 1, \quad k \geq 1 + \max \{2l + 1/(c_0(r_1 + l_0)), 8p/(c_0(r_1 + l_0))\},
\end{align*}
and \begin{align*}
A := e^{-c\Psi(r_1+l_0)} \left[ c^{-1}p(1 + ((k+1)(r_1 + l_0))^{p-2}e^{ck(r_1+l_0)} + p(1 + l_0)^{p-1}\right]^{-1},
\end{align*}
and
\begin{align*}
B := e^{-2p-1} A ((k+1)(r_1 + l_0))^{p-2}e^{ck(r_1+l_0)}.
\end{align*}
We can check that $f_4 \in C([0, \infty)) \cap C^1((0, \infty))$ is such that $f_4' > 0$ on $(0, \infty)$, $f_4''$ is decreasing on $(0, (k+1)(r_1 + l_0))$, $f_4'' > 0$ on $(0, r_1 + l_0)$, $f_4''(1 + ((k+1)(r_1 + l_0)) = 0$ and $f_4'' > 0$ on $((k+1)(r_1 + l_0), \infty)$. Note that $f_4''(r)$ exists for all $r \neq r_1 + l_0$. Moreover, there is a constant $\tilde{c} \geq 1$ such that for all $r > 0$,
\begin{align*}
\tilde{c}^{-1}(r \vee r^p) \leq f_4(r) \leq \widetilde{c}(r \vee r^p).
\end{align*}
In particular, for any $q \in [1, p]$, there is a constant $c_q > 0$ such that for all $r > 0$, $c_q r^q \leq f_4(r)$. Thus, with the choice of $f_4(r)$ defined by (2.13), one can obtain a contraction of the Markov coupling $((X, Y), \mathbb{P}_{x,y})$ in a distance based on the metric $\rho(x, y) = f_4(d(x, y))$, which controls the $L^q$-Wasserstein distance with $q \in [1, p]$ from above. More specifically, the contraction of the Markov coupling $((X, Y), \mathbb{P}_{x,y})$ in the Wasserstein distance with $f_4$ can yield convergence rates of the corresponding Markov chain in terms of the $L^q$-Wasserstein distance, see [25, Corollary 2.4].

Theorem 2.7. Suppose that Assumption (B) is satisfied, and there is a constant $l > 0$ such that $r_1 \geq l + 1$ and the Markov coupling $((X, Y), \mathbb{P}_{x,y})$ satisfies $d(X, Y) \leq d(x, y) + l$ for all $x, y \in S$. Let $\rho(x, y) = f_4(d(x, y))$, where $f_4$ is defined by (2.13). Then for all $x, y \in S$,
\begin{align*}
E_{x,y}[\rho(X, Y)] \leq (1 - c_*) \rho(x, y),
\end{align*}
where
\begin{align*}
c_* := \min \left\{ \Psi'(r_1 + l_0)e^{-c\Psi(r_1+l_0)} \inf_{r \in (0, r_1]} \frac{\alpha_0(r)}{r}, \frac{c_0 p A}{1 + A + B(r_1 + l_0)^{-p}}, \frac{c_0 p A}{24^{p-1} A + B((k+1)(r_1 + l_0))^{-(p-1)}} \right\},
\end{align*}
and $\alpha_0(r)$ is defined in (2.11).
3. Applications: Euler Schemes of SDEs

There are numerous works devoted to Euler discretizations of the following stochastic differential equation (SDE)

\[ dX_t = b(X_t) \, dt + dZ_t, \]

where \( b : \mathbb{R}^d \to \mathbb{R}^d \) is Lipschitz continuous, and \( (Z_t)_{t \geq 0} \) is a stochastic process on \( \mathbb{R}^d \). In particular, when \( (Z_t)_{t \geq 0} \) is a rotationally invariant symmetric \( \alpha \)-stable process with \( \alpha \in (0, 2) \) (when \( \alpha = 2 \), \( (Z_t)_{t \geq 0} \) is a \( d \)-dimensional standard Brownian motion), the transitions of the Markov chain for the Euler schemes corresponding to the SDE (3.1) with step size \( h > 0 \) are given by

\[ x \mapsto x + h b(x) + h^{1/\alpha} \xi, \quad x \in \mathbb{R}^d, \]

where \( \xi \) is a random variable with symmetric \( \alpha \)-stable distribution (when \( \alpha = 2 \), \( \xi \) is a random variable with the standard \( d \)-dimensional normal distribution). The purpose of this section is to apply the results in the previous section to study the strict contraction of Markov chains including the system (3.2).

Let \( h > 0 \), \( g \) be a continuous and strictly increasing function on \([0, \infty)\) with \( g(0) = 0\), and \( \xi \) be a random variable whose distribution is given by \( \mu \). We will consider the Markov chain \( X \) on \( \mathbb{R}^d \) whose transitions are of the following form:

\[ x \mapsto x + h b(x) + g(h) \xi, \quad x \in \mathbb{R}^d. \]

Throughout this section, we always assume that the coefficient \( b(x) \) satisfies the following assumption.

**Assumption (C):**

1. There is a constant \( L \geq 0 \) such that \( |b(x) - b(y)| \leq L|x - y| \) for \( x, y \in \mathbb{R}^d \);
2. There are constants \( K > 0 \) and \( \mathfrak{R} \geq 0 \) such that \( \langle x - y, b(x) - b(y) \rangle \leq -K|x - y|^2 \) for \( x, y \in \mathbb{R}^d \) with \( |x - y| \geq \mathfrak{R} \).

Note that it must hold that \( K \leq L \). We will construct three explicit Markov couplings of the chain \( X \) according to different conditions on the distribution \( \mu \) of the random variable \( \xi \).

3.1. General case. For any \( x \in \mathbb{R}^d \) and \( \kappa > 0 \), set

\[ (x)_\kappa := \left( 1 - \frac{\kappa}{|x|} \right) x, \quad \mu_x(dz) := (\mu \wedge (\delta_x \ast \mu))(dz). \]

In this part, we suppose that the following condition holds for the measure \( \mu \).

1. There is a constant \( \kappa_0 > 0 \) such that

\[ J_{\kappa_0} := \inf_{|x| \leq \kappa_0} (\mu \wedge (\delta_x \ast \mu))(\mathbb{R}^d) > 0. \]

Note that \( (x)_\kappa \) is just a truncation of the vector \( x \) and \( \mu_x \) is the common part of the measure \( \mu \) and its translation by shifting the support by \( x \). Condition (c3) indicates that there is sufficient mass in the overlap of the measure \( \mu \) and its (small) translation. It can be interpreted as a very weak non-degeneracy condition on \( \mu \) (see also Example 3.6).

Write \( \hat{x} = \hat{x}_h := x + h b(x) \). We will adopt the following notation:

\[ X = \begin{cases} f_1(z), & \mu_1(dz); \\ f_2(z), & \mu(dz) - \mu_1(dz) \end{cases} \]

for some functions \( f_1, f_2 : \mathbb{R}^d \to \mathbb{R}^d \) and a sub-probability measure \( \mu_1 \ll \mu \) means that the value of \( z \) is drawn according to \( \mu \) and then, conditional on the value of \( z \), the random variable \( X \) takes value \( f_1(z) \) with probability \( (d\mu_1/d\mu)(z) \) and value \( f_2(z) \) with probability \( 1 - (d\mu_1/d\mu)(z) \). Then
we define the following Markov coupling \((X, Y), \mathbb{P}_{x,y}\) of the chain \(X\):

\[
\begin{align*}
X &= \hat{x} + g(h)z, & \mu(dz); \\
Y &= \begin{cases} 
\hat{y} + g(h)(z + g(h)^{-1}(\hat{x} - \hat{y}), & \frac{1}{2}\mu_{g(h)^{-1}(\hat{\hat{y}} - \hat{\hat{x}})}(dz), \\
\hat{y} + g(h)z, & \frac{1}{2}\mu_{g(h)}(\hat{\hat{x}} - \hat{\hat{y}})^{(1)}(dz), \end{cases}
\end{align*}
\]

Note that when \(x = y\), \(\hat{x} = \hat{y}\) and so (3.6) naturally degenerates into the synchronous coupling, i.e.,

\[
\begin{align*}
X &= \hat{x} + g(h)z, & \mu(dz); \\
Y &= \hat{y} + g(h)z, & \mu(dz).
\end{align*}
\]

To check that \((X, Y), \mathbb{P}_{x,y}\) is a Markov coupling, we apply the fact that \((\delta_x \ast \mu_x)(dz) = \mu_x(dz)\) for all \(x \in \mathbb{R}^d\) (see Lemma 5.1 in the appendix of this paper). Let us discuss the intuition behind the construction of (3.6). A natural idea for constructing couplings that are useful for studying convergence of Markov chains is the maximal coupling, where the two marginals are moved to the same point with the maximal possible probability (which would be \(\mu_{g(h)}(\hat{\hat{x}} - \hat{\hat{y}})^{(1)}(\mathbb{R}^d)\) in the setting of (3.6)). However, an important question then is what can be done with the remaining probability mass. If the original distribution \(\mu\) of the noise is rotationally symmetric, a natural choice is the reflection coupling, and that idea was used in [13], which dealt with the Gaussian noise, see also Subsection 3.2 below. However, for more general distributions \(\mu\) that do not exhibit any specific geometric properties such as symmetry, the choice is much less clear. One can always couple the remaining mass independently, which is the idea in the usual basic coupling (i.e., a combination of the maximal coupling and the independent coupling, see [12] Example 2.10), although this choice does not seem to produce sharp upper bounds on the corresponding Wasserstein-type distances between time-marginal distributions of the coupled processes. From the technical point of view, in calculations one would want the remaining mass to be coupled synchronously, however, a combination of the maximal coupling and the synchronous coupling does not produce a coupling (since the marginals end up having different distributions). A solution to this problem has been found in [22], which introduced the refined basic coupling that moves the marginals to the same point only with half the maximal possible probability. As it turns out, if we then couple the other half of that probability in the way as it is done in the second line for \(Y\) in (3.6), we then can couple the remaining mass synchronously. This produces a coupling that works very well in computations related to bounding Wasserstein-type distances between time-marginals of stochastic processes with non-symmetric noise. It was first introduced in the context of Lévy-driven SDEs in [22] and in the present paper we apply it for the first time to study processes in discrete time.

Note that the parameter \(\kappa\) in (3.4) is introduced so that the coupling (3.6) has a positive probability of being successful even when the jump distribution \(\mu\) has finite support. If we did not introduce \(\kappa\) in (3.6), then for \(\mu\) with finite support it could happen that if the coupled processes started too far apart, the probability of jumping to the same point would be zero, all the transitions would be just coupled synchronously and the coupling could never be successful. This explains the reason why we use \(\mu_{g(h)}(\hat{\hat{x}} - \hat{\hat{y}})^{(1)}(\mathbb{R}^d)\) instead of \(\mu_{g(h)}(\hat{\hat{x}} - \hat{\hat{y}})^{(1)}(\mathbb{R}^d)\) in the setting of (3.6). Due to this modification, even if the marginals cannot jump to the same point, they can at least get closer to each other by \(\kappa\). However, for distributions \(\mu\) with full support, the value of \(\kappa\) can be chosen arbitrarily large, in particular we can just take \((x)_\kappa = x\) (that formally corresponds to taking \(\kappa = \infty\)).

Recall that in Section 2, Assumptions (A) and (B) were formulated in terms of quantities \(\beta, \alpha\) and \(\pi\), cf. (2.1) and (2.2), which in the case of the metric space \((S,d)\) being \(\mathbb{R}^d\) with the Euclidean metric, are defined by

\[
\beta(x, y) = E_{x,y}[R - r], \quad \alpha(x, y) = \frac{1}{2}E[(R - r)^21_{\{R < r\}}], \quad \pi(x, y) = \mathbb{P}_{x,y}(R = 0),
\]
where \( r = |x - y| \) and \( R = |X - Y| \). We further set \( \hat{r} = |\hat{x} - \hat{y}| \). We will now present a result that explains how Assumptions (A) and (B) from Section 2 can be verified for Markov chains of the form (3.3) and the Markov coupling \(((X, Y), P_{x,y})\) defined by (3.6) under Assumption (C).

**Lemma 3.1.** Under Assumption (C) and condition (c3), for the Markov coupling \(((X, Y), P_{x,y})\) defined by (3.6) and for \( h > 0 \) we have

(i) \( \beta(x, y) \leq hL|x - y| \) for any \( x, y \in \mathbb{R}^d \);

(ii) \( \beta(x, y) \leq -(K - hL^2/2)h|x - y| \) for any \( x, y \in \mathbb{R}^d \) with \( |x - y| \geq \mathcal{R} \) and any \( h \leq 2KL^{-2} \);

(iii) \( \pi(x, y) \geq \frac{1}{2} J_{g(h)^{-1}\kappa} I_{\{\hat{r} \leq \kappa\}} \) for any \( x, y \in \mathbb{R}^d \) and \( 0 < \kappa \leq g(h)\kappa_0 \). In particular, for any \( x, y \in \mathbb{R}^d \) with \( |x - y| \leq \mathcal{R} \) and \( 0 < \kappa \leq g(h)\kappa_0 \),

\[
\pi(x, y) \geq \frac{1}{2} J_{g(h)^{-1}\kappa} > 0;
\]

(iv) \( \alpha(x, y) \geq \frac{1}{2} J_{g(h)^{-1}\kappa} (\hat{r} - \hat{r} \wedge \kappa) - r)^{2} I_{\{\hat{r} - \hat{r} \wedge \kappa < r\}} \) for any \( x, y \in \mathbb{R}^d \) and \( 0 < \kappa \leq g(h)\kappa_0 \). In particular, for any \( x, y \in \mathbb{R}^d \) with \( |x - y| \leq \mathcal{R} \), \( 0 < \kappa \leq g(h)\kappa_0 \) and any \( h \in (0, L^{-1} \wedge (\kappa(2L\mathcal{R})^{-1})) \),

\[
\alpha(x, y) \geq \left( \frac{\kappa^2}{16} \right)^{2} J_{g(h)^{-1}\kappa} > 0.
\]

**Remark 3.2.** Lemma 3.1 and its proof in Section 4 show that, for the Markov coupling \(((X, Y), P_{x,y})\) given by (3.6), \( \beta(x, y) = \mathbb{E}_{x,y}[R - \hat{r}] = 0 \) for any \( x, y \in \mathbb{R}^d \), and so \( \beta(x, y) = \hat{r} - r \) is only determined by the coefficient \( b(x) \); on the other hand, the lower bound on \( \alpha(x, y) \) depends only on the stochastic noise \( \xi \) involved in transitions.

Now, according to Theorem 2.1 and Lemma 3.1, we have the following statement in this part.

**Theorem 3.3.** Suppose that Assumption (C) and condition (c3) hold, and that \( \lim_{h \to 0^+} h/g(h) = 0 \). Consider any \( h \in (0, 2KL^{-2} \wedge L^{-1}) \) such that \( h/g(h) \leq \kappa_0(2L\mathcal{R})^{-1} \). Let \( \rho = f(\{ \cdot \}) \) be defined by (2.3) with \( r_1 = \mathcal{R} \),

\[
c = \frac{64hL\mathcal{R}}{g(h)^2\kappa_0^3J_{\kappa_0}} + 1, \quad a = \frac{4c(1 + e^{-c\xi})h(g(h)\kappa_0L)}{J_{\kappa_0}} + 1.
\]

Then, for all \( x, y \in \mathbb{R}^d \),

\[
\mathbb{E}_{x,y}[\rho(X, Y)] \leq (1 - c_*) \rho(x, y),
\]

where

\[
c_* := \min \left\{ \frac{aJ_{\kappa_0}}{4(a + 1 + cr_0 e^{-c\xi})}, \frac{c^2 e^{-c\xi}g(h)^2\kappa_0^2 J_{\kappa_0}}{32(a + 1 + c\mathcal{R}e^{-c\xi})}, \frac{(K - hL^2/2)h}{(1 + a)(1 + a)\mathcal{R}^{-1}e^{-c\xi}} \right\}
\]

with \( r_0 := g(h)\kappa_0/(1 + hL) \).

**Proof.** Let \( \kappa = g(h)\kappa_0 \), \( r_0 = g(h)\kappa_0/(1 + hL) \) and \( r_1 = \mathcal{R} \). Then it follows from Lemma 3.1 that for any \( h \in (0, 2KL^{-2} \wedge L^{-1}) \),

(i) \( \inf_{r \in (0, r_0]} \overline{\beta}(r) \geq \frac{1}{2} J_{\kappa_0} > 0 \);

(ii) \( \inf_{r \in (r_0, r_1]} \alpha(r) \geq \left( \frac{\kappa^2}{16} \right)^{2} J_{g(h)^{-1}\kappa_0} \) for any \( r \in (r_1, \infty) \),

(iii) \( \overline{\beta}(r) \leq -(K - hL^2/2)hr \) for any \( r \in (r_1, \infty) \).

In particular,

\[
\sup_{r \in (r_0, r_1]} \frac{4\overline{\beta}(r)}{\alpha(r)} \leq \frac{64hL\mathcal{R}}{g(h)^2\kappa_0^3J_{\kappa_0}} \quad \text{and} \quad \sup_{r \in (0, r_0]} \frac{\overline{\beta}(r)}{\pi(r)} \leq \frac{hLr_0}{J_{\kappa_0}/2} \leq \frac{2h(g(h)\kappa_0L)}{J_{\kappa_0}}.
\]

With those two estimates at hand, the required assertion follows from Theorem 2.1. \( \square \)

**Remark 3.4.** Note that \( c \geq \frac{64hL\mathcal{R}}{g(h)^2\kappa_0^3J_{\kappa_0}} \) and \( a \geq \frac{4c(1 + e^{-c\xi})g(h)\kappa_0L}{J_{\kappa_0}} \) and hence, in order to obtain more explicit dependence on \( h \), we can also write

\[
\mathbb{E}_{x,y}[\rho(X, Y)] \leq (1 - c_*) \rho(x, y),
\]
where
\[
c_{ss} := \min \left\{ \frac{c(1 + e^{-cR})g(h)\kappa_0 L}{a + 1 + cR e^{-cR}} \frac{2e^{-cR}L^2 R^2 h}{(a + 1 + cR e^{-cR})g(h)^2 \kappa_0^2 J_0'(1 + (1 + a)R^{-1} e^{-cR})} \right\}.
\]

**Remark 3.5.** In many applications, the support of the distribution \( \mu \) of the noise random variable \( \xi \) in the chain given by (3.3) is unbounded. Hence we can choose \( (x)_\xi = x \) for all \( x \in \mathbb{R}^d \) in (3.6), cf. the discussion below (1.4). Moreover, condition (c3) is satisfied for any \( \kappa_0 > 0 \). Then points (iii) and (iv) in the statement of Lemma 3.1 can be modified to state that for any \( h > 0 \) and any \( x, y, z \in \mathbb{R}^d \) we have
\[
\pi(x, y) \geq \frac{1}{2} \mu_{g(h^{-1}z)}(\mathbb{R}^d) \quad \text{and} \quad \alpha(x, y) \geq \frac{1}{4} r_0^2 \mu_{g(h^{-1}z)}(\mathbb{R}^d).
\]

Then, in the proof of Theorem 3.3 we can use the fact that \( |\hat{y} - \hat{x}| \leq (1 + hL)|x - y| \leq 2|x - y| \) for any \( x, y \in \mathbb{R}^d \) with \( |x - y| \leq \mathcal{R} \) and \( h < L^{-1} \), and get that
\[
\sup_{r \in [r_0, r_1]} \frac{4\beta(r)_+}{A(r)} \leq \frac{16hL\mathcal{R}}{r_0^2 J_{2g(h)^{-1}r_1}} \quad \text{and} \quad \sup_{r \in (0, r_0]} \frac{\beta(r)_+}{\alpha(r)} \leq \frac{2hL r_0}{J_{2g(h)^{-1}r_0}},
\]
which in turn yields that in Theorem 3.3 one can take
\[
c = \frac{16hL\mathcal{R}}{r_0^2 J_{2g(h)^{-1}r_1}} + 1, \quad a = \frac{4c(1 + e^{-cR})hL r_0}{J_{2g(h)^{-1}r_0}} + 1.
\]

Indeed, in this special case we may further apply the simplified version of Theorem 2.1 as explained in Remark 2.2. In particular, we have \( r_0 = r_1 = \mathcal{R} \), and we obtain \( c = 1 \) and
\[
a = \frac{4(1 + e^{-cR})hL \mathcal{R}}{J_{2g(h)^{-1}r_0}} + 1 \geq \frac{4hL \mathcal{R}}{J_{2g(h)^{-1}r_0}}.
\]

Then, based on (2.6), we obtain the contractivity constant
\[
c_\ast = \min \left\{ \frac{hL \mathcal{R}}{a + 1 + r_1 e^{-cR_1}} : 1 + (1 + a) r_1^{-1} e^{cR_1} \right\}.
\]

Hence if we want to track the dependence of \( c_\ast \) on the parameters such as \( h \) and \( \mathcal{R} \), then we have to analyse the quantity \( J_{2g(h)^{-1}r_0} \), which depends on the noise distribution \( \mu \). This will be explained on the examples of \( \alpha \)-stable and Gaussian noises in Remark 3.11.

Discretisations of SDEs driven by general non-isotropic Lévy processes have been extensively studied in the literature, see e.g. [6, 20] and the references therein. As demonstrated by the following example, our results apply to very general non-isotropic distributions \( \mu \).

**Example 3.6.** Suppose that the distribution \( \mu \) of the random variable \( \xi \) is given by
\[
\mu(dz) = M^{-1} \mathbf{1}_{\{0 < z_1 \leq 1\}} \frac{1}{(1 + |z|)^{d+\alpha}} dz,
\]
where \( z := (z_1, \ldots, z_d) \in \mathbb{R}^d \), \( \alpha \in (0, 2) \) and \( M := \int_{\{0 < z_1 \leq 1\}} \frac{1}{(1 + |z|)^{d+\alpha}} dz \). Then, we can get from the proof of [23, Example 1.2] that for any \( \kappa \in (0, 1) \),
\[
J_\kappa := \inf_{|z| \leq \kappa} (\mu \wedge (\delta_z \ast \mu))(\mathbb{R}^d) \geq c((1 + \kappa)^{-\alpha} - 2^{-\alpha})
\]
with some constant \( c := c(\alpha, M) > 0 \). In particular, taking \( \kappa_0 = 1/2 \) in (3.5), we obtain from (i) and (ii) in the proof of Theorem 3.3 that
\[
\inf_{r \in \{0, r_0\}} \frac{\alpha(r)}{\alpha} \geq J_{\kappa_0}/2 \geq \frac{c}{2}((3/2)^{-\alpha} - 2^{-\alpha}) \quad \text{and} \quad \inf_{r \in \{0, r_1\}} \frac{\alpha(r)}{\alpha} \geq \frac{cg(h)^2}{64}((3/2)^{-\alpha} - 2^{-\alpha}).
\]

Hence, under Assumption (C) we can apply Theorem 3.3 to this example.

**Remark 3.7.** Let us briefly discuss an extension of Theorem 3.3 in which condition (c2) in Assumption (C) is replaced by the following weaker condition.

(c2') There are constants \( M_1, M_2 > 0 \) such that \( \langle x, b(x) \rangle \leq M_1 - M_2|x|^2 \) for all \( x \in \mathbb{R}^d \).
Suppose that random variable $\xi$ has finite $\theta$-th moment for some $\theta \in (0, 2]$, i.e., $\mathbb{E}[|\xi|^\theta] < \infty$. One can check that under conditions (c1) and (c2*), condition (a3*) is satisfied for $V(x) = |x|^{\theta}$. Indeed, letting $h < 2M_2/L_0$ and $L_0 := 2 \max\{L^2, |b(0)|^2\}$, we have

$$\int_{\mathbb{R}^d} V(z)p(x, dz) = \mathbb{E}[|x + hb(x) + g(h)|^{\theta}] = \mathbb{E}[|x + hb(x) + g(h)|^{\theta/2}]$$

$$\leq (1 - 2hM_2 + h^2L_0)^{\theta/2}|x|^{\theta} + (2hM_1)^{\theta/2} + g(h)^{\theta} \mathbb{E}[|\xi|^{\theta}] + (h^2L_0)^{\theta/2}$$

$$+ \left[(2g(h)^{\theta/2} + (2hM_1)^{\theta/2}L_0^{1/2}) |x|^{\theta/2} + (2hM_1)^{\theta/2}b(0)|^{\theta/2}\right] \mathbb{E}[|\xi|^{\theta/2}],$$

where in the inequality above we factored that

$$|x + hb(x) + g(h)|^2 = |x|^2 + 2h(x, b(x)) + h^2|b(x)|^2 + (g(h))^2|x|^2 + 2g(h)(x, \xi) + 2h(g)(b(x), \xi)$$

$$\leq |x|^2 + 2h(M_1 - M_2|x|^2) + h^2L_0|x|^2 + (g(h))^2|x|^2 + h^2L_0$$

$$+ 2g(h)(x, \xi) + 2h(g)(b(x), \xi)$$

$$= (1 - 2hM_2 + h^2L_0)|x|^2 + 2hM_1 + h^2L_0 + (g(h))^2|x|^2$$

$$+ 2g(h)(x, \xi) + 2h(g)(b(x), \xi).$$

This implies that (a3*) holds for some suitable constants $\lambda$ and $C_0$. In particular, if $\theta = 2$ and $\mathbb{E}[\xi] = 0$, one can easily check (cf. [13, Example 6.2]) that (a3*) holds with

$$\lambda = 2hM_2 - h^2L_0, \quad C_0 = h^2L_0 + 2hM_1 + (g(h))^2 \mathbb{E}[|\xi|^2].$$

Moreover, under conditions (c1), (c2*) and (c3), points (i), (iii) and (iv) in Lemma 3.1 remain unchanged. Hence under (c1), (c2*) and (c3), we can apply Theorem 2.3 to prove an analogue of Theorem 3.3 under weaker conditions. We leave the details to the reader.

### 3.2. Special case: $\mu$ is rotationally invariant.

In this part, we are concerned with the case where the distribution $\mu$ of the random variable $X$ has a density function $m(x)$ with respect to the Lebesgue measure such that $m(x) = m(|x|)$ for all $x \in \mathbb{R}^d$. For this special case, we use the following Markov coupling $((X, Y^{(1)}), P_{x,y})$ of the chain $X$:

$$(3.7) \begin{align*}
X &= \hat{x} + g(h)z, \\
Y^{(1)} &= \begin{cases}
\hat{y} + g(h)(z + g(h)^{-1}(\hat{x} - \hat{y}))^\kappa, & m(z) dz, \\
\hat{y} + g(h)R_{x,y}(z), & m(z) - m(z) \wedge m(z - g(h)^{-1}(\hat{x} - \hat{y})) dz.
\end{cases}
\end{align*}$$

Here, $\kappa > 0$ is a constant fixed later, $(\kappa)_{x}$ is defined in (3.4), and for any $x, y \in \mathbb{R}^d$, $R_{x,y}$ is a reflection matrix defined by

$$R_{x,y} z := R_{x,y}(z) := \begin{cases}
z - \frac{2(x-y,z)(x-y)}{|x-y|^2}(x-y), & x \neq y, \\
z, & x = y.
\end{cases}$$

Note that in this setting, $m(z) \wedge m(z - g(h)^{-1}(\hat{x} - \hat{y})) dz = \mu_{g(h)^{-1}(\hat{y} - \hat{x})}(dz)$, where $\mu_{x}$ is defined in (3.4). See Lemma 5.2 in the appendix for the proof that $((X, Y^{(1)}), P_{x,y})$ is a Markov coupling of $X$. Note also that this coupling is a generalisation of the coupling by reflection, which was used for the Gaussian noise in [13, Section 2.4], to arbitrary rotationally invariant distributions (possibly with compact support).

We first make the following assumption on the density function $m(x)$:

(c4) The density function $m(x) = m(|x|)$ of the distribution $\mu$ satisfies that $m(r)$ is non-increasing in $(0, \infty)$, and has finite first moment, i.e., $\int_{\mathbb{R}^d} |z|m(|z|) dz < \infty$.

Note that under condition (c4), there is a constant $\kappa_0 > 0$ such that $J_{\kappa_0} > 0$, where $J_{\kappa_0}$ is defined by (3.3). Indeed, without loss of generality, we still assume that $d = 1$. Then, for any $z \in \mathbb{R}$,

$$\mu \wedge (\delta_z \ast \mu)(R) = \int_R m(u) \wedge m(u - z) du = 2 \int_{|z|/2}^\infty m(u) du.$$

This yields the desired assertion.
Similarly as in Lemma 3.1 in Section 3.1, we will now show that in the setting of the present Section 3.2, under Assumption (C) we can verify the conditions from Assumptions (A) and (B) in Section 2.

Lemma 3.8. Suppose that Assumption (C) and (c4) hold. Then, for the Markov coupling given by (3.7) and \( h > 0 \), we have

(i) \( \beta(x,y) \leq hL|x-y| \) for any \( x,y \in \mathbb{R}^d \);
(ii) \( \beta(x,y) \leq -(K - hL^2/2)h|x-y| \) for any \( x,y \in \mathbb{R}^d \) with \( |x-y| \geq \mathcal{R} \) and any \( h < 2KL^{-2} \);
(iii) \( \pi(x,y) \geq J_{g(h)^{-1}c}^c \mathbb{1}_{\{x \leq y\}} \) for any \( x,y \in \mathbb{R}^d \). In particular, for any \( x,y \in \mathbb{R}^d \) with \( |x-y| \leq \kappa/(1 + hL) \),

\[
\pi(x,y) \geq J_{g(h)^{-1}c}^c
\]

(iv) \( \alpha(x,y) \geq \frac{1}{2}J_{g(h)^{-1}c}(\hat{r} - \hat{r}^2 \wedge \kappa) - r^2 \mathbb{1}_{\{r > \kappa\}} \) for any \( x,y \in \mathbb{R}^d \). In particular, for any \( x,y \in \mathbb{R}^d \) with \( |x-y| \leq \mathcal{R} \) and any \( h \in (0, L^{-1} \wedge (\kappa(2L\mathcal{R})^{-1})) \),

\[
\alpha(x,y) \geq \frac{1}{2}J_{g(h)^{-1}c}(\hat{r} - \hat{r}^2 \wedge \kappa^2)
\]

Combining this with Lemma 3.8 and following the proof of Theorem 3.3, we have the statement below.

Theorem 3.9. Suppose Assumption (C) and condition (c4) hold, and that \( \lim_{h \to 0^+} h/g(h) = 0 \). Then there exists a constant \( \kappa_0 > 0 \) such that for any \( h \in (0, 2KL^{-2} \wedge L^{-1}) \) with \( h/g(h) \leq \kappa_0(2L\mathcal{R})^{-1} \) and for all \( x,y \in \mathbb{R}^d \),

\[
\mathbb{E}_{x,y} [\rho(X, Y^{(1)})] \leq (1 - c_s)\rho(x,y),
\]

where \( \rho = f(\cdot | \cdot) \) is defined by (2.3) with \( r_1 = \mathcal{R} \),

\[
c = \frac{32hL\mathcal{R}}{g(h)^2 \kappa_0^2 J_{\kappa_0}} + 1, \quad a = \frac{2c(1 + e^{-cg})hL\kappa_0 L}{J_{\kappa_0}} + 1
\]

and the constant \( c_s \in (0, 1) \) given in Theorem 3.3 with the constants \( r_1, c, a \) above and \( r_0 = g(h)\kappa_0/(1 + hL) \).

Remark 3.10. As explained in Remark 3.5 for Theorem 3.3, when the support of the distribution \( \mu \) of the random variable \( \xi \) is unbounded, we can modify our bounds for \( \pi(x,y) \) and \( \alpha(x,y) \). Points (iii) and (iv) in Lemma 3.8 become

\[
\pi(x,y) \geq J_{g(h)^{-1}(\hat{y} - \hat{z})}(\mathbb{R}^d) \quad \text{and} \quad \alpha(x,y) \geq \frac{1}{2}J_{g(h)^{-1}(\hat{y} - \hat{z})}(\mathbb{R}^d)
\]

for any \( h > 0 \) and any \( x, y \in \mathbb{R}^d \), i.e., they only differ from the bounds considered in Remark 3.5 by the absence of factor \( \frac{1}{2} \). As a consequence, one can take

\[
c = \frac{8hL\mathcal{R}}{r_0^2 J_{2g(h)^{-1}r_1}} + 1, \quad a = \frac{2c(1 + e^{-cg})hLr_0}{J_{2g(h)^{-1}r_0}} + 1
\]

in Theorem 2.1. Furthermore, by applying its simplified version as explained in Remark 2.2 with \( r_0 = r_1 = \mathcal{R} \), we have \( c = 1 \) and

\[
a = \frac{2(1 + e^{-c})(1 + e^{-cg})hL\mathcal{R}}{J_{2g(h)^{-1}\mathcal{R}}} + 1 \geq \frac{2hL\mathcal{R}}{J_{2g(h)^{-1}\mathcal{R}}}
\]

In particular, we observe that for isotropic noise distributions \( \mu \) (for which both couplings discussed above are applicable), the result based on the reflection coupling (3.7) can lead to a slightly better contractivity constant than the result based on the refined basic coupling (3.6).

Remark 3.11. Let us now consider two different noise distributions \( \mu \), namely, the normal distribution and the rotationally invariant \( \alpha \)-stable distribution, with the aim of tracking the dependence of the quantity \( J_{2g(h)^{-1}\mathcal{R}} \) (and hence of the contractivity constant for the corresponding chains) on parameters \( h \) and \( \mathcal{R} \). To this end, we will use the bounds discussed above in Remark 3.10 and combine them with the simplified version of Theorem 2.1 as discussed in Remark 2.2.
is well-known that for a one-dimensional random variable \( Z \) with the standard normal distribution one has the tail estimate
\[
P(Z > x) \approx \exp(-x^2/2)
\]
for any \( x > 0 \), whereas if \( Z \) has the rotationally invariant \( \alpha \)-stable distribution, one has
\[
P(Z > x) \approx (1 + x)^{-\alpha}
\]
for all \( x > 0 \). Note that, for a rotationally invariant distribution \( \mu \) with a density \( m(z) = m(|z|) \) for \( z \in \mathbb{R}^d \), we have for any \( x \in \mathbb{R}^d \),
\[
(\mu \wedge (\delta_x \ast \mu))(\mathbb{R}^d) = \int_{\mathbb{R}^d} m(z) \wedge m(z + |x|e_1)dz
= \int_{\mathbb{R}^d} m((|z_1|^2 + |\tilde{z}|^2)^{1/2}) \wedge m((|z_1 + |x||^2 + |\tilde{z}|^2)^{1/2}) dz_1 \tilde{d}
= \int_{|z_1| \geq |x|/2} \left( \int_{\mathbb{R}^d-1} m((|z_1|^2 + |\tilde{z}|^2)^{1/2}) \tilde{d} \right) dz_1,
\]
where \( z = (z_1, \tilde{z}) \) and \( \tilde{z} := (z_2, \cdots, z_d) \). Furthermore, observe that
\[
d \int_{|z_1| \geq |x|/2} \left( \int_{\mathbb{R}^d-1} m(z) \tilde{d} \right) dz_1 \geq \int_{|z_1| \geq |x|/2 \text{ or } |z_2| \geq |x|/2 \text{ or } \cdots \text{ or } |z_d| \geq |x|/2} m(z) dz
\geq \int_{|z| \geq \sqrt{d}|x|/2} m(z) \tilde{d}.
\]
This shows that \( (\mu \wedge (\delta_x \ast \mu))(\mathbb{R}^d) \geq \frac{1}{d} \int_{|z| \geq \sqrt{d}|x|/2} m(z) \tilde{d} \) and hence, by applying the tail estimates above,
\[
J_{2g(h)^{-1}R} = \begin{cases} 
\Omega(d^{-1} \exp(-4dR^2/h)) & \text{when } \mu \text{ is Gaussian,} \\
\Omega(d^{-1} (1 + d^{1/2}h^{-1/\alpha}R)^{-\alpha}) & \text{when } \mu \text{ is } \alpha\text{-stable,}
\end{cases}
\]
since \( g(h) = h^{1/2} \) when \( \mu \) is Gaussian and \( g(h) = h^{1/\alpha} \) when \( \mu \) is \( \alpha \)-stable. Following Remark 3.10, we need to have
\[
(3.8)
\]
hence, choosing \( a = \frac{2hLR}{J_{2g(h)^{-1}R}} \), we see that it is of order
\[
a = \begin{cases} 
\Theta(hR \exp(4dR^2/h)) & \text{when } \mu \text{ is Gaussian,} \\
\Theta(hR(1 + d^{1/2}h^{-1/\alpha}R)^\alpha) & \text{when } \mu \text{ is } \alpha\text{-stable.}
\end{cases}
\]
Moreover, in the simplified version of Theorem 2.1 (cf. Remark 2.2) the contractivity constant given by \( (2.6) \) becomes
\[
c_* = \min \left\{ \frac{hLR}{a + 1 + \Re^{-R} \cdot \frac{(K - hL^2/2)h}{1 + (1+a)\Re^{-1}e^R}} \right\} =: \min\{c_{*,1}, c_{*,2}\},
\]
where for the term \( c_{*,1} \) we used \( (3.8) \) and the bound on \( \pi \) from Remark 3.10. This shows that for large \( \Re \) or for small \( h \), the contractivity constant in the \( \alpha \)-stable case can remain much larger than the corresponding constant in the Gaussian case. In particular, \( c_{*,1} = \Omega(\Re^{-\alpha}) \) as \( \Re \to \infty \) in the \( \alpha \)-stable case, whereas \( c_{*,1} = \Omega(\exp(-R^2)) \) as \( \Re \to \infty \) in the Gaussian case.

Next, we will consider the contraction of the Markov coupling defined by \( (3.7) \) in terms of the \( L^1 \)-Wasserstein distance. Recall again that
\[
\alpha_1(x, y) = \frac{1}{2} \mathbb{E}_{x,y}[(R - r)^2 1_{R < r + t}] \quad \text{and} \quad \alpha_1(r) = \inf_{|x-y|=r} \alpha_1(x, y).
\]
Lemma 3.12. Suppose that Assumption (C) and condition (c4) hold, and that \( \lim_{h \to 0^+} h/g(h) = 0 \). Then there exist constants \( \varepsilon \in (0, 1/4) \), \( \gamma > 0 \) large enough and \( c^* > 0 \) (which is independent of \( \gamma \) but depends on \( \varepsilon \)) such that for any \( \kappa, h > 0 \) with \( \kappa \leq \varepsilon g(h)/4 \) and \( h/g(h) \leq \varepsilon (2L\mathcal{R})^{-1} \), and for any \( x, y \in \mathbb{R}^d \) with \( |x - y| \in (0, \mathcal{R}] \),
\[ \alpha_{g(h)}(x, y) \geq c^* g(h)(\hat{r} \wedge \kappa). \]

Combining Lemmas 3.8 and 3.12 with Theorem 2.5, we can obtain the following statement.

Theorem 3.13. Suppose that Assumption (C) and condition (c4) hold, and that \( \lim_{h \to 0^+} h/g(h) = 0 \). Let \( h \in (0, 2KL^{-2} \wedge (2L)^{-1}) \) and \( \kappa \leq \varepsilon g(h)/4 \) such that \( h/g(h) \leq \varepsilon (2L\mathcal{R})^{-1} \) and
\[ \left(2hL\mathcal{R} \frac{c^* g(h)(2\mathcal{R}/2 \wedge \kappa)}{\varepsilon g(h)(2\mathcal{R}/2 \wedge \kappa)} + 1\right) \gamma g(h) \leq \log 2 \]
with \( \varepsilon, \gamma \) and \( c^* \) being the constants given in Lemma 3.12. Let \( \rho = f(| \cdot |) \) be the function defined by (2.11) with \( \Psi(r) = r \), \( r_1 = \mathcal{R} \), \( l_0 = \gamma g(h) \) and \( c = \frac{2hL\mathcal{R}}{c^* g(h)(2\mathcal{R}/2 \wedge \kappa)} + 1 \).

Then, for all \( x, y \in \mathbb{R}^d \),
\[ \mathbb{E}_{x,y}[\rho(X, Y^{(1)})] \leq (1 - c_* \rho(x, y), \]
where \( c_* := \min \{ c^* g(h)e^{-c(\mathcal{R} + t_b)}(2 \wedge \kappa \mathcal{R}^{-1}), 2^{-1}(K - hL^2/2)he^{-c(\mathcal{R} + t_b)}\} \).

Proof. According to Lemma 3.8, \( \beta(r) \leq hLr \) for all \( r > 0 \), and \( \beta(r) \leq -(K - hL^2/2)hr \) for all \( r \in (\mathcal{R}, \infty) \). On the other hand, by Lemma 3.12,
\[ \alpha_{g(h)}(r) \geq c^* g(h)(\hat{r} \wedge \kappa) \geq c^* g(h)(\hat{r} \wedge \kappa) \]
for all \( r \in (0, \mathcal{R}] \), where in the last inequality we used the fact that \( |\hat{r} - r| \leq hLr \leq r/2 \) due to \( h \leq (2L)^{-1} \). In particular, (b1) and (b3) in Assumption (B) are satisfied with \( l_0 = \gamma g(h) \), \( r_1 = \mathcal{R} \) and \( c_0 = (K - hL^2/2)h \). Furthermore,
\[ \sup_{r \in (0, \mathcal{R}]} \frac{2\beta(r)_+}{\alpha_{g(h)}(r)} \leq \frac{2hL\mathcal{R}}{c^* g(h)(2\mathcal{R}/2 \wedge \kappa)}. \]
Hence, (b2) in Assumption (B) is satisfied because of (3.9). Then, the desired assertion follows from Theorem 2.5. \( \square \)

Remark 3.14. Similarly as in Remark 3.11 for Theorem 3.9, we can now analyse the difference in the contractivity constant \( c_0 \) in Theorem 3.13 between the Gaussian and the \( \alpha \)-stable case. To this end, note that we can refine the estimate from the proof of Lemma 3.12 by writing
\[ 2\alpha_{g(h)}(x, y) = \mathbb{E}_{x,y}[(R - r)^21_{(R \leq r + \gamma g(h))}] =: I_1 + I_2, \]
where \( I_1 = r^2P_{x,y}(R = 0) \) and \( I_2 \) is bounded from below exactly as in the proof of Lemma 3.12. Note that \( I_1 \) corresponds to the coalescence behaviour of the coupling, while \( I_2 \) corresponds to the reflection. In the case of unbounded support, if we assume that \( h \leq 1/L \), using \( \hat{r} \leq (1 + hL)r \leq 2\mathcal{R} \), we get
\[ I_1 \geq r^2J_{g(h)}(h)^{-1}R. \]
From our bounds in Remark 3.11, we see that the right hand side, as a function of \( h \), is of order \( \Omega(\exp(-1/h)) \) in the Gaussian case, and of order \( \Omega(h^{1/\alpha}) \) in the \( \alpha \)-stable case. On the other hand, the bound on \( I_2 \) obtained in the proof of Lemma 3.12 is always of order \( \Omega(g(h)) = \Omega(h^{1/\alpha}) \), even for \( \alpha = 2 \). Hence we see that in the Gaussian case, for small \( h \), \( I_1 \) becomes negligibly small, but for the \( \alpha \)-stable case it can be taken into account, and hence can lead to an improved lower bound on \( \alpha_{g(h)}(x, y) \), and, in consequence, on \( c^* \) and \( c_* \) in Theorem 3.13. We leave the details to the reader. Moreover, observe that the bound from Lemma 3.12 requires the assumption \( h/g(h) \leq \varepsilon (2L\mathcal{R})^{-1} \), which for large \( \mathcal{R} \) can become restrictive. However, in the \( \alpha \)-stable case we can just bound \( I_2 \geq 0 \) and use the bound on \( I_1 \) presented above, obtained just under the
assumption $h \leq 1/L$. This leads to a lower bound on $\alpha_{g(x,y)}(x,y)$ of the same order in $h$ as in Lemma 3.12 under a more relaxed condition on $h$. In the Gaussian case, as mentioned above, bounding $I_2 \geq 0$ would result in a bound on $\alpha_{g(x,y)}(x,y)$ of a much worse order $\Omega(\exp(-1/h))$, hence the condition $h/g(h) \leq \varepsilon(2LR)^{-1}$ becomes necessary.

Finally, we consider the $L^p$-Wasserstein distance with $p > 1$. For this purpose, we need to do some modification to the Markov coupling $((X,Y^{(1)})$, $P_{x,y})$. For $h, s, l' > 0$, we consider the following Markov coupling $((X,Y^{(2)})$:

(i) When $r = |x - y| \in (0,s]$,

$$X = \hat{x} + g(h)z, \quad Y^{(2)} = \begin{cases} \hat{y} + g(h)(z + g(h)^{-1}(\hat{x} - \hat{y})_s), & \mu(dz); \\ \hat{y} + g(h)R_{x,y}(z), & I_{|z| \leq l'}|z + g(h)^{-1}(\hat{x} - \hat{y})_s| \leq l'\mu(g(h)^{-1}(\hat{y} - \hat{z})_s)(dz), \\ \hat{y} + g(h)z, & I_{|z| > l' \text{ or } |z + g(h)^{-1}(\hat{x} - \hat{y})_s| > l'}\mu(dz). \\
\end{cases}$$  \hspace{1cm} (3.10)

(ii) When $r \in (s, \infty)$,

$$X = \hat{x} + g(h)z, \quad Y^{(2)} = \hat{y} + g(h)z, \quad \mu(dz).$$  \hspace{1cm} (3.11)

See Lemma 5.3 in appendix for the proof that $((X,Y^{(2)})$, $P_{x,y})$ is a Markov coupling of chain $X$. Note that this coupling behaves like the reflection coupling (3.7) when the distance between the marginals before the jump is small (smaller than $s$) and both jump sizes are also small (smaller than $l'$), and otherwise behaves like the synchronous coupling. It is a generalisation of the coupling that was used in [25, (2.7)] to obtain $L^2$ bounds in the case of the Gaussian noise. Furthermore, we would like to point out that under Assumption (C) and for the step size $h < 2KL^{-2}$, it is easy to see that the Markov coupling $((X,Y^{(2)})$, $P_{x,y})$ defined by (3.10) and (3.11) satisfies $|X - Y^{(2)}| \leq |x - y| + l$, where

$$l = hLs + \kappa \lor (2g(h)l').$$

**Lemma 3.15.** Suppose that Assumption (C) and condition (c4) hold, and that $\lim_{h \to 0^+} h/g(h) = 0$. Consider the Markov coupling $((X,Y^{(2)})$, $P_{x,y})$ with $s = R$ and $l' \geq g(h)^{-1}(1 + hL)R + 1$. It holds that for any $0 < h < 2KL^{-2}$,

(i) $\beta(x,y) \leq hL|x - y|/\alpha_{g(h)}(x,y)$ for any $x, y \in \mathbb{R}^d$;

(ii) $\beta(x,y) \leq -(K - hL^2/2)h|x - y|$ for any $x, y \in \mathbb{R}^d$ with $|x - y| \geq R$.

Moreover, there exist constants $\varepsilon \in (0, 0.1/4)$, $\gamma > 0$ large enough and $c^* > 0$ (which is independent of $\gamma$ but depends on $\varepsilon$) such that for any $l' \geq g(h)^{-1}(1 + hL)R + \gamma/2 + 1$, any $\kappa > 0$ with $\kappa \leq \varepsilon g(h)/4$ and $h/g(h) \leq \varepsilon(2LR)^{-1}$ and any $x, y \in \mathbb{R}^d$ with $|x - y| \in (0, R)$,

$$\alpha_{g(h)}(x,y) \geq c^* g(h)(\hat{r} \land \kappa).$$

With this estimate and Lemma 3.15 at hand, we can follow the proof of Theorem 2.7 to get the following assertion.

**Theorem 3.16.** Consider the Markov coupling $((X,Y^{(2)})$, $P_{x,y})$ with $s = R$ and $l' \geq (2g(h))^{-1}(1 + hL)R + \gamma/2 + 1$, where $\gamma$ is the constant given in Lemma 3.15. Suppose that Assumption (C) and condition (c4) hold, and that $\lim_{h \to 0^+} h/g(h) = 0$. Let $\varepsilon \in (0, 2KL^{-2} \land (2L)^{-1})$ and $\kappa \leq \varepsilon g(h)/4$ such that $h/g(h) \leq \varepsilon(2LR)^{-1}$ and

$$\gamma g(h) \leq \log 2 \left[\frac{2hL \mathbb{R}}{c^* g(h)((\mathbb{R}/2) \land \kappa)} + 1\right] \gamma g(h) \leq \log 2$$

with $\varepsilon$, $\gamma$ and $c^*$ being the constants given in Lemma 3.15. Let $\rho = f(| \cdot |)$ be the function defined by (2.13) with $\Psi(r) = r$, $r_1 = R$, $l_0 = \gamma g(h)$, $l$ given by (3.12) and

$$c = \frac{2hL \mathbb{R}}{c^* g(h)((\mathbb{R}/2) \land \kappa)} + 1.$$
Then, for all $x, y \in \mathbb{R}^d$,

$$E_{x,y} [\rho(X, Y^{(2)})] \leq (1 - c_*) \rho(x, y),$$

where

$$c_* : = \min \left\{ e^c (\beta + (x + (x + (x + 1)) (2^{-1} \wedge \kappa R^{-1})) \frac{c_0 p A}{1 + A + B(R + l_0) - p} \frac{c_0 p A}{2(4^p A + R((k + 1)(R + l_0)) - (p - 1))} \right\}$$

with $c_0 : = (K - h L^2/2) h$, $k \geq 1 + \max \left\{ 2l e^{c} / (c_0(R + l_0)), 4pl(c_0(R + l_0)) \right\}$,

$$A : = e^{-c(R + l_0)} \left( \frac{1 - p - 1((k + 1)(R + l_0))^{-2} e^{(R + l_0)}}{p(R + l_0)^{-1}} \right)$$

and

$$B : = e^{-2(p - 1) A((k + 1)(R + l_0))^{-2} e^{(R + l_0)}}.$$

**Remark 3.17.** Note that Theorem 3.16 is an extension of [25, Theorem 2.1], where a similar contraction result was proven, but only in the $L^2$-Wasserstein distance and only for chains with the Gaussian noise. We would like to point out that our result in Theorem 3.16 is consistent with in [25] Theorem 2.1 in the sense that applying Theorem 3.16 with $p = 2$ leads to a contractivity constant $c_*$ of the same order in model parameters as the constant from [25]. For instance, it can be easily checked that both constants are of order $\Omega(\mathbb{R}^2 \exp(-R^2))$. One of the main motivations for considering such contractions in [25] was the analysis of Multi-level Monte Carlo (MLMC) methods based on chains in the Gaussian case, for approximating integrals of Lipschitz functions with respect to invariant measures of Langevin SDEs, see Theorem 1.7 therein. By following the analysis of MLMC in [25] Subsection 2.3, it is easy to see that by employing our Theorem 3.16 it is possible to extend [25] Theorem 1.7 from Lipschitz functions to all functions with a polynomial growth. Another possible extension would be the analysis of MLMC methods based on discretisations of SDEs with Lévy noises. We leave the details for future work.

4. Proofs

**Proof of Theorem 2.1** When $x = y$, the assertion holds trivially by our assumption on the Markov coupling $((X, Y), P_{x,y})$ that $P_{x,y}(d(X, Y) = 0) = 1$ and the fact that $\rho(x, x) = 0$ for all $x \in S$.

Next, let $x, y \in S$ with $r : = d(x, y) > 0$. Denote $R = d(X, Y)$. By the mean value theorem and the definition of $f_0$,

$$\rho(X, Y) - \rho(x, y) = -a 1_{\{R = 0\}} + f_0(R) - f_0(r)$$

$$\leq -a 1_{\{R = 0\}} + f_0'(r)(R - r) + \frac{1}{2} \left( \max_{\xi \in [r, R \wedge R]} f_0''(\xi) \right) (R - r)^2$$

$$\leq -a 1_{\{R = 0\}} + f_0'(r)(R - r) + \frac{1}{2} f_0''(r)(R - r)^2 1_{\{R < r\}},$$

where the last inequality follows from the facts that $f_0^{(3)} > 0$ and $f_0'' < 0$ on $(0, \infty)$. By taking expectation and using the facts that $f_0' > 0$ and $f_0'' < 0$ on $(0, \infty)$, we get that

$$E_{x,y}[\rho(X, Y)] - \rho(x, y) \leq -a \pi(r) + B(r) f_0'(r) + \alpha(r) f_0''(r).$$

(1) Suppose that $r \in (0, r_0]$. Then, it follows from (a1), (a2) and the definition of $a$ given in 2.4 that

$$E_{x,y}[\rho(X, Y)] - \rho(x, y) \leq -a \pi(r) + (c e^{-cr} + c e^{-cr}) B(r) \leq -a \pi(r) + c(1 + e^{-cr}) B(r)$$

$$\leq -a \frac{1}{2} \pi(r) \leq -c_1 \rho(x, y),$$

where

$$c_1 : = \frac{a}{2(a + 1 + cr e^{-cr})} \inf_{r \in (0, r_0]} \pi(r).$$
(2) Suppose that \( r \in (r_0, r_1] \). Then, by (a2), (a3) and the definition of \( c \) given in (2.4),
\[
E_{x,y}[\rho(X,Y)] - \rho(x,y) \leq \bar{\beta}(r) f'_0(r) + \alpha(r) f''_0(r) = (\alpha e^{-\sigma} + c e^{-cr}) \bar{\beta}(r) - c^2 e^{-c r} \alpha(r)
\]
\[
\leq 2\alpha e^{-\sigma} \bar{\beta}(r)_+ - c^2 e^{-c r} \alpha(r) \leq -\frac{1}{2} c^2 e^{-c r} \alpha(r) \leq -c_2 \rho(x,y),
\]
where
\[
c_2 := \frac{c^2 e^{-c r}}{2(a + 1 + cr_1 e^{-cr})} \inf_{r \in (r_0, r_1]} \alpha(r).
\]

(3) Suppose that \( r \in (r_1, \infty) \). Then, according to (a3),
\[
E_{x,y}[\rho(X,Y)] - \rho(x,y) \leq (\alpha e^{-\sigma} + \alpha e^{-c r}) \bar{\beta}(r) \leq \alpha e^{-c r} \bar{\beta}(r)
\]
\[
\leq -c_0 \alpha e^{-c r} \leq -c_3 (a + 1 + e^{-c r} \rho(x,y) \leq -c_3 \rho(x,y),
\]
where
\[
c_3 := c_0 [1 + (1 + a) r_1^{-1} e^{-c r}]^{-1}.
\]

Therefore, taking \( c_* = \min\{c_1, c_2, c_3\} \) gives us the desired assertion. \( \Box \)

**Proof of Theorem 2.3** Let \( x, y \in S \) with \( r = d(x,y) > 0 \). According to the argument for (4.1) and (i) in condition (a3*),
\[
E_{x,y}[\rho(X,Y)] - \rho(x,y) \leq -a \bar{\pi}(r) + f'_1(r) \bar{\beta}(r) + f''_1(r) \alpha(r) - \lambda e(V(x) + V(y)) + 2cC_0.
\]

(1) When \( r \in (0, r_0] \), it follows from (4.2) and the definition of the constant \( a \) given in (2.9) that
\[
E_{x,y}[\rho(X,Y)] - \rho(x,y) \leq -a \bar{\pi}(r) + \alpha e^{-c r} \bar{\beta}(r) - \lambda e(V(x) + V(y)) + 2cC_0
\]
\[
\leq -a \bar{\pi}(r) + a \bar{\beta}(r)_+ - \lambda e(V(x) + V(y)) + 2cC_0
\]
\[
\leq -\frac{1}{2} \bar{\pi}(r) - \lambda e(V(x) + V(y))
\]
\[
\leq -c_1 \rho(x,y),
\]
where
\[
c_1 := \min \left\{ \frac{a}{2(a + 1)} \inf_{r \in (0, r_0]} \bar{\pi}(r), \lambda \right\}.
\]

(2) When \( r \in (r_0, r_1] \), we get from (4.2) and the definitions of the constants \( c \) and \( \epsilon \) given in (2.9) and (2.10) respectively that
\[
E_{x,y}[\rho(X,Y)] - \rho(x,y) \leq \alpha e^{-c r} \bar{\beta}(r) - c^2 e^{-c r} \alpha(r) - \lambda e(V(x) + V(y)) + 2cC_0
\]
\[
\leq -\frac{1}{2} c^2 e^{-c r} \alpha(r) - \lambda e(V(x) + V(y)) + 2cC_0
\]
\[
\leq -\frac{1}{4} c^2 e^{-c r} \alpha(r) - \lambda e(V(x) + V(y))
\]
\[
\leq -\frac{1}{4} c^2 e^{-c r} \inf_{r \in (r_0, r_1]} \alpha(r) - \lambda e(V(x) + V(y))
\]
\[
\leq -c_2 \rho(x,y),
\]
where
\[
c_2 := \min \left\{ \frac{c^2 e^{-c r} \inf_{r \in (r_0, r_1]} \alpha(r)}{4(a + 1)}, \lambda \right\}.
\]

(3) When \( r \in (r_1, \infty) \), by (4.2) and the definition of \( r_1 \) given by (2.7), we get that
\[
E_{x,y}[\rho(X,Y)] - \rho(x,y) \leq f'_1(r) \bar{\beta}(r) - \lambda e(V(x) + V(y)) + 2cC_0
\]
\[
\leq \left( \frac{K \alpha e^{-c r_1} + c r_1}{2} \right) (V(x) + V(y)) \leq -c_3 (V(x) + V(y)),
\]
where
\[
c_3 := \frac{\lambda}{16C_0} e^{-c r_1} \left( \inf_{r \in (r_0, r_1]} \alpha(r) \right) \left( \sup_{r \in (r_0, r_1]} \frac{2 \bar{\beta}(r)_+}{\alpha(r)} + c_0 + 1 \right).
\]
due to the definitions of constants $A$, $c$ and $\epsilon$ given in (2.9) and (2.10) respectively. This along with the definition of $r_1$ again further yields that

$$E_{x,y}[\rho(X,Y)] - \rho(x,y) \leq -\frac{2C_0c_3}{\lambda} - \frac{c_2}{2}(V(x) + V(y)) \leq -c_4\rho(x,y),$$

where

$$c_4 := \min \left\{ \frac{2C_0c_3}{\lambda(a+1)}, \frac{c_3}{2c} \right\}.$$ 

Combining with all the estimates above, we can obtain the desired assertion with $c_* = \min\{c_1, c_2, c_4\}$. 

\[ \square \]

**Proof of Theorem 2.5.** It is clear that $f_3 \in C([0, \infty)) \cap C^2((0, \infty))$ such that

$$f_3'(r) = \begin{cases} e^{-c\Psi(r)}, & 0 < r \leq r_1 + l_0, \\ f_3'(r_1 + l_0) \left[ 1 + \exp \left( \frac{2f_3''(r_1 + l_0)}{f_3'(r_1 + l_0)}(r - (r_1 + l_0)) \right) \right], & r > r_1 + l_0, \end{cases}$$

and

$$f_3''(r) = \begin{cases} -c\Psi'(r)e^{-c\Psi(r)}, & 0 < r \leq r_1 + l_0, \\ f_3''(r_1 + l_0) \exp \left( \frac{2f_3''(r_1 + l_0)}{f_3'(r_1 + l_0)}(r - (r_1 + l_0)) \right), & r > r_1 + l_0. \end{cases}$$

In particular, for all $r \geq 0$,

$$\frac{f_3'(r_1 + l_0)r}{2} \leq f_3(r) \leq \max \left\{ 1, \frac{f_3(r_1 + l_0)}{r_1 + l_0} \right\} r,$$

and $f_3' > 0$, $f_3'' \leq 0$ and $f_3''$ is increasing on $(0, \infty)$.

Let $x, y \in S$ with $r = d(x, y) > 0$. If $r \in (0, r_1]$, then, by (2.12) and (b1)–(b2),

$$E_{x,y}[f_3(R) - f_3(r)] \leq f_3'(r_1 + l_0)\beta(r) + f_3''(r + l_0)\alpha_{l_0}(r)$$

$$= e^{-c\Psi(r)}\beta(r) - c\Psi'(r + l_0)e^{-c(\Psi(r + l_0) - \Psi(r))}\alpha_{l_0}(r)$$

$$\leq e^{-c\Psi(r)}\beta(r) - 2e^{-c(\Psi(r + l_0) - \Psi(r))}\beta(r)_+ - \Psi'(r + l_0)e^{-c(\Psi(r + l_0) - \Psi(r))}\alpha_{l_0}(r)$$

$$\leq -\Psi'(r + l_0)e^{-c(\Psi(r + l_0) - \Psi(r))}\alpha_{l_0}(r) - c_1f_3(r),$$

where the third inequality holds true since $\sup_{r \in [0, r_1]}(\Psi(r + l_0) - \Psi(r)) \leq \Psi(l_0)$, due to the fact that $\Psi(0) = 0$ and $\Psi'' \leq 0$, and hence, by (b2) and the definition of $c$ given in (2.12),

$$\beta(r) - 2e^{-c(\Psi(r + l_0) - \Psi(r))}\beta(r)_+ \leq \beta(r) - 2e^{-c\Psi(l_0)}\beta(r)_+ \leq \beta(r) - \beta(r)_+ \leq 0,$$

whereas in the last inequality

$$c_1 := \Psi'(r_1 + l_0)e^{-c\Psi(r_1 + l_0)} \inf_{r \in (0, r_1]} \frac{\alpha_{l_0}(r)}{r} > 0,$$

thanks to (b1) and $\Psi' > 0$ on $(0, \infty)$.

If $r \in (r_1, \infty)$, then, by (b3),

$$E_{x,y}[f_3(R) - f_3(r)] \leq f_3'(r_1 + l_0)\beta(r) \leq -c_0f_3'(r_1 + l_0)\beta(r)/2 \leq -c_2f_3(r),$$

where

$$c_2 := \frac{c_0}{2}e^{-c(r_1 + l_0)} \max \left\{ 1, \frac{f_3(r_1 + l_0)}{r_1 + l_0} \right\}^{-1}.$$

The proof is completed by taking $c_* = c_1 \wedge c_2$. 

\[ \square \]
Proof of Theorem 2.7. Let \( x, y \in S \) with \( r = d(x, y) > 0 \). Suppose that \( r \in (0, r_1) \). Then, by using the fact that \( f''_4 \) is increasing on \((0, r_1 + l_0)\), and following the arguments in the proof of Theorem 2.5, we can obtain that

\[
\mathbb{E}_{x,y}[f_4(R) - f_4(r)] \leq f'_4(r)\beta(r) + f''_4(r + l_0)\alpha_0(r)
\]

\[
= e^{-c\Psi'(r)}\beta(r) - e^{-c\Psi'(r + l_0)}\alpha_0(r)
\]

\[
\leq -c_1 f_4(r),
\]

where

\[
c_1 := \Psi'(r_1 + l_0)e^{-c\Psi(r_1 + l_0)} \inf_{r \in [0, r_1]} \frac{\alpha_0(r)}{r} > 0.
\]

Suppose that \( r \in [r_1,(k + 1)(r_1 + l_0) - l] \). By (b3) and the fact that \( f'_4 \) is decreasing on \((0,(k + 1)(r_1 + l_0) - l)\), we get

\[
\mathbb{E}_{x,y}[f_4(R) - f_4(r)] \leq f'_4(r)\beta(r) \leq -c_0 f'_4(r)r \leq -c_2 f'_4(r),
\]

where

\[
c_2 := \min \left\{ c_0e^{-c\Psi(r_1+l_0)}, \frac{c_0pA}{1 + A + B(r_1 + l_0)^{-p}} \right\}.
\]

Here in the last inequality we used the facts that when \( r \in [r_1, r_1 + l_0) \),

\[
\frac{f''_4(r)r}{f'_4(r)} = \frac{re^{-c\Psi(r)}}{\int_0^r e^{-c\Psi(s)}ds} \geq e^{-c\Psi(r)} \geq e^{-c\Psi(r_1+l_0)};
\]

and that when \( r \in [r_1 + l_0, (k + 1)(r_1 + l_0) - l] \),

\[
\frac{f''_4(r)r}{f'_4(r)} \geq \frac{pAr^p}{(r_1 + l_0) + Ar^p + B} \geq \frac{pA}{1 + A + B(r_1 + l_0)^{-p}}.
\]

Now consider the case where \( r \in ((k + 1)(r_1 + l_0) - l, 4k(r_1 + l_0)] \). It follows from the properties of the Markov coupling \((X, Y), \mathbb{P}_{x,y}\) and the definition of the function \( f_4 \) that

\[
f_4(R) - f_4(r) = (f_4(R) - f_4(r))1_{\{R \leq r\}} + (f_4(R) - f_4(r))1_{\{r < R \leq r + l\}}
\]

\[
\leq \left[ \inf_{\xi \in (r, R)} f'_4(\xi) \right] (R - r)1_{\{R \leq r\}} + \left[ \sup_{\xi \in (r, R)} f'_4(\xi) \right] (R - r)1_{\{r < R \leq r + l\}}
\]

\[
\leq f'_4((k + 1)(r_1 + l_0))(R - r)1_{\{R \leq r\}} + [f'_4(r) \vee f'_4(r + l)](R - r)1_{\{r < R \leq r + l\}}
\]

\[
\leq f'_4((k + 1)(r_1 + l_0))(R - r)
\]

\[
+ [f'_4(r) \vee f'_4(r + l) - f'_4((k + 1)(r_1 + l_0))](R - r)1_{\{r < R \leq r + l\}}.
\]

Taking expectation and (b3) give us that

\[
\mathbb{E}_{x,y}[f_4(R) - f_4(r)] \leq -c_0 f'_4((k + 1)(r_1 + l_0))r + l[f'_4(r) \vee f'_4(r + l) - f'_4((k + 1)(r_1 + l_0))]
\]

\[
\leq -\frac{c_0}{2} f'_4((k + 1)(r_1 + l_0))r \leq -c_3 f_4(r)
\]

with

\[
c_3 := \frac{c_0pA}{2(4^{p-1}A + B((k + 1)(r_1 + l_0))^{-p-1})}.
\]

Here we used the fact that for all \( r \in ((k + 1)(r_1 + l_0) - l, 4k(r_1 + l_0)] \),

\[
\frac{c_0}{2} f'_4((k + 1)(r_1 + l_0))r \geq l(f'_4(r) \vee f'_4(r + l) - f'_4((k + 1)(r_1 + l_0))).
\]

Indeed, by the definition of \( f'_4 \), (4.3) is a consequence of

\[
\frac{c_0}{2} e^{-c(k+1)(r_1+l_0)}r \geq le^{-cr}, \quad \frac{c_0}{2} ((k + 1)(r_1 + l_0))^{p-1}r \geq l(r + l)^{p-1}
\]

for all \( r \in ((k + 1)(r_1 + l_0) - l, 4k(r_1 + l_0)] \), thanks to the fact that \( k \geq \max\{2le^d/(c_0(r_1 + l_0)), 8p/(c_0(r_1 + l_0))\} \).
Finally, suppose that \( r \in (4k(r_1 + l_0), \infty) \). Since \( f_4'' > 0 \) on \( ((k + 1)(r_1 + l_0), \infty) \),
\[
  f_4(R) - f_4(r) = (f_4(R) - f_4(r))1_{\{R \leq r/2\}} + (f_4(R) - f_4(r))1_{\{r/2 < R \leq r\}}
  + (f_4(R) - f_4(r))1_{\{r < R \leq r + l\}}
  \leq (f_4(r/2) - f_4(r))1_{\{R \leq r/2\}} + \left[ \inf_{\xi \in (r,R)} f'_4(\xi) \right] (R - r)1_{\{r/2 < R \leq r\}}
  + \left[ \sup_{\xi \in (r,R)} f'_4(\xi) \right] (R - r)1_{\{r < R \leq r + l\}}
  \leq \left[ \inf_{\xi \in (r/2,r)} f'_4(\xi) \right] (r/2 - r)1_{\{R \leq r/2\}} + f'_4(r/2) (R - r)1_{\{r/2 < R \leq r\}}
  + f'_4(r + l)1_{\{r < R \leq r + l\}}
  \leq \frac{1}{2} f'_4(r/2) (R - r)1_{\{R \leq r\}} + f'_4(r + l) (R - r)1_{\{r < R \leq r + l\}}
  \leq \frac{1}{2} f'_4(r/2) (R - r) + \left( f'_4(r + l) - \frac{1}{2} f'_4(r/2) \right) (R - r)1_{\{r < R \leq r + l\}}.
\]
Thus, by (b3) we have
\[
  E_{x,y}[f_4(R) - f_4(r)] \leq -\frac{c_0}{2} f'_4(r/2) r + l \left( f'_4(r + l) - \frac{1}{2} f'_4(r/2) \right) \leq -\frac{c_0}{4} f'_4(r/2) r \leq -c_4 f_4(r)
\]
with
\[
  c_4 := \frac{2^{-(p+1)}c_0 p A}{A + B(4k(r_1 + l_0))^{-p}}.
\]
Here we used the fact that
\[
  \frac{c_0}{4} f'_4(r/2) r \geq l \left( f'_4(r + l) - \frac{1}{2} f'_4(r/2) \right), \quad r > 4k(r_1 + l_0),
\]
which follows from
\[
  \frac{c_0}{4} e^{-cr^2/2} r \geq l e^{-cr}, \quad \frac{c_0}{4} (r/2)^{p-1} r > l(r + l)^{p-1} \quad \text{for all } r > 4k(r_1 + l_0)
\]
and the fact that \( k \geq \max\{2le^l/(c_0(r_1 + l_0)), 8p l/(c_0(r_1 + l_0))\} \).

Combining all the estimates above, we prove the desired assertion with \( c_* = \min\{c_1, c_2, c_3, c_4\} \), where \( c_i > 0, 1 \leq i \leq 4 \), are given above. □

4.2. Proofs of results for the Euler scheme.

Proof of Lemma 3.1. Recall that \((\delta_x \ast \mu_x)(dz) = \mu_x(dz)\) for all \( x \in \mathbb{R}^d \). In particular, this implies that
\[
  \mu_x(\mathbb{R}^d) = \mu_x(\mathbb{R}^d).
\]
Let \( \tilde{\beta}(x,y) = E_{x,y}[R - \hat{r}] \). We first claim that for all \( x, y \in \mathbb{R}^d \), \( \tilde{\beta}(x,y) = 0 \). Indeed, for fixed \( x, y \in \mathbb{R}^d \) with \( r = |x - y| > 0 \), by (3.6) and (4.4), we get
\[
  E_{x,y}[R] = \frac{1}{2} \int_{\mathbb{R}^d} (\hat{r} - \hat{r} \wedge \kappa) \mu_{g(h)}^{-1}(\hat{y} - \hat{x})(dz) + \frac{1}{2} \int_{\mathbb{R}^d} (\hat{r} + \hat{r} \wedge \kappa) \mu_{g(h)}^{-1}(\hat{x} - \hat{y})(dz)
  + \int_{\mathbb{R}^d} \hat{r} \left[ \mu(dz) - \frac{1}{2} \mu_{g(h)}^{-1}(\hat{y} - \hat{x})(dz) - \frac{1}{2} \mu_{g(h)}^{-1}(\hat{x} - \hat{y})(dz) \right]
  = \hat{r} + \frac{1}{2} (\hat{r} \wedge \kappa)(\mu_{g(h)}^{-1}(\mathbb{R}^d) - \mu_{g(h)}^{-1}(\hat{y} - \hat{x})(\mathbb{R}^d)) = \hat{r},
\]
which implies that \( \hat{\beta}(x, y) = 0 \) for all \( x, y \in \mathbb{R}^d \). Hence, for any \( x, y \in \mathbb{R}^d \),

\[
\hat{\beta}(x, y) = \hat{\beta}(x, y) + \hat{r} - r = \hat{r} - r.
\]

This along with Assumption (c1) yields that for any \( x, y \in \mathbb{R}^d \),

\[
\beta(x, y) \leq |\hat{r} - r| \leq |(\hat{x} - \hat{y}) - (x - y)| = h|b(x) - b(y)| \leq hLr,
\]

proving the assertion (i).

Next, we suppose that \( r = |x - y| \geq \mathcal{R} \). As we mentioned above, \( K \leq L \) and so \( 1 - 2hK + h^2L^2 \geq 0 \). Combining this fact with (c2) and the elementary inequality \( \sqrt{1 + x} \leq 1 + x/2 \) for \( x \geq 0 \), we arrive at that for any \( h \leq 2KL^{-2} \),

\[
\hat{r} = \sqrt{|x - y|^2 + 2h(x - y, b(x) - b(y)) + h^2|b(x) - b(y)|^2}
\leq r\sqrt{1 - 2hK + h^2L^2} \leq r(1 - hK + h^2L^2/2).
\]

This proves (ii) due to \( \beta(x, y) = \hat{r} - r \).

(iii) immediately follows from the definition (3.6) for the Markov coupling \( ((X, Y), \mathbb{P}_{x,y}) \), while (iv) is also a consequence of (3.6). Indeed, for any \( x, y \in \mathbb{R}^d \),

\[
2\alpha(x, y) = E[(R - r)^2 \mathbb{1}_{\{R < r\}}] \geq \frac{1}{2} ((\hat{r} - r \wedge \kappa) - r)^2 \mathbb{1}_{\{\hat{r} - r \wedge \kappa < r\}}(\mathbb{R}^d)
\geq \frac{1}{2} J_{g(h)^{-1}\kappa}((\hat{r} - r \wedge \kappa) - r)^2 \mathbb{1}_{\{\hat{r} - r \wedge \kappa < r\}}.
\]

Now, suppose that \( x, y \in \mathbb{R}^d \) satisfies \( |x - y| \leq \mathcal{R} \). If \( \hat{r} \leq \kappa \), then \( \hat{r} - \hat{r} \wedge \kappa = 0 \) and so

\[
2\alpha(x, y) \geq \frac{1}{2} r^2 J_{g(h)^{-1}\kappa}.
\]

If \( \hat{r} \geq \kappa \), then, due to the fact that \( h \leq \frac{\kappa}{2\mathcal{R}} \) implies \( \hat{r} - r \leq hLr \leq \frac{\kappa}{2\mathcal{R}} Lr \leq \kappa/2 \), it holds that \( \hat{r} - \hat{r} \wedge \kappa = \hat{r} - \kappa \leq r - \kappa/2 \), and so

\[
2\alpha(x, y) \geq \frac{\kappa^2}{8} J_{g(h)^{-1}\kappa} = \left( \frac{r^2}{2} \wedge \frac{\kappa^2}{8} \right) J_{g(h)^{-1}\kappa},
\]

where in the equality above we used Assumption (c1) and the fact that \( \hat{r} \leq r + h|b(x) - b(y)| \leq r + hLr \leq 2r \) for \( h \leq L^{-1} \) and so \( \hat{r} \geq \kappa \) implies that \( \kappa \leq 2r \). The proof is complete. \( \square \)

**Proof of Lemma 3.8**. The proofs of (iii) and (iv) mainly follow from the definition of \( ((X, Y^{(1)}), \mathbb{P}_{x,y}) \) and the argument for Lemma 3.1. So, we only need to verify (i) and (ii). The proof is similar in spirit to the proof of [13, Lemma 2.7].

Recall that \( \hat{r} = |\hat{x} - \hat{y}| \) and \( R = |X - Y| \). According to (3.7), for any \( x, y \in \mathbb{R}^d \),

\[
E_{x,y}[R] = (\hat{r} - \hat{r} \wedge \kappa) \mu_{g(h)^{-1}(\hat{y} - \hat{x})}(\mathbb{R}^d) + \int_{\mathbb{R}^d} \hat{r} \left| 1 + 2g(h) \frac{(\hat{x} - \hat{y}, z)}{|\hat{x} - \hat{y}|^2} (\mu - \mu_{g(h)^{-1}(\hat{y} - \hat{x})}) (dz)ight|
\]

Due to the rotational invariance of the measure \( \mu \), it suffices to assume that \( \hat{x} = 0 \) and \( \hat{y} = \hat{r}e_1 \), where \( e_1, \ldots, e_d \) is the canonical basis of \( \mathbb{R}^d \). Hence, without loss of generality, we can carry out the argument to the one-dimensional case. In particular, when \( d = 1 \),

\[
(4.5) \quad E_{x,y}[R] = (\hat{r} - \hat{r} \wedge \kappa) \mu_{g(h)^{-1}(\hat{r} \wedge \kappa)}(\mathbb{R}) + \int_{\mathbb{R}} |\hat{r} - 2g(h)z| (\mu - \mu_{g(h)^{-1}(\hat{r} \wedge \kappa)})(dz)
\]

For the integration in the right-hand side of the equality above, using the assumptions that \( r \mapsto m(r) \) is non-increasing in \( (0, \infty) \) and \( z \mapsto m(z) \) is symmetric on \( \mathbb{R} \), we have

\[
\int_{\mathbb{R}} |\hat{r} - 2g(h)z| (\mu - \mu_{g(h)^{-1}(\hat{r} \wedge \kappa)})(dz)
= \int_{\mathbb{R}} |\hat{r} - 2g(h)z| (m(z) - m(z) \wedge m(z - g(h)^{-1}(\hat{r} \wedge \kappa))) \, dz
= \int_{-\infty}^{(2g(h)^{-1}(\hat{r} \wedge \kappa))} (\hat{r} - 2g(h)z) \left[ m(z) - m(z - g(h)^{-1}(\hat{r} \wedge \kappa)) \right] \, dz
\]
\[
\begin{align*}
&= (\hat{r} - \hat{r} \land \kappa) \mu(\{ |z| \leq (2g(h))^{-1}(\hat{r} \land \kappa) \}) \\
&\quad + \int_{-\infty}^{(2g(h))^{-1}(\hat{r} \land \kappa)} ((\hat{r} \land \kappa) - 2g(h)z) \left[ m(z) - m(z - g(h)^{-1}(\hat{r} \land \kappa)) \right] dz \\
&= (\hat{r} - \hat{r} \land \kappa) \mu(\{ |z| \leq (2g(h))^{-1}(\hat{r} \land \kappa) \}) \\
&\quad + (2g(h))^{-1} \int_{0}^{\infty} u \left[ m(2g(h)^{-1}((\hat{r} \land \kappa) - u)) - m(2g(h)^{-1}((\hat{r} \land \kappa) + u)) \right] du \\
&= (\hat{r} - \hat{r} \land \kappa) \mu(\{ |z| \leq (2g(h))^{-1}(\hat{r} \land \kappa) \}) \\
&\quad + (4g(h))^{-1} \int_{-\infty}^{\infty} u \left[ m(2g(h)^{-1}((\hat{r} \land \kappa) - u)) - m(2g(h)^{-1}((\hat{r} \land \kappa) + u)) \right] du \\
&= (\hat{r} - \hat{r} \land \kappa) \mu(\{ |z| \leq (2g(h))^{-1}(\hat{r} \land \kappa) \}) \\
&\quad + \frac{1}{2} \int_{-\infty}^{+\infty} ((\hat{r} \land \kappa) - 2g(h)z) \left[ m(z) - m(z - g(h)^{-1}(\hat{r} \land \kappa)) \right] dz
\end{align*}
\]

where in the last equality we used the fact that \( \int_{\mathbb{R}} zm(z) \, dz = 0 \) thanks to the symmetry property of \( m(z) \) and \( \int_{\mathbb{R}} |z|m(|z|) \, dz < \infty. \)

On the other hand, also due to the assumptions that \( r \mapsto m(r) \) is non-increasing in \((0, \infty)\) and \( z \mapsto m(z) \) is symmetric on \( \mathbb{R} \),

\[ \mu_{g(h)^{-1}(\hat{r} \land \kappa)}(\mathbb{R}) = \mu(\{ |z| > (2g(h))^{-1}(\hat{r} \land \kappa) \}). \]

Hence, putting both estimates above into (4.5), we get

\[ E_{x,y}[R] = (\hat{r} - \hat{r} \land \kappa) \mu(\{ |z| > (2g(h))^{-1}(\hat{r} \land \kappa) \}) + (\hat{r} - \hat{r} \land \kappa) \mu(\{ |z| \leq (2g(h))^{-1}(\hat{r} \land \kappa) \}) + (\hat{r} \land \kappa) = \hat{r}, \]

which implies that \( \beta(x, y) = \hat{r} - r \). With this at hand, the assertions (i) and (ii) then also follow from the proof of Lemma 3.1.

\[ \square \]

**Proof of Lemma 3.12** Similarly to the proof of Lemma 3.8, it suffices to consider the case \( d = 1 \). Let \( x, y \in \mathbb{R} \) with \( r = |x - y| \in (0, \mathcal{R}] \). Without loss of generality, we assume that \( \hat{x} = x + h\hat{b}(x) = 0 \) and \( \hat{r} = \hat{y} > 0 \). For \( \varepsilon \in (0, 1/4) \) and \( \gamma > 1 \) large enough, by (3.7) and the assumptions that \( r \mapsto m(r) \) is non-increasing in \((0, \infty)\) and \( z \mapsto m(z) \) is symmetric on \( \mathbb{R} \),

\[ 2\alpha_{g(h)}(x, y) = E_{x,y}[(R - r)^2 I_{(R \leq r + \gamma g(h))}] \]

\[
\geq \varepsilon^2 g(h)^2 \int_{\{r + \varepsilon g(h) \leq |\hat{r} - 2g(h)z| \leq \gamma g(h)\}} (\mu - \mu_{g(h)^{-1}(\hat{r} \land \kappa)}) \, (dz) \\
= \varepsilon^2 g(h)^2 \int_{\{z \leq (2g(h))^{-1}(\hat{r} \land \kappa), (2g(h))^{-1}(\hat{r} - r) - \gamma/2 \leq z \leq (2g(h))^{-1}(\hat{r} - r) - \varepsilon/2\}} (m(z) - m(z - g(h)^{-1}(\hat{r} \land \kappa))) \, dz.
\]

Note that, according to (c1) in Assumption (C) and \( h/g(h) \leq \varepsilon(2L\mathcal{R})^{-1} \),

\[ (2g(h))^{-1}|\hat{r} - r| \leq (2g(h))^{-1}hLr \leq \varepsilon/4. \]
This, along with the condition that $\kappa \leq \varepsilon g(h)/4$, yields that
\[
\begin{align*}
\alpha_{g(h)}(x, y) &\geq \frac{\varepsilon^2 g(h)^{2}}{2} \int_{-\gamma/2}^{\gamma/2} (m(z) - m(z - g(h)^{-1}(\hat{r} \wedge \kappa))) \, dz \\
&\geq \frac{\varepsilon^2 g(h)^{2}}{2} \int_{-\gamma/2}^{\gamma/2} (m(z) - m(z - g(h)^{-1}(\hat{r} \wedge \kappa))) \, dz \\
&= \frac{\varepsilon^2 g(h)^{2}}{2} \int_{\gamma/2}^{3\varepsilon/4} (m(z) - m(z + g(h)^{-1}(\hat{r} \wedge \kappa))) \, dz \\
&= \frac{\varepsilon^2 g(h)^{2}}{2} \int_{\gamma/2}^{\gamma/2 - \varepsilon/4} \left[ m(z) - m(x) + g(h)^{-1}(\hat{r} \wedge \kappa) \right] \, dz \\
&\geq \frac{\varepsilon^2 g(h)^{2}}{2} \int_{\gamma/2 - \varepsilon/4}^{\gamma/2} \left[ m(z) - m(x) + g(h)^{-1}(\hat{r} \wedge \kappa) \right] \, dz \\
&= \frac{\varepsilon^2 g(h)^{2}}{2} (m(\varepsilon) - m(\gamma/2 - \varepsilon/4)) \\
&\geq \varepsilon^2 g(h)(\hat{r} \wedge \kappa),
\end{align*}
\]
where in the first equality we used the symmetry of $z \mapsto m(z)$ on $\mathbb{R}$ and the change of variable, in the third inequality we used Fubini’s lemma and $\kappa \leq \varepsilon g(h)/4$, and the last inequality follows from the fact $\lim_{s \to \infty} m(s) = 0$ (which is implied by condition (c4)) and by choosing $\varepsilon \in (0, 1/4)$ small and $\gamma > 0$ large enough so that $m(\varepsilon) - m(\gamma/2 - \varepsilon/4) \geq m(\varepsilon)/2 > 0$. Hence, the proof is complete.

**Proof of Lemma 3.15** Similarly to the proof of Lemma 3.8, we only need to consider the case $d = 1$. Without loss of generality, we assume that $\hat{y} \geq \hat{x}$. According to (3.10), we know that when $r \in (0, \mathcal{R}]$,
\[
E_{x, y} | R = (\hat{r} - \hat{r} \wedge \kappa) \mu_{g(h)^{-1}(\hat{r} \wedge \kappa)}(\{ |z| \leq l', |z + g(h)^{-1}(\hat{x} - \hat{y})_n| \leq l' \}) \\
+ \int_{\mathbb{R}} |\hat{r} - 2g(h)z| 1_{\{ |z| \leq l', |z + g(h)^{-1}(\hat{x} - \hat{y})_n| \leq l' \}}(\mu - \mu_{g(h)^{-1}(\hat{r} \wedge \kappa)})(dz) \\
+ \hat{r} \mu(\{ |z| > l' \text{ or } |z + g(h)^{-1}(\hat{x} - \hat{y})_n| \leq l' \}) = : (I) + (II) + (III).
\]
According to the properties of the function $m(z)$ (i.e., $r \mapsto m(r)$ is non-increasing in $(0, \infty)$ and $z \mapsto m(z)$ is symmetric on $\mathbb{R}$), and $l' \geq g(h)^{-1}(1 + hL)\mathcal{R}$, we arrive at
\[
(II) = \int_{-\infty}^{\infty} (\hat{r} - 2g(h)z) 1_{\{ |z| \leq l', |z + g(h)^{-1}(\hat{x} - \hat{y})_n| \leq l' \}}(m(z) - m(z - g(h)^{-1}(\hat{r} \wedge \kappa))) \, dz \\
= \int_{\mathbb{R}} (\hat{r} - 2g(h)z) 1_{\{ |z| \leq l', |z + g(h)^{-1}(\hat{x} - \hat{y})_n| \leq l' \}}(\mu - \mu_{g(h)^{-1}(\hat{r} \wedge \kappa)})(dz) \\
= (\hat{r} - \hat{r} \wedge \kappa)(\mu - \mu_{g(h)^{-1}(\hat{r} \wedge \kappa)})(\{ |z| \leq l', |z + g(h)^{-1}(\hat{x} - \hat{y})_n| \leq l' \}) \\
+ \int_{-l' + g(h)^{-1}(\hat{r} \wedge \kappa)}^{\hat{r} + g(h)^{-1}(\hat{r} \wedge \kappa)} (\hat{r} \wedge \kappa - 2g(h)z)(m(z) - m(z - g(h)^{-1}(\hat{r} \wedge \kappa))) \, dz \\
= (\hat{r} - \hat{r} \wedge \kappa)(\mu - \mu_{g(h)^{-1}(\hat{r} \wedge \kappa)})(\{ |z| \leq l', |z + g(h)^{-1}(\hat{x} - \hat{y})_n| \leq l' \}) \\
+ (2g(h))^{-1} \int_{0}^{2g(h)(\hat{r} \wedge \kappa - u)} u \left[ m((2g(h))^{-1}(\hat{r} \wedge \kappa - u)) - m((2g(h))^{-1}(\hat{r} \wedge \kappa + u)) \right] \, du \\
= (\hat{r} - \hat{r} \wedge \kappa)(\mu - \mu_{g(h)^{-1}(\hat{r} \wedge \kappa)})(\{ |z| \leq l', |z + g(h)^{-1}(\hat{x} - \hat{y})_n| \leq l' \}) \\
+ (4g(h))^{-1} \int_{-2g(h)(\hat{r} + \hat{r} \wedge \kappa)}^{2g(h)(\hat{r} \wedge \kappa - u)} u \left[ m((2g(h))^{-1}(\hat{r} \wedge \kappa - u)) - m((2g(h))^{-1}(\hat{r} \wedge \kappa + u)) \right] \, du \\
= (\hat{r} - \hat{r} \wedge \kappa)(\mu - \mu_{g(h)^{-1}(\hat{r} \wedge \kappa)})(\{ |z| \leq l', |z + g(h)^{-1}(\hat{x} - \hat{y})_n| \leq l' \})
\]
Lemma 5.1. According to (3.11), it is clear that $R = \tilde{r}$. Therefore, we have $\beta(x, y) \leq \tilde{r} - r$ for all $x, y \in \mathbb{R}$. This proves the first assertion.

Next, we turn to the proof of the second assertion. From condition (c4) and the fact that $\lim_{s \to \infty} m(s) = 0$, we can choose $\varepsilon \in (0, 1/4)$ small enough and $\gamma > 1$ large enough so that $m(\varepsilon) - m(\varepsilon/2 - \varepsilon/4) \geq m(\varepsilon)/2 > 0$. Since $h/g(h) \leq \varepsilon (2L\mathcal{R})^{-1}$ and $l' \geq g(h)^{-1}(1 + h\mathcal{L})\mathcal{R} + \gamma/2 + 1$, for all $r \in (0, \mathcal{R})$, 

$$[-\gamma/2 - 1, 0] \subset [-l' + g(h)^{-1}(\tilde{r} \land \kappa), l'] = \{z \in \mathbb{R} : |z| \leq l', |z + g(h)^{-1}(\tilde{x} - \tilde{y})| \leq l'\}$$

and

$$[(2g(h))^{-1}(\tilde{r} - r) - \gamma/2, (2g(h))^{-1}(\tilde{r} - r) + \varepsilon/2] \subset [-\gamma/2 - \varepsilon/4, -\varepsilon/4] \subset [-\gamma/2 - 1, 0].$$

Then, following the proof of Lemma 3.12, we have

$$\alpha_{g(h)}(x, y) = \frac{1}{2} \mathbb{E}_{x, y} [ (R - r)^2 \mathbb{1}_{\{R \leq r + \gamma g(h)\}}]$$

$$\geq \frac{\varepsilon^2 g(h)^2}{2} \int_{\{r + \varepsilon g(h) \leq |\tilde{r} - 2g(h)\} \leq r + \gamma g(h)\} (\mu - \mu_{g(h)^{-1}(\tilde{r} \land \kappa)})(dz)$$

$$= \frac{\varepsilon^2 g(h)^2}{2} \int_{(2g(h))^{-1}(\tilde{r} - r) - \varepsilon/2}^{(2g(h))^{-1}(\tilde{r} - r) + \varepsilon/2} (m(z) - m(z - g(h)^{-1}(\tilde{r} \land \kappa)))(dz)$$

$$\geq e^* g(h) (\tilde{r} \land \kappa).$$

The proof is complete. \hfill \square

5. APPENDIX

Lemma 5.1. $(X, Y), P_{x, y}$ defined by (3.6) is a Markov coupling of the chain $X$.

Proof. Fix $x, y \in \mathbb{R}^d$, and recall that $\hat{x} = x + hB(x)$ and $\hat{y} = y + hB(y)$. By (3.6), it suffices to prove that the distribution of the random variable $g(h)^{-1}(Y - \hat{y})$ is $\mu$. Indeed, for any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$P_{x, y}(g(h)^{-1}(Y - \hat{y}) \in A) = \mathbb{1}_{\{A - g(h)^{-1}(\hat{x} - \hat{y}) \}}(A - g(h)^{-1}(\hat{x} - \hat{y})) + \frac{1}{2} \mu_{g(h)^{-1}(\hat{y} - \hat{x})}(A - g(h)^{-1}(\hat{y} - \hat{x}))$$

$$+ \left( \mu - \frac{1}{2} \mu_{g(h)^{-1}(\hat{y} - \hat{x})} - \frac{1}{2} \mu_{g(h)^{-1}(\hat{x} - \hat{y})} \right)(A).$$

According to

$$(\delta_{-v} \ast \mu_v)(dz) = \mu_{-v}(dz) \quad \text{for all } v \in \mathbb{R}^d;$$

we find that

$$\mu_{g(h)^{-1}(\hat{y} - \hat{x})}(A - g(h)^{-1}(\hat{x} - \hat{y})) = \delta_{g(h)^{-1}(\hat{x} - \hat{y})} \ast \mu_{g(h)^{-1}(\hat{y} - \hat{x})}(A) = \mu_{g(h)^{-1}(\hat{x} - \hat{y})}(A).$$

Hence, $g(h)^{-1}(X - Y) = \hat{y} - \hat{x}$ almost surely, and $X - Y$ is a Markov coupling of $X$. \hfill \square
Similarly,
\[ \mu_{g(h)^{-1} (\hat{x} - \hat{y})}(A - g(h)^{-1}(\hat{y} - \hat{x})) = \mu_{g(h)^{-1} (\hat{x} - \hat{y})}(A). \]
Hence, \( P_{x,y}(g(h)^{-1}(Y^{-1} - \hat{y}) \in A) = \mu(A) \) for all \( A \in \mathcal{B}(\mathbb{R}^d) \). This completes the proof. \( \square \)

**Lemma 5.2.** \( ((X, Y^{(1)}), P_{x,y}) \) defined by (3.7) is a Markov coupling of the chain \( X \).

**Proof.** Similarly as in the proof of Lemma 5.1, we only need to verify that the distribution of the random variable \( g(h)^{-1}(Y^{(1)} - \hat{y}) \) is \( \mu \). For any \( A \in \mathcal{B}(\mathbb{R}^d) \), by (3.7),
\[ P_{x,y}(g(h)^{-1}(Y^{(1)} - \hat{y}) \in A) = \mu_{g(h)^{-1}(\hat{x} - \hat{y})}\mu(A - g(h)^{-1}(\hat{y} - \hat{x})) + (\mu - \mu_{g(h)^{-1}(\hat{x} - \hat{y})})(R_{\hat{x},\hat{y}}^{-1}(A)), \]
where \( R_{\hat{x},\hat{y}}^{-1}(A) = \{ z \in \mathbb{R}^d : R_{\hat{x},\hat{y}}(z) \in A \} \). Since \( R_{\hat{x},\hat{y}}(z) = R_{\hat{x},\hat{y}}^{-1}(z) \), \( |R_{\hat{x},\hat{y}}(z)| = |z| \) and \( m(z) = m(|z|) \) for all \( z \in \mathbb{R}^d \), we have \( \mu(R_{\hat{x},\hat{y}}^{-1}(A)) = \mu(R_{\hat{x},\hat{y}}(A)) = \mu(A) \). Moreover,
\[ \mu_{g(h)^{-1}(\hat{x} - \hat{y})}(A) = \int_A m(R_{\hat{x},\hat{y}}^{-1}(z)) dA = \int_A m(z) \cup m(z - g(h)^{-1}(\hat{y} - \hat{x})) dA = \mu_{g(h)^{-1}(\hat{x} - \hat{y})}(A), \]
where in the third equality we used the fact that
\[ (3.3) \quad R_{\hat{x},\hat{y}}(g(h)^{-1}(\hat{x} - \hat{y})) = g(h)^{-1}(\hat{x} - \hat{y}). \]
Therefore, \( P_{x,y}(g(h)^{-1}(Y^{(1)} - \hat{y}) \in A) = \mu(A) \) by (5.2).

**Lemma 5.3.** \( ((X, Y^{(2)}), P_{x,y}) \) defined by (3.10) and (3.11) is a Markov coupling of the chain \( X \).

**Proof.** Fix \( h, s, t, \kappa > 0 \). When \( r = |x - y| \in (s, \infty) \), \( ((X, Y^{(2)}), P_{x,y}) \) defined by (3.11) is a synchronous coupling. Hence, we only need to consider the case where \( r \in (0, s) \).

For any \( A \in \mathcal{B}(\mathbb{R}^d) \), by (3.10),
\[ P_{x,y}(g(h)^{-1}(Y^{(2)} - \hat{y}) \in A) = \int_{A - g(h)^{-1}(\hat{x} - \hat{y})} 1_{\{|z| \leq t', |z + g(h)^{-1}(\hat{x} - \hat{y})| \leq t\}} \mu_{g(h)^{-1}(\hat{x} - \hat{y})}(dz) + \int_{R_{\hat{x},\hat{y}}^{-1}(A)} 1_{\{|z| \leq t', |z + g(h)^{-1}(\hat{x} - \hat{y})| \leq t\}} (\mu - \mu_{g(h)^{-1}(\hat{x} - \hat{y})})(dz) + \int_{A} 1_{\{|z| > t' \text{ or } |z + g(h)^{-1}(\hat{x} - \hat{y})| > t\}} \mu(dz) =: (I) + (II) + (III). \]

It follows from (3.1) that
\[ (I) = \int_A 1_{\{|u - g(h)^{-1}(\hat{x} - \hat{y})| \leq t', |u| \leq t\}} \mu_{g(h)^{-1}(\hat{x} - \hat{y})}(du). \]
On the other hand, due to the rotational invariance of \( \mu \), the properties of \( R_{\hat{x},\hat{y}} \) and (3.3), we have
\[ (II) = \int_A 1_{\{|R_{\hat{x},\hat{y}}^{-1}(u)| \leq t', |R_{\hat{x},\hat{y}}^{-1}(u) + g(h)^{-1}(\hat{x} - \hat{y})| \leq t\}} (\mu - \mu_{g(h)^{-1}(\hat{x} - \hat{y})})(dR_{\hat{x},\hat{y}}^{-1}(u)) = \int_A 1_{\{|u| \leq t', |u - g(h)^{-1}(\hat{x} - \hat{y})| \leq t\}} \mu(dR_{\hat{x},\hat{y}}^{-1}(u)) \]
\[ = \int_A 1_{\{|u| \leq t', |u - g(h)^{-1}(\hat{x} - \hat{y})| \leq t\}} [\mu(dR_{\hat{x},\hat{y}}^{-1}(u)) - \mu(d(R_{\hat{x},\hat{y}}^{-1}(u) - g(h)^{-1}(\hat{y} - \hat{x})))]. \]
\[ = \int_A 1_{\{|u| \leq t', |u - g(h)^{-1}(\hat{x} - \hat{y})| \leq t\}} \mu(du) - \int_A 1_{\{|u| \leq t', |u - g(h)^{-1}(\hat{x} - \hat{y})| \leq t\}} \mu_{g(h)^{-1}(\hat{x} - \hat{y})}(du). \]
Therefore,
\[ P_{x,y}(g(h)^{-1}(Y(2) - \hat{y}) \in A) = \mu(A) \text{ for all } A \in \mathcal{B}(\mathbb{R}^d). \]

The proof is complete. \qed

**Acknowledgement.** We thank the two anonymous referees for their helpful comments and valuable suggestions that have led to significant improvements of the presentation in the article. The research of Lu-Jing Huang is supported by the National Natural Science Foundation of China (No. 11901096) and the National Natural Science Foundation of Fujian (No. 2020J05036). The research of Jian Wang is supported by the National Natural Science Foundation of China (Nos. 11831014 and 12071076), and the Education and Research Support Program for Fujian Provincial Agencies.

**REFERENCES**

[1] Bou-Rabee, Z., Eb erle, A. and Zimmer, R.: Coupling and convergence for Hamiltonian Monte Carlo, *Ann. Appl. Probab.* 30 (2020), 1200–1250.
[2] Bou-Rabee, Z. and Schuh, K.: Convergence of unadjusted Hamiltonian Monte Carlo for mean-field models, arXiv:200908735.
[3] Chen, M.-F.: *From Markov Chains to Non-Equilibrium Particle Systems*, Second ed., World Scientific, 2004.
[4] Chen, M.-F.: *Eigenvalues, Inequalities, and Ergodic Theory*, Springer-Verlag London, Ltd., London, 2005.
[5] Chen, M.-F. and Wang, F.-Y.: Application of coupling method to the first eigenvalue on manifold, *Sci. China Ser. A* 37 (1994), 114.
[6] Dareiotis, K., Kumar, C. and Sabanis, S.: On tamed Euler approximations of SDEs driven by Lévy noise with applications to delay equations, *SIAM J. Numer. Anal.*, 54 (2016), 1840–1872.
[7] De Bortoli, V. and Durmus, A.: Convergence of diffusions and their discretizations: from continuous to discrete processes and back, arXiv:1904.00808.
[8] Djellout, H., Guillin, A. and Wu, L.: Transportation cost-information inequalities and applications to random dynamical systems and diffusions, *Ann. Probab.*, 32 (2004), 2702–2723.
[9] Dobrushin, R.L.: Prescribing a system of random variables by conditional distributions, *Theory Probab. Appl.*, 15 (1970), 458–486.
[10] Durmus, A., Eberle, A., Enfroy, A., Guillin, A. and Moumarché, P.: Discrete sticky couplings of functional autoregressive processes, arXiv:2104.06771.
[11] Eberle, A.: Reflection coupling and Wasserstein contractivity without convexity, *C. R. Math. Acad. Sci. Paris*, 349 (2011), 1101–1104.
[12] Eberle, A.: Reflection couplings and contraction rates for diffusions, *Probab. Theory Related Fields*, 166 (2016), 851–886.
[13] Eberle, A. and Majka, M.B.: Quantitative contraction rates for Markov chains on general state spaces, *Electron. J. Probab.*, 24 (2019), 1–36.
[14] Grafakos, L. and Teschl, G.: On Fourier transforms of radial functions and distributions, *J. Fourier Anal. Appl.*, 19 (2013), 167–179.
[15] Hairer, M. and Mattingly, J.C.: Yet another look at Harris’ ergodic theorem for Markov chains. In: *Seminar on Stochastic Analysis, Random Fields and Applications VI*, pages 109–117, 2011, Springer.
[16] Hairer, M., Stuart, A.M. and Vollmer, S.J.: Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions, *Ann. Appl. Probab.*, 24 (2014), 2455–2490.
[17] Huang, L.-J., Majka, M.B. and Wang, J.: Approximation of heavy-tailed distributions via stable-driven SDEs, *Bernoulli*, 27 (2021), 2040–2068.
[18] Johndrow, J.E. and Mattingly, C.: Error bounds for approximations of Markov chains used in Bayesian sampling, arXiv:1711.05382.
[19] Joulin, A. and Ollivier, Y.: Curvature, concentration and error estimates for Markov chain Monte Carlo, *Ann. Probab.*, 38 (2010), 2418–2442.
[20] Kühn, F. and Schilling, R. L.: Strong convergence of the Euler-Maruyama approximation for a class of Lévy-driven SDEs, *Stoch. Proc. Appl.*, 129 (2019), 2654–2680.
[21] Liang, M. and Wang, J.: Gradient estimates and ergodicity for SDEs driven by multiplicative Lévy noises via coupling, *Stochastic Process. Appl.*, 130 (2020), 3053–3094.
[22] Luo, D. and Wang, J.: Exponential convergence in $L^p$-Wasserstein distance for diffusion processes without uniformly dissipative drift, *Math. Nachr.*, 289 (2016), 1909–1926.
[23] Luo, D. and Wang, J.: Refined basic couplings and Wasserstein-type distances for SDEs with Lévy noises, *Stochastic Process. Appl.*, 129 (2019), 3129–3173.
[24] Majka, M.B.: Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, *Stochastic Process. Appl.*, 127 (2017), 4083–4125.
[25] Majka, M.B., Mijatović, A. and Szpruch, L.: Non-asymptotic bounds for sampling algorithms without log-concavity, *Ann. Appl. Probab.*, 30 (2020), 1534–1581.

[26] Meyn, S.P. and Tweedie, R.T.: Stability of Markovian processes I: discrete time chains, *Adv. Appl. Probab.*, 24 (1992), 542–574.

[27] Nguyen, T.H., Şimşekli, U. and Richard, G.: Non-asymptotic analysis of fractional Langevin Monte Carlo for non-convex optimization. In: Chaudhuri, K. and Salakhutdinov, R. editors, *Proceedings of the 36th International Conference on Machine Learning*, vol. 97 of *Proceedings of Machine Learning Research*, pages 4810–4819, Long Beach, California, USA, June 09–15, 2019.

[28] Ollivier, Y.: Ricci curvature of Markov chains on metric spaces, *J. Funct. Anal.*, 256 (2009), 810–864.

[29] Paulin, D.: Mixing and concentration by Ricci curvature, *J. Funct. Anal.*, 270 (2016), 1623–1662.

[30] Pillai, N.S. and Smith, A.: Ergodicity of approximate MCMC chains with applications to large data sets, arXiv:1405.0182

[31] Qin, Q. and Hobert, J.P.: Geometric convergence bounds for Markov chains in Wasserstein distance based on generalized drift and contraction conditions, to appear in *Annales de l’Institut Henri Poincaré*

[32] Rudolf, D. and Schweizer, N.: Perturbation theory for Markov chains via Wasserstein distance, *Bernoulli*, 24 (2018), 2610–2639.

[33] Şimşekli, U.: Fractional Langevin Monte Carlo: exploring Lévy driven stochastic differential equations for Markov chain Monte Carlo. In: Precup, D. and Teh, Y.W. editors, *Proceedings of the 34th International Conference on Machine Learning*, vol. 70 of *Proceedings of Machine Learning Research*, pages 3200–3209, International Convention Centre, Sydney, Australia, August 06–11, 2017.

[34] Wang, J.: $L_p$-Wasserstein distance for stochastic differential equations driven by Lévy processes, *Bernoulli*, 22 (2016), 1598–1616.

[35] Ye, N. and Zhu, Z.: Stochastic fractional Hamiltonian Monte Carlo. In: Lang, J. editor, *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 3019–3023, International Joint Conference on Artificial Intelligence, Stockholm, Sweden, July 13–19, 2018.