Bergische Universität Wuppertal

Fachbereich Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM January 05, 2019

Martin Friesen and Peng Jin and Barbara Rüdiger

Stochastic equation and exponential ergodicity in Wasserstein distances for affine processes

January 23, 2019

http://www.math.uni-wuppertal.de

Stochastic equation and exponential ergodicity in Wasserstein distances for affine processes

Martin Friesen*
Peng Jin†
Barbara Rüdiger‡

January 23, 2019

Abstract: This work is devoted to the study of conservative affine processes on the canonical state space $D = \mathbb{R}_+^m \times \mathbb{R}^n$, where $m + n > 0$. We show that each affine process can be obtained as the pathwise unique strong solution to a stochastic equation driven by Brownian motions and Poisson random measures. This extends and unifies known results for multi-type CBI processes [BLP15a] and affine diffusions [FM09]. In the second part of this work we study the long-time behavior of affine processes, i.e., we show that under first moment condition on the state-dependent and log-moment conditions on the state-independent jump measures, respectively, each subcritical affine process is exponentially ergodic in a suitably chosen Wasserstein distance. Moments of affine processes are studied as well.

AMS Subject Classification: 37A25; 60H10; 60J25
Keywords: affine process; ergodicity; Wasserstein distance; coupling; stochastic differential equation

1 Introduction and statement of the result

1.1 General introduction

An affine process is a time-homogeneous Markov processes $(X_t)_{t\geq 0}$ whose characteristic function satisfies

$$\mathbb{E}_x\left(e^{i\langle u, X_t\rangle}\right) = \exp\left(\phi(t, iu) + \langle x, \psi(t, iu)\rangle\right),$$

where $t \geq 0$ is the time and $X_0 = x$ the starting point of the process. The general theory of affine processes, including a full characterization on the canonical state space $D = \mathbb{R}_+^m \times \mathbb{R}^n$ where

*Fakultät für Mathematik und Naturwissenschaften, Bergische Universität Wuppertal, Gaußstraße 20, 42119 Wuppertal, Germany, friesen@math.uni-wuppertal.de
†Department of Mathematics, Shantou University, Shantou, Guangdong 515063, China, pjin@stu.edu.cn
‡Fakultät für Mathematik und Naturwissenschaften, Bergische Universität Wuppertal, Gaußstraße 20, 42119 Wuppertal, Germany, ruediger@uni-wuppertal.de

$m,n \in \mathbb{N}_0$ and $m+n > 0$, was discussed in [DFS03]. In particular, it is shown that the functions ϕ and ψ should satisfy certain generalized Riccati equations. Common applications of affine processes in mathematical finance are interest rate models (e.g., the Cox-Ingersoll-Ross, Vašiček or general affine term structure short rate models), option pricing (e.g., the Heston model) and credit risk models, see also [Alf15] and the references therein. After [DFS03], the mathematical theory of affine processes was developed in various directions. Regularity of affine processes was studied in [KRST11] and [KRST13]. Based on a Hörmander-type condition, existence and smoothness of transition densities were obtained in [FMS13]. Exponential moments for affine processes were studied in [JKX12] and [KRM15]. The theory of affine diffusions, i.e., processes without jumps, was developed in [FM09], while its application to large deviations for affine diffusions was studied in [KK14]. The possibility to obtain affine processes as multi-parameter time changes of Lévy processes was recently discussed in [CPGUB17]. It is worthwhile to mention that the above list is, by far, not complete. For further references and additional details on the general theory of affine processes we refer to the book [Alf15].

Below we describe two important sub-classes of affine processes. *Continuous-state branching processes with immigration* (shorted as CBI processes) are affine processes with state space $D = \mathbb{R}_+^m$. Such processes have been first introduced in 1958 by Jiřina [Ji58] and then studied in [Wat69, KW71, SW73], where it was also shown that CBI processes arise as scaling limits of Galton-Watson processes (see also [Li06]). For a general introduction to CBI processes (and more generally superprocesses) we refer to [Li11]. Various properties of one-dimensional CBI processes were studied in [Gre74, FFS85, CPGUB13, KRM12, FUB14, DFM14] and [CLP18]. For results applicable in general dimension we refer to [BLP15a], [BLP16] and [FJR18]. Another important class of affine processes corresponds to the state space $D = \mathbb{R}^n$ and is consisted of processes of Ornstein-Uhlenbeck (OU) type. These processes include also Lévy processes as a particular case.

In this work we study the long-time behavior of affine processes. Such analysis includes existence, uniqueness and convergence to the invariant distribution. It is worthwhile to mention that previous results in this direction have been used for parameter estimation of particular affine models, see, e.g., [BDLP13], [BDLP14], [LM15] and [BBAKP18]. The long-time behavior of OU-type processes (i.e. $D = \mathbb{R}^n$) was studied in [SY84], while exponential ergodicity in the total variation distance was obtained in [Wan12]. The latter result was based on suitable coupling techniques. Existence, uniqueness and some properties of the invariant distribution was studied in [Li11] for one-dimensional CBI processes, see also [Pin72] and [KRM12]. A general ergodicity result for one-dimensional subcritical CBI processes was obtained in [LM15]. Supercritical multi-type CBI processes have been also subject of recent developments in [KPR17, BPP18b, BPP18a], while multi-dimensional subcritical affine processes on cones were considered in [MSV18]. Particular affine models have been considered in [BDLP13, JMRT13, JRT16, JKR17b]. The first result for the stationarity of general affine processes (without convergence rate) was recently obtained in [JKR18]. Independently, under additional conditions, ergodicity in total variation distance was also recently obtained in [GZ18]. In this work we continue our research and prove, under the same conditions as in [JKR18], that affine processes converge exponentially fast to the invariant distribution in a Wasserstein distance. While mathematical techniques used in [JKR18] are mainly of analytical nature relying on a detailed study of the generalized Riccati equations, this work provides a probabilistic approach to ergodicity of affine processes. It is worthwhile to mention that it is the first time (up to our knowledge) a convergence rate is obtained solely under a log-moment condition on the state-independent jump measure. This seems to be new even for one-dimensional CBI processes.

In order to prove our result we first show that each affine process can be obtained as a unique strong solution to a stochastic equation driven by Brownian motions and Poisson random measures. Although such a result is not surprising, in the literature it was only stated for multi-type CBI processes (see [Ma13, BLP15a, BPP18a]) and affine diffusions (see [FM09]). Hence we extend and unify these known results now to general conservative affine processes. The precise formulations and proofs are given in Section 3 and Section 4.

1.2 Affine processes

Let us describe affine processes in more detail. For $m, n \in \mathbb{N}_0$ let $d = n + m$, and suppose that $d > 0$. In this work we study affine processes on the canonical state space $D = \mathbb{R}_+^m \times \mathbb{R}^n$. Let

$$I = \{1, \dots, m\}, \quad J = \{m+1, \dots, d\}.$$

If $x \in D$, then let $x_I = (x_i)_{i \in I}$ and $x_J = (x_j)_{j \in J}$. Denote by $\mathbb{R}^{d \times d}$ the space of $d \times d$-matrices. For $A \in \mathbb{R}^{d \times d}$ we write

$$A = \begin{pmatrix} A_{II} & A_{IJ} \\ A_{JI} & A_{JJ} \end{pmatrix},$$

where $A_{II} = (a_{ij})_{i,j \in I}$, $A_{IJ} = (a_{ij})_{i \in I,\, j \in J}$, $A_{JI} = (a_{ij})_{i \in J,\, j \in I}$, and $A_{JJ} = (a_{ij})_{i,j \in J}$. Denote by S_d^+ the space of symmetric and positive semidefinite $d \times d$-matrices. Finally, let δ_{kl}, $k, l \in \{1, \dots, d\}$, stand for the Kronecker-Delta.

Definition 1.1. *We call a tuple $(a, \alpha, b, \beta, m, \mu)$ admissible parameters, if they satisfy the following conditions:*

(i) $a \in S_d^+$ *with* $a_{II} = 0$, $a_{IJ} = 0$ *and* $a_{JI} = 0$.

(ii) $\alpha = (\alpha_1, \dots, \alpha_m)$ *with* $\alpha_i = (\alpha_{i,kl})_{1 \le k,l \le d} \in S_d^+$ *and* $\alpha_{i,kl} = 0$ *if* $k \in I \backslash \{i\}$ *or* $l \in I \backslash \{i\}$.

(iii) $b \in D$.

(iv) $\beta \in \mathbb{R}^{d \times d}$ *is such that* $\beta_{ki} - \int_D \xi_k \mu_i(d\xi) \ge 0$ *for all* $i \in I$ *and* $k \in I \backslash \{i\}$, *and* $\beta_{IJ} = 0$.

(v) m *is a Borel measure on* D *such that* $m(\{0\}) = 0$ *and*

$$\int_D \left(1 \wedge |\xi|^2 + \sum_{i \in I} (1 \wedge \xi_i) \right) m(d\xi) < \infty.$$

(vi) $\mu = (\mu_1, \dots, \mu_m)$ *where* $\mu_1, \dots, \mu_m$ *are Borel measures on* D *such that*

$$\mu_i(\{0\}) = 0, \qquad \int_D \left(|\xi| \wedge |\xi|^2 + \sum_{k \in I \backslash \{i\}} \xi_k \right) \mu_i(d\xi) < \infty, \quad i \in I.$$

In contrast to [DFS03], we do not consider killing for affine processes and, moreover, we suppose that $\mu_1, \ldots, \mu_m$ integrate $\mathbb{1}_{\{|\xi|>1\}}|\xi|$, i.e., the first moment for big jumps is finite. It is well-known that without killing and under first moment condition for the big jumps of $\mu_1, \ldots, \mu_d$, the corresponding affine process (introduced below) is conservative (see [DFS03, Lemma 9.2]). In this paper we work with Definition 1.1 and thus restrict our study to conservative affine processes. In order to simplify the notation, we have also set $m(\{0\}) = 0$ and $\mu_i(\{0\}) = 0$, for $i \in I$. Hence all integrals with respect to the measures $\mu_1, \ldots, \mu_d, m$ can be taken over D instead of $D\backslash\{0\}$.

Denote by $B_b(D)$ the Banach space of bounded measurable functions over D. This space is equipped with the supremum norm $\|f\|_\infty = \sup_{x\in D} |f(x)|$. Define

$$\mathcal{U} = \mathbb{C}^m_{\leq 0} \times i\mathbb{R}^n = \{u = (u_1, u_2) \in \mathbb{C}^m \times \mathbb{C}^n \mid \mathrm{Re}(u_1) \leq 0, \quad \mathrm{Re}(u_2) = 0\}.$$

Note that $D \ni x \longmapsto e^{\langle u,x\rangle}$ is bounded for any $u \in \mathcal{U}$. Here $\langle\cdot,\cdot\rangle$ denotes the Euclidean scalar product on $\mathbb{R}^d$. By abuse of notation, we later also use $\langle\cdot,\cdot\rangle$ for the scalar product on $\mathbb{R}^m$ or $\mathbb{R}^n$. The following is due to [DFS03].

Theorem 1.2. *Let $(a, \alpha, b, \beta, m, \mu)$ be admissible parameters. Then there exists a unique conservative Feller semigroup $(P_t)_{t\geq 0}$ on $B_b(D)$ with generator $(L, D(L))$ such that $C_c^2(D) \subset D(L)$ and, for $f \in C_c^2(D)$ and $x \in D$,*

$$\begin{aligned}(Lf)(x) &= \langle b + \beta x, \nabla f(x)\rangle + \sum_{k,l=1}^{d}\left(a_{kl} + \sum_{i=1}^{m}\alpha_{i,kl}x_i\right)\frac{\partial^2 f(x)}{\partial x_k \partial x_l} \\ &\quad + \int_D \left(f(x+\xi) - f(x) - \langle \xi_J, \nabla_J f(x)\rangle \mathbb{1}_{\{|\xi|\leq 1\}}\right) m(d\xi) \\ &\quad + \sum_{i=1}^{m} x_i \int_D \left(f(x+\xi) - f(x) - \langle \xi, \nabla f(x)\rangle\right)\mu_i(d\xi),\end{aligned}$$

where $\nabla_J = (\frac{\partial}{\partial x_j})_{j\in J}$. Moreover, $C_c^\infty(D)$ is a core for the generator. Let $P_t(x, dx')$ be the transition probabilities. Then

$$\int_D e^{\langle u,x'\rangle} P_t(x, dx') = \exp\left(\phi(t,u) + \langle x, \psi(t,u)\rangle\right), \qquad u \in \mathcal{U}, \tag{1.1}$$

where $\phi : \mathbb{R}_+ \times \mathcal{U} \longrightarrow \mathbb{C}$ and $\psi : \mathbb{R}_+ \times \mathcal{U} \longrightarrow \mathbb{C}^d$ are uniquely determined by the generalized Riccati differential equations: for $u = (u_1, u_2) \in \mathbb{C}^m_{\leq 0} \times i\mathbb{R}^n$,

$$\begin{aligned}\partial_t \phi(t,u) &= F(\phi(t,u)), \quad \phi(0,u) = 0, \\ \partial_t \psi_I(t,u) &= R(\psi_I(t,u), e^{t\beta_{JJ}^\top}u_2), \quad \psi_I(0,u) = u_1, \\ \psi_J(t,u) &= e^{t\beta_{JJ}^\top}u_2,\end{aligned} \tag{1.2}$$

and F, R are of Lévy-Khintchine form

$$F(u) = \langle u, au\rangle + \langle b, u\rangle + \int_D \left(e^{\langle u,\xi\rangle} - 1 - \mathbb{1}_{\{|\xi|\leq 1\}}\langle \xi_J, u_J\rangle \right) m(d\xi),$$

$$R_i(u) = \langle u, \alpha_i u\rangle + \sum_{k=1}^{d} \beta_{ki}u_k + \int_D \left(e^{\langle u,\xi\rangle} - 1 - \langle u, \xi\rangle \right) \mu_i(d\xi), \qquad i \in I.$$

Consequently, there exists a unique Feller process X with generator L. This process is called affine process with admissible parameters $(a, \alpha, b, \beta, m, \mu)$.

Remark 1.3. *Let $(a, \alpha, b, \beta, m, \mu)$ be admissible parameters. According to [DFS03, Lemma 10.1 and Lemma 10.2], the martingale problem with generator L and domain $C_c^2(D)$ is well-posed in the Skorokhod space over D equipped with the usual Skorokhod topology. Hence, we can characterise an affine process with admissible parameters $(a, \alpha, b, \beta, m, \mu)$ as the unique solution to the martingale problem with generator L and domain $C_c^2(D)$. In any case, it can be constructed as a Markov process on the Skorokhod space over D.*

1.3 Ergodicity in Wasserstein distance for affine processes

Let $\mathcal{P}(D)$ be the space of all Borel probability measures over D. By abuse of notation, we extend the transition semigroup $(P_t)_{t\geq 0}$ (given by Theorem 1.2) onto $\mathcal{P}(D)$ via

$$(P_t\rho)(dx) = \int_D P_t(\widetilde{x}, dx)\rho(d\widetilde{x}), \qquad t \geq 0, \quad \rho \in \mathcal{P}(D). \tag{1.3}$$

Then $P_t\rho$ describes the distribution of the affine process at time $t \geq 0$ such that it has at initial time $t = 0$ law ρ. Note that $P_t\delta_x = P_t(x, \cdot)$, and $(P_t)_{t\geq 0}$ is a semigroup on $\mathcal{P}(D)$ in the sense that $P_{t+s}\rho = P_tP_s\rho$, for any $t, s \geq 0$ and $\rho \in \mathcal{P}(D)$. Such semigroup property is simply a compact notation for the Chapman-Kolmogorov equations satisfied by $P_t(x, \cdot)$. Since the martingale problem with generator L and domain $C_c^\infty(D)$ is well-posed, and $C_c^\infty(D) \subset D(L)$ is a core (see Theorem 1.2 and Remark 1.3), it follows from [EK86, Proposition 9.2] that, for some given $\pi \in \mathcal{P}(D)$, the following properties are equivalent:

(i) $P_t\pi = \pi$, for all $t \geq 0$.

(ii) $\int_D (Lf)(x)\pi(dx) = 0$, for all $f \in C_c^\infty(D)$.

(iii) $\int_D (P_tf)(x)\pi(dx) = \int_D f(x)\pi(dx)$, for all $t \geq 0$ and all $f \in B(D)$.

A distribution $\pi \in \mathcal{P}(D)$ which satisfies one of these properties (i) – (iii) is called invariant distribution for the semigroup $(P_t)_{t\geq 0}$. In this work we will prove that, under some appropriate assumptions, $(P_t)_{t\geq 0}$ has a unique invariant distribution π, this distribution has some finite log-moment and, moreover, $P_t(x, \cdot) \longrightarrow \pi$ with exponential rate. For this purpose we use the Wasserstein distance on $\mathcal{P}(D)$ introduced below. Given $\rho, \widetilde{\rho} \in \mathcal{P}(D)$, a coupling H of $(\rho, \widetilde{\rho})$

is a Borel probability measure on $D \times D$ which has marginals ρ and $\widetilde{\rho}$, respectively, i.e., for $f, g \in B(D)$ it holds that

$$\int_{D\times D} (f(x) + g(\widetilde{x}))\, H(dx, d\widetilde{x}) = \int_D f(x)\rho(dx) + \int_D g(x)\widetilde{\rho}(dx).$$

Denote by $\mathcal{H}(\rho, \widetilde{\rho})$ the collection of all such couplings. Let us now introduce two different metrics on D as follows:

(a) Define, for $\kappa \in (0,1]$, $d_\kappa(x, \widetilde{x}) = \left(\mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2} + |x - \widetilde{x}|\right)^\kappa$, $x = (y, z), \widetilde{x} = (\widetilde{y}, \widetilde{z}) \in \mathbb{R}^m_+ \times \mathbb{R}^n$, and let

$$\mathcal{P}_{d_\kappa}(D) = \left\{ \rho \in \mathcal{P}(D) \mid \int_D |x|^\kappa \rho(dx) < \infty \right\}.$$

(b) Introduce $d_{\log}(x, \widetilde{x}) = \log(1 + \mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2} + |x - \widetilde{x}|)$, $x = (y, z), \widetilde{x} = (\widetilde{y}, \widetilde{z}) \in \mathbb{R}^m_+ \times \mathbb{R}^n$, and let

$$\mathcal{P}_{d_{\log}}(D) = \left\{ \rho \in \mathcal{P}(D) \mid \int_D \log(1 + |x|)\rho(dx) < \infty \right\}.$$

Let $d \in \{d_{\log}, d_\kappa\}$. The Wasserstein distance on $\mathcal{P}_d(D)$ is defined by

$$W_d(\rho, \widetilde{\rho}) = \inf\left\{ \int_{D\times D} d(x, \widetilde{x}) H(dx, d\widetilde{x}) \mid H \in \mathcal{H}(\rho, \widetilde{\rho}) \right\}. \tag{1.4}$$

The appearance of the additional factor $\mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2}$ is purely technical and is a consequence of the estimates proved in Section 6. By general theory of Wasserstein distances we see that $(\mathcal{P}_d(D), W_d)$ is a complete seperable metric space, see, e.g., [Vil09, Theorem 6.18]. Convergence with respect to this distances is explained in the following remark, see also [Vil09, Theorem 6.9].

Remark 1.4. *Let $d \in \{d_{\log}, d_\kappa\}$, $(\rho_n)_{n\in\mathbb{N}} \subset \mathcal{P}_d(D)$ and $\rho \in \mathcal{P}_d(D)$. The following are equivalent*

(i) $W_d(\rho_n, \rho) \longrightarrow 0$ *as* $n \to \infty$.

(ii) For each continuous function $f : D \longrightarrow \mathbb{R}$ *with* $|f(x)| \le C_f(1 + d(x, 0))$, *it holds that*

$$\int_D f(x)\rho_n(dx) \longrightarrow \int_D f(x)\rho(dx), \qquad n \to \infty.$$

(iii) $\rho_n \longrightarrow \rho$ *weakly as* $n \to \infty$, *and*

$$\int_D d(x, 0)\rho_n(dx) \longrightarrow \int_D d(x, 0)\rho(dx), \qquad n \to \infty.$$

(iv) $\rho_n \longrightarrow \rho$ *weakly as* $n \to \infty$*, and*

$$\lim_{R\to\infty} \limsup_{n\to\infty} \int\limits_D d(x,0)\mathbb{1}_{\{d(x,0)\geq R\}}\rho_n(dx) = 0.$$

For simplicity of notation, we let $\mathcal{P}_\kappa(D) = \mathcal{P}_{d_\kappa}(D)$, $\mathcal{P}_{\log}(D) = \mathcal{P}_{d_{\log}}(D)$, $W_\kappa = W_{d_\kappa}$, and $W_{\log} = W_{d_{\log}}$. Then it is easy to see that $\mathcal{P}_\kappa(D) \subset \mathcal{P}_{\log}(D)$ and $W_{\log} \leq C_\kappa W_\kappa$, for some constant $C_\kappa > 0$, i.e., W_κ is stronger then $W_{\log}$. The following is our main result.

Theorem 1.5. *Let* $(a, \alpha, b, \beta, m, \mu)$ *be admissible parameters. Suppose that* β *has only eigenvalues with negative real parts, and*

$$\int\limits_{|\xi|>1} \log(|\xi|)m(d\xi) < \infty. \tag{1.5}$$

Then $(P_t)_{t\geq 0}$ *has a unique invariant distribution* π *and the following assertions hold:*

(a) $\pi \in \mathcal{P}_{\log}(D)$ *and there exist constants* $K, \delta > 0$ *such that, for all* $\rho \in \mathcal{P}_{\log}(D)$,

$$W_{\log}(P_t\rho, \pi) \leq K \min\left\{e^{-\delta t}, W_{\log}(\rho,\pi)\right\} + Ke^{-\delta t}W_{\log}(\rho,\pi), \qquad t \geq 0. \tag{1.6}$$

(b) If there exists $\kappa \in (0, 1]$ *satisfying*

$$\int\limits_{|\xi|>1} |\xi|^\kappa m(d\xi) < \infty, \tag{1.7}$$

then $\pi \in \mathcal{P}_\kappa(D)$ *and there exists constants* $K', \delta' > 0$ *such that, for all* $\rho \in \mathcal{P}_\kappa(D)$,

$$W_\kappa(P_t\rho, \pi) \leq K' W_\kappa(\rho, \pi)e^{-\delta' t}, \qquad t \geq 0. \tag{1.8}$$

In order that $W_{\log}(P_t\rho, \pi)$ and $W_\kappa(P_t\rho, \pi)$ are well-defined, we need to show that $P_t\rho$ belongs to $\mathcal{P}_{\log}(D)$ or $\mathcal{P}_\kappa(D)$, respectively. This will be shown in Section 5, where general moment estimates for affine processes are studied. Using $P_t\delta_x = P_t(x, \cdot)$ combined with Remark 1.4 we conclude the following.

Remark 1.6. *Under the conditions of Theorem 1.5, there exist constants* $\delta, K > 0$ *such that*

$$W_d(P_t(x, \cdot), \pi) \leq Ke^{-\delta t}\left(1 + W_d(\delta_x, \pi)\right), \qquad t \geq 0,\ x \in D, \tag{1.9}$$

where $d \in \{d_\kappa, d_{\log}\}$*. Let* $W_{d\wedge 1}$ *be the Wasserstein distance given by (1.4) with* d *replaced by* $d \wedge 1$*. Then similarly to Remark 1.4, convergence with respect to* $W_{d\wedge 1}$ *is equivalent to weak convergence of probability measures on* $\mathcal{P}(D)$*. Since* $W_{d\wedge 1} \leq W_d$*, we conclude from (1.9) that* $P_t(x, \cdot) \longrightarrow \pi$ *weakly as* $t \to \infty$ *with exponential rate.*

Let $X = (X_t)_{t\geq 0}$ be an affine process. For the parameter estimation of affine models, see, e.g., [BDLP14], [LM15] and [BBAKP18], it is often necessary to prove a Birkhoff ergodic theorem, i.e.,

$$\frac{1}{t}\int_0^t f(X_s)ds \longrightarrow \int_D f(x)\pi(dx), \quad t \to \infty \tag{1.10}$$

holds almost surely for sufficiently many test functions f. Using classical theory, see, e.g., [MT09, Theorem 17.1.7] and [San17], such convergence is implied by the ergodicity in the total variation distance, i.e., by

$$\lim_{t\to\infty} \|P_t(x,\cdot) - \pi\|_{\mathrm{TV}} = 0, \qquad x \in D, \tag{1.11}$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation distance. Unfortunately, it is typically a very difficult mathematical task to prove (1.11) even for particular models. An extension of (1.10) applicable in the case where $P_t(x,\cdot) \longrightarrow \pi$ holds in the Wasserstein distance generated by the metric $d(x,\widetilde{x}) = 1 \wedge |x - \widetilde{x}|$ was recently studied in [San17]. Applying the main result of [San17] to the case of affine processes and using the fact that each affine process can be obtained as a pathwise unique strong solution to some stochastic equation with jumps (see Section 4), yields the following corollary.

Corollary 1.7. *Let $(a, \alpha, b, \beta, m, \mu)$ be admissible parameters. Suppose that β has only eigenvalues with negative real parts, and (1.5) is satisfied. Let $(X_t)_{t\geq 0}$ be the corresponding affine process constructed as the pathwise unique strong solution on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ in Section 4. Let $f \in L^p(D,\pi)$ for some $p \in [1,\infty)$, then (1.10) holds in $L^p(\Omega, \mathbb{P})$.*

Although we have formulated (1.10) in continuous time, the discrete-time analog can be obtained in the same manner.

1.4 Comparison with related works

Consider an affine process on state space $D = \mathbb{R}^n$ with admissible parameters $(a, \alpha, b, \beta, m, \mu)$ such that $\alpha = 0$, $b = 0$ and $\mu = 0$, i.e., an Ornstein-Uhlenbeck process on $\mathbb{R}^n$. If β has only eigenvalues with negative real parts and (1.5) is satisfied, then [SY84] is applicable and hence the corresponding Ornstein-Uhlenbeck process satisfies, for all $x \in \mathbb{R}^n$, $P_t(x,\cdot) \longrightarrow \pi$ weakly as $t \to \infty$. Under additional technical conditions on the measure m, it follows that the corresponding process also satisfies (1.11) with exponential rate, see [Wan12]. Since in view of Theorem 1.5 the convergence (in the Wasserstein distance) has already exponential rate, we conclude that the additional restriction on m imposed in [Wan12] is only used to guarantee that convergence takes place in the stronger total variation distance, i.e., it is not necessary for the speed of convergence.

In [Li11] it was shown for general one-dimensional subcritical CBI processes (i.e., $D = \mathbb{R}_+$) that $P_t(x,\cdot) \longrightarrow \pi$ weakly is equivalent to (1.5). A partial extension of this result applicable to general affine processes was also recently studied in [JKR18] where the following result was obtained.

Theorem 1.8. *[[JKR18]] Let $(a, \alpha, b, \beta, m, \mu)$ be admissible parameters. Suppose that β has only eigenvalues with negative real parts and* (1.5) *is satisfied. Then there exists a unique invariant distribution π for $(P_t)_{t\geq 0}$. Moreover, π has Laplace transform*

$$\int_D e^{\langle u,x\rangle}\pi(dx) = \exp\left(\int_0^\infty F(\phi(t,u))dt\right), \qquad u \in \mathcal{U}, \tag{1.12}$$

and one has, for all $x \in D$, $P_t(x,\cdot) \longrightarrow \pi$ weakly as $t \to \infty$.

The proof of Theorem 1.8 is based on a fine stability analysis of the Riccati equations (1.2). Comparing with our main result Theorem 1.5, the authors have, in addition, established a formula for the Laplace transform of π, but have not studied any convergence rate. We emphasize that the main aim of our Theorem 1.5 is to establish the exponential convergence speed (1.6) and (1.8) with respect to the corresponding Wasserstein metrics. However, in the process of proving (1.6) we also obtain the existence and uniqueness of an invariant distribution as a natural by-product. Moreover, in Theorem 1.5 and Theorem 1.8 existence and uniqueness of an invariant distribution is shown by essentially different techniques.

In [LM15] exponential ergodicity in total variation distance, see (1.11), was established for one-dimensional subcritical CBI processes with $m = 0$. An extension of their results to higher dimensions does not seem to be straightforward, while our results can be applied in arbitrary dimension. Recently, in [MSV18] another approach for the exponential ergodicity in the total variation distance for affine processes on cones, including multi-type CBI processes, was provided. Their techniques were closely related to stochastic stability of Markov chains in the sense of Meyn and Tweedie [MT09], see also the references therein. More precisely, it was shown that each subcritical CBI process X which is ν-irreducible, aperiodic and has finite second moments, where ν is a reference measure with its support having non-empty interior, is exponentially ergodic in the total variation distance. Although such result is formulated in a very general way, it is still a delicate mathematical task to show that such conditions are satisfied for CBI processes with jumps of infinite activity or with degenerate diffusion components. Moreover, assuming that X has at least finite second moments rules out some natural examples as studied in [LM15] for $d = 1$ and in Section 2 of this work. Contrary to this our (weaker) ergodicity result does not require any of these conditions.

Another recent work on this topic is [GZ18], where ergodic properties and functional limit theorems for affine processes with non-degenerate diffusion components are studied. For this purpose the authors assumed that ν and $\mu_1, \dots, \mu_d$ are probability measures, i.e., the corresponding affine process has only jumps of finite variation. However, our result also applies to affine processes with jumps of infinite variation and, moreover, we can also treat cases where the diffusion components are degenerate (or even absent). It is worthwile to mention that in [GZ18] affine processes were studied as strong solutions to a stochastic equation with jumps with random compensators. In this case it is a nontrivial task to find sufficient conditions for the existence and uniqueness of such strong solutions. In Section 4 we provide a simpler stochastic equation for affine processes where the jumps are described by certain Poisson random measures. Consequently we are able to prove that each affine process can be obtained as the pathwise unique strong solution to such type of equations.

1.5 Main idea of proof and structure of the work

The proof of Theorem 1.5 is divided in 4 steps as explained below.

Step 1. Provide a stochastic description of conservative affine processes. More precisely, in Section 3 we discuss a stochastic equation for multi-type CBI processes and recall a comparison principle due to [BLP15a]. In Section 4 we prove that each affine process can be obtained as the pathwise unique strong solution to a certain stochastic equation. Although such a result is not surprising, in the literature it was only stated for multi-type CBI processes (see [Ma13, BLP15a, BPP18a]) and affine diffusions (see [FM09]).

Step 2. Let $(X_t)_{t\geq 0}$ be an affine process. Based on the stochastic equation from the first step, we study in Section 5 finiteness of the moments $\mathbb{E}(|X_t|^\kappa)$ and $\mathbb{E}(\log(1+|X_t|))$. Since the proofs in this section are rather standard, we only outline the main steps, while technical details are postponed to the appendix.

Step 3. Let $(X_t(x))_{t\geq 0}$ and $(X_t(\widetilde{x}))_{t\geq 0}$ be the affine processes with initial states $x,\widetilde{x} \in \mathbb{R}_+^m \times \mathbb{R}^n$, respectively, obtained as the unique strong solutions to the stochastic equation discussed in Section 4. Suppose that (1.7) is satisfied for $\kappa = 1$. The following key estimate is proved in Section 6:

$$\mathbb{E}(|X_t(x) - X_t(\widetilde{x})|) \leq K e^{-\delta t} \left(\mathbb{1}_{\{n>0\}} |y - \widetilde{y}|^{1/2} + |x - \widetilde{x}| \right), \qquad t \geq 0,$$

where $K, \delta > 0$ are some constants.

Step 4. The results obtained in Steps 1 – 3 provide us all necessary tools to give a full proof of Theorem 1.5 in Section 7. For the sake of simplicity, we explain below how (1.8) is shown. Estimate (1.6) can be obtained in the same way. Using classical arguments, we may deduce assertion (1.8) from the contraction estimate

$$W_\kappa(P_t\rho, P_t\widetilde{\rho}) \leq K e^{-\delta t} W_\kappa(\rho, \widetilde{\rho}), \qquad t \geq 0. \tag{1.13}$$

Next observe that, by the convexity of the Wasserstein distance (see Lemma 8.4) combined with (1.3), property (1.13) is implied by

$$W_\kappa(P_t\delta_x, P_t\delta_{\widetilde{x}}) \leq K e^{-\delta t} \left(\mathbb{1}_{\{n>0\}} |y - \widetilde{y}|^{1/2} + |x - \widetilde{x}| \right)^\kappa, \qquad t \geq 0. \tag{1.14}$$

Let $(P_t^0)_{t\geq 0}$ be the transition semigroup for the affine process with admissible parameters $(a = 0, \alpha, b = 0, \beta, m = 0, \mu)$. In view of (1.1) one can show that $P_t(x,\cdot) = P_t^0(x,\cdot) * P_t(0,\cdot)$, where $*$ denotes the usual convolution of measures. A similar decomposition for affine processes was also used in [JKR18]. Applying now Lemma 8.3 and the Jensen inequality gives

$$\begin{aligned} W_\kappa(P_t\delta_x, P_t\delta_{\widetilde{x}}) &\leq W_\kappa(P_t^0\delta_x, P_t\delta_{\widetilde{x}}) \\ &\leq (W_1(P_t^0\delta_x, P_t^0\delta_{\widetilde{x}}))^\kappa \leq K^\kappa e^{-\delta\kappa t} \left(\mathbb{1}_{\{n>0\}} |y - \widetilde{y}|^{1/2} + |x - \widetilde{x}| \right)^\kappa, \end{aligned}$$

where the last inequality follows from Step 3 applied to $(P_t^0)_{t\geq 0}$.

2 Examples

2.1 Anisotropic (γ_1, γ_2)-root process

Let Z_1, Z_2 be independent one-dimensional Lévy processes with symbols

$$\Psi_j(\xi) = \int_0^\infty \left(e^{-\xi z} - 1 + \xi z \right) \frac{dz}{z^{1+\gamma_j}}, \qquad \xi \geq 0,\ j = 1, 2,$$

where $\gamma_1, \gamma_2 \in (1, 2)$. Let $S = (S_1, S_2)$ be another 2-dimensional Lévy process with symbol

$$\Psi_m(\xi) = \int_{\mathbb{R}^2_+} \left(e^{-\langle \xi, z \rangle} - 1 \right) m(dz), \qquad \xi \in \mathbb{R}^2_+,$$

where m is a measure on $\mathbb{R}^2_+$ with $m(\{0\}) = 0$ and

$$\int_{\mathbb{R}^2_+} (1 \wedge |z|)\, m(dz) < \infty.$$

Suppose that Z and S are independent. Applying the results of [BLP15a] to this particular case shows that, for each $x \in \mathbb{R}^2_+$, there exists a pathwise unique strong solution to

$$dX_1(t) = (b_1 + \beta_{11} X_1(t) + \beta_{12} X_2(t))\, dt + X_1(t-)^{1/\gamma_1} dZ_1(t) + dS_1(t),$$
$$dX_2(t) = (b_2 + \beta_{21} X_1(t) + \beta_{22} X_2(t))\, dt + X_2(t-)^{1/\gamma_2} dZ_2(t) + dS_2(t),$$

This process is an affine process on $D = \mathbb{R}^2_+$ with admissible parameters

$$a = 0, \quad \alpha_1 = \alpha_2 = 0, \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}$$

and corresponding Lévy measures m,

$$\mu_1(d\xi) = \frac{d\xi_1}{\xi_1^{1+\gamma_1}} \otimes \delta_0(d\xi_2), \quad \mu_2(d\xi) = \delta_0(d\xi_1) \otimes \frac{d\xi_2}{\xi_2^{1+\gamma_2}}.$$

Applying our main result to this particular case gives the following.

Corollary 2.1. *If β has only eigenvalues with negative real parts and m satisfies*

$$\int_{|\xi|>1} \log(|\xi|) m(d\xi) < \infty,$$

then the assertions of Theorem 1.5 are true.

Convergence in total variation distance for a similar one-dimensional model was studied in [LM15]. Similar two-dimensional processes were also studied in [BDLP14] and [JKR17a]. In view of our main result Theorem 1.5, it is straightforward to extend this model to arbitrary dimension $d \geq 1$, with possibly non-vanishing diffusion part and more general driving noise of Lévy type.

2.2 Stochastic volatility model

Let $D = \mathbb{R}_+ \times \mathbb{R}$, i.e., $m = n = 1$. Let (V, Y) be the unique strong solution to

$$dV(t) = (b_1 + \beta_{11}V(t))dt + \sqrt{V(t)}dB_1(t) + dJ_1(t),$$
$$dY(t) = (b_2 + \beta_{22}Y(t))dt + \sqrt{V(t)}\left(\rho dB_1(t) + \sqrt{1-\rho^2}dB_2(t)\right) + dJ_2(t)$$

where $b_1 \geq 0$, $b_2 \in \mathbb{R}$, $\beta_{11}, \beta_{22} \in \mathbb{R}$, $\rho \in (-1, 1)$ is the correlation coefficient, $B = (B_1, B_2)$ is a two-dimensional Brownian motion, J_1 is a one-dimensional Lévy subordinator with Lévy measure m_1, and J_2 a one-dimensional Lévy process with Lévy measure m_2. Suppose that B, J_1 and J_2 are mutually independent. It is not difficult to see that (V, Y) is an affine process with admissible parameters

$$a = 0, \quad \alpha_1 = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_{11} & 0 \\ 0 & \beta_{22} \end{pmatrix}$$

and measures

$$m(d\xi) = m_1(d\xi_1) \otimes \delta_0(d\xi_2) + \delta_0(d\xi_1) \otimes m_2(d\xi_2), \quad \mu_1 = \mu_2 = 0.$$

Then we obtain the following.

Corollary 2.2. *If $\beta_{11}, \beta_{22} < 0$ and*

$$\int_{(1,\infty)} \log(\xi_1) m_1(d\xi_1) + \int_{|\xi_2|>1} \log(|\xi_2|) m_2(d\xi_2) < \infty,$$

then the assertions of Theorem 1.5 are true.

It is straightforward to extend this model to higher dimensions and more general driving noises.

3 Stochastic equation for multi-type CBI processes

In this section we recall some results for the particular case of multi-type CBI processes, i.e. affine processes on state space $D = \mathbb{R}_+^m$ (that is, $n = 0$). For further references and additional explanations we refer to [BLP15a] and [BPP18a]. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space rich enough to support the following objects:

(B1) A m-dimensional Brownian motion $(W_t)_{t\geq 0} := (W_{t,1}, \ldots, W_{t,m})_{t\geq 0}$.

(B2) A Poisson random measure $M_I(ds, d\xi)$ on $\mathbb{R}_+ \times \mathbb{R}_+^m$ with compensator $\widehat{M}_I(ds, d\xi) = ds m_I(d\xi)$, where m_I is a Borel measure supported on $\mathbb{R}_+^m$ satisfying

$$m_I(\{0\}) = 0, \qquad \int_{\mathbb{R}_+^m} (1 \wedge |\xi|) m_I(d\xi) < \infty.$$

(B3) Poisson random measures $N_1^I(ds, d\xi, dr), \ldots, N_m^I(ds, d\xi, dr)$ on $\mathbb{R}_+ \times \mathbb{R}_+^m \times \mathbb{R}_+$ with compensators $\widehat{N}_i^I(ds, d\xi, dr) = ds\mu_i^I(d\xi)dr$, $i \in I$, where $\mu_1^I, \ldots, \mu_m^I$ are Borel measures supported on $\mathbb{R}_+^m$ satisfying

$$\mu_i^I(\{0\}) = 0, \quad \int\limits_{\mathbb{R}_+^m} \left(|\xi| \wedge |\xi|^2 + \sum_{j \in \{1,\ldots,m\} \backslash \{i\}} \xi_j \right) \mu_i^I(d\xi) < \infty, \qquad i \in I.$$

The objects $W, M_I, N_1^I, \ldots, N_m^I$ are supposed to be mutually independent. Let $\widetilde{M}_I(ds, d\xi) = M_I(ds, d\xi) - \widehat{M}_I(ds, d\xi)$ and $\widetilde{N}_i^I(ds, d\xi, dr) = N_i^I(ds, d\xi, dr) - \widehat{N}_i^I(ds, d\xi, dr)$ be the corresponding compensated Poisson random measures. Here and below we consider the natural augmented filtration generated by $W, M_I, N_1^I, \ldots, N_m^I$. Finally let

(a) $b \in \mathbb{R}_+^m$.

(b) $\beta = (\beta_{ij})_{i,j \in I}$ such that $\beta_{ji} - \int_{\mathbb{R}_+^m} \xi_j \mu_i^I(d\xi) \geq 0$, for $i \in I$ and $j \in I \backslash \{i\}$.

(c) A matrix $\sigma(y) = \mathrm{diag}(\sqrt{2c_1 y_1}, \cdots, \sqrt{2c_m y_m}) \in \mathbb{R}^{m \times m}$, where $c_1, \ldots, c_m \geq 0$.

For $y \in \mathbb{R}_+^m$, consider the stochastic equation

$$\begin{aligned} Y_t = y + \int\limits_0^t \left(b + \widetilde{\beta} Y_s \right) ds + \int\limits_0^t \sigma(Y_s) dW_s + \int\limits_0^t \int\limits_{\mathbb{R}_+^m} \xi M_I(ds, d\xi) \qquad (3.1) \\ + \sum_{i=1}^m \int\limits_0^t \int\limits_{|\xi| \leq 1} \int\limits_{\mathbb{R}_+} \xi \mathbb{1}_{\{r \leq Y_{s-,i}\}} \widetilde{N}_i^I(ds, d\xi, dr) + \sum_{i=1}^m \int\limits_0^t \int\limits_{|\xi| > 1} \int\limits_{\mathbb{R}_+} \xi \mathbb{1}_{\{r \leq Y_{s-,i}\}} N_i^I(ds, d\xi, dr), \end{aligned}$$

where $\widetilde{\beta}_{ji} = \beta_{ji} - \int_{|\xi|>1} \xi_j \mu_i^I(d\xi)$. Pathwise uniqueness for a slightly more complicated equation was recently obtained in [BLP15a], while (3.1) in this form appeared first in [BPP18a]. The following is essentially due to [BLP15a].

Proposition 3.1. *Let (b, β, σ) be as in (a) – (c), and consider objects $W, M_I, N_1^I, \ldots, N_m^I$ that are given in (B1) – (B3). Then the following assertions hold:*

(a) For each $y \in \mathbb{R}_+^m$, there exists a pathwise unique strong solution $Y = (Y_t)_{t \geq 0}$ to (3.1).

(b) Let Y be any solution to (3.1). Then Y is a multi-type CBI process starting from y, and the generator L_Y of Y is of the following form: for $f \in C_c^2(\mathbb{R}_+^m)$,

$$\begin{aligned} (L_Y f)(y) = (b + \beta y, \nabla f(y)) + \sum_{i=1}^m c_i y_i \frac{\partial^2 f(y)}{\partial y_i^2} + \int\limits_{\mathbb{R}_+^m} (f(y+\xi) - f(y)) \, m_I(d\xi) \\ + \sum_{i=1}^m y_i \int\limits_{\mathbb{R}_+^m} (f(y+\xi) - f(y) - (\xi, \nabla f(y))) \, \mu_i^I(d\xi). \end{aligned}$$

Conversely, given any multi-type CBI process $\widetilde{Y}$ *with generator* L_Y *and starting point* y*, we can find a solution* Y *to* (3.1) *such that* Y *and* $\widetilde{Y}$ *have the same law.*

The proof of the pathwise uniqueness is based on a comparison principle for multi-type CBI processes, see [BLP15a, Lemma 4.2]. This comparison principle is stated below.

Lemma 3.2. *[BLP15a, Lemma 4.2] Let* $(Y_t)_{t\geq 0}$ *be a weak solution to* (3.1) *with parameters* (b,β,σ)*, let* $(Y'_t)_{t\geq 0}$ *be another weak solution to* (3.1) *with parameters* (b',β,σ)*, where* (b,β,σ) *and* (b',β,σ) *satisfy (a) – (c). Both solutions are supposed to be defined on the same probability space and with respect to the same noises* $W, M_I, N_1^I, \dots, N_m^I$ *that satisfy (B1) – (B3). Suppose that, for all* $j \in \{1,\dots,m\}$*,* $y_j \leq y'_j$ *and* $b_j \leq b'_j$*. Then*

$$\mathbb{P}(Y_{j,t} \leq Y'_{j,t}, \quad \forall j \in \{1,\dots,m\},\ \forall t \geq 0) = 1.$$

4 Stochastic equation for affine processes

Below we show that any affine process can also be obtained as the pathwise unique strong solution to a certain stochastic equation. Let (a,α,b,β,m,μ) be admissible parameters. For the parameters a and $\alpha = (\alpha_1,\dots,\alpha_m)$ consider the following objects:

(A0) An $n \times n$-matrix σ_a such that $\sigma_a \sigma_a^\top = a_{JJ}$.

(A1) Matrices $\sigma_1,\dots,\sigma_m \in \mathbb{R}^{d\times d}$ such that, for all $j \in I$, $\sigma_j\sigma_j^\top = \alpha_j$ and

$$\sigma_j = \begin{pmatrix} \sigma_{j,II} & 0 \\ \sigma_{j,JI} & \sigma_{j,JJ} \end{pmatrix}, \qquad (\sigma_{j,II})_{kl} = \delta_{kj}\delta_{lj}\alpha_{j,jj}^{1/2}. \tag{4.1}$$

Let us remark the following.

Remark 4.1. *(i) The first condition is simple to check. Indeed, by definition, one has* $a = \begin{pmatrix} 0 & 0 \\ 0 & a_{JJ} \end{pmatrix} \in S_d^+$*, thus* a_{JJ} *is symmetric and positive semidefinite. Hence* σ_a *denotes the non-negative square root of* a_{JJ}*.*

(ii) Concerning the second condition, recall that $\alpha_j \in S_d^+$ *and hence* $\alpha_{j,II}$ *is positive semidefinite. Moreover, by definition of admissible parameters,* $\alpha_{j,II}$ *is everywhere zero except at the entry* (j,j)*. Hence* $\alpha_{j,jj}^{1/2}$ *is well-defined. Existence of* σ_j *satisfying* (4.1) *follows from the characterization of positive semidefiniteness for symmetric block matrices, see, e.g., [Gal11, Theorem 16.1]. The latter result is based on pseudo-inverses and properties of the Schur-complement for block matrices.*

Below we describe the noises appearing in the stochastic equation for affine processes. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a complete probability space rich enough to support the following objects:

(A2) A n-dimensional Brownian motion $B = (B_t)_{t\geq 0}$.

(A3) For each $i \in I$, a d-dimensional Brownian motion $W^i = (W^i_t)_{t\geq 0}$.

(A4) A Poisson random measure $M(ds, d\xi)$ with compensator $\widehat{M}(ds, d\xi) = dsm(d\xi)$ on $\mathbb{R}_+ \times D$.

(A5) For each $i \in I$, a Poisson random measure $N_i(ds, d\xi, dr)$ with compensator $\widehat{N}_i(ds, d\xi, dr) = ds\mu_i(d\xi)dr$ on $\mathbb{R}_+ \times D \times \mathbb{R}_+$.

We suppose that all objects $B, W^1, \dots, W^m, M, N_1, \dots, N_m$ are mutually independent. Denote by $\widetilde{M}(ds, d\xi) = M(ds, d\xi) - \widehat{M}(ds, d\xi)$ and $\widetilde{N}_i(ds, d\xi, dr) = N_i(ds, d\xi, dr) - \widehat{N}_i(ds, d\xi, dr)$, $i \in I$, the corresponding compensated Poisson random measures. Here and below we consider the natural augmented filtration generated by these noise terms. For $x \in D$, consider the stochastic equation

$$\begin{aligned} X_t = x &+ \int_0^t \left(\widetilde{b} + \widetilde{\beta} X_s\right) ds + \sqrt{2}\begin{pmatrix} 0 \\ \sigma_a B_t \end{pmatrix} + \sum_{i \in I} \int_0^t \sqrt{2X_{s,i}}\sigma_i dW_s^i \\ &+ \int_0^t \int_{|\xi| \le 1} \xi \widetilde{M}(ds, d\xi) + \int_0^t \int_{|\xi|>1} \xi M(ds, d\xi) \\ &+ \sum_{i \in I} \int_0^t \int_{|\xi|\le 1} \int_{\mathbb{R}_+} \xi \mathbb{1}_{\{r \le X_{s-,i}\}} \widetilde{N}_i(ds, d\xi, dr) + \sum_{i \in I} \int_0^t \int_{|\xi|>1} \int_{\mathbb{R}_+} \xi \mathbb{1}_{\{r \le X_{s-,i}\}} N_i(ds, d\xi, dr), \end{aligned} \tag{4.2}$$

where $\widetilde{b}$ and $\widetilde{\beta} = (\widetilde{b}_{ki})_{k,i \in \{1,\dots,d\}}$ are, for $i, k \in \{1, \dots, d\}$, given by

$$\widetilde{b}_i = b_i + \mathbb{1}_I(i) \int_{|\xi|\le 1} \xi_i m(d\xi), \qquad \widetilde{\beta}_{ki} = \beta_{ki} - \mathbb{1}_I(i) \int_{|\xi|>1} \xi_k \mu_i(d\xi). \tag{4.3}$$

Note that we have changed the drift coefficients to $\widetilde{b}$ and $\widetilde{\beta}$ in order to change the compensators in the stochastic integrals. Such change is, under the given moment conditions on $\mu = (\mu_1, \dots, \mu_m)$, always possible and does not affect our results. Concerning existence and uniqueness for (4.2), we obtain the following.

Theorem 4.2. *Let $(a, \alpha, b, \beta, m, \mu)$ be admissible parameters. Then, for each $x \in D$, there exists a pathwise unique D-valued strong solution $X = (X_t)_{t\ge 0}$ to (4.2).*

This result will be proved later in this Section. Let us first relate (4.2) to affine processes.

Proposition 4.3. *Let $(a, \alpha, b, \beta, m, \mu)$ be admissible parameters. Then each solution X to (4.2) is an affine process with admissible parameters $(a, \alpha, b, \beta, m, \mu)$ and starting point x.*

Proof. Let X be a solution to (4.2) and $f \in C_c^2(D)$. Applying the Itô formula shows that

$$M_f(t) := f(X_t) - f(x) - \int_0^t (Lf)(X_s) ds, \quad t \ge 0$$

is a local martingale. Note that Lf is bounded. Hence

$$\mathbb{E}(\sup_{s\in[0,t]}|M_f(t)|) \leq 2\|f\|_\infty + \int_0^t \mathbb{E}(|Lf(X_s)|)ds \leq 2\|f\|_\infty + t\|Lf\|_\infty < \infty, \qquad t \geq 0,$$

and we conclude that $(M_f(t))_{t\geq 0}$ is a true martingale. It follows from Remark 1.3 that X is an affine process with admissible parameters $(a, \alpha, b, \beta, m, \mu)$. □

The rest of this section is devoted to the proof of Theorem 4.2. As often in the theory of stochastic equations, existence of weak solutions is the easy part.

Lemma 4.4. *Let $(a, \alpha, b, \beta, m, \mu)$ be admissible parameters. Then, for each $x \in D$, there exists a weak solution X to* (4.2).

Proof. Since existence of a solution to the martingale problem with sample paths in the Skorokhod space over D is known, the assertion is a consequence of [Kur11], namely, the equivalence between weak solutions to stochastic equations and martingale problems. Alternatively, following [DFS03, p.993] we can show that each solution to the martingale problem with generator L and domain $C_c^2(D)$ is a semimartingale and compute its semimartingale characteristics (see [DFS03, Theorem 2.12]). The assertion is then a consequence of the equivalence between weak solutions to stochastic equations and semimartingales (see [JS03, Chapter III, Theorem 2.26]). □

In view of the Yamada-Watanabe Theorem (see [BLP15b]), Theorem 4.2 is proved, provided we can show pathwise uniqueness for (4.2). For this purpose we rewrite (4.2) into its components $X = (Y, Z)$, where $Y \in \mathbb{R}_+^m$ and $Z \in \mathbb{R}^n$. Introduce the notation $\xi = (\xi_I, \xi_J) \in D$, where $\xi_I = (\xi_i)_{i\in I}$ and $\xi_J = (\xi_j)_{j\in J}$. Moreover, let $W_s^i = (W_{s,I}^i, W_{s,J}^i)$ and write for the initial condition $x = (y, z) \in D$. Finally, let $e_1, \ldots, e_d$ denote the canonical basis vectors in $\mathbb{R}^d$. Then

(4.2) is equivalent to the system of equations

$$Y_t = y + \int_0^t \left(b_I + \widetilde{\beta}_{II} Y_s \right) ds + \sum_{i \in I} e_i \int_0^t \sqrt{2\alpha_{i,ii} Y_{s,i}} dW_{s,i}^i + \int_0^t \int_D \xi_I M(ds, d\xi) \tag{4.4}$$
$$+ \sum_{i \in I} \int_0^t \int_{|\xi| \leq 1} \int_{\mathbb{R}_+} \xi_I \mathbb{1}_{\{r \leq Y_{s-,i}\}} \widetilde{N}_i(ds, d\xi, dr) + \sum_{i \in I} \int_0^t \int_{|\xi| > 1} \int_{\mathbb{R}_+} \xi_I \mathbb{1}_{\{r \leq Y_{s-,i}\}} N_i(ds, d\xi, dr),$$

$$Z_t = z + \int_0^t \left(b_J + \widetilde{\beta}_{JI} Y_s + \widetilde{\beta}_{JJ} Z_s \right) ds + \sqrt{2}\sigma_a B_t + \sum_{i \in I} \int_0^t \sqrt{2 Y_{s,i}} \left(\sigma_{i,JI} dW_{s,I}^i + \sigma_{i,JJ} dW_{s,J}^i \right) \tag{4.5}$$
$$+ \int_0^t \int_{|\xi| \leq 1} \xi_J \widetilde{M}(ds, d\xi) + \int_0^t \int_{|\xi| > 1} \xi_J M(ds, d\xi)$$
$$+ \sum_{i \in I} \int_0^t \int_{|\xi| \leq 1} \int_{\mathbb{R}_+} \xi_J \mathbb{1}_{\{r \leq Y_{s-,i}\}} \widetilde{N}_i(ds, d\xi, dr) + \sum_{i \in I} \int_0^t \int_{|\xi| > 1} \int_{\mathbb{R}_+} \xi_J \mathbb{1}_{\{r \leq Y_{s-,i}\}} N_i(ds, d\xi, dr).$$

Observe that the first equation for Y does not involve Z. We will show that (4.4) is precisely (3.1), i.e., Y is a multi-type CBI process and pathwise uniqueness holds for Y. The second equation for Z describes an OU-type process with random coefficients depending on Y. If we regard Y as conditionally fixed, then pathwise uniqueness for (4.5) is obvious. These ideas are summarized in the next lemma.

Lemma 4.5. *Let $(a, \alpha, b, \beta, m, \mu)$ be admissible parameters. Then pathwise uniqueness holds for* (4.4) *and* (4.5), *and hence for* (4.2).

Proof. Let $X = (Y, Z)$ and $X' = (Y', Z')$ be two solutions to (4.2) with the same initial condition $x = (y, z) \in D$ both defined on the same probability space. Then Y and Y' both satisfy (4.4). Let us show that (4.4) is precisely (3.1), from which we deduce $\mathbb{P}(Y_t = Y_t', \quad t \geq 0) = 1$. Set $\mathrm{pr}_I : D \longrightarrow \mathbb{R}_+^m$, $\mathrm{pr}_I(x) = (x_i)_{i \in I}$, and define

- A m-dimensional Brownian motion $W_t := (W_{t,1}^1, \ldots, W_{t,m}^m)$.
- A Poisson random measure $M_I(ds, d\xi)$ on $\mathbb{R}_+ \times \mathbb{R}_+^m$ by
$$M_I([s,t] \times A) = M([s,t] \times \mathrm{pr}_I^{-1}(A)),$$
where $0 \leq s < t$ and $A \subset \mathbb{R}_+^m$ is a Borel set.
- Poisson random measures $N_1^I, \ldots, N_m^I$ on $\mathbb{R}_+ \times \mathbb{R}_+^m \times \mathbb{R}_+$ by
$$N_i^I([s,t] \times A \times [c,d]) = N_i([s,t] \times \mathrm{pr}_I^{-1}(A) \times [c,d]), \qquad i \in I,$$
where $0 \leq s < t$, $0 \leq c < d$ and $A \subset \mathbb{R}_+^m$ is a Borel set.

Note that the random objects $W, M_I, N_1^I, \dots, N_m^I$ are mutually independent. Moreover, it is not difficult to see that M_I and $N_1^I, \dots, N_m^I$ have compensators

$$\widehat{M}_I(ds, d\xi) = ds m_I(d\xi), \quad \widehat{N}_i^I(ds, d\xi, dr) = ds\mu_i^I(d\xi)dr, \qquad i \in I,$$

where $m_I = m \circ \mathrm{pr}_I^{-1}$ and $\mu_i^I = \mu_i \circ \mathrm{pr}_I^{-1}$. Finally let $c_j = \alpha_{j,jj}$, $j \in \{1, \dots, m\}$, and

$$\sigma(y) = \mathrm{diag}(\sqrt{2c_1 y_1}, \cdots, \sqrt{2c_m y_m}) \in \mathbb{R}^{m \times m}.$$

Then (4.4) is precisely (3.1), and it follows from Proposition 3.1.(a) that $\mathbb{P}(Y_t = Y_t', \ \ t \geq 0) = 1$.

It remains to prove pathwise uniqueness for (4.5). Define, for $l \geq 1$, a stopping time $\inf\{t > 0 \mid \max\{|Z_t|, |Z_t'|\} > l\}$. Since Z and Z' both satisfy (4.5) for the same Y, we obtain

$$Z_{t \wedge \tau_l} - Z'_{t \wedge \tau_l} = \int_0^{t \wedge \tau_l} \widetilde{\beta}_{JJ}(Z_s - Z_s')ds \tag{4.6}$$

and hence, for some constant $C > 0$,

$$\mathbb{E}(|Z_{t \wedge \tau_l} - Z'_{t \wedge \tau_l}|) \leq C \int_0^t \mathbb{E}(|Z_{s \wedge \tau_l} - Z'_{s \wedge \tau_l}|)ds.$$

The Grownwall lemma gives $\mathbb{P}(Z_{t \wedge \tau_l} = Z'_{t \wedge \tau_l}) = 1$, for all $t \geq 0$ and $l \geq 1$. Note that Z and Z' have no explosion. Taking $l \to \infty$ proves the assertion. □

5 Moments of affine processes

The stochastic equation introduced in Section 4 can be used to provide a simple proof for the finiteness of moments of affine processes. The following is our main result for this section.

Proposition 5.1. *Let $(a, \alpha, b, \beta, m, \mu)$ be admissible parameters. For $x \in D$, let X be the unique solution to (4.2).*

(a) Suppose that there exists $\kappa > 0$ such that

$$\int_{|\xi|>1} |\xi|^\kappa \mu_i(d\xi) + \int_{|\xi|>1} |\xi|^\kappa m(d\xi) < \infty, \qquad i \in I.$$

Then there exists a constant $C_\kappa > 0$ (independent of x and X) such that

$$\mathbb{E}(|X_t|^\kappa) \leq (1 + |x|^\kappa)e^{C_\kappa t}, \qquad t \geq 0.$$

(b) Suppose that (1.5) is satisfied. Then there exists a constant $C > 0$ (independent of x and X) such that

$$\mathbb{E}(\log(1 + |X_t|)) \leq (1 + \log(1 + |x|))e^{Ct}, \qquad t \geq 0.$$

Proof. Define $V_1(h) = (1 + |h|^2)^{\kappa/2}$ and $V_2(h) = \log(1 + |h|^2)$, where $h \in D$. Applying the Itô formula for V_j, $j \in \{1, 2\}$, gives

$$V_j(X_t) = V_j(x) + \int_0^t \mathcal{A}_j(X_s)ds + \mathcal{M}_j(t), \tag{5.1}$$

where $(\mathcal{M}_j(t))_{t\geq 0}$ and $\mathcal{A}_j(\cdot)$ are given by

$$\begin{aligned}
\mathcal{A}_j(h) &= \langle \widetilde{b} + \beta h, \nabla V_j(h)\rangle + \sum_{k,l=1}^{d} \left(a_{kl} + \sum_{i=1}^{m} \alpha_{i,kl} x_i \right) \frac{\partial^2 V_j(h)}{\partial h_k \partial h_l} \\
&\quad + \int_D \left(V_j(h+\xi) - V_j(h) - \langle \xi, \nabla V_j(h)\rangle \mathbb{1}_{\{|\xi|\leq 1\}} \right) m(d\xi) \\
&\quad + \sum_{i=1}^{m} h_i \int_D \left(V_j(h+\xi) - V_j(h) - \langle \xi, \nabla V_j(h)\rangle \right) \mu_i(d\xi), \\
\mathcal{M}_j(t) &= \sqrt{2} \int_0^t \langle \nabla_J V_j(X_s), \sigma_a dB_{s,J}\rangle + \sum_{i=1}^{m} \int_0^t \sqrt{2X_{s,i}}\, \langle \nabla V_j(X_s), \sigma_i dW_s^i \rangle \\
&\quad + \int_0^t \int_D \left(V_j(X_{s-} + \xi) - V_j(X_{s-}) \right) \widetilde{M}(ds, d\xi) \\
&\quad + \sum_{i=1}^{m} \int_0^t \int_D \int_{\mathbb{R}_+} \left(V_j(X_{s-} + \xi \mathbb{1}_{\{r \leq X_{s-,i}\}}) - V_j(X_{s-}) \right) \widetilde{N}_i(ds, d\xi, dr),
\end{aligned}$$

where $\widetilde{b}$ was defined in (4.3). Define, for $l \geq 1$, a stopping time $\tau_l = \inf\{t \geq 0 \mid |X_t| > l\}$. Then it is not difficult to see that $(\mathcal{M}_j(t \wedge \tau_l))_{t\geq 0}$ is a martingale, for any $l \geq 1$. Moreover, we will prove in the appendix that there exists a constant $C > 0$ such that

$$\mathcal{A}_j(h) \leq C(1 + V_j(h)), \qquad h \in D. \tag{5.2}$$

Hence taking expectations in (5.1) gives

$$\mathbb{E}(V_j(X_{t\wedge\tau_l})) \leq V_j(x) + C \int_0^t \left(1 + \mathbb{E}(V_j(X_{s\wedge\tau_l}))\right) ds.$$

Applying the Gronwall lemma gives $\mathbb{E}(V_j(X_{t\wedge\tau_l})) \leq (V_j(x) + Ct)e^{Ct} \leq (1 + V_j(x))e^{C't}$, for all $t \geq 0$ and some constant $C' > 0$. Since $(X_t)_{t\geq 0}$ has cádlág paths and C' is independent of l, we may take the limit $l \to \infty$ and conclude the assertion by the lemma of Fatou. □

We close this section with a formula for the first moment of general affine processes. The particular case of multi-type CBI processes was treated in [BLP15a, Lemma 3.4], while recursion formulas for higher-order moments of multi-type CBI processes were provided in [BLP16].

Lemma 5.2. *Let $(a, \alpha, b, \beta, \nu, \mu)$ be admissible parameters and suppose that*

$$\int\limits_{|\xi|>1} |\xi| m(d\xi) < \infty. \tag{5.3}$$

Let $(X_t)_{t\geq 0}$ be an affine process obtained from (4.2) with $X_0 = x \in D$. Then

$$\mathbb{E}(X_t) = e^{t\beta}x + \int\limits_0^t e^{s\beta}\bar{b}ds,$$

where $\bar{b}_i = b_i + \int_{|\xi|>1} \xi_i m(d\xi) + \mathbb{1}_I(i) \int_{|\xi|\leq 1} \xi_i m(d\xi)$. $x = (y, z) \in \mathbb{R}^m_+ \times \mathbb{R}^n$ and $X = (Y, Z) \in \mathbb{R}^m_+ \times \mathbb{R}^n$, then

$$\mathbb{E}(Y_t) = e^{t\beta_{II}}y + \int\limits_0^t e^{s\beta_{II}}\bar{b}_I ds,$$

$$E(Z_t) = e^{t\beta_{JJ}}z + \int\limits_0^t e^{s\beta_{JJ}}\bar{b}_J ds + \int\limits_0^t e^{(t-s)\beta_{JJ}}\beta_{JI}e^{s\beta_{II}}yds + \int\limits_0^t\int\limits_0^s e^{(t-s)\beta_{JJ}}\beta_{JI}e^{u\beta_{II}}\bar{b}_I duds.$$

Proof. First observe that, by definition of admissible parameters and (5.3), we may apply Proposition 5.1 (a) and deduce that X_t has finite first moment. Taking expectations in (4.2) gives

$$\mathbb{E}(X_t) = x + \int\limits_0^t \left(\bar{b} + \beta\mathbb{E}(X_s)\right) ds.$$

Solving this equation gives the desired formula for $\mathbb{E}(X_t)$. Taking expectations in (3.1) (or (4.4)) gives

$$\mathbb{E}(Y_t) = y + \int\limits_0^t \left(\bar{b}_I + \beta_{II}\mathbb{E}(Y_s)\right) ds,$$

which implies the desired formula for $\mathbb{E}(Y_t)$. Finally, taking expectations in (4.5) gives

$$\mathbb{E}(Z_t) = z + \int\limits_0^t \left(\bar{b}_J + \beta_{JI}\mathbb{E}(Y_s) + \beta_{JJ}\mathbb{E}(Z_s)\right) ds.$$

Solving this equation and using previous formula for $\mathbb{E}(Y_s)$, we obtain the assertion. □

6 Contraction estimate for trajectories of affine processes

The following is our main estimate for this section.

Proposition 6.1. *Let $(a, \alpha, b, \beta, m, \mu)$ be admissible parameters, suppose that (5.3) is satisfied, and assume that β has only eigenvalues with negative real parts. Let $x = (y, z), \widetilde{x} = (\widetilde{y}, \widetilde{z}) \in \mathbb{R}^m_+ \times \mathbb{R}^n$, and let $X(x) = (Y(y), Z(x))$ and $X(\widetilde{x}) = (Y(\widetilde{y}), Z(\widetilde{x}))$ be the unique strong solutions to (4.2) with initial condition x and $\widetilde{x}$, respectively. Then there exist constants $K, \delta, \delta' > 0$ independent of $X(x)$ and $X(\widetilde{x})$ such that, for all $t \geq 0$,*

$$\mathbb{E}(|Y_t(y) - Y_t(\widetilde{y})|) \leq d|y - \widetilde{y}|e^{-\delta' t}, \tag{6.1}$$

$$\mathbb{E}(|X_t(x) - X_t(\widetilde{x})|) \leq Ke^{-\delta t}\left(\mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2} + |x - \widetilde{x}|\right). \tag{6.2}$$

Proof. Let us first prove (6.1). Note that $Y(y)$ and $Y(\widetilde{y})$ are multi-type CBI processes with the same parameters. If $\widetilde{y}_j \leq y_j$ for all $j \in \{1, \dots, m\}$, then we obtain from Lemma 3.2 and Lemma 5.2

$$\begin{aligned}
\mathbb{E}(|Y_t(y) - Y_t(\widetilde{y})|) &\leq \sum_{j=1}^m \mathbb{E}(|Y_{t,j}(y) - Y_{t,j}(\widetilde{y})|) \\
&= \sum_{j=1}^m \mathbb{E}(Y_{t,j}(y) - Y_{t,j}(\widetilde{y})) \\
&= \sum_{j=1}^m \left(e^{t\beta_{II}}(y - \widetilde{y})\right)_j \leq \sqrt{d}|e^{t\beta_{II}}(y - \widetilde{y})| \leq \sqrt{d}e^{-\delta' t}|y - \widetilde{y}|,
\end{aligned}$$

where we have used that β_{II} has only eigenvalues with negative real parts (since β has this property and $\beta_{IJ} = 0$). For general $y, \widetilde{y}$, let $y^0, \dots, y^m \in \mathbb{R}^m_+$ be such that

$$y^0 := y, \quad y^m = \widetilde{y}, \quad y^j = \sum_{k=1}^j e_k \widetilde{y}_k + \sum_{k=j+1}^m e_k y_k, \quad j \in \{1, \dots, m-1\},$$

where $e_1, \dots, e_m$ denote the canonical basis vectors in $\mathbb{R}^m$. Then, for each $j \in \{0, \dots, m-1\}$, the elements y^j, y^{j+1} are comparable in the sense that $y^j_k = y^{j+1}_k$ if $k \neq j+1$, and either $y^j_{j+1} \leq y^{j+1}_{j+1}$ or $y^j_{j+1} \geq y^{j+1}_{j+1}$. In any case, we obtain from the previous consideration

$$\begin{aligned}
\mathbb{E}(|Y_t(y) - Y_t(\widetilde{y})|) &\leq \sum_{j=0}^{m-1} \mathbb{E}(|Y_t(y^j) - Y_t(y^{j+1})|) \\
&\leq \sqrt{d}e^{-\delta' t} \sum_{j=0}^{m-1} |y^j - y^{j+1}| \\
&= \sqrt{d}e^{-\delta' t} \sum_{j=0}^{m-1} |y_{j+1} - \widetilde{y}_{j+1}| \leq de^{-\delta' t}|y - \widetilde{y}|,
\end{aligned}$$

where we have used $|y^j - y^{j+1}| = |y_{j+1} - \widetilde{y}_{j+1}|$. This completes the proof of (6.1).

If $n = 0$, then (6.2) is trivial. Suppose that $n > 0$. Applying the Itô formula to $e^{-t\beta} X_t(x)$ and $e^{-t\beta} X_t(\widetilde{x})$, and then taking the difference, gives

$$X_t(x) - X_t(\widetilde{x}) = e^{t\beta}(x - \widetilde{x}) + \sum_{i \in I} \int_0^t e^{(t-s)\beta} \left(\sqrt{2X_{s,i}(x)} - \sqrt{2X_{s,i}(\widetilde{x})} \right) \sigma_i dW_s^i$$
$$+ \sum_{i \in I} \int_0^t \int_D \int_{\mathbb{R}_+} e^{(t-s)\beta} \xi \left(\mathbb{1}_{\{r \leq X_{s-,i}(x)\}} - \mathbb{1}_{\{r \leq X_{s-,i}(\widetilde{x})\}} \right) \widetilde{N}_i(ds, d\xi, dr).$$

Here and below we denote by $K > 0$ a generic constant which may vary from line to line. Moreover, we find $\delta_0 > 0$ and $\delta \in (0, \delta')$ such that

$$|e^{t\beta}|^2 \leq e^{-\delta_0 t} \text{ and } \int_0^t e^{-(t-s)\frac{\delta_0}{2}} e^{-\delta' s} ds \leq K e^{-2\delta t}, \qquad t \geq 0. \tag{6.3}$$

The stochastic integral against the Brownian motion is estimated by the BDG-inequality as follows

$$\begin{aligned}
&\mathbb{E}\left(\left| \int_0^t e^{(t-s)\beta} \left(\sqrt{2X_{s,i}(x)} - \sqrt{2X_{s,i}(\widetilde{x})} \right) \sigma_i dW_s^i \right| \right) \\
&\leq K \left(\int_0^t \mathbb{E}\left(\left| e^{(t-s)\beta} \left(\sqrt{2X_{s,i}(x)} - \sqrt{2X_{s,i}(\widetilde{x})} \right) \sigma_i \right|^2 \right) ds \right)^{1/2} \\
&\leq K \left(\int_0^t e^{-\delta_0(t-s)} \mathbb{E}(|X_{s,i}(x) - X_{s,i}(\widetilde{x})|) ds \right)^{1/2} \\
&\leq K \left(\int_0^t e^{-\delta_0(t-s)} e^{-\delta' s} ds \right)^{1/2} |y - \widetilde{y}|^{1/2} \leq K e^{-\delta t} |y - \widetilde{y}|^{1/2},
\end{aligned}$$

where we have used (6.1) and (6.3). For the stochastic integral against $\widetilde{N}_i$ we consider the cases $|\xi| \leq 1$ and $|\xi| > 1$ separately. For $|\xi| \leq 1$ we apply first the BDG-inequality and then the

Jensen inequality to obtain, for each $i \in I$,

$$\begin{aligned}
&\mathbb{E}\left(\left|\int_0^t \int_{|\xi|\leq 1} \int_{\mathbb{R}_+} e^{(t-s)\beta}\xi\left(\mathbb{1}_{\{r\leq X_{s-,i}(x)\}} - \mathbb{1}_{\{r\leq X_{s-,i}(\widetilde{x})\}}\right)\widetilde{N}_i(ds,d\xi,dr)\right|\right)\\
&\leq K\mathbb{E}\left(\left|\int_0^t \int_{|\xi|\leq 1} \int_{\mathbb{R}_+} |e^{(t-s)\beta}\xi|^2|\mathbb{1}_{\{r\leq X_{s-,i}(x)\}} - \mathbb{1}_{\{r\leq X_{s-,i}(\widetilde{x})\}}|^2 N_i(dr,d\xi,ds)\right|^{1/2}\right)\\
&\leq K\left(\int_0^t \int_{|\xi|\leq 1} \int_{\mathbb{R}_+} |e^{(t-s)\beta}\xi|^2\mathbb{E}(|\mathbb{1}_{\{r\leq X_{s-,i}(x)\}} - \mathbb{1}_{\{r\leq X_{s-,i}(\widetilde{x})\}}|^2) dr\mu_i(d\xi)ds\right)^{1/2}\\
&\leq K\left(\int_0^t e^{-(t-s)\delta_0}\mathbb{E}(|X_{s,i}(x) - X_{s,i}(\widetilde{x})|)ds\right)^{1/2}\\
&\leq K|y-\widetilde{y}|^{1/2}\left(\int_0^t e^{-(t-s)\delta_0}e^{-\delta' s}ds\right)^{1/2} \leq Ke^{-\delta t}|y-\widetilde{y}|^{1/2}.
\end{aligned}$$

For $|\xi| > 1$, we apply first the BDG-inequality and then use the sub-additivity of $a \longmapsto a^{1/2}$ to obtain

$$\begin{aligned}
&\mathbb{E}\left(\left|\int_0^t \int_{|\xi|>1} \int_{\mathbb{R}_+} e^{(t-s)\beta}\xi\left(\mathbb{1}_{\{r\leq X_{s-,i}(x)\}} - \mathbb{1}_{\{r\leq X_{s-,i}(\widetilde{x})\}}\right)\widetilde{N}_i(ds,d\xi,dr)\right|\right)\\
&\leq K\mathbb{E}\left(\left|\int_0^t \int_{|\xi|>1} \int_{\mathbb{R}_+} |e^{(t-s)\beta}\xi|^2|\mathbb{1}_{\{r\leq X_{s-,i}(x)\}} - \mathbb{1}_{\{r\leq X_{s-,i}(\widetilde{x})\}}|^2 N_i(dr,d\xi,ds)\right|^{1/2}\right)\\
&\leq K\int_0^t \int_{|\xi|>1} \int_{\mathbb{R}_+} \mathbb{E}\left(|e^{(t-s)\beta}\xi||\mathbb{1}_{\{r\leq X_{s-,i}(x)\}} - \mathbb{1}_{\{r\leq X_{s-,i}(\widetilde{x})\}}|\right) dr\mu_i(d\xi)ds\\
&\leq K\int_0^t e^{-(t-s)\frac{\delta_0}{2}}\mathbb{E}(|X_{s,i}(x) - X_{s,i}(\widetilde{x})|)ds\\
&\leq K|y-\widetilde{y}|\int_0^t e^{-(t-s)\frac{\delta_0}{2}}e^{-\delta' s}ds \leq Ke^{-2\delta t}|x-\widetilde{x}|,
\end{aligned}$$

where we have used $|y-\widetilde{y}| \leq |x-\widetilde{x}|$. Collecting all estimates proves the assertion. □

7 Proof of Theorem 1.5

7.1 The log-Wasserstein estimate

Based on the results of Section 6, we first deduce the following estimate with respect to the log-Wasserstein distance.

Proposition 7.1. *Let $(P_t)_{t\geq 0}$ be the transition semigroup with admissible parameters $(a, \alpha, b, \beta, m, \mu)$, suppose that β has only eigenvalues with negative real parts, and (1.5) is satisfied. Then there exist constants $K, \delta > 0$ such that, for any $\rho, \widetilde{\rho} \in \mathcal{P}_{\log}(D)$, one has*

$$W_{\log}(P_t\rho, P_t\widetilde{\rho}) \leq K \min\left\{e^{-\delta t}, W_{\log}(\rho, \widetilde{\rho})\right\} + Ke^{-\delta t} W_{\log}(\rho, \widetilde{\rho}), \qquad t \geq 0.$$

Proof. Let $\left(P_t^0(x, \cdot)\right)_{t\geq 0}$ be the transition semigroup with admissible parameters $(a, \alpha, b = 0, \beta, m = 0, \mu)$ given by Theorem 1.2. Take $x = (y, z), \widetilde{x} = (\widetilde{y}, \widetilde{z}) \in \mathbb{R}_+^m \times \mathbb{R}^n$ and let $X^0(x) = (Y^0(y), Z^0(x))$ and $X^0(\widetilde{x}) = (Y^0(\widetilde{y}), Z^0(\widetilde{x}))$, respectively, be the corresponding affine processes obtained from (4.2) with admissible parameters $(a = 0, \alpha, b = 0, \beta, m = 0, \mu)$. Since $X_t^0(x)$ has law $P_t^0(x, \cdot)$ and $X_t^0(\widetilde{x})$ has law $P_t^0(\widetilde{x}, \cdot)$, there exist by Proposition 6.1 constants $K, \delta > 0$ such that

$$\begin{aligned} W_1(P_t^0(x, \cdot), P_t^0(\widetilde{x}, \cdot)) &\leq \mathbb{E}\left(\mathbb{1}_{\{n>0\}}|Y_t^0(y) - Y_t^0(\widetilde{y})|^{1/2} + |X_t^0(x) - X_t^0(\widetilde{x})|\right) \\ &\leq \mathbb{1}_{\{n>0\}}\left(\mathbb{E}(|Y_t^0(y) - Y_t^0(\widetilde{y})|)\right)^{1/2} + \mathbb{E}(|X_t^0(x) - X_t^0(\widetilde{x})|) \\ &\leq Ke^{-\delta t}\left(\mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2} + |x - \widetilde{x}|\right). \end{aligned}$$

Next observe that, for $u \in \mathcal{U}$, one has

$$\int_D e^{\langle u, x'\rangle} P_t^0(x, dx') = \exp\left(\langle x, \psi(t, u)\rangle\right), \qquad \int_D e^{\langle u, x'\rangle} P_t(0, dx') = \exp\left(\phi(t, u)\right).$$

Combining this with (1.1) proves $P_t(x, \cdot) = P_t^0(x, \cdot) * P_t(0, \cdot)$, where $*$ denotes the convolution of measures on D. Let us now prove the desired log-estimate. Using Lemma 8.3 from the appendix and then the Jensen inequality for the concave function $\mathbb{R}_+ \ni a \longmapsto \log(1 + a)$, gives for some generic constant $K > 0$

$$\begin{aligned} W_{\log}(P_t\delta_x, P_t\delta_{\widetilde{x}}) &\leq W_{\log}(P_t^0\delta_x, P_t^0\delta_{\widetilde{x}}) \\ &\leq \log(1 + W_1(P_t^0\delta_x, P_t^0\delta_{\widetilde{x}})) \\ &\leq \log\left(1 + Ke^{-\delta t}\left(\mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2} + |x - \widetilde{x}|\right)\right) \\ &\leq K\min\{e^{-\delta t}, \log(1 + \mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2} + |x - \widetilde{x}|)\} \\ &\quad + Ke^{-\delta t}\log\left(1 + \mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2} + |x - \widetilde{x}|\right), \end{aligned} \tag{7.1}$$

where we have used, for $a, b \geq 0$, the elementary inequality

$$\begin{aligned} \log(1 + ab) &\leq K\min\{\log(1 + a), \log(1 + b)\} + K\log(1 + a)\log(1 + b) \\ &\leq K\min\{a, \log(1 + b)\} + Ka\log(1 + b), \end{aligned}$$

which is proved in the appendix. Applying now Lemma 8.4 from the appendix gives for any $H \in \mathcal{H}(\rho, \widetilde{\rho})$

$$\begin{aligned} W_{\log}(P_t\rho, P_t\widetilde{\rho}) &\le \int_{D\times D} W_{\log}(P_t\delta_x, P_t\delta_{\widetilde{x}})H(dx, d\widetilde{x}) \\ &\le K \int_{D\times D} \min\left\{e^{-\delta t}, \log(1 + \mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2} + |x - \widetilde{x}|)\right\} H(dx, d\widetilde{x}) \\ &\quad + Ke^{-\delta t} \int_{D\times D} \log(1 + \mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2} + |x - \widetilde{x}|)H(dx, d\widetilde{x}) \\ &\le K \min\left\{e^{-\delta t}, \int_{D\times D} \log(1 + \mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2} + |x - \widetilde{x}|)H(dx, d\widetilde{x})\right\} \\ &\quad + Ke^{-\delta t} \int_{D\times D} \log(1 + \mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2} + |x - \widetilde{x}|)H(dx, d\widetilde{x}). \end{aligned}$$

Choosing H as the optimal coupling of $(\rho, \widetilde{\rho})$, i.e.,

$$W_{\log}(\rho, \widetilde{\rho}) = \int_{D\times D} \log(1 + \mathbb{1}_{\{n>0\}}|y - \widetilde{y}|^{1/2} + |x - \widetilde{x}|)H(dx, d\widetilde{x}),$$

proves the assertion. □

Based on previous proposition, the proof of Theorem 1.5 is easy. It is given below.

Lemma 7.2. *Let $(P_t)_{t\ge 0}$ be the transition semigroup with admissible parameters $(a, \alpha, b, \beta, m, \mu)$. Suppose that β has only eigenvalues with negative real parts, and (1.5) is satisfied. Then $(P_t)_{t\ge 0}$ has a unique invariant distribution π. Moreover, this distribution belongs to $\mathcal{P}_{\log}(D)$ and, for any $\rho \in \mathcal{P}_{\log}(D)$, one has (1.6).*

Proof. Let us first prove existence of an invariant distribution $\widetilde{\pi} \in \mathcal{P}_{\log}(D)$. Observe that, by Proposition 5.1, we easily deduce that $P_t\mathcal{P}_{\log}(D) \subset \mathcal{P}_{\log}(D)$, for any $t \ge 0$. Fix any $\rho \in \mathcal{P}_{\log}(D)$ and let $k, l \in \mathbb{N}$ with $k > l$. Then

$$\begin{aligned} W_{\log}(P_k\rho, P_l\rho) &\le \sum_{s=l}^{k-1} W_{\log}(P_sP_1\rho, P_s\rho) \\ &\le K\sum_{s=l}^{k-1} \min\left\{e^{-\delta s}, W_{\log}(P_1\rho, \rho)\right\} + K\sum_{s=l}^{k-1} e^{-s\delta}W_{\log}(P_1\rho, \rho). \end{aligned}$$

Since the right-hand side tends to zero as $k, l \to \infty$, we conclude that $(P_k\rho)_{k\in\mathbb{N}}$ is a Cauchy sequence in $(\mathcal{P}_{\log}(D), W_{\log})$. In particular, there exists a limit $\pi \in \mathcal{P}_{\log}(D)$, i.e., $W_{\log}(P_k\rho, \pi) \longrightarrow$

0 as $k \to \infty$. Let us show that π is an invariant distribution for P_t. Indeed, take $h \geq 0$, then

$$\begin{aligned} W_{\log}(P_h\pi,\pi) &\leq W_{\log}(P_h\pi, P_hP_k\rho) + W_{\log}(P_kP_h\rho, P_k\rho) + W_{\log}(P_k\rho,\pi) \\ &\leq K\min\left\{e^{-\delta h}, W_{\log}(\pi, P_k\rho)\right\} + Ke^{-\delta h}W_{\log}(\pi,P_k\rho) \\ &\quad + K\min\left\{e^{-\delta k}, W_{\log}(P_h\rho,\rho)\right\} + Ke^{-\delta k}W_{\log}(P_h\rho,\rho) + W_{\log}(P_k\rho,\pi). \end{aligned}$$

Since $W_{\log}(P_k\rho,\pi) \longrightarrow 0$ as $k \to \infty$, we conclude that all terms tend to zero. Hence $W_{\log}(P_h\pi,\pi) = 0$, i.e., $P_h\pi = \pi$, for all $h \geq 0$. Next we prove that π is the unique invariant distribution. Let π_0, π_1 be any two invariant distributions and define $W_{\log}^{\leq 1}$ as in (1.4) with $d_{\log}$ replaced by $d_{\log} \wedge 1$. Then we obtain, for any $t \geq 0$ and all $x, \widetilde{x} \in D$, by the proof of Proposition 7.1 (see (7.1))

$$\begin{aligned} W_{\log}^{\leq 1}(P_t(x,\cdot), P_t(\widetilde{x},\cdot)) &\leq 1 \wedge W_{\log}(P_t(x,\cdot), P_t(\widetilde{x},\cdot)) \\ &\leq 1 \wedge \log\left(1 + Ke^{-\delta t}\left(\mathbb{1}_{\{n>0\}}|y-\widetilde{y}| + |x-\widetilde{x}|\right)\right). \end{aligned}$$

Fix any $H \in \mathcal{H}(\pi_0,\pi_1)$, then using the invariance of π_0, π_1 together with the convexity of the Wasserstein distance gives

$$\begin{aligned} W_{\log}^{\leq 1}(\pi_0,\pi_1) &= W_{\log}^{\leq 1}(P_t\pi_0, P_t\pi_1) \\ &\leq \int\limits_{D\times D} W_{\log}^{\leq 1}(P_t(x,\cdot), P_t(\widetilde{x},\cdot))H(dx,d\widetilde{x}) \\ &\leq \int\limits_{D\times D} \min\{1, \log(1+2Ke^{-\delta t}|x-\widetilde{x}|)H(dx,d\widetilde{x}). \end{aligned}$$

By dominated convergence we deduce that the right-hand side tends to zero as $t \to \infty$ and hence $\pi_0 = \pi_1$. The last assertion can now be deduced from

$$W_{\log}(P_t\rho,\pi) = W_{\log}(P_t\rho, P_t\pi) \leq K\min\left\{e^{-\delta t}, W_{\log}(\rho,\pi)\right\} + Ke^{-\delta t}W_{\log}(\rho,\pi),$$

where we have first used the invariance of π and then Proposition 7.1. □

7.2 The κ-Wasserstein estimate

As before, we start with an estimate with respect to the Wasserstein distance W_κ.

Proposition 7.3. *Let $(P_t)_{t\geq 0}$ be the transition semigroup with admissible parameters (a,α,b,β,m,μ). Suppose that β has only eigenvalues with negative real parts, and (1.7) is satisfied for some $\kappa \in (0,1]$. Then there exist constants $K,\delta > 0$ such that, for any $\rho,\widetilde{\rho} \in \mathcal{P}_\kappa(D)$, one has*

$$W_\kappa(P_t\rho, P_t\widetilde{\rho}) \leq Ke^{-\delta t}W_\kappa(\rho,\widetilde{\rho}), \qquad t \geq 0.$$

Proof. Let $\left(P_t^0(x,\cdot)\right)_{t\geq 0}$ be the transition semigroup with admissible parameters $(a = 0, \alpha, b = 0, \beta, m = 0, \mu)$ given by Theorem 1.2. Arguing as in the proof of Proposition 7.1, we obtain

$$W_1(P_t^0(x,\cdot), P_t^0(\widetilde{x},\cdot)) \leq Ke^{-\delta t}\left(\mathbb{1}_{\{n>0\}}|y-\widetilde{y}|^{1/2} + |x-\widetilde{x}|\right), \tag{7.2}$$

and $P_t(x,\cdot) = P_t^0(x,\cdot) * P_t(0,\cdot)$. Then we obtain from Lemma 8.3 from the appendix

$$\begin{aligned} W_\kappa(P_t\delta_x, P_t\delta_{\widetilde{x}}) &\leq W_\kappa(P_t^0\delta_x, P_t^0\delta_{\widetilde{x}}) \\ &\leq \left(W_1(P_t^0\delta_x, P_t^0\delta_{\widetilde{x}})\right)^\kappa \leq K^\kappa e^{-\delta\kappa t}\left(\mathbb{1}_{\{n>0\}}|y-\widetilde{y}|^{1/2} + |x-\widetilde{x}|\right)^\kappa, \end{aligned}$$

where the second inequality follows from the Jensen inequality and the third is a consequence of (7.2). Using now Lemma 8.4 from the appendix, we conclude that

$$\begin{aligned} W_\kappa(P_t\rho, P_t\widetilde{\rho}) &\leq \inf_{H\in\mathcal{H}(\rho,\widetilde{\rho})} \int_{D\times D} W_\kappa(P_t\delta_x, P_t\delta_{\widetilde{x}})H(dx,d\widetilde{x}) \\ &\leq K^\kappa e^{-\delta\kappa t} \inf_{H\in\mathcal{H}(\rho,\widetilde{\rho})} \int_{D\times D} \left(\mathbb{1}_{\{n>0\}}|y-\widetilde{y}| + |x-\widetilde{x}|\right)^\kappa H(dx,d\widetilde{x}) \\ &= K^\kappa e^{-\delta\kappa t} W_\kappa(\rho,\widetilde{\rho}). \end{aligned}$$

This proves the assertion. □

Based on previous proposition, the proof of the W_κ-estimate in Theorem 1.5 can be deduced by exactly the same arguments as in Lemma 7.2. So Theorem 1.5 is proved.

8 Appendix

8.1 Moment estimates for V_1 and V_2

In this section we prove (5.2).

Lemma 8.1. *Suppose that the same conditions as in Proposition 5.1 (a) are satisfied. Then there exists a constant $C = C_\kappa > 0$ such that*

$$\mathcal{A}_1(x) \leq CV_1(x), \qquad x = (y,z) \in \mathbb{R}_+^m \times \mathbb{R}^n.$$

Proof. Observe that $\nabla V_1(x) = \kappa x(1+|x|^2)^{\frac{\kappa-2}{2}}$. Using $|x| \leq (1+|x|^2)^{1/2}$ gives $|\nabla V_1(x)| \leq \kappa(1+|x|^2)^{\frac{\kappa-1}{2}}$, and hence we obtain for some generic constant $C = C_\kappa > 0$

$$(\widetilde{b} + \beta x, \nabla V_1(x)) \leq C\,(1+|x|)\,|\nabla V_1(x)| \leq CV_1(x).$$

For the second order term we first observe that, for $k,l \in \{1,\dots,d\}$,

$$\frac{\partial^2 V_1(x)}{\partial x_k \partial x_l} = \kappa(\kappa-2)x_k x_l(1+|x|^2)^{\frac{\kappa-4}{2}} + \delta_{kl}\kappa(1+|x|^2)^{\frac{\kappa-2}{2}},$$

where δ_{kl} denotes the Kronecker-Delta symbol. Using $x_k x_l \leq \frac{x_k^2+x_l^2}{2} \leq |x|^2 \leq (1+|x|^2)$ gives $\left|\frac{\partial^2 V_1(x)}{\partial x_k \partial x_l}\right| \leq C(1+|x|^2)^{\frac{\kappa-2}{2}}$. This implies that

$$\sum_{k,l=1}^d \left(a_{kl} + \sum_{i=1}^m \alpha_{i,kl}x_i\right)\frac{\partial^2 V_1(x)}{\partial x_k \partial x_l} \leq C(1+|x|)(1+|x|^2)^{\frac{\kappa-2}{2}} \leq CV_1(x).$$

Let us now estimate the integrals against m and $\mu_1, \dots, \mu_m$. Consider first the case $|\xi| > 1$. The mean value theorem gives

$$\begin{aligned} V_1(x+\xi) - V_1(x) &= \int_0^1 \langle \xi, \nabla V_1(x+t\xi) \rangle \, dt \\ &= \kappa \int_0^1 \langle \xi, x+t\xi \rangle \left(1+|x+t\xi|^2\right)^{\frac{\kappa-2}{2}} dt \leq \kappa |\xi| \int_0^1 \left(1+|x+t\xi|^2\right)^{\frac{\kappa-1}{2}} dt, \end{aligned}$$

where we have used $\langle \xi, x+t\xi \rangle \leq |\xi||x+t\xi| \leq |\xi|(1+|x+t\xi|^2)^{1/2}$ in the last inequality. If $\kappa > 1$, then

$$\begin{aligned} |\xi|(1+|x+t\xi|^2)^{\frac{\kappa-1}{2}} &\leq C|\xi|(1+|x|^2+|\xi|^2)^{\frac{\kappa-1}{2}} \\ &\leq C|\xi|(1+|\xi|^2)^{\frac{\kappa-1}{2}}(1+|x|^2)^{\frac{\kappa-1}{2}} \leq C(1+|\xi|^2)^{\kappa/2}(1+|x|^2)^{\frac{\kappa-1}{2}}. \end{aligned}$$

If $\kappa \in (0,1]$, then $|\xi|(1+|x+t\xi|^2)^{\frac{\kappa-1}{2}} \leq |\xi|$. In any case, we obtain, for $|\xi| > 1$,

$$\begin{aligned} V_1(x+\xi) - V_1(x) &\leq \mathbb{1}_{(0,1]}(\kappa) C|\xi| + \mathbb{1}_{(1,\infty)}(\kappa)(1+|\xi|^2)^{\kappa/2}(1+|x|^2)^{\frac{\kappa-1}{2}} \\ &\leq C\left(1+|\xi|+|\xi|^\kappa\right)\left(1+|x|^2\right)^{\frac{\kappa-1}{2}}. \end{aligned}$$

Using $\langle \xi, \nabla V_1(x) \rangle \leq |\xi||\nabla V_1(x)| \leq C|\xi|(1+|x|^2)^{\frac{\kappa-1}{2}}$ and

$$V_1(x+\xi) - V_1(x) \leq V_1(x+\xi) \leq C(1+|x|^2+|\xi|^2)^{\kappa/2} \leq CV_1(x)(1+|\xi|^2)^{\kappa/2},$$

for the integral against m, gives

$$\begin{aligned} &\int_{|\xi|>1} \left(V_1(x+\xi) - V_1(x)\right) m(d\xi) + \sum_{i=1}^m x_i \int_{|\xi|>1} \left(V_1(x+\xi) - V_1(x) - \langle \xi, \nabla V_1(x) \rangle\right) \mu_i(d\xi) \\ &\leq CV_1(x) \int_{|\xi|>1} (1+|\xi|^2)^{\kappa/2} m(d\xi) + C(1+|x|^2)^{\frac{\kappa-1}{2}} \sum_{i=1}^m x_i \int_{|\xi|>1} \left(1+|\xi|+|\xi|^\kappa\right) \mu_i(d\xi) \\ &\leq CV_1(x) \left(\int_{|\xi|>1} \left(1+|\xi|^\kappa\right) m(d\xi) + \sum_{i=1}^m \int_{|\xi|>1} \left(1+|\xi|+|\xi|^\kappa\right) \mu_i(d\xi) \right), \end{aligned}$$

where we have used $x_i \leq |x| \leq (1+|x|^2)^{1/2}$, $i \in \{1, \dots, m\}$. It remains to estimate the corresponding integrals for $|\xi| \leq 1$. Applying twice the mean value theorem gives

$$\begin{aligned}
V_1(x+\xi) - V_1(x) - \langle \xi, \nabla V_1(x) \rangle &= \int_0^1 \left\{ \langle \xi, \nabla V_1(x+t\xi) \rangle - \langle \xi, \nabla V_1(x) \rangle \right\} dt \\
&= \int_0^1 \int_0^t \sum_{k,l=1}^{d} \frac{\partial^2 V_1(x+s\xi)}{\partial x_k \partial x_l} \xi_k \xi_l ds dt \\
&\leq C|\xi|^2 \int_0^1 \int_0^t (1+|x+s\xi|^2)^{\frac{\kappa-2}{2}} ds dt, \qquad (8.1)
\end{aligned}$$

where we have used $\xi_k \xi_l \leq \frac{\xi_k^2 + \xi_l^2}{2} \leq |\xi|^2$. Using, for $i \in I$ and $|\xi| \leq 1$,

$$\begin{aligned}
(1+x_i)(1+|x+s\xi|^2)^{\frac{\kappa-2}{2}} &\leq (1+|y+s\xi_I|^2)^{1/2}(1+|x+s\xi|^2)^{\frac{\kappa-2}{2}} \\
&\leq (1+|x+s\xi|^2)^{\frac{\kappa-1}{2}} \\
&\leq (1+|x+s\xi|^2)^{\kappa/2} \leq C V_1(x),
\end{aligned}$$

we conclude that

$$\begin{aligned}
&\int_{|\xi|\leq 1} \left(V_1(x+\xi) - V_1(x) - \langle \xi, \nabla V_1(x) \rangle \right) m(d\xi) \\
&\quad + \sum_{i=1}^{m} x_i \int_{|\xi|\leq 1} \left(V_1(x+\xi) - V_1(x) - \langle \xi, \nabla V_1(x) \rangle \right) \mu_i(d\xi) \\
&\leq C V_1(x) \left(\int_{|\xi|\leq 1} |\xi|^2 m(d\xi) + \int_{|\xi|\leq 1} |\xi|^2 \mu_i(d\xi) \right).
\end{aligned}$$

Collecting all estimates proves the desired estimate for $\mathcal{A}_1$. □

Let us now prove the desired estimate for $\mathcal{A}_2$.

Lemma 8.2. *Suppose that the same conditions as in Proposition 5.1 (b) are satisfied. Then there exists a constant $C > 0$ such that*

$$\mathcal{A}_2(x) \leq C\left(1 + V_2(x)\right), \qquad x \in D.$$

Proof. Observe that $\nabla V_2(x) = \frac{2x}{1+|x|^2}$. Hence we obtain for some generic constant $C > 0$

$$\left\langle \widetilde{b} + \beta x, \nabla V_2(x) \right\rangle \leq C\left(1+|x|\right)|\nabla V_2(x)| \leq C \frac{(1+|x|)|x|}{1+|x|^2} \leq C.$$

Observe that, for $k, l \in \{1, \dots, d\}$,

$$\frac{\partial^2 V_2(x)}{\partial x_k \partial x_l} = \frac{2\delta_{kl}}{1+|x|^2} - \frac{4x_k x_l}{(1+|x|^2)^2}.$$

Using $x_k x_l \le C(1+|x|^2)$ gives $\left|\frac{\partial^2 V_2(x)}{\partial x_k \partial x_l}\right| \le \frac{C}{1+|x|^2}$. This implies that

$$\sum_{k,l=1}^{d} \left(a_{kl} + \sum_{i=1}^{m} \alpha_{i,kl} x_i \right) \frac{\partial^2 V_2(x)}{\partial x_k \partial x_l} \le C \frac{1+|x|}{1+|x|^2} \le C.$$

Let us estimate the integrals against m and $\mu_1, \dots, \mu_m$. Consider first the case $|\xi| > 1$. Then

$$V_2(x+\xi) - V_2(x) \le V_2(x+\xi) \le C\log(1+|x|^2+|\xi|^2) \le C\log(1+|x|^2) + C\log(1+|\xi|^2),$$

and hence we obtain

$$\int_{|\xi|>1} (V_2(x+\xi) - V_2(x))\, m(d\xi) \le C \int_{|\xi|>1} (V_2(x) + V_2(\xi))\, m(d\xi) \le C(1+V_2(x)).$$

From the mean value theorem we obtain

$$V_2(x+\xi) - V_2(x) = \int_0^1 \langle \xi, \nabla V_2(x+t\xi)\rangle\, dt = 2\int_0^1 \frac{\langle \xi, x+t\xi\rangle}{1+|x+t\xi|^2} dt \le 2|\xi| \int_0^1 \frac{|x+t\xi|}{1+|x+t\xi|^2} dt.$$

In view of $x_i \le x_i + t\xi_i \le |x_I + t\xi_I| \le |x+t\xi|$ for $i \in I$, we obtain $x_i(V_2(x+\xi) - V_2(x)) \le 2|\xi|$. Using $\langle \xi, \nabla V_2(x)\rangle \le |\xi||\nabla V_2(x)| \le C|\xi|$ gives

$$\sum_{i=1}^{m} x_i \int_{|\xi|>1} (V_2(x+\xi) - V_2(x) - \langle \xi, \nabla V_2(x)\rangle)\, \mu_i(d\xi) \le C \sum_{i=1}^{m} \int_{|\xi|>1} |\xi| \mu_i(d\xi).$$

It remains to estimate the corresponding integrals for $|\xi| \le 1$. As in (8.1), we get

$$V_2(x+\xi) - V_2(x) - \langle \xi, \nabla V_2(x)\rangle \le C|\xi|^2 \int_0^1 \int_0^t \frac{1}{1+|x+s\xi|^2} ds dt.$$

This implies

$$\int_{|\xi|\le 1} (V_2(x+\xi) - V_2(x) - \langle \xi, \nabla V_2(x)\rangle)\, m(d\xi) \le C \int_{|\xi|\le 1} |\xi|^2 m(d\xi).$$

For $i \in I$, by $x_i \le |x+s\xi|$, we get $\frac{x_i}{1+|x+s\xi|^2} \le 1$ and hence

$$\sum_{i=1}^{m} x_i \int_{|\xi|\le 1} (V_2(x+\xi) - V_2(x) - \langle \xi, \nabla V_2(x)\rangle)\, \mu_i(d\xi) \le C \sum_{i=1}^{m} \int_{|\xi|\le 1} |\xi|^2 \mu_i(d\xi).$$

Collecting all estimates proves the desired estimate for $\mathcal{A}_2$. □

8.2 Some estimate on the Wasserstein distance

Here and below we let $d \in \{d_\kappa, d_{\log}\}$. Below we provide two simple and known estimates for Wasserstein distances.

Lemma 8.3. *Let $f, \widetilde{f}, g \in \mathcal{P}_d(D)$. Then*

$$W_d(f * g, \widetilde{f} * g) \le W_d(f, \widetilde{f}).$$

Proof. Using the Kantorovich duality (see [Vil09, Theorem 5.10, Case 5.16], we obtain

$$W_d(f * g, \widetilde{f} * g) = \sup_{\|h\| \le 1} \left(\int_D h(x)(f * g)(dx) - \int_D h(x)(\widetilde{f} * g)(dx) \right),$$

where $\|h\| = \sup_{x \neq x'} \frac{|h(x) - h(x')|}{d(x,x')}$. Using now the definition of the convolution on the right-hand side gives

$$\begin{aligned}
&\int_D h(x)(f * g)(dx) - \int_D h(x)(\widetilde{f} * g)(dx) \\
&= \int_D \int_D h(x + x') f(dx) g(dx') - \int_D \int_D h(x + x') \widetilde{f}(dx) g(dx') \\
&= \int_D \widetilde{h}(x) f(dx) - \int_D \widetilde{h}(x) \widetilde{f}(dx),
\end{aligned}$$

where $\widetilde{h}(x) = \int_D h(x + x') g(dx')$. Since $\|\widetilde{h}\| \le 1$, we conclude that

$$\begin{aligned}
W_d(f * g, \widetilde{f} * g) &= \sup_{\|h\| \le 1} \left(\int_D \widetilde{h}(x) f(dx) - \int_D \widetilde{h}(x) \widetilde{f}(dx) \right) \\
&\le \sup_{\|h\| \le 1} \left(\int_D h(x) f(dx) - \int_D h(x) \widetilde{f}(dx) \right) = W_d(f, \widetilde{f}),
\end{aligned}$$

where we have used again the Kantorovich duality. This completes the proof. □

The next estimate shows that the Wasserstein distance is convex. For additional details we refer to [Vil09, Theorem 4.8].

Lemma 8.4. *Let $P(x, \cdot)$ be a Markov transition function on $D \times \mathcal{P}_d(D)$. Then, for any $f, g \in \mathcal{P}_d(D)$ and any coupling H of (f, g), it holds that*

$$W_d \left(\int_D P(x, \cdot) f(dx), \int_D P(x, \cdot) g(dx) \right) \le \int_{D \times D} W_d(P(x, \cdot), P(\widetilde{x}, \cdot)) H(dx, d\widetilde{x}).$$

8.3 Proof of the elementary inequality with respect to log

Below we prove the following inequality.

Lemma 8.5. *For any* $a, b \geq 0$ *one has*

$$\log(1+ab) \leq \log(2e-1)\min\{\log(1+a), \log(1+b)\} + \log(2e-1)\log(1+a)\log(1+b).$$

Proof. Using the elementary inequality $\log(e+ab) \leq \log(e+a)\log(e+b)$, see [GMP89], we easily obtain

$$\begin{aligned}\log(1+ab) &= \log(e^{-1}) + \log(e+eab) \\ &\leq \log(e+a)\big(\log(e^{-1}) + \log(e+eb)\big) \leq \log(e+a)\log(1+b)\end{aligned}$$

from which we readily deduce

$$\log(1+ab) \leq \min\{\log(e+a)\log(1+b), \log(e+b)\log(1+a)\}.$$

Fix any $\varepsilon > 0$. If $a \geq \varepsilon$, then we obtain

$$\log(1+ab) \leq \log(e+a)\log(1+b) \leq \frac{\log(e+\varepsilon)}{\log(1+\varepsilon)}\log(1+a)\log(1+b).$$

The case $b \geq \varepsilon$ can be treated in the same way. Finally, if $0 \leq a, b \leq \varepsilon$, then we obtain

$$\begin{aligned}\log(1+ab) &\leq \min\{\log(e+a)\log(1+b), \log(e+b)\log(1+a)\} \\ &\leq \log(e+\varepsilon)\min\left\{\log(e+\varepsilon), \frac{\log(e+\varepsilon)}{\log(1+\varepsilon)}\right\}.\end{aligned}$$

Collecting both estimates gives, for all $a, b \geq 0$, the estimate

$$\log(1+ab) \leq g(\varepsilon)\min\{\log(1+a), \log(1+b)\} + g(\varepsilon)\log(1+a)\log(1+b),$$

where $g(\varepsilon) = \min\left\{\log(e+\varepsilon), \frac{\log(e+\varepsilon)}{\log(1+\varepsilon)}\right\}$. A simple extreme value analysis shows that g attains its maximum at $\varepsilon = e-1$ which gives $\inf_{\varepsilon>0} g(\varepsilon) = g(e-1) = \log(2e-1)$. □

Acknowledgments

The authors would like to thank Jonas Kremer for several discussions on affine processes and pointing out some interesting references on this topic. Peng Jin is partially supported by the STU Scientific Research Foundation for Talents (No. NTF18023).

References

[Alf15] Aurélien Alfonsi. *Affine diffusions and related processes: simulation, theory and applications*, volume 6 of *Bocconi & Springer Series*. Springer, Cham; Bocconi University Press, Milan, 2015.

[BBAKP18] Mátyás Barczy, Mohamed Ben Alaya, Ahmed Kebaier, and Gyula Pap. Asymptotic properties of maximum likelihood estimator for the growth rate for a jump-type CIR process based on continuous time observations. *Stochastic Process. Appl.*, 128(4):1135–1164, 2018.

[BDLP13] Mátyás Barczy, Leif Döring, Zenghu Li, and Gyula Pap. On parameter estimation for critical affine processes. *Electron. J. Stat.*, 7:647–696, 2013.

[BDLP14] Mátyás Barczy, Leif Döring, Zenghu Li, and Gyula Pap. Stationarity and ergodicity for an affine two-factor model. *Adv. in Appl. Probab.*, 46(3):878–898, 2014.

[BLP15a] Mátyás Barczy, Zenghu Li, and Gyula Pap. Stochastic differential equation with jumps for multi-type continuous state and continuous time branching processes with immigration. *ALEA Lat. Am. J. Probab. Math. Stat.*, 12(1):129–169, 2015.

[BLP15b] Mátyás Barczy, Zenghu Li, and Gyula Pap. Yamada-Watanabe results for stochastic differential equations with jumps. *Int. J. Stoch. Anal.*, pages Art. ID 460472, 23, 2015.

[BLP16] Mátyás Barczy, Zenghu Li, and Gyula Pap. Moment formulas for multitype continuous state and continuous time branching process with immigration. *J. Theoret. Probab.*, 29(3):958–995, 2016.

[BPP18a] Mátyás Barczy, Sandra Palau, and Gyula Pap. Almost sure, L_1- and L_2-growth behavior of supercritical multi-type continuous state and continuous time branching processes with immigration. *arXiv:1803.10176 [math.PR]*, 2018.

[BPP18b] Mátyás Barczy, Sandra Palau, and Gyula Pap. Asymptotic behavior of projections of supercritical multi-type continuous state and continuous time branching processes with immigration. *arXiv:1806.10559 [math.PR]*, 2018.

[CLP18] Marie Chazal, Ronnie Loeffen, and Pierre Patie. Smoothness of continuous state branching with immigration semigroups. *J. Math. Anal. Appl.*, 459(2):619–660, 2018.

[CPGUB13] M. Emilia Caballero, José Luis Pérez Garmendia, and Gerónimo Uribe Bravo. A Lamperti-type representation of continuous-state branching processes with immigration. *Ann. Probab.*, 41(3A):1585–1627, 2013.

[CPGUB17] M. Emilia Caballero, José Luis Pérez Garmendia, and Gerónimo Uribe Bravo. Affine processes on $\mathbb{R}_+^m \times \mathbb{R}^n$ and multiparameter time changes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(3):1280–1304, 2017.

[DFM14] Xan Duhalde, Clément Foucart, and Chunhua Ma. On the hitting times of continuous-state branching processes with immigration. *Stochastic Process. Appl.*, 124(12):4182–4201, 2014.

[DFS03] Darrell Duffie, Damir Filipović, and Walter Schachermayer. Affine processes and applications in finance. *Ann. Appl. Probab.*, 13(3):984–1053, 2003.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[FFS85] P. J. Fitzsimmons, Bert Fristedt, and L. A. Shepp. The set of real numbers left uncovered by random covering intervals. *Z. Wahrsch. Verw. Gebiete*, 70(2):175–189, 1985.

[FJR18] Martin Friesen, Peng Jin, and Barbara Rüdiger. Existence of densities for multi-type CBI processes. *arXiv:1810.00400 [math.PR]*, 2018.

[FM09] Damir Filipović and Eberhard Mayerhofer. Affine diffusion processes: theory and applications. In *Advanced financial modelling*, volume 8 of *Radon Ser. Comput. Appl. Math.*, pages 125–164. Walter de Gruyter, Berlin, 2009.

[FMS13] Damir Filipović, Eberhard Mayerhofer, and Paul Schneider. Density approximations for multivariate affine jump-diffusion processes. *J. Econometrics*, 176(2):93–111, 2013.

[FUB14] Clément Foucart and Gerónimo Uribe Bravo. Local extinction in continuous-state branching processes with immigration. *Bernoulli*, 20(4):1819–1844, 2014.

[Gal11] Jean Gallier. *Geometric methods and applications: For computer science and engineering*, volume 38 of *Texts in Applied Mathematics*. Springer, New York, second edition, 2011.

[GMP89] Gustavsson, Jan and Maligranda, Lech and Peetre, Jaak. A submultiplicative function. *Nederl. Akad. Wetensch. Indag. Math.*, 51(4):435–442, 1989

[Gre74] D. R. Grey. Asymptotic behaviour of continuous time, continuous state-space branching processes. *J. Appl. Probability*, 11:669–677, 1974.

[GZ18] Peter W. Glynn and Xiaowei Zhang. Affine jump-diffusions: Stochastic stability and limit theorems. *arXiv:1811.00122 [q-fin.MF]*, 2018.

[Ji58] Miloslav Jiřina. Stochastic branching processes with continuous state space. *Czechoslovak Math. J.*, 8 (83):292–313, 1958.

[JKR17a] Peng Jin, Jonas Kremer, and Barbara Rüdiger. Exponential ergodicity of an affine two-factor model based on the α-root process. *Adv. in Appl. Probab.*, 49(4):1144–1169, 2017.

[JKR17b] Peng Jin, Jonas Kremer, and Barbara Rüdiger. Moments and ergodicity of the jump-diffusion CIR process. *arXiv:1709.00969v2 [math.PR]*, 2017.

[JKR18] Peng Jin, Jonas Kremer, and Barbara Rüdiger. Existence of limiting distribution for affine processes. *arXiv:1812.05402 [math.PR]*, 2018.

[JKX12] Rudra P. Jena, Kyoung-Kuk Kim, and Hao Xing. Long-term and blow-up behaviors of exponential moments in multi-dimensional affine diffusions. *Stochastic Process. Appl.*, 122(8):2961–2993, 2012.

[JMRT13] Peng Jin, Vidyadhar Mandrekar, Barbara Rüdiger, and Chiraz Trabelsi. Positive Harris recurrence of the CIR process and its applications. *Commun. Stoch. Anal.*, 7(3):409–424, 2013.

[JRT16] Peng Jin, Barbara Rüdiger, and Chiraz Trabelsi. Exponential ergodicity of the jump-diffusion CIR process. In *Stochastics of environmental and financial economics—Centre of Advanced Study, Oslo, Norway, 2014–2015*, volume 138 of *Springer Proc. Math. Stat.*, pages 285–300. Springer, Cham, 2016.

[JS03] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[KK14] Wanmo Kang and Chulmin Kang. Large deviations for affine diffusion processes on $\mathbb{R}_+^m \times \mathbb{R}^n$. *Stochastic Process. Appl.*, 124(6):2188–2227, 2014.

[KPR17] Andreas Kyprianou, Sandra Palau, and Yan-Xia Ren. Almost sure growth of supercritical multi-type continuous state branching process. *arXiv:1707.04955 [math.PR]*, 2017.

[KRM12] Martin Keller-Ressel and Aleksandar Mijatović. On the limit distributions of continuous-state branching processes with immigration. *Stochastic Process. Appl.*, 122(6):2329–2345, 2012.

[KRM15] Martin Keller-Ressel and Eberhard Mayerhofer. Exponential moments of affine processes. *Ann. Appl. Probab.*, 25(2):714–752, 2015.

[KRST11] Martin Keller-Ressel, Walter Schachermayer, and Josef Teichmann. Affine processes are regular. *Probab. Theory Related Fields*, 151(3-4):591–611, 2011.

[KRST13] Martin Keller-Ressel, Walter Schachermayer, and Josef Teichmann. Regularity of affine processes on general state spaces. *Electron. J. Probab.*, 18:no. 43, 17, 2013.

[Kur11] Thomas G. Kurtz. Equivalence of stochastic equations and martingale problems. In *Stochastic analysis 2010*, pages 113–130. Springer, Heidelberg, 2011.

[KW71] Kiyoshi Kawazu and Shinzo Watanabe. Branching processes with immigration and related limit theorems. *Teor. Verojatnost. i Primenen.*, 16:34–51, 1971.

[Li06] Zenghu Li. A limit theorem for discrete Galton-Watson branching processes with immigration. *J. Appl. Probab.*, 43(1):289–295, 2006.

[Li11] Zenghu Li. *Measure-valued branching Markov processes.* Probability and its Applications (New York). Springer, Heidelberg, 2011.

[LM15] Zenghu Li and Chunhua Ma. Asymptotic properties of estimators in a stable Cox-Ingersoll-Ross model. *Stochastic Process. Appl.*, 125(8):3196–3233, 2015.

[Ma13] Ru Gang Ma. Stochastic equations for two-type continuous-state branching processes with immigration. *Acta Math. Sin. (Engl. Ser.)*, 29(2):287–294, 2013.

[MSV18] Eberhard Mayerhofer, Robert Stelzer, and Johanna Vestweber. Geometric ergodicity of affine processes on cones. *arXiv:1811.10542 [math.PR]*, 2018.

[MT09] Sean Meyn and Richard L. Tweedie. *Markov chains and stochastic stability.* Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.

[Pin72] Mark A. Pinsky. Limit theorems for continuous state branching processes with immigration. *Bull. Amer. Math. Soc.*, 78:242–244, 1972.

[San17] Nikola Sandrić. A Note on the Birkhoff Ergodic Theorem. *Results in Mathematics*, 72(1-2):715–730, 2017.

[SW73] Tokuzo Shiga and Shinzo Watanabe. Bessel diffusions as a one-parameter family of diffusion processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 27:37–46, 1973.

[SY84] Ken-iti Sato and Makoto Yamazato. Operator-self-decomposable distributions as limit distributions of processes of Ornstein-Uhlenbeck type. *Stochastic Process. Appl.*, 17(1):73–100, 1984.

[Vil09] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[Wan12] Jian Wang. On the exponential ergodicity of Lévy-driven Ornstein-Uhlenbeck processes. *J. Appl. Probab.*, 49(4):990–1004, 2012.

[Wat69] Shinzo Watanabe. On two dimensional Markov processes with branching property. *Trans. Amer. Math. Soc.*, 136:447–466, 1969.