

# STOCHASTIC EQUATION AND EXPONENTIAL ERGODICITY IN WASSERSTEIN DISTANCES FOR AFFINE PROCESSES

BY MARTIN FRIESEN<sup>1,\*</sup>, PENG JIN<sup>2</sup> AND BARBARA RÜDIGER<sup>1,†</sup>

<sup>1</sup>*School of Mathematics and Natural Sciences, University of Wuppertal, \*friesen@math.uni-wuppertal.de;*

*†ruediger@uni-wuppertal.de*

<sup>2</sup>*Department of Mathematics, Shantou University, pjjin@stu.edu.cn*

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. Then we study the long-time behavior of affine processes, that is, 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.

## 1. Introduction and statement of the result.

1.1. *General introduction.* A time-homogeneous Markov processes  $X = (X_t)_{t \geq 0}$  is called affine process, if its characteristic function satisfies

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

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  $m, n \in \mathbb{N}_0$  and  $m + n > 0$ , was discussed in [16]. In particular, it is shown that the functions  $\phi$  and  $\psi$  should satisfy certain generalized Riccati equations. Common applications of affine processes in mathematical finance are interest rate models (e.g., the Cox–Ingersoll–Ross [13], Vašíček [56] or general affine term structure short rate models), option pricing (e.g., the Heston model [29]) and credit risk models, see also [1] and the references therein. After [16], the mathematical theory of affine processes was developed in various directions. Regularity of affine processes was studied in [40] and [41]. Based on a Hörmander-type condition, existence and smoothness of transition densities were obtained in [20]. Exponential moments for affine processes were studied in [31] and [38]. The theory of affine diffusions, that is, processes without jumps, was developed in [19], while its application to large deviations for affine diffusions was studied in [36]. The possibility to obtain affine processes as multiparameter time changes of Lévy processes was recently discussed in [11]. 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 [1].

Below we briefly comment on 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 [35] and then studied in [37, 55, 59], where it was also shown that these processes arise as scaling limits of Galton–Watson processes. Various properties of one-dimensional CBI processes

---

Received March 2019; revised October 2019.

*MSC2020 subject classifications.* Primary 37A25, 60H10; secondary 60J25.

*Key words and phrases.* Affine process, ergodicity, Wasserstein distance, coupling, stochastic differential equation.

were studied in [10, 17, 21, 22, 27, 39] and [12]. For results applicable in arbitrary dimension we refer to [5, 7] and [25]. Let us mention that CBI processes are also measure-valued Markov processes as studied in [46]. Another important class of affine processes corresponds to the state-space  $D = \mathbb{R}^n$  and consists of processes of Ornstein–Uhlenbeck (OU) type. These processes include also Lévy processes as a particular case.

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  $\delta_{kl}$ ,  $k, l \in \{1, \dots, d\}$ , stand for the Kronecker–Delta.

DEFINITION 1.1. We call a tuple  $(a, \alpha, b, \beta, \nu, \mu)$  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  $\alpha_i = (\alpha_{i,kl})_{1 \leq k,l \leq d} \in S_d^+$  and  $\alpha_{i,kl} = 0$  if  $k \in I \setminus \{i\}$  or  $l \in I \setminus \{i\}$ .
- (iii)  $b \in D$ .
- (iv)  $\beta \in \mathbb{R}^{d \times d}$  is such that  $\beta_{ki} - \int_D \xi_k \mu_i(d\xi) \geq 0$  for all  $i \in I$  and  $k \in I \setminus \{i\}$ , and  $\beta_{IJ} = 0$ .
- (v)  $\nu$  is a Borel measure on  $D$  such that  $\nu(\{0\}) = 0$  and

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

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

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

In contrast to [16], we do not consider killing for affine processes and, moreover, we suppose that  $\mu_1, \dots, \mu_m$  integrate  $\mathbb{1}_{\{|\xi|>1\}}|\xi|$ , that is, 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  $\mu_1, \dots, \mu_m$ , the corresponding affine process (introduced below) is conservative (see [16], Lemma 9.2). Moreover, it was shown in [49], Example 3.6, that such a moment condition is sufficient but not necessary for an affine process to be conservative. In this paper we work with Definition 1.1 and thus restrict our study to conservative affine processes satisfying a mild moment condition on the measures  $\mu_1, \dots, \mu_m$ . In order to simplify the notation, we have also set  $\nu(\{0\}) = 0$  and  $\mu_i(\{0\}) = 0$ , for  $i \in I$ . Hence all integrals with respect to the measures  $\mu_1, \dots, \mu_m, \nu$  can be taken over  $D$  instead of  $D \setminus \{0\}$ .

Denote by  $B_b(D)$  the Banach space of bounded measurable functions over  $D$ . This space is equipped with the supremum norm  $\|f\|_\infty = \sup_{x \in D} |f(x)|$ . Similarly, let  $C_0(D)$  be the Banach space of continuous functions  $f : D \rightarrow \mathbb{R}$  vanishing at infinity. Define

$$\mathcal{U} = \mathbb{C}_{\geq 0}^m \times i\mathbb{R}^n = \{u = (u_I, u_J) \in \mathbb{C}^m \times \mathbb{C}^n \mid \operatorname{Re}(u_I) \leq 0, \operatorname{Re}(u_J) = 0\}.$$

Note that  $D \ni x \mapsto e^{\langle u, x \rangle}$  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 [16].

**THEOREM 1.2.** *Let  $(a, \alpha, b, \beta, \nu, \mu)$  be admissible parameters. Then there exists a unique conservative transition semigroup  $(P_t)_{t \geq 0}$  on  $B_b(D)$  which is  $C_0(D)$ -Feller and its generator  $(L, D(L))$  satisfies  $C_c^2(D) \subset D(L)$  and, for  $f \in C_c^2(D)$  and  $x \in D$ ,*

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

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

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

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

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

and  $F, R$  are of Lévy–Khintchine form

$$\begin{aligned} F(u) &= \langle u, au \rangle + \langle b, u \rangle + \int_D (e^{\langle u, \xi \rangle} - 1 - \mathbb{1}_{\{|\xi| \leq 1\}} \langle \xi_J, u_J \rangle) \nu(d\xi), \\ R_i(u) &= \langle u, \alpha_i u \rangle + \sum_{k=1}^d \beta_{ki} u_k + \int_D (e^{\langle u, \xi \rangle} - 1 - \langle u, \xi \rangle) \mu_i(d\xi), \quad i \in I. \end{aligned}$$

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

**REMARK 1.3.** Let  $(a, \alpha, b, \beta, \nu, \mu)$  be admissible parameters. According to [16], Lemma 10.1 and Lemma 10.2, the martingale problem with generator  $L$  and domain  $C_c^\infty(D)$  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  $(a, \alpha, b, \beta, \nu, \mu)$  as the unique solution to the martingale problem with generator  $L$  and domain  $C_c^\infty(D)$ . In any case, it can be constructed as a Markov process on the Skorokhod space over  $D$ .

Affine processes are thus constructed on the canonical state space. In order to prove the main result of this work, we provide in Section 4 a pathwise construction of affine processes. The latter one extends previous cases from the literature such as [15, 19, 48] and [5].

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

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

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  $\rho$ . 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  $P_{t+s} \rho = P_t P_s \rho$ , 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 [18], 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)  $\int_D (P_t f)(x) \pi(dx) = \int_D f(x) \pi(dx)$ , 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  $\pi$ , this distribution has some finite log-moment and, moreover,  $P_t(x, \cdot) \rightarrow \pi$  with exponential rate. For this purpose we use the Wasserstein distance on  $\mathcal{P}(D)$  introduced below. Given  $\rho, \tilde{\rho} \in \mathcal{P}(D)$ , a coupling  $H$  of  $(\rho, \tilde{\rho})$  is a Borel probability measure on  $D \times D$  which has marginals  $\rho$  and  $\tilde{\rho}$ , respectively, that is, for  $f, g \in B(D)$  it holds that

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

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

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

and let

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

- (b) Introduce, for  $x = (y, z), \tilde{x} = (\tilde{y}, \tilde{z}) \in \mathbb{R}_+^m \times \mathbb{R}^n$ ,

$$d_{\log}(x, \tilde{x}) = \log(1 + \mathbb{1}_{\{n>0\}} |y - \tilde{y}|^{1/2} + |x - \tilde{x}|),$$

and let

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

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

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

The appearance of the additional factor  $\mathbb{1}_{\{n>0\}} |y - \tilde{y}|^{1/2}$  is purely technical, it 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 separable metric space, see, for example, [57], Theorem 6.18. A characterization of convergence with respect to  $W_{d_\kappa}$  and  $W_{d_{\log}}$  is given in the following remark, see also [57], 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)  $W_d(\rho_n, \rho) \rightarrow 0$  as  $n \rightarrow \infty$ .

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

$$\int_D f(x)\rho_n(dx) \rightarrow \int_D f(x)\rho(dx), \quad n \rightarrow \infty.$$

(iii)  $\rho_n \rightarrow \rho$  weakly as  $n \rightarrow \infty$ , and

$$\int_D d(x, 0)\rho_n(dx) \rightarrow \int_D d(x, 0)\rho(dx), \quad n \rightarrow \infty.$$

(iv)  $\rho_n \rightarrow \rho$  weakly as  $n \rightarrow \infty$ , and

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

It is easy to see that  $\mathcal{P}_\kappa(D) \subset \mathcal{P}_{\log}(D)$  and  $W_{d_{\log}} \leq C_\kappa W_{d_\kappa}$ , for some constant  $C_\kappa > 0$ , that is,  $W_{d_\kappa}$  is stronger than  $W_{d_{\log}}$ . The following is our main result.

THEOREM 1.5. Let  $(a, \alpha, b, \beta, \nu, \mu)$  be admissible parameters. Suppose that  $\beta$  has only eigenvalues with strictly negative real parts, and

$$(1.5) \quad \int_{|\xi| > 1} \log(|\xi|)\nu(d\xi) < \infty.$$

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

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

$$(1.6) \quad W_{d_{\log}}(P_t \rho, \pi) \leq K \min\{e^{-\delta t}, W_{d_{\log}}(\rho, \pi)\} + K e^{-\delta t} W_{d_{\log}}(\rho, \pi).$$

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

$$(1.7) \quad \int_{|\xi| > 1} |\xi|^\kappa \nu(d\xi) < \infty,$$

then  $\pi \in \mathcal{P}_\kappa(D)$  and there exists constants  $K', \delta' > 0$  such that, for all  $\rho \in \mathcal{P}_\kappa(D)$  and  $t \geq 0$ ,

$$(1.8) \quad W_{d_\kappa}(P_t \rho, \pi) \leq K' W_{d_\kappa}(\rho, \pi) e^{-\delta' t}.$$

It is worthwhile to mention that to our knowledge a convergence rate solely under a log-moment condition on the state-independent jump measure was not even obtained for one-dimensional CBI processes. In order that  $W_{d_{\log}}(P_t \rho, \pi)$  and  $W_{d_\kappa}(P_t \rho, \pi)$  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

$$(1.9) \quad W_d(P_t(x, \cdot), \pi) \leq K e^{-\delta t} (1 + W_d(\delta_x, \pi)), \quad t \geq 0, x \in D,$$

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  $P_t(x, \cdot) \rightarrow \pi$  weakly as  $t \rightarrow \infty$  with exponential rate.

Let  $X = (X_t)_{t \geq 0}$  be an affine process. For the parameter estimation of affine models, see, for example, [3, 47] and [2], it is often necessary to prove a Birkhoff ergodic theorem, that is,

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

holds almost surely for sufficiently many test functions  $f$ . Using classical theory, see, for example, [51], Theorem 17.1.7, and [53], such convergence is implied by the ergodicity in the total variation distance, that is, by

$$(1.11) \quad \lim_{t \rightarrow \infty} \|P_t(x, \cdot) - \pi\|_{TV} = 0, \quad x \in D,$$

where  $\|\cdot\|_{TV}$  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  $d(x, \tilde{x}) = 1 \wedge |x - \tilde{x}|$  was recently studied in [53]. Applying the main result of [53] 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  $(a, \alpha, b, \beta, \nu, \mu)$  be admissible parameters. Suppose that  $\beta$  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 literature.** Consider an Ornstein–Uhlenbeck process on  $\mathbb{R}^n$ , that is, an affine process on state space  $D = \mathbb{R}^n$  with admissible parameters  $(a, \alpha = 0, b, \beta, \nu, \mu = 0)$ . If  $\beta$  has only eigenvalues with strictly negative real parts and (1.5) is satisfied, then [54] 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  $t \rightarrow \infty$ . Under additional technical conditions on the measure  $\nu$ , it follows that the corresponding process also satisfies (1.11) with an exponential rate, see [58]. Since in view of Theorem 1.5 the convergence (in the Wasserstein distance) has already exponential rate, we conclude that the additional restriction on  $\nu$  imposed in [58] is only used to guarantee that convergence takes place in the stronger total variation distance, that is, it is not necessary for the speed of convergence.

Consider a subcritical multitype CBI process on  $\mathbb{R}_+^m$ , that is, an affine process on state space  $D = \mathbb{R}_+^m$  for which the parameter  $\beta$  has only eigenvalues with strictly negative real parts. In dimension  $m = 1$ , Pinsky [52] announced (without proof) the existence of a limiting distribution under condition (1.5). A proof of this fact was then given in [42], Theorem 3.16, while in [46], Theorem 3.20 and Corollary 3.21, it was shown that  $P_t(x, \cdot) \longrightarrow \pi$  is equivalent to (1.5). Some properties of the invariant distribution  $\pi$  have been studied in [39]. In [47] exponential ergodicity in total variation distance, see (1.11), was established for one-dimensional subcritical CBI processes with  $\nu = 0$ , while some other related results for stochastic equations on  $\mathbb{R}_+$  have been recently considered in [24]. An extension of the techniques from [47] to arbitrary dimension  $m \geq 2$  is still an interesting open problem. Recently, in [50] another approach for the exponential ergodicity in the total variation distance for affine processes on cones, which include multitype CBI processes as well as matrix-valued affine processes such as the important case of the Wishart process, was provided. Their techniques were closely re-

lated to stochastic stability properties of Markov processes in the sense of Meyn and Tweedie [51], see also the references therein. More precisely, it was shown that each subcritical affine process  $X$  on a proper closed convex cone which is  $\rho$ -irreducible, aperiodic and has finite second moments, where  $\rho$  is a reference measure with its support having nonempty interior, is exponentially ergodic in the total variation distance. As such a result is formulated in a very general way, it becomes a delicate mathematical task to show that such conditions are satisfied for affine 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 [47] for  $m = 1$  and in Section 2 of this work. In contrast, our results can be applied in arbitrary dimension without the need to prove irreducibility or aperiodicity, paying the price that we use the Wasserstein distance instead. Let us mention that recently also asymptotic results for supercritical CBI processes have been obtained in [8, 9, 45].

Consider now the general case of an *affine process* on the canonical state space  $D = \mathbb{R}_+^m \times \mathbb{R}^n$ . Based on the stability theory of Markov chains in the sense of Meyn and Tweedie the long-time behavior of some particular two-dimensional models on state-space  $D = \mathbb{R}_+ \times \mathbb{R}$  was studied in [4, 33]. These results have been further developed in [60] for arbitrary dimensions, where also functional limit theorems were obtained. In order to prove irreducibility and aperiodicity, the authors supposed that the diffusion component is nondegenerate and that  $\nu$  and  $\mu_1, \dots, \mu_m$  are probability measures, that is, the corresponding affine process has only finitely many jumps on bounded time intervals  $[0, T]$ ,  $T > 0$ . Independently in [34], the following result was obtained.

**THEOREM 1.8 ([34]).** *Let  $(a, \alpha, b, \beta, \nu, \mu)$  be admissible parameters. Suppose that  $\beta$  has only eigenvalues with negative real parts and (1.5) is satisfied. Then there exists a unique invariant distribution  $\pi$  for  $(P_t)_{t \geq 0}$ . Moreover,  $\pi$  has Laplace transform*

$$(1.12) \quad \int_D e^{(u, x)} \pi(dx) = \exp\left(\int_0^\infty F(\psi(t, u)) dt\right), \quad u \in \mathcal{U},$$

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

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  $\pi$ , 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.

**1.5. Main idea of proof and structure of the work.** The proof of Theorem 1.5 is divided in four steps as explained below.

*Step 1.* Provide a stochastic description of conservative affine processes. More precisely, in Section 3 we recall a stochastic equation for multitype CBI processes and a comparison principle due to [5]. In Section 4 we prove that each affine process can be obtained as the pathwise unique strong solution  $(X_t(x))_{t \geq 0}$  to a certain stochastic equation, where  $x = (y, z) \in \mathbb{R}_+^m \times \mathbb{R}^n$  denotes the initial condition. The particular structure of this equation shows that the process takes the form  $X_t(x) = (Y_t(y), Z_t(x))$ , where  $(Y_t(y))_{t \geq 0} \subset \mathbb{R}_+^m$  is a CBI process with initial condition  $y$  and  $(Z_t(x))_{t \geq 0}$  is an OU-type process with initial condition  $z$  whose coefficients depend on the process  $(Y_t(y))_{t \geq 0}$ .

*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  $(X_t(\tilde{x}))_{t \geq 0}$  be the affine processes with initial states  $x, \tilde{x} \in \mathbb{R}_+^m \times \mathbb{R}^n$ , 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:

$$(1.13) \quad \mathbb{E}(|X_t(x) - X_t(\tilde{x})|) \leq K e^{-\delta t} (\mathbb{1}_{\{n>0\}} |y - \tilde{y}|^{1/2} + |x - \tilde{x}|), \quad t \geq 0,$$

where  $K, \delta > 0$  are some constants. Indeed, write  $X_t(x) = (Y_t(y), Z_t(x))$  and  $X_t(\tilde{x}) = (Y_t(\tilde{y}), Z_t(\tilde{x}))$ , respectively. Using the comparison principle for the CBI component we prove that

$$(1.14) \quad \mathbb{E}(|Y_t(x) - Y_t(\tilde{x})|) \leq d |y - \tilde{y}| e^{-\delta' t},$$

where  $\delta' > 0$  is some constant. From this and the particular structure of the stochastic equation solved by  $(X_t(x))_{t \geq 0}$  and  $(X_t(\tilde{x}))_{t \geq 0}$  we then easily deduce (1.13). In the literature the proof of similar inequalities to (1.13) and (1.14) is often based on the construction of a successful coupling being typically a difficult task. In the framework of affine processes a surprisingly simple proof of such estimates is given in Section 6 by using monotone couplings as explained above.

*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

$$(1.15) \quad W_{d_\kappa}(P_t \rho, P_t \tilde{\rho}) \leq K e^{-\delta t} W_{d_\kappa}(\rho, \tilde{\rho}), \quad t \geq 0.$$

Next observe that, by the convexity of the Wasserstein distance (see Lemma A.4) combined with (1.3), property (1.15) is implied by

$$(1.16) \quad W_{d_\kappa}(P_t \delta_x, P_t \delta_{\tilde{x}}) \leq K e^{-\delta t} (\mathbb{1}_{\{n>0\}} |y - \tilde{y}|^{1/2} + |x - \tilde{x}|)^\kappa, \quad t \geq 0.$$

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 has  $P_t(x, \cdot) = P_t^0(x, \cdot) * P_t(0, \cdot)$ , where  $*$  denotes the usual convolution of measures. A similar decomposition for affine processes was also used in [34]. Applying now Lemma A.3 and the Jensen inequality gives

$$\begin{aligned} W_{d_\kappa}(P_t \delta_x, P_t \delta_{\tilde{x}}) &\leq W_{d_\kappa}(P_t^0 \delta_x, P_t^0 \delta_{\tilde{x}}) \\ &\leq (W_{d_1}(P_t^0 \delta_x, P_t^0 \delta_{\tilde{x}}))^\kappa \\ &\leq K^\kappa e^{-\delta \kappa t} (\mathbb{1}_{\{n>0\}} |y - \tilde{y}|^{1/2} + |x - \tilde{x}|)^\kappa, \end{aligned}$$

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

## 2. Examples.

*2.1. Anisotropic stable JCIR process.* Let  $Z_1, Z_2$  be independent one-dimensional Lévy processes with characteristic function

$$\Psi_j(\xi) = \int_0^\infty (e^{i\xi z} - 1 - i\xi z) \frac{dz}{z^{1+\gamma_j}}, \quad \xi \in \mathbb{R}, j = 1, 2,$$

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

$$\Psi_\nu(\xi) = \int_{\mathbb{R}_+^2} (e^{-i\langle \xi, z \rangle} - 1) \nu(dz), \quad \xi \in \mathbb{R}_+^2,$$

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

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

Suppose that  $Z$  and  $S$  are independent. Applying the results of [5] to this particular case shows that, for each  $x \in \mathbb{R}_+^2$ , there exists a pathwise unique strong solution to

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

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  $\nu$ ,

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

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

**COROLLARY 2.1.** *If  $\beta$  has only eigenvalues with negative real parts and  $\nu$  satisfies*

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

*then the assertions of Theorem 1.5 are true.*

Convergence in total variation distance for a similar one-dimensional model was studied in [47]. Similar two-dimensional processes were also studied in [4] and [32]. In view of our main result Theorem 1.5, it is straightforward to extend this model to arbitrary dimension  $d \geq 2$ , with possibly nonvanishing diffusion part and more general driving noise of Lévy type. It is also worth mentioning that the analogue of this model for dimension  $d \geq 2$  was recently studied in [23], where the strong Feller property and, combined with the results of this work, exponential ergodicity in total variation were shown.

**2.2. Stochastic volatility model.** Let  $D = \mathbb{R}_+ \times \mathbb{R}$ , that is,  $m = n = 1$ . Let  $(V, Y)$  be the unique strong solution to

$$dV(t) = (b_1 + \beta_{11}V(t)) dt + \sqrt{V(t)} dB_1(t) + dJ_1(t),$$

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

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  $\nu_1$ , and  $J_2$  a one-dimensional Lévy process with Lévy measure  $\nu_2$ . 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

$$\nu(d\xi) = \nu_1(d\xi_1) \otimes \delta_0(d\xi_2) + \delta_0(d\xi_1) \otimes \nu_2(d\xi_2), \quad \mu_1 = \mu_2 = 0.$$

Then we obtain the following.

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

$$\int_{(1,\infty)} \log(\xi_1) \nu_1(d\xi_1) + \int_{|\xi_2|>1} \log(|\xi_2|) \nu_2(d\xi_2) < \infty,$$

*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 multitype CBI processes.** In this section we recall some results for the particular case of multitype CBI processes, that is, affine processes on state space  $D = \mathbb{R}_+^m$  (i.e.,  $n = 0$ ). For further references and additional explanations we refer to [5] and [8]. Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a complete probability space rich enough to support the following objects:

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

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

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

(B3) Poisson random measures  $N_1^I(ds, d\xi, dr), \dots, N_m^I(ds, d\xi, dr)$  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  $\mu_1^I, \dots, \mu_m^I$  are Borel measures supported on  $\mathbb{R}_+^m$  satisfying, for  $i \in I$ ,

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

The objects  $W, M_I, N_1^I, \dots, N_m^I$  are supposed to be mutually independent. The corresponding compensated Poisson random measures are denoted by

$$\widetilde{M}_I(ds, d\xi) = M_I(ds, d\xi) - \widehat{M}_I(ds, d\xi)$$

and

$$\widetilde{N}_i^I(ds, d\xi, dr) = N_i^I(ds, d\xi, dr) - \widehat{N}_i^I(ds, d\xi, dr), \quad i \in I.$$

Here and below we consider the natural augmented filtration generated by  $W, M_I, N_1^I, \dots, N_m^I$ . 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  $j \in I \setminus \{i\}$ .
- (c) Elements  $c_1, \dots, c_m \geq 0$  and a matrix

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

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

$$\begin{aligned}
 (3.1) \quad Y_t &= y + \int_0^t (b + \tilde{\beta} Y_s) ds + \int_0^t \sigma(Y_s) dW_s + \int_0^t \int_{\mathbb{R}_+^m} \xi M_I(ds, d\xi) \\
 &+ \sum_{i=1}^m \int_0^t \int_{|\xi| \leq 1} \int_{\mathbb{R}_+} \xi \mathbb{1}_{\{r \leq Y_{s-,i}\}} \tilde{N}_i^I(ds, d\xi, dr) \\
 &+ \sum_{i=1}^m \int_0^t \int_{|\xi| > 1} \int_{\mathbb{R}_+} \xi \mathbb{1}_{\{r \leq Y_{s-,i}\}} N_i^I(ds, d\xi, dr),
 \end{aligned}$$

where  $\tilde{\beta}_{ji} = \beta_{ji} - \int_{|\xi| > 1} \xi_j \mu_i^I(d\xi)$ . Pathwise uniqueness for a slightly more complicated equation was recently obtained in [5], while (3.1) in this form appeared first in [8]. The following is essentially due to [5].

**PROPOSITION 3.1.** *Let  $(b, \beta, \sigma)$  be as in (a)–(c), and consider objects  $W, M_I, N_1^I, \dots, N_m^I$  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 multitype CBI process starting from  $y$ , and the generator  $L_Y$  of  $Y$  is, for  $f \in C_c^2(\mathbb{R}_+^m)$ , of the following form*

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

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

The proof of the pathwise uniqueness is based on a comparison principle for multitype CBI processes, see [5], Lemma 4.2. This comparison principle is stated below.

**LEMMA 3.2** ([5], Lemma 4.2). *Let  $(Y_t)_{t \geq 0}$  be a weak solution to (3.1) with parameters  $(b, \beta, \sigma)$ , let  $(Y'_t)_{t \geq 0}$  be another weak solution to (3.1) with parameters  $(b', \beta, \sigma)$ , where  $(b, \beta, \sigma)$  and  $(b', \beta, \sigma)$  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*

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

**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. In the two-dimensional case  $D = \mathbb{R}_+ \times \mathbb{R}$  such a result was first obtained in [15]. Independently, the case of affine diffusions on the canonical state space  $D = \mathbb{R}_+^m \times \mathbb{R}^n$  (i.e., processes without jumps) was studied in [19]. The main obstacle there is related with the diffusion component which is degenerate at the boundary but also has a nontrivial structure in higher dimensions. In order to take this into account we use, compared to [19], another representation of the diffusion matrix (see (A0) and (A1) below). Such a representation is used to decompose

the corresponding affine process into a CBI and an OU component which are then treated separately. Consequently, based on the available results for CBI processes, the proofs in this section become relatively simple.

Let  $(a, \alpha, b, \beta, \nu, \mu)$  be admissible parameters. For the parameters  $a$  and  $\alpha = (\alpha_1, \dots, \alpha_m)$  consider the following objects:

(A0) An  $n \times n$ -matrix  $\sigma_a$  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

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

The first condition is simple to check. Indeed, by definition, one has  $a \in S_d^+$ ,  $a_{II} = a_{JJ} = a_{JI} = 0$ . Thus  $a_{JJ}$  is symmetric and positive semidefinite. Hence  $\sigma_a$  denotes the nonnegative square root of  $a_{JJ}$ . Concerning the second condition we use the following lemma.

LEMMA 4.1. *Let  $\alpha_1, \dots, \alpha_m \in S_d^+$  be such that  $(\alpha_j)_{kl} = \alpha_{j,jj} \delta_{kj} \delta_{lj}$  for  $1 \leq k, l \leq m$ . Then there exist matrices  $\sigma_1, \dots, \sigma_m \in \mathbb{R}^{d \times d}$  such that condition (A1) is satisfied.*

PROOF. Fix  $j \in I$ . Since  $\alpha_j \in S_d^+$  has the block structure

$$\alpha_j = \begin{pmatrix} \alpha_{j,II} & \alpha_{j,IJ} \\ \alpha_{j,JI} & \alpha_{j,JJ} \end{pmatrix},$$

the characterization of positive semidefinite block matrices (see [26], Theorem 16.1) yields

$$(4.2) \quad \alpha_{j,II} \in S_m^+, \quad \alpha_{j,JJ} - \alpha_{j,JI} \alpha_{j,II}^{-1} \alpha_{j,IJ} \in S_n^+, \quad \alpha_{j,IJ} = \alpha_{j,II} \alpha_{j,II}^{-1} \alpha_{j,IJ},$$

where  $\alpha_{j,II}^{-1}$  denotes the pseudoinverse of  $\alpha_{j,II}$ . Using the diagonal structure of  $\alpha_{j,II}$  we find that

$$(\alpha_{j,II}^{-1})_{kl} = \delta_{kj} \delta_{lj} \begin{cases} \alpha_{j,jj}^{-1}, & \alpha_{j,jj} > 0, \\ 0, & \alpha_{j,jj} = 0, \end{cases} \quad k, l \in I.$$

If  $\alpha_{j,jj} = 0$ , then  $\alpha_{j,II} = \alpha_{j,II}^{-1} = 0$  and hence, by (4.2),  $\alpha_{j,IJ} = 0$  and  $\alpha_{j,JI} = 0$ . Letting

$$\sigma_j = \begin{pmatrix} \sigma_{j,II} & 0 \\ \sigma_{j,JI} & \sigma_{j,JJ} \end{pmatrix}$$

be given such that  $\sigma_{j,JJ} \sigma_{j,JJ}^\top = \alpha_{j,JJ}$ ,  $\sigma_{j,II} = 0$  and  $\sigma_{j,JI} = 0$ , we find that  $\sigma_j \sigma_j^\top = \alpha_j$  has the desired form (4.1).

Suppose now that  $\alpha_{j,jj} > 0$ . Using the last equality in (4.2) we find that

$$(4.3) \quad (\alpha_{j,IJ})_{kl} = 0 = (\alpha_{j,JI})_{lk}, \quad k \in I \setminus \{j\}, l \in J.$$

Put  $(\sigma_{j,II})_{k,l} = \alpha_{j,jj}^{1/2} \delta_{kj} \delta_{lj}$ ,  $\sigma_{j,IJ} = 0$ ,  $(\sigma_{j,JI})_{k,l} = \alpha_{kj} \alpha_{j,jj}^{-1/2} \delta_{jl}$ ,  $k \in J, l \in I$  and let  $\sigma_{j,JJ}$  be given such that

$$\sigma_{j,JJ} \sigma_{j,JJ}^\top = \alpha_{j,JJ} - \alpha_{j,JI} \alpha_{j,II}^{-1} \alpha_{j,IJ}.$$

Note that the existence of such  $\sigma_{j,JJ}$  follows from (4.2). By direct computation one finds that  $\alpha_{j,IJ} = \sigma_{j,II} \sigma_{j,JI}^\top$ ,  $\alpha_{j,JI} = \sigma_{j,JI} \sigma_{j,II}^\top$  and  $\alpha_{j,JI} \alpha_{j,II}^{-1} \alpha_{j,IJ} = \sigma_{j,JI} \sigma_{j,II}^{-1} \alpha_{j,IJ}$ , from which we deduce  $\alpha_{j,JJ} = \sigma_{j,JI} \sigma_{j,JI}^\top + \sigma_{j,JJ} \sigma_{j,JJ}^\top$ . This shows that  $\sigma_j \sigma_j^\top = \alpha_j$  and, moreover, it is clear that  $\sigma_j$  has the desired form (4.1). Since  $j \in I$  was arbitrary, the assertion is proved. □

Note that (4.3) is already assumed in the definition of admissible parameters. The proof shows that it is a simple consequence of the particular structure of  $\alpha_{j,II}$  and (4.2).

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  $W^i = (W_t^i)_{t \geq 0}$ .

(A4) A Poisson random measure  $M(ds, d\xi)$  with compensator  $\widehat{M}(ds, d\xi) = ds\nu(d\xi)$  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. The corresponding compensated Poisson random measures are denoted by

$$\widetilde{M}(ds, d\xi) = M(ds, d\xi) - \widehat{M}(ds, d\xi)$$

and

$$\widetilde{N}_i(ds, d\xi, dr) = N_i(ds, d\xi, dr) - \widehat{N}_i(ds, d\xi, dr), \quad i \in I.$$

Here and below we consider the natural augmented filtration generated by these noise terms. For  $x \in D$ , consider the stochastic equation

$$\begin{aligned} X_t = x &+ \int_0^t (\widetilde{b} + \widetilde{\beta}X_s) ds + \sqrt{2} \begin{pmatrix} 0 \\ \sigma_a B_t \end{pmatrix} + \sum_{i \in I} \int_0^t \sqrt{2X_{s,i}} \sigma_i dW_s^i \\ &+ \int_0^t \int_{|\xi| \leq 1} \xi \widetilde{M}(ds, d\xi) + \int_0^t \int_{|\xi| > 1} \xi M(ds, d\xi) \\ (4.4) \quad &+ \sum_{i \in I} \int_0^t \int_{|\xi| \leq 1} \int_{\mathbb{R}_+} \xi \mathbb{1}_{\{r \leq X_{s-,i}\}} \widetilde{N}_i(ds, d\xi, dr) \\ &+ \sum_{i \in I} \int_0^t \int_{|\xi| > 1} \int_{\mathbb{R}_+} \xi \mathbb{1}_{\{r \leq X_{s-,i}\}} N_i(ds, d\xi, dr), \end{aligned}$$

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

$$(4.5) \quad \widetilde{b}_i = b_i + \mathbb{1}_I(i) \int_{|\xi| \leq 1} \xi_i \nu(d\xi), \quad \widetilde{\beta}_{ki} = \beta_{ki} - \mathbb{1}_I(i) \int_{|\xi| > 1} \xi_k \mu_i(d\xi).$$

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 (4.4), we obtain the following.

**THEOREM 4.2.** *Let  $(a, \alpha, b, \beta, \nu, \mu)$  be admissible parameters. Then, for each  $x \in D$ , there exists a pathwise unique  $D$ -valued strong solution  $X = (X_t)_{t \geq 0}$  to (4.4).*

This result will be proved later in this section. Let us first relate (4.4) to affine processes.

PROPOSITION 4.3. *Let  $(a, \alpha, b, \beta, \nu, \mu)$  be admissible parameters. Then each solution  $X$  to (4.4) is an affine process with admissible parameters  $(a, \alpha, b, \beta, \nu, \mu)$  and starting point  $x$ .*

PROOF. Let  $X$  be a solution to (4.4) and  $f \in C_c^2(D)$ . Applying the Itô formula shows that

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

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

$$\begin{aligned} \mathbb{E}\left(\sup_{s \in [0, t]} |M_f(t)|\right) &\leq 2\|f\|_\infty + \int_0^t \mathbb{E}(|Lf(X_s)|) ds \\ &\leq 2\|f\|_\infty + t\|Lf\|_\infty < \infty, \end{aligned}$$

for  $t \geq 0$ , 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  $(a, \alpha, b, \beta, \nu, \mu)$ .  $\square$

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

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 [44], namely, the equivalence between weak solutions to stochastic equations and martingale problems. Alternatively, following [16], page 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 [16], Theorem 2.12). The assertion is then a consequence of the equivalence between weak solutions to stochastic equations and semimartingales (see [30], Chapter III, Theorem 2.26).  $\square$

In view of the Yamada–Watanabe Theorem (see [6] or [43] for a more general account on this topic), Theorem 4.2 is proved once we have shown pathwise uniqueness for (4.4). For this purpose we rewrite (4.4) 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  $e_1, \dots, e_d$  denote the canonical basis vectors in  $\mathbb{R}^d$ . Then (4.4) is equivalent to the system of equations

$$\begin{aligned} (4.6) \quad Y_t &= y + \int_0^t (b_I + \tilde{\beta}_{II} Y_s) ds + \sum_{i \in I} e_i \int_0^t \sqrt{2\alpha_{i,ii} Y_{s,i}} dW_{s,i}^i \\ &+ \int_0^t \int_D \xi_I M(ds, d\xi) + \sum_{i \in I} \int_0^t \int_{|\xi| \leq 1} \int_{\mathbb{R}_+} \xi_I \mathbb{1}_{\{r \leq Y_{s-,i}\}} \tilde{N}_i(ds, d\xi, dr) \\ &+ \sum_{i \in I} \int_0^t \int_{|\xi| > 1} \int_{\mathbb{R}_+} \xi_I \mathbb{1}_{\{r \leq Y_{s-,i}\}} N_i(ds, d\xi, dr), \end{aligned}$$

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

Observe that the first equation for  $Y$  does not involve  $Z$ . We will show that (4.6) is precisely (3.1), that is,  $Y$  is a multitype 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 (4.7) is obvious. These ideas are summarized in the next lemma.

LEMMA 4.5. *Let  $(a, \alpha, b, \beta, \nu, \mu)$  be admissible parameters. Then pathwise uniqueness holds for (4.6) and (4.7), and hence for (4.4).*

PROOF. Let  $X = (Y, Z)$  and  $X' = (Y', Z')$  be two solutions to (4.4) with the same initial condition  $x = (y, z) \in D$  both defined on the same probability space. Then  $Y$  and  $Y'$  both satisfy (4.6). Let us show that (4.6) is precisely (3.1). Set  $\text{pr}_I : D \rightarrow \mathbb{R}_+^m$ ,  $\text{pr}_I(x) = (x_i)_{i \in I}$ , and define:

- A  $m$ -dimensional Brownian motion  $W_t := (W_{t,1}^1, \dots, W_{t,m}^m)$ .
- A Poisson random measure  $M_I(ds, d\xi)$  on  $\mathbb{R}_+ \times \mathbb{R}_+^m$  by

$$M_I([s, t] \times A) = M([s, t] \times \text{pr}_I^{-1}(A)),$$

where  $0 \leq s < t$  and  $A \subset \mathbb{R}_+^m$  is a Borel set.

- Poisson random measures  $N_1^I, \dots, N_m^I$  on  $\mathbb{R}_+ \times \mathbb{R}_+^m \times \mathbb{R}_+$  by

$$N_i^I([s, t] \times A \times [c, d]) = N_i([s, t] \times \text{pr}_I^{-1}(A) \times [c, d]), \quad i \in I,$$

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

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

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

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

Then (4.6) 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 (4.7). Define, for  $l \geq 1$ , a stopping time  $\inf\{t > 0 \mid \max\{|Z_t|, |Z'_t|\} > l\}$ . Since  $Z$  and  $Z'$  both satisfy (4.7) for the same  $Y$ , we obtain

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

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 Gronwall 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  $l \rightarrow \infty$  proves the assertion.  $\square$

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

**PROPOSITION 5.1.** *Let  $(a, \alpha, b, \beta, \nu, \mu)$  be admissible parameters. For  $x \in D$ , let  $X$  be the unique solution to (4.4).*

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

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

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

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

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

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

**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

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

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

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

where  $\tilde{b}$  was defined in (4.5). 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

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

Hence taking expectations in (5.1) gives

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

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 càdlàg paths and  $C'$  is independent of  $l$ , we may take the limit  $l \rightarrow \infty$  and conclude the assertion by the lemma of Fatou.  $\square$

We close this section with a formula for the first moment of general affine processes. The particular case of multitype CBI processes was treated in [5], Lemma 3.4, while recursion formulas for higher-order moments of multitype CBI processes were provided in [7]. At this point it is worthwhile to mention [14] where polynomial processes (which include affine processes on the canonical state space as a special case) are investigated. These processes are characterized by the property that the moments up to a given order  $p \in \mathbb{N}$  can be computed from an  $p$ -dimensional system of ordinary differential equations. For affine processes on the canonical state space, such equations can be derived from the particular form of the extended generator. Alternatively, the same equations could also be obtained from the Itô formula using the stochastic equation presented in Section 4. The next lemma is a particular case where  $p = 1$ .

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

$$(5.3) \quad \int_{|\xi|>1} |\xi| \nu(d\xi) < \infty.$$

*Let  $(X_t)_{t \geq 0}$  be an affine process obtained from (4.4) with  $X_0 = x \in D$ . Then*

$$\mathbb{E}(X_t) = e^{t\beta} x + \int_0^t e^{s\beta} \bar{b} ds,$$

*where  $\bar{b}_i = b_i + \int_{|\xi|>1} \xi_i \nu(d\xi) + \mathbb{1}_I(i) \int_{|\xi| \leq 1} \xi_i \nu(d\xi)$  holds. Letting  $x = (y, z) \in \mathbb{R}_+^m \times \mathbb{R}^n$  and  $X = (Y, Z) \in \mathbb{R}_+^m \times \mathbb{R}^n$  we find that*

$$\begin{aligned} \mathbb{E}(Y_t) &= e^{t\beta_{II}} y + \int_0^t e^{s\beta_{II}} \bar{b}_I ds, \\ \mathbb{E}(Z_t) &= e^{t\beta_{JJ}} z + \int_0^t e^{s\beta_{JJ}} \bar{b}_J ds + \int_0^t e^{(t-s)\beta_{JJ}} \beta_{JI} e^{s\beta_{II}} y ds \\ &\quad + \int_0^t \int_0^s e^{(t-s)\beta_{JJ}} \beta_{JI} e^{u\beta_{II}} \bar{b}_I du ds. \end{aligned}$$

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 (4.4) gives

$$\mathbb{E}(X_t) = x + \int_0^t (\bar{b} + \beta \mathbb{E}(X_s)) ds.$$

Solving this equation gives the desired formula for  $\mathbb{E}(X_t)$ . Taking expectations in (3.1) (or (4.6)) gives

$$\mathbb{E}(Y_t) = y + \int_0^t (\bar{b}_I + \beta_{II} \mathbb{E}(Y_s)) ds,$$

which implies the desired formula for  $\mathbb{E}(Y_t)$ . Finally, taking expectations in (4.7) gives

$$\mathbb{E}(Z_t) = z + \int_0^t (\bar{b}_J + \beta_{JJ} \mathbb{E}(Y_s) + \beta_{JJ} \mathbb{E}(Z_s)) ds.$$

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

**6. Contraction estimate for trajectories of affine processes.** The following is our main estimate for this section.

**PROPOSITION 6.1.** *Let  $(a, \alpha, b, \beta, v, \mu)$  be admissible parameters, suppose that (5.3) is satisfied, and assume that  $\beta$  has only eigenvalues with strictly negative real parts. Let  $x = (y, z), \tilde{x} = (\tilde{y}, \tilde{z}) \in \mathbb{R}_+^m \times \mathbb{R}^n$ , and let  $X(x) = (Y(y), Z(x))$  and  $X(\tilde{x}) = (Y(\tilde{y}), Z(\tilde{x}))$  be the unique strong solutions to (4.4) with initial condition  $x$  and  $\tilde{x}$ , respectively. Then there exist constants  $K, \delta, \delta' > 0$  independent of  $X(x)$  and  $X(\tilde{x})$  such that, for all  $t \geq 0$ ,*

$$(6.1) \quad \mathbb{E}(|Y_t(y) - Y_t(\tilde{y})|) \leq dK|y - \tilde{y}|e^{-\delta't},$$

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

**PROOF.** Here and below we denote by  $K > 0$  a generic constant which may vary from line to line. Let us first prove (6.1). Note that  $Y(y)$  and  $Y(\tilde{y})$  are multi-type CBI processes with the same parameters. If  $\tilde{y}_j \leq y_j$  for all  $j \in \{1, \dots, m\}$ , then we obtain from Lemma 3.2 and Lemma 5.2

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

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

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

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  $y_k^j = y_k^{j+1}$  if  $k \neq j+1$ , and either  $y_{j+1}^j \leq y_{j+1}^{j+1}$  or  $y_{j+1}^j \geq y_{j+1}^{j+1}$ . In any case, we obtain from the previous consideration

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

where we have used  $|y^j - y^{j+1}| = |y_{j+1} - \tilde{y}_{j+1}|$ . 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  $e^{-t\beta} X_t(\tilde{x})$ , and then taking the difference, gives

$$\begin{aligned} X_t(x) - X_t(\tilde{x}) &= e^{t\beta}(x - \tilde{x}) \\ &\quad + \sum_{i \in I} \int_0^t e^{(t-s)\beta} (\sqrt{2Y_{s,i}(x)} - \sqrt{2Y_{s,i}(\tilde{x})}) \sigma_i dW_s^i \\ &\quad + \sum_{i \in I} \int_0^t \int_D \int_{\mathbb{R}_+} e^{(t-s)\beta} \xi \gamma_i(r, s) \tilde{N}_i(ds, d\xi, dr), \end{aligned}$$

where  $\gamma_i(r, s) = \mathbb{1}_{\{r \leq Y_{s,-i}(x)\}} - \mathbb{1}_{\{r \leq Y_{s,-i}(\tilde{x})\}}$ . We find  $\delta_0 > 0$  and  $\delta \in (0, \delta')$  such that

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

The stochastic integral against the Brownian motion is estimated by the Burkholder–Davis–Gundy inequality (shortened to BDG-inequality) also making use of Jensen inequality, Fubini Theorem and using the  $\frac{1}{2}$ -Hölder continuity of  $\sqrt{\cdot}$ , which gives

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

where we have used (6.1) and (6.3). For the stochastic integral against  $\tilde{N}_i$  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 combined with the Fubini Theorem to find, for each  $i \in I$ ,

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

where we have used (6.1), (6.3) and the identity

$$\int_0^\infty |\mathbb{1}_{\{r \leq x\}} - \mathbb{1}_{\{r \leq y\}}|^2 dr = \max(x, y) - \min(x, y) = |x - y|.$$

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

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

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

**7. Proof of Theorem 1.5.**

7.1. *The  $W_{d_{\log}}$ -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  $(a, \alpha, b, \beta, \nu, \mu)$ , suppose that  $\beta$  has only eigenvalues with negative real parts, and (1.5) is satisfied. Then there exist constants  $K, \delta > 0$  such that, for any  $\rho, \tilde{\rho} \in \mathcal{P}_{\log}(D)$ , one has*

$$W_{d_{\log}}(P_t \rho, P_t \tilde{\rho}) \leq K \min\{e^{-\delta t}, W_{d_{\log}}(\rho, \tilde{\rho})\} + K e^{-\delta t} W_{d_{\log}}(\rho, \tilde{\rho}), \quad t \geq 0.$$

PROOF. Let  $(P_t^0(x, \cdot))_{t \geq 0}$  be the transition semigroup with admissible parameters  $(a, \alpha, b = 0, \beta, m = 0, \mu)$  given by Theorem 1.2. Take  $x = (y, z), \tilde{x} = (\tilde{y}, \tilde{z}) \in \mathbb{R}_+^m \times \mathbb{R}^n$  and let  $X^0(x) = (Y^0(y), Z^0(x))$  and  $X^0(\tilde{x}) = (Y^0(\tilde{y}), Z^0(\tilde{x}))$ , respectively, be the corresponding affine processes obtained from (4.4) 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  $X_t^0(\tilde{x})$  has law  $P_t^0(\tilde{x}, \cdot)$ , there exist by Proposition 6.1 constants  $K, \delta > 0$  such that

$$\begin{aligned} W_{d_1}(P_t^0(x, \cdot), P_t^0(\tilde{x}, \cdot)) & \leq \mathbb{E}(\mathbb{1}_{\{n>0\}} |Y_t^0(y) - Y_t^0(\tilde{y})|^{1/2} + |X_t^0(x) - X_t^0(\tilde{x})|) \\ & \leq \mathbb{1}_{\{n>0\}} (\mathbb{E}(|Y_t^0(y) - Y_t^0(\tilde{y})|))^{1/2} \\ & \quad + \mathbb{E}(|X_t^0(x) - X_t^0(\tilde{x})|) \\ & \leq K e^{-\delta t} (\mathbb{1}_{\{n>0\}} |y - \tilde{y}|^{1/2} + |x - \tilde{x}|). \end{aligned}$$

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

$$\begin{aligned} \int_D e^{(u, x')} P_t^0(x, dx') & = \exp(\langle x, \psi(t, u) \rangle), \\ \int_D e^{(u, x')} P_t(0, dx') & = \exp(\phi(t, u)). \end{aligned}$$

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 A.3 from

the [Appendix](#) and then the Jensen inequality for the concave function  $\mathbb{R}_+ \ni a \mapsto \log(1 + a)$ , gives for some generic constant  $K > 0$

$$\begin{aligned}
 W_{d_{\log}}(P_t \delta_x, P_t \delta_{\tilde{x}}) &\leq W_{d_{\log}}(P_t^0 \delta_x, P_t^0 \delta_{\tilde{x}}) \\
 &\leq \log(1 + W_{d_1}(P_t^0 \delta_x, P_t^0 \delta_{\tilde{x}})) \\
 (7.1) \quad &\leq \log(1 + K e^{-\delta t} (\mathbb{1}_{\{n>0\}} |y - \tilde{y}|^{1/2} + |x - \tilde{x}|)) \\
 &\leq K \min\{e^{-\delta t}, \log(1 + \mathbb{1}_{\{n>0\}} |y - \tilde{y}|^{1/2} + |x - \tilde{x}|)\} \\
 &\quad + K e^{-\delta t} \log(1 + \mathbb{1}_{\{n>0\}} |y - \tilde{y}|^{1/2} + |x - \tilde{x}|),
 \end{aligned}$$

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

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

which is proved in the [Appendix](#). Applying now Lemma A.4 from the [Appendix](#) gives for any  $H \in \mathcal{H}(\rho, \tilde{\rho})$

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

Choosing  $H$  as the optimal coupling of  $(\rho, \tilde{\rho})$ , that is,

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

proves the assertion.  $\square$

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

LEMMA 7.2. *Let  $(P_t)_{t \geq 0}$  be the transition semigroup with admissible parameters  $(a, \alpha, b, \beta, \nu, \mu)$ . Suppose that  $\beta$  has only eigenvalues with negative real parts, and (1.5) is satisfied. Then  $(P_t)_{t \geq 0}$  has a unique invariant distribution  $\pi$ . 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  $\tilde{\pi} \in \mathcal{P}_{\log}(D)$ . Observe that, by Proposition 5.1, we easily deduce that  $P_t \mathcal{P}_{\log}(D) \subset \mathcal{P}_{\log}(D)$ , for any  $t \geq 0$ . Fix any

$\rho \in \mathcal{P}_{\log}(D)$  and let  $k, l \in \mathbb{N}$  with  $k > l$ . Then

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

Since the right-hand side tends to zero as  $k, l \rightarrow \infty$ , we conclude that  $(P_k \rho)_{k \in \mathbb{N}}$  is a Cauchy sequence in  $(\mathcal{P}_{\log}(D), W_{d_{\log}})$ . In particular, there exists a limit  $\pi \in \mathcal{P}_{\log}(D)$ , that is,  $W_{d_{\log}}(P_k \rho, \pi) \rightarrow 0$  as  $k \rightarrow \infty$ . Let us show that  $\pi$  is an invariant distribution for  $P_t$ . Indeed, take  $h \geq 0$ , then

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

Since  $W_{d_{\log}}(P_k \rho, \pi) \rightarrow 0$  as  $k \rightarrow \infty$ , we conclude that all terms tend to zero. Hence  $W_{d_{\log}}(P_h \pi, \pi) = 0$ , that is,  $P_h \pi = \pi$ , for all  $h \geq 0$ . Next we prove that  $\pi$  is the unique invariant distribution. Let  $\pi_0, \pi_1$  be any two invariant distributions and define  $W_{d_{\log}}^{\leq 1}$  as in (1.4) with  $d_{\log}$  replaced by  $d_{\log} \wedge 1$ . Then we obtain, for any  $t \geq 0$  and all  $x, \tilde{x} \in D$ , by the proof of Proposition 7.1 (see (7.1))

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

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

$$\begin{aligned} W_{d_{\log}}^{\leq 1}(\pi_0, \pi_1) &= W_{d_{\log}}^{\leq 1}(P_t \pi_0, P_t \pi_1) \\ &\leq \int_{D \times D} W_{d_{\log}}^{\leq 1}(P_t(x, \cdot), P_t(\tilde{x}, \cdot)) H(dx, d\tilde{x}) \\ &\leq \int_{D \times D} \min\{1, \log(1 + 2K e^{-\delta t} |x - \tilde{x}|\}) H(dx, d\tilde{x}). \end{aligned}$$

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

$$\begin{aligned} W_{d_{\log}}(P_t \rho, \pi) &= W_{d_{\log}}(P_t \rho, P_t \pi) \\ &\leq K \min\{e^{-\delta t}, W_{d_{\log}}(\rho, \pi)\} + K e^{-\delta t} W_{d_{\log}}(\rho, \pi), \end{aligned}$$

where we have first used the invariance of  $\pi$  and then Proposition 7.1.  $\square$

7.2. *The  $W_{d_\kappa}$ -Wasserstein estimate.* As before, we start with an estimate with respect to the Wasserstein distance  $W_{d_\kappa}$ .

PROPOSITION 7.3. *Let  $(P_t)_{t \geq 0}$  be the transition semigroup with admissible parameters  $(a, \alpha, b, \beta, \nu, \mu)$ . Suppose that  $\beta$  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  $\rho, \tilde{\rho} \in \mathcal{P}_\kappa(D)$ , one has*

$$W_{d_\kappa}(P_t \rho, P_t \tilde{\rho}) \leq K e^{-\delta t} W_{d_\kappa}(\rho, \tilde{\rho}), \quad t \geq 0.$$

PROOF. Let  $(P_t^0(x, \cdot))_{t \geq 0}$  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

$$(7.2) \quad W_{d_1}(P_t^0(x, \cdot), P_t^0(\tilde{x}, \cdot)) \leq K e^{-\delta t} (\mathbb{1}_{\{m > 0\}} |y - \tilde{y}|^{1/2} + |x - \tilde{x}|),$$

and  $P_t(x, \cdot) = P_t^0(x, \cdot) * P_t(0, \cdot)$ . Then we obtain from Lemma A.3 from the Appendix

$$\begin{aligned} W_{d_\kappa}(P_t \delta_x, P_t \delta_{\tilde{x}}) &\leq W_{d_\kappa}(P_t^0 \delta_x, P_t^0 \delta_{\tilde{x}}) \\ &\leq (W_{d_1}(P_t^0 \delta_x, P_t^0 \delta_{\tilde{x}}))^\kappa \\ &\leq K^\kappa e^{-\delta \kappa t} (\mathbb{1}_{\{m > 0\}} |y - \tilde{y}|^{1/2} + |x - \tilde{x}|)^\kappa, \end{aligned}$$

where the second inequality follows from the Jensen inequality and the third is a consequence of (7.2). Using now Lemma A.4 from the Appendix, we conclude for each  $H \in \mathcal{H}(\rho, \tilde{\rho})$

$$\begin{aligned} W_{d_\kappa}(P_t \rho, P_t \tilde{\rho}) &\leq \int_{D \times D} W_{d_\kappa}(P_t \delta_x, P_t \delta_{\tilde{x}}) H(dx, d\tilde{x}) \\ &\leq K^\kappa e^{-\delta \kappa t} \int_{D \times D} (\mathbb{1}_{\{m > 0\}} |y - \tilde{y}| + |x - \tilde{x}|)^\kappa H(dx, d\tilde{x}). \end{aligned}$$

Letting  $H$  to be the optimal coupling gives

$$W_{d_\kappa}(P_t \rho, P_t \tilde{\rho}) \leq K^\kappa e^{-\delta \kappa t} W_{d_\kappa}(\rho, \tilde{\rho}).$$

This proves the assertion.  $\square$

Based on the previous proposition, the proof of the  $W_{d_\kappa}$ -estimate in Theorem 1.5 can be deduced by exactly the same arguments as in Lemma 7.2. So Theorem 1.5 is proved.

## APPENDIX

### A.1. Moment estimates for $V_1$ and $V_2$ .

In this section we prove (5.2).

LEMMA A.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*

$$\mathcal{A}_1(x) \leq C V_1(x), \quad x = (y, z) \in \mathbb{R}_+^m \times \mathbb{R}^n.$$

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  $C = C_\kappa > 0$

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

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  $\delta_{kl}$  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  $|\frac{\partial^2 V_1(x)}{\partial x_k \partial x_l}| \leq C(1 + |x|^2)^{\frac{\kappa-2}{2}}$ . This implies that

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

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

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

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  $|\xi| > 1$ ,

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

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

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

for the integral against  $\nu$ , gives

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

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

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

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

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

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

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

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

LEMMA A.2. *Suppose that the same conditions as in Proposition 5.1(b) are satisfied. Then there exists a constant  $C > 0$  such that*

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

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

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

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

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

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

$$\begin{aligned} V_2(x + \xi) - V_2(x) &\leq V_2(x + \xi) \\ &\leq C \log(1 + |x|^2 + |\xi|^2) \\ &\leq C \log(1 + |x|^2) + C \log(1 + |\xi|^2), \end{aligned}$$

and hence we obtain

$$\begin{aligned} \int_{|\xi|>1} (V_2(x + \xi) - V_2(x)) \nu(d\xi) &\leq C \int_{|\xi|>1} (V_2(x) + V_2(\xi)) \nu(d\xi) \\ &\leq C(1 + V_2(x)). \end{aligned}$$

From the mean value theorem we obtain

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

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

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

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

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

This implies

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

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

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

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

**A.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 A.3. *Let  $f, \tilde{f}, g \in \mathcal{P}_d(D)$ . Then*

$$W_d(f * g, \tilde{f} * g) \leq W_d(f, \tilde{f}).$$

PROOF. Using the Kantorovich duality (see [57], Theorem 5.10, Case 5.16), we obtain

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

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

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

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

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

where we have used again the Kantorovich duality. This completes the proof.  $\square$

The next estimate shows that the Wasserstein distance is convex. For additional details we refer to [57], Theorem 4.8.

LEMMA A.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*

$$\begin{aligned} & W_d \left( \int_D P(x, \cdot)f(dx), \int_D P(x, \cdot)g(dx) \right) \\ & \leq \int_{D \times D} W_d(P(x, \cdot), P(\tilde{x}, \cdot))H(dx, d\tilde{x}). \end{aligned}$$

**A.3. Proof of the elementary inequality with respect to log.** Below we prove the following inequality.

LEMMA A.5. *For any  $a, b \geq 0$  one has*

$$\begin{aligned} \log(1 + ab) &\leq \log(2e - 1) \min\{\log(1 + a), \log(1 + b)\} \\ &\quad + \log(2e - 1) \log(1 + a) \log(1 + b). \end{aligned}$$

PROOF. Using the elementary inequality  $\log(e + ab) \leq \log(e + a) \log(e + b)$ , see [28], we easily obtain

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

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

**Acknowledgments.** The authors would like to thank Jonas Kremer for several discussions on affine processes and pointing out some interesting references on this topic. The authors would like to thank one of the referees for pointing out the reference [14].

Peng Jin is supported by the STU Scientific Research Foundation for Talents (No. NTF18023).

## REFERENCES

- [1] ALFONSI, A. (2015). *Affine Diffusions and Related Processes: Simulation, Theory and Applications*. *Bocconi & Springer Series* **6**. Springer, Cham; Bocconi Univ. Press, Milan. MR3363174
- [2] BARCZY, M., BEN ALAYA, M., KEBAIER, A. and PAP, G. (2018). 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** 1135–1164. MR3769658 <https://doi.org/10.1016/j.spa.2017.07.004>
- [3] BARCZY, M., DÖRING, L., LI, Z. and PAP, G. (2013). On parameter estimation for critical affine processes. *Electron. J. Stat.* **7** 647–696. MR3035268 <https://doi.org/10.1214/13-EJS786>
- [4] BARCZY, M., DÖRING, L., LI, Z. and PAP, G. (2014). Stationarity and ergodicity for an affine two-factor model. *Adv. in Appl. Probab.* **46** 878–898. MR3254346 <https://doi.org/10.1239/aap/1409319564>
- [5] BARCZY, M., LI, Z. and PAP, G. (2015). 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** 129–169. MR3340375
- [6] BARCZY, M., LI, Z. and PAP, G. (2015). Yamada–Watanabe results for stochastic differential equations with jumps. *Int. J. Stoch. Anal.* **2015** Art. ID 460472, 23. MR3298537 <https://doi.org/10.1155/2015/460472>
- [7] BARCZY, M., LI, Z. and PAP, G. (2016). Moment formulas for multitype continuous state and continuous time branching process with immigration. *J. Theoret. Probab.* **29** 958–995. MR3540486 <https://doi.org/10.1007/s10959-015-0605-0>
- [8] BARCZY, M., PALAU, S. and PAP, G. (2018). Almost sure,  $L_1$ - and  $L_2$ -growth behavior of supercritical multi-type continuous state and continuous time branching processes with immigration. Preprint. Available at [arXiv:1803.10176](https://arxiv.org/abs/1803.10176) [math.PR].

- [9] BARCZY, M., PALAU, S. and PAP, G. (2018). Asymptotic behavior of projections of supercritical multi-type continuous state and continuous time branching processes with immigration. Preprint. Available at [arXiv:1806.10559](https://arxiv.org/abs/1806.10559) [math.PR].
- [10] CABALLERO, M. E., PÉREZ GARMENDIA, J. L. and URIBE BRAVO, G. (2013). A Lamperti-type representation of continuous-state branching processes with immigration. *Ann. Probab.* **41** 1585–1627. [MR3098685 https://doi.org/10.1214/12-AOP766](https://doi.org/10.1214/12-AOP766)
- [11] CABALLERO, M. E., PÉREZ GARMENDIA, J. L. and URIBE BRAVO, G. (2017). Affine processes on  $\mathbb{R}_+^n \times \mathbb{R}^n$  and multiparameter time changes. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 1280–1304. [MR3689968 https://doi.org/10.1214/16-AIHP755](https://doi.org/10.1214/16-AIHP755)
- [12] CHAZAL, M., LOEFFEN, R. and PATIE, P. (2018). Smoothness of continuous state branching with immigration semigroups. *J. Math. Anal. Appl.* **459** 619–660. [MR3732548 https://doi.org/10.1016/j.jmaa.2017.10.071](https://doi.org/10.1016/j.jmaa.2017.10.071)
- [13] COX, J. C., INGERSOLL, J. E. JR. and ROSS, S. A. (1985). A theory of the term structure of interest rates. *Econometrica* **53** 385–407. [MR0785475 https://doi.org/10.2307/1911242](https://doi.org/10.2307/1911242)
- [14] CUCHIERO, C., KELLER-RESSEL, M. and TEICHMANN, J. (2012). Polynomial processes and their applications to mathematical finance. *Finance Stoch.* **16** 711–740. [MR2972239 https://doi.org/10.1007/s00780-012-0188-x](https://doi.org/10.1007/s00780-012-0188-x)
- [15] DAWSON, D. A. and LI, Z. (2006). Skew convolution semigroups and affine Markov processes. *Ann. Probab.* **34** 1103–1142. [MR2243880 https://doi.org/10.1214/009117905000000747](https://doi.org/10.1214/009117905000000747)
- [16] DUFFIE, D., FILIPOVIĆ, D. and SCHACHERMAYER, W. (2003). Affine processes and applications in finance. *Ann. Appl. Probab.* **13** 984–1053. [MR1994043 https://doi.org/10.1214/aoap/1060202833](https://doi.org/10.1214/aoap/1060202833)
- [17] DUHALDE, X., FOUCART, C. and MA, C. (2014). On the hitting times of continuous-state branching processes with immigration. *Stochastic Process. Appl.* **124** 4182–4201. [MR3264444 https://doi.org/10.1016/j.spa.2014.07.019](https://doi.org/10.1016/j.spa.2014.07.019)
- [18] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Wiley, New York. [MR0838085 https://doi.org/10.1002/9780470316658](https://doi.org/10.1002/9780470316658)
- [19] FILIPOVIĆ, D. and MAYERHOFER, E. (2009). Affine diffusion processes: Theory and applications. In *Advanced Financial Modelling. Radon Ser. Comput. Appl. Math.* **8** 125–164. de Gruyter, Berlin. [MR2648460 https://doi.org/10.1515/9783110213140.125](https://doi.org/10.1515/9783110213140.125)
- [20] FILIPOVIĆ, D., MAYERHOFER, E. and SCHNEIDER, P. (2013). Density approximations for multivariate affine jump-diffusion processes. *J. Econometrics* **176** 93–111. [MR3084047 https://doi.org/10.1016/j.jeconom.2012.12.003](https://doi.org/10.1016/j.jeconom.2012.12.003)
- [21] FITZSIMMONS, P. J., FRISTEDT, B. and SHEPP, L. A. (1985). The set of real numbers left uncovered by random covering intervals. *Z. Wahrsch. Verw. Gebiete* **70** 175–189. [MR0799145 https://doi.org/10.1007/BF02451427](https://doi.org/10.1007/BF02451427)
- [22] FOUCART, C. and URIBE BRAVO, G. (2014). Local extinction in continuous-state branching processes with immigration. *Bernoulli* **20** 1819–1844. [MR3263091 https://doi.org/10.3150/13-BEJ543](https://doi.org/10.3150/13-BEJ543)
- [23] FRIESEN, M. and JIN, P. (2019). On the anisotropic stable JCIR process. Preprint. Available at [arXiv:1908.05473](https://arxiv.org/abs/1908.05473) [math.PR].
- [24] FRIESEN, M., JIN, P., KREMER, J. and RÜDIGER, B. (2019). Exponential ergodicity for stochastic equations of nonnegative processes with jumps. Preprint. Available at [arXiv:1902.02833](https://arxiv.org/abs/1902.02833) [math.PR].
- [25] FRIESEN, M., JIN, P. and RÜDIGER, B. (2018). Existence of densities for multi-type continuous-state branching processes with immigration. Preprint. Available at [arXiv:1810.00400](https://arxiv.org/abs/1810.00400) [math.PR].
- [26] GALLIER, J. (2011). *Geometric Methods and Applications: For Computer Science and Engineering*, 2nd ed. *Texts in Applied Mathematics* **38**. Springer, New York. [MR2663906 https://doi.org/10.1007/978-1-4419-9961-0](https://doi.org/10.1007/978-1-4419-9961-0)
- [27] GREY, D. R. (1974). Asymptotic behaviour of continuous time, continuous state-space branching processes. *J. Appl. Probab.* **11** 669–677. [MR0408016 https://doi.org/10.2307/3212550](https://doi.org/10.2307/3212550)
- [28] GUSTAVSSON, J., MALIGRANDA, L. and PEETRE, J. (1989). A submultiplicative function. *Nederl. Akad. Wetensch. Indag. Math.* **51** 435–442. [MR1041496](https://doi.org/10.1007/978-1-4419-9961-0)
- [29] HESTON, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.* **6** 327–343. [MR3929676 https://doi.org/10.1093/rfs/6.2.327](https://doi.org/10.1093/rfs/6.2.327)
- [30] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **288**. Springer, Berlin. [MR1943877 https://doi.org/10.1007/978-3-662-05265-5](https://doi.org/10.1007/978-3-662-05265-5)
- [31] JENA, R. P., KIM, K.-K. and XING, H. (2012). Long-term and blow-up behaviors of exponential moments in multi-dimensional affine diffusions. *Stochastic Process. Appl.* **122** 2961–2993. [MR2931348 https://doi.org/10.1016/j.spa.2012.05.007](https://doi.org/10.1016/j.spa.2012.05.007)

- [32] JIN, P., KREMER, J. and RÜDIGER, B. (2017). Exponential ergodicity of an affine two-factor model based on the  $\alpha$ -root process. *Adv. in Appl. Probab.* **49** 1144–1169. MR3732190 <https://doi.org/10.1017/apr.2017.37>
- [33] JIN, P., KREMER, J. and RÜDIGER, B. (2019). Moments and ergodicity of the jump-diffusion CIR process. *Stochastics* **91** 974–997. MR4018158 <https://doi.org/10.1080/17442508.2019.1576686>
- [34] JIN, P., KREMER, J. and RÜDIGER, B. (2020). Existence of limiting distribution for affine processes. *J. Math. Anal. Appl.* **486** 123912. MR4060087 <https://doi.org/10.1016/j.jmaa.2020.123912>
- [35] JIŘINA, M. (1958). Stochastic branching processes with continuous state space. *Czechoslovak Math. J.* **8** (83) 292–313. MR0101554
- [36] KANG, W. and KANG, C. (2014). Large deviations for affine diffusion processes on  $\mathbb{R}_+^n \times \mathbb{R}^n$ . *Stochastic Process. Appl.* **124** 2188–2227. MR3188353 <https://doi.org/10.1016/j.spa.2014.02.002>
- [37] KAWAZU, K. and WATANABE, S. (1971). Branching processes with immigration and related limit theorems. *Teor. Veroyatn. Primen.* **16** 34–51. MR0290475
- [38] KELLER-RESSEL, M. and MAYERHOFER, E. (2015). Exponential moments of affine processes. *Ann. Appl. Probab.* **25** 714–752. MR3313754 <https://doi.org/10.1214/14-AAP1009>
- [39] KELLER-RESSEL, M. and MIJATOVIĆ, A. (2012). On the limit distributions of continuous-state branching processes with immigration. *Stochastic Process. Appl.* **122** 2329–2345. MR2922631 <https://doi.org/10.1016/j.spa.2012.03.012>
- [40] KELLER-RESSEL, M., SCHACHERMAYER, W. and TEICHMANN, J. (2011). Affine processes are regular. *Probab. Theory Related Fields* **151** 591–611. MR2851694 <https://doi.org/10.1007/s00440-010-0309-4>
- [41] KELLER-RESSEL, M., SCHACHERMAYER, W. and TEICHMANN, J. (2013). Regularity of affine processes on general state spaces. *Electron. J. Probab.* **18** no. 43, 17. MR3040553 <https://doi.org/10.1214/EJP.v18-2043>
- [42] KELLER-RESSEL, M. and STEINER, T. (2008). Yield curve shapes and the asymptotic short rate distribution in affine one-factor models. *Finance Stoch.* **12** 149–172. MR2390186 <https://doi.org/10.1007/s00780-007-0059-z>
- [43] KURTZ, T. G. (2007). The Yamada–Watanabe–Engelbert theorem for general stochastic equations and inequalities. *Electron. J. Probab.* **12** 951–965. MR2336594 <https://doi.org/10.1214/EJP.v12-431>
- [44] KURTZ, T. G. (2011). Equivalence of stochastic equations and martingale problems. In *Stochastic Analysis 2010* 113–130. Springer, Heidelberg. MR2789081 [https://doi.org/10.1007/978-3-642-15358-7\\_6](https://doi.org/10.1007/978-3-642-15358-7_6)
- [45] KYPRIANOU, A. E., PALAU, S. and REN, Y.-X. (2018). Almost sure growth of supercritical multi-type continuous-state branching process. *ALEA Lat. Am. J. Probab. Math. Stat.* **15** 409–428. MR3795454 <https://doi.org/10.30757/alea.v15-17>
- [46] LI, Z. (2011). *Measure-Valued Branching Markov Processes. Probability and Its Applications (New York)*. Springer, Heidelberg. MR2760602 <https://doi.org/10.1007/978-3-642-15004-3>
- [47] LI, Z. and MA, C. (2015). Asymptotic properties of estimators in a stable Cox–Ingersoll–Ross model. *Stochastic Process. Appl.* **125** 3196–3233. MR3343292 <https://doi.org/10.1016/j.spa.2015.03.002>
- [48] MA, R. G. (2013). Stochastic equations for two-type continuous-state branching processes with immigration. *Acta Math. Sin. (Engl. Ser.)* **29** 287–294. MR3016529 <https://doi.org/10.1007/s10114-012-1378-3>
- [49] MAYERHOFER, E., MUHLE-KARBE, J. and SMIRNOV, A. G. (2011). A characterization of the martingale property of exponentially affine processes. *Stochastic Process. Appl.* **121** 568–582. MR2763096 <https://doi.org/10.1016/j.spa.2010.11.015>
- [50] MAYERHOFER, E., STELZER, R. and WESTWEBER, J. (2019). Geometric ergodicity of affine processes on cones. *Stochastic Process. Appl.* <https://doi.org/10.1016/j.spa.2019.11.012>
- [51] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253 <https://doi.org/10.1017/CBO9780511626630>
- [52] PINSKY, M. A. (1972). Limit theorems for continuous state branching processes with immigration. *Bull. Amer. Math. Soc.* **78** 242–244. MR0295450 <https://doi.org/10.1090/S0002-9904-1972-12938-0>
- [53] SANDRIĆ, N. (2017). A note on the Birkhoff ergodic theorem. *Results Math.* **72** 715–730. MR3684455 <https://doi.org/10.1007/s00025-017-0681-9>
- [54] SATO, K. and YAMAZATO, M. (1984). Operator-self-decomposable distributions as limit distributions of processes of Ornstein–Uhlenbeck type. *Stochastic Process. Appl.* **17** 73–100. MR0738769 [https://doi.org/10.1016/0304-4149\(84\)90312-0](https://doi.org/10.1016/0304-4149(84)90312-0)
- [55] SHIGA, T. and WATANABE, S. (1973). Bessel diffusions as a one-parameter family of diffusion processes. *Z. Wahrsch. Verw. Gebiete* **27** 37–46. MR0368192 <https://doi.org/10.1007/BF00736006>
- [56] VASICEK, O. (2012). An equilibrium characterization of the term structure [reprint of J. Financ. Econ. **5** (1977), no. 2, 177–188]. In *Financial Risk Measurement and Management. Internat. Lib. Crit. Writ. Econ.* **267** 724–735. Edward Elgar, Cheltenham. MR3235239

- [57] VILLANI, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **338**. Springer, Berlin. MR2459454 <https://doi.org/10.1007/978-3-540-71050-9>
- [58] WANG, J. (2012). On the exponential ergodicity of Lévy-driven Ornstein–Uhlenbeck processes. *J. Appl. Probab.* **49** 990–1004. MR3058984 <https://doi.org/10.1017/s0021900200012833>
- [59] WATANABE, S. (1969). On two dimensional Markov processes with branching property. *Trans. Amer. Math. Soc.* **136** 447–466. MR0234531 <https://doi.org/10.2307/1994726>
- [60] ZHANG, X. and GLYNN, P. W. (2018). Affine jump-diffusions: Stochastic stability and limit theorems. Preprint. Available at [arXiv:1811.00122](https://arxiv.org/abs/1811.00122) [q-fin.MF].