Affine Jump-Diffusions: Stochastic Stability and Limit Theorems

Xiaowei Zhang[†] Peter W. Glynn[*]

Abstract. Affine jump-diffusions constitute a large class of continuous-time stochastic models that are particularly popular in finance and economics due to its analytical tractability. Many methods for parameter estimation of this type of processes generally assume ergodicity in order establish consistency and asymptotic normality of the estimator. In this paper, we study stochastic stability of affine jump-diffusions, which provides the large-sample theoretical support for many estimation procedures for such processes. We establish ergodicity of this class of models by imposing a "strong mean reversion" condition and a mild condition on the distribution of the jumps, i.e. the finiteness of a logarithmic moment. Exponential ergodicity is proved if the jumps have a finite moment of a positive order. In addition, we prove strong laws of large numbers and functional central limit theorems for additive functionals of this class of models.

Key words. affine jump-diffusion; ergodicity; Lyapunov inequality; strong law of large numbers; functional central limit theorem

1 Introduction

Affine jump-diffusion (AJD) processes constitute an important class of continuous time stochastic models that are widely used in finance and econometrics. This class of models is flexible enough to capture various empirical attributes such as stochastic volatility, leverage effects; see, e.g., Barndorff-Nielsen and Shephard (2001). Furthermore, the affine structure permits efficient computation, as a consequence of the fact that the characteristic function of its transient distribution is of an exponential affine form. The transform can then be computed by solving a system of ordinary differential equations (ODEs) of generalized Riccati type; see Duffie et al. (2000). The ability to efficiently compute such characteristic functions then leads to significant tractability both for computing various expectations and probabilities, and for the use of "method of moments" methods for calibrating such models; see Singleton (2001), Bates (2006), Filipović et al. (2013). In addition,

[†]Corresponding author. Department of Management Sciences, College of Business, City University of Hong Kong, Hong Kong. Email: xiaowei.w.zhang@cityu.edu.hk
[*]Department of Management Science and Engineering, Stanford University, CA 94305, U.S.

AJD processes include a number of important special cases: the Ornstein-Uhlenbeck (OU) process, also known as the Vasicek model in Vasicek (1977); the square-root diffusion process, also known as the Cox-Ingersoll-Ross (CIR) model in Cox et al. (1985); and the Heston stochastic volatility model in Heston (1993). Discussion of related extensions can be found in Bates (1996), Duffie and Kan (1996), Bates (2000), Dai and Singleton (2000), Duffie et al. (2000), Barndorff-Nielsen and Shephard (2001), Cheridito et al. (2007), and Collin-Dufresne et al. (2008).

As just noted, the ability to estimate and calibrate AJD models plays a key role in the popularity of such models in finance and economics. Of course, the large-sample theoretical support for estimation procedures for such AJD processes relies upon strong laws of large numbers (SLLNs), functional central limit theorems (FCLTs), and related ergodic theory. Our goal in this paper is to provide the first comprehensive development of the mathematical conditions under which such SLLNs and FCLTs hold for AJD processes. After stating our main results in Section 2, we prove in Section 3 that a canonical AJD is ergodic under mild conditions having to do with mean reversion and the distribution of the jumps. We further study conditions guaranteeing exponential ergodicity. Finally, in Section 4 we develop SLLNs and FCLTs for AJDs.

In terms of related literature, the study of large time "moment explosions" for AJD processes, and its close connection to implied volatility asymptotics, has been studied by, for example, Lee (2004). A class of two-dimensional stochastic volatility models, of which the Heston model is a special case, is analyzed in Andersen and Piterbarg (2007) to identify the "explosion time" at which the moment of a given positive order becomes infinite. Glasserman and Kim (2010) study the exponential moments of multidimensional affine diffusions (ADs) having no jumps. They characterize the domain of finiteness of the moment generating function of the diffusion process as well as the behavior of this domain in the "large time" limit. The existence of a non-degenerate limit of such moments implies tightness of the marginals of the AD, and is closely connected to existence of a stationary distribution. Their approach is based on the stability analysis of the limiting behavior of the Riccati equations that define the AD. Their results are extended in Jena et al. (2012), using a similar approach. See also Keller-Ressel (2011) for related analysis of a two-dimensional affine stochastic volatility model that permits jumps. Our work differs from the above papers primarily in two aspects. First, our model is a general AJD of arbitrary dimension, whereas the models in the above papers are special cases of ours. In particular, the critical conditions for stochastic stability in these papers are special cases of ours. For instance, our stability condition is reduced to those in Glasserman and Kim (2010) and Jena et al. (2012) in the absence of jumps; the stability condition of the so-called "variance" process in Keller-Ressel (2011) is a one-dimensional special case of ours. Second, their work focuses on existence of the limiting distribution, whereas we further establish ergodicity/exponential ergodicity results. The richer results on stochastic stability stem from our different approach in this paper. Our analysis relies on Lyapunov criteria for Markov processes; see, e.g., Meyn and Tweedie (1993c). Related stability theory can be found in Masuda (2004), Barczy et al. (2014), Jin et al. (2016) and Jin et al. (2017), but the AJDs studied there all involve a state-independent jump intensity.

2 Model Formulation and Main Results

We will adopt the following notation throughout the paper.

- We write $\mathbb{R}^d_+ := \{v \in \mathbb{R}^d : v_i \geq 0, i = 1, \ldots, d\}$ and $\mathbb{R}^d_- := \{v \in \mathbb{R}^d : v_i \leq 0, i = 1, \ldots, d\}$.
- A vector $v \in \mathbb{R}^d$ is treated as a column vector, $v^\intercal$ denotes its transpose, $\|v\|$ denotes its Euclidean norm.
- For a matrix A, $A \succeq 0$ means that A is symmetric positive semidefinite and $A \succ 0$ means that A is symmetric positive definite.
- We write $v_{\mathcal{I}} = (v_i : i \in I)$ and $A_{\mathcal{I}\mathcal{J}} = (A_{ij} : i \in I, j \in J)$, where $v \in \mathbb{R}^d$ is a vector, $A \in \mathbb{R}^{d\times d}$ is a matrix, and $\mathcal{I}, \mathcal{J} \subseteq \{1, \ldots, d\}$ are two index sets.
- We use $\mathbf{0}$ to denote a zero vector or a zero matrix, and $\mathrm{Id}(i)$ to denote a matrix with all zero entries except the i-th diagonal entry is 1, regardless of dimension.
- For a set $K \subseteq \mathbb{R}^d$, $\mathbb{I}_K(x)$ denotes the indicator function associated with K, i.e. $\mathbb{I}_K(x) = 1$ if $x \in K$ and 0 otherwise.

We assume as given a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$ equipped with a filtration $\{\mathscr{F}_t : t \geq 0\}$ that satisfies the *usual hypotheses*; see Protter (2003, p.3). Suppose that a stochastic process $X = (X(t) : t \geq 0)$ with state space $\mathcal{X} \subseteq \mathbb{R}^d$ satisfies the following stochastic differential equation (SDE)

$$\begin{aligned} \mathrm{d}X(t) &= \mu(X(t))\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}W(t) + \int_{\mathbb{R}^d} zN(\mathrm{d}t, \mathrm{d}z), \\ X(0) &= x \in \mathcal{X}, \end{aligned} \tag{1}$$

where $W = (W(t) : t \geq 0)$ is a d-dimensional Wiener process and $N(\mathrm{d}t, \mathrm{d}z)$ is a random counting measure on $[0, \infty) \times \mathbb{R}^d$ with compensator measure $\Lambda(X(t\text{-}))\mathrm{d}t\nu(\mathrm{d}z)$; moreover, $\mu : \mathbb{R}^d \mapsto \mathbb{R}^d$, $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d\times d}$, and $\Lambda : \mathbb{R}^d \mapsto \mathbb{R}$ are measurable functions, and ν is a Borel measure on $\mathbb{R}^d$. In the sequel, we will write $\mathbb{P}_x(\cdot) = \mathbb{P}(\cdot|X(0) = x)$ and $\mathbb{P}_\theta(\cdot) = \int_{\mathcal{X}} \mathbb{P}(\cdot|X(0) = x)\theta(\mathrm{d}x)$ for an initial distribution θ; $\mathbb{E}_x$ and $\mathbb{E}_\theta$ denote the corresponding expectation operators.

We call X an *affine jump-diffusion* (AJD) if the drift $\mu(x)$, diffusion matrix $\sigma(x)\sigma(x)^\intercal$, and jump intensity $\Lambda(x)$ are all affine in x, namely,

$$\begin{aligned} \mu(x) &= b + \beta x, \qquad b \in \mathbb{R}^d,\ \beta \in \mathbb{R}^{d\times d} \\ \sigma(x)\sigma(x)^\intercal &= a + \sum_{i=1}^d x_i \alpha_i, \qquad a \in \mathbb{R}^{d\times d},\ \alpha_i \in \mathbb{R}^{d\times d},\ i = 1, \ldots, d \\ \Lambda(x) &= \lambda + \kappa^\intercal x, \qquad \lambda \in \mathbb{R},\ \kappa \in \mathbb{R}^d. \end{aligned} \tag{2}$$

This paper is largely motivated by statistical calibration of AJDs. Most calibration procedures that have been applied to AJDs are based on some estimating equation as follows. Let Ξ denote the collection of unknown parameters. For simplicity, we assume that the process X is discretely sampled at time epochs $\{k\Delta : k = 0, 1, \ldots, n\}$ for some $\Delta > 0$. To estimate Ξ, one judiciously selects a tractable function $h(x, y; \Xi)$ for which $\mathbb{E}[h(X(0), X(\Delta); \Xi)] = 0$, and then solves the equation

$$\frac{1}{n}\sum_{k=1}^{n} h(X((k-1)\Delta), X(k\Delta); \hat{\Xi}_n) = 0,$$

to compute the estimate $\hat{\Xi}_n$. In a situation where the dimension of h is greater than the dimension of Ξ, one can use the generalized method of moments (see Hansen (1982)). Typical choices of h include the marginal characteristic function of the conditional distribution of $X(k\Delta)$ given $X((k-1)\Delta)$ as in Singleton (2001), or $\mathscr{A}g(x)$ for some tractable function g with enough smoothness, where $\mathscr{A}$ is the operator defined in (10) as in Hansen and Scheinkman (1995). See also Duffie and Glynn (2004) for a choice of h that also utilizes the operator $\mathscr{A}$ but in a context where X is sampled at random times rather than deterministic times.

In order to establish consistency and asymptotic normality of $\hat{\Xi}_n$, it is standard to assume positive Harris recurrence as well as certain moment conditions on the function h; see, e.g., Hansen (1982). The SLLNs and FCLTs that we present as part of Theorem 1 and Theorem 2 provide a large-sample theoretical support for establishing these asymptotic properties of the estimator. We refer interested readers to Aït-Sahalia (2007) for an extensive survey on various statistical calibration methods for general jump-diffusions and related assumptions for statistical validity.

2.1 Main Assumptions

The following three assumptions are universal throughout the paper.

Assumption 1. *Let $\mathcal{X} = \mathbb{R}_+^m \times \mathbb{R}^{d-m}$. For each $x \in \mathcal{X}$, there exists a unique $\mathcal{X}$-valued strong solution to SDE* (1) *with coefficients* (2)*.*

Assumption 2. *Let $\mathcal{I} = \{1, \ldots, m\}$ and $\mathcal{J} = \{m+1, \ldots, d\}$ for some $0 \le m \le d$.*

(i) $a \succeq 0$ with $a_{\mathcal{I}\mathcal{I}} = \mathbf{0}$

(ii) $\alpha_i \succeq 0$ and $\alpha_{i,\mathcal{I}\mathcal{I}} = \alpha_{i,ii} \cdot \mathrm{Id}(i)$ for $i \in \mathcal{I}$; $\alpha_i = \mathbf{0}$ for $i \in \mathcal{J}$;

(iii) $b \in \mathbb{R}_+^m \times \mathbb{R}^{d-m}$;

(iv) $\beta_{\mathcal{I}\mathcal{J}} = \mathbf{0}$ and $\beta_{\mathcal{I}\mathcal{I}}$ has non-negative off-diagonal elements;

(v) $\lambda \in \mathbb{R}_+$, $\kappa_{\mathcal{I}} \in \mathbb{R}_+^m$ and $\kappa_{\mathcal{J}} = \mathbf{0}$;

(vi) ν is a probability distribution on $\mathcal{X}$.

Assumption 3. *$a_{\mathcal{J}\mathcal{J}} \succ 0$ and $2b_i > \alpha_{i,ii} > 0$ for $i = 1, \ldots, m$.*

In this paper, we focus on AJDs with *canonical* state space (Assumption 1) and *admissible* parameters (Assumption 2). In the absence of jumps (i.e., $\lambda = 0$ and $\kappa = \mathbf{0}$), the existence and uniqueness of a strong solution to the SDE (1) with coefficients (2) is established in Filipović and Mayerhofer (2009). They first prove the existence of a weak solution, then prove pathwise uniqueness of the solution, and finally apply the Yamada–Watanabe theorem (Karatzas and Shreve 1991, Corollary 5.3.23). The same approach can be used to prove the case of AJDs with a state-independent jump intensity (i.e., $\lambda > 0$ and $\kappa = \mathbf{0}$); see, e.g., Fu and Li (2010).

Clearly, under Assumption 2 both the diffusion matrix and the jump intensity are independent of $x_{\mathcal{J}}$, i.e. $\sigma(x)\sigma(x)^\intercal = a + \sum_{i=1}^m x_i\alpha_i$ and the jump intensity $\Lambda(x) = \lambda + \sum_{i=1}^m x_i\kappa_i$. In financial applications, the first m components $(X_1, \dots, X_m)$ are often used to model volatility processes and thus are referred to as *volatility factors*, whereas the other $(d - m)$ components are referred to as *dependent factors*.

The jumps of the AJDs we study here have finite activity, a consequence of the fact that ν is assumed to be a probability distribution rather than a σ-finite measure. Nevertheless, this restriction is imposed merely for mathematical simplicity; the main results could also be proved for the case of infinite activity at the cost of more involved analysis. One may recognize that SDE (1) with finite activity jumps is the precisely the model proposed in Duffie et al. (2000), which already covers a substantial number of financial and economic applications.

For a one-dimensional AJD such as the CIR model, the condition $2b_i > \alpha_{i,ii} > 0$ in Assumption 3 is known as the Feller condition, which guarantees that the process stays positive. On the other hand, as detailed in Meyn and Tweedie (1993b,c), irreducibility is usually required in order to apply Lyapunov criteria to a continuous time Markov process. For instance, positive Harris recurrence is shown to be equivalent to ergodicity if some skeleton chain is irreducible in Meyn and Tweedie (1993b). Assumption 3 serves this purpose (Proposition 1). Specifically, it is used to prove that X admits a positive transition density. Note that the existence of a transition density for AJDs is established in Filipović et al. (2013) but their proof requires $b_i > \alpha_{i,ii} > 0$ for $i = 1, \dots, m$, which is stronger than our Assumption 3.

2.2 Main Results

Prior to presenting the main results of the paper, let us review several concepts regarding stochastic stability of a Markov process.

Definition 1. A Markov process X with state space $\mathcal{X}$ is called *Harris recurrent* if there exists a non-trivial σ-finite measure φ on $\mathcal{X}$ for which

$$\mathbb{P}_x\left(\int_0^\infty \mathbb{I}_K(X(t))\,\mathrm{d}t = \infty\right) = 1,$$

for all $x \in \mathcal{X}$ and any measurable set K with $\varphi(K) > 0$.

Definition 2. A Harris recurrent Markov process X is called *positive Harris recurrent* if it admits a unique (up to a multiplicative constant) finite invariant measure, which can be normalized to a probability measure that is called the *stationary distribution* of X.

Definition 3. For any measurable function $f : \mathcal{X} \mapsto [1, \infty)$ and any signed-measure φ on $\mathcal{X}$, define the f-*norm* of φ by $\|\varphi\|_f := \sup_{|h|\le f} |\varphi(h)|$, where

$$\varphi(h) := \int_{\mathcal{X}} h(x)\, \varphi(\mathrm{d}x).$$

When $f \equiv 1$, $\|\cdot\|_f$ is called the *total variation norm* and is denoted by $\|\cdot\|$.

The following notation facilitates the presentation of the main theorems and discussions thereafter. In the sequel, we let Z denote an $\mathbb{R}^d$-valued random variable with distribution ν. For $q > 0$, set $f_q(x) := 1 + \|x\|^q$. For measurable functions $f : \mathcal{X} \mapsto [1, \infty)$ and $h : \mathcal{X} \mapsto \mathbb{R}$, set

$$\|h\|_f := \sup\left\{ \frac{|h(x)|}{f(x)} : x \in \mathcal{X} \right\}.$$

Moreover, let $\mathcal{D}[0, 1]$ denote the space of right continuous functions $x : [0, 1] \mapsto \mathbb{R}$ with left limits, endowed with the Skorohod topology.

A distinctive feature of AJDs, besides the affine structure, relative to other jump-diffusion models is that its jump intensity is state-dependent. This property endows AJDs with greater flexibility in financial modeling but creates technical difficulties for analyzing the dynamics of the process. Indeed, differing theoretical treatments are needed, depending on whether the jump intensity is state-dependent, when we establish Lyapunov inequalities in Section 3. We therefore present our main results in two separate theorems. Theorem 1 covers only AJDs with state-independent jump intensities ($\kappa = \mathbf{0}$), whereas Theorem 2 allows state-dependent jump intensities.

Theorem 1. *If Assumptions 1–3 hold,* $\kappa = \mathbf{0}$*,* β *is a stable matrix, and* $\mathbb{E} \log(1 + \|Z\|) < \infty$*, then:*

(i) X *is positive Harris recurrent and*

$$\lim_{t\to\infty} \|\mathbb{P}_x(X(t) \in \cdot) - \pi(\cdot)\| = 0, \quad x \in \mathcal{X}, \tag{3}$$

where π *is the stationary distribution of* X*.*

If, in addition, $\mathbb{E}\, \|Z\|^p < \infty$ *for some* $p > 0$*, then:*

(ii) For each $q \in (0, p]$*, there exist positive finite constants* c_q *and* ρ_q *such that*

$$\|\mathbb{P}_x(X(t) \in \cdot) - \pi(\cdot)\|_{f_q} \le c_q f_q(x) e^{-\rho_q t}, \quad t \ge 0,\ x \in \mathcal{X}. \tag{4}$$

(iii) For any measurable function $h : \mathcal{X} \mapsto \mathbb{R}$ *with* $\|h\|_{f_p} < \infty$*,*

$$\mathbb{P}_x\left(\lim_{t\to\infty} \frac{1}{t} \int_0^t h(X(s))\, \mathrm{d}s = \pi(h) \right) = 1, \quad x \in \mathcal{X}, \tag{5}$$

and

$$\mathbb{P}_x\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}h(X(i\Delta))=\pi(h)\right)=1,\quad x\in\mathcal{X}. \tag{6}$$

(iv) For any measurable function $h:\mathcal{X}\mapsto\mathbb{R}$ with $\|h^q\|_{f_p}<\infty$ for some $q>2$, there exists non-negative finite constants σ_h and γ_h such that

$$n^{1/2}\left(\frac{1}{n}\int_0^{n\cdot}h(X(s))\,\mathrm{d}s-\pi(h)\right)\Rightarrow\sigma_h W(\cdot), \tag{7}$$

and

$$n^{1/2}\left(\frac{1}{n}\sum_{i=1}^{\lfloor n\cdot\rfloor}h(X(i\Delta))-\pi(h)\right)\Rightarrow\gamma_h W(\cdot), \tag{8}$$

as $n\to\infty$ $\mathbb{P}_x$-weakly in $\mathcal{D}[0,1]$ for all $x\in\mathcal{X}$, where W is a one-dimensional Wiener process.

Theorem 2. *If Assumptions 1–3 hold, $\beta+\mathbb{E}(Z)\kappa^{\intercal}$ is a stable matrix, and $\mathbb{E}\|Z\|<\infty$, then:*

(i) X is positive Harris recurrent and (3) holds.

If, in addition, $\mathbb{E}\|Z\|^p<\infty$ for some $p\geq 1$. Then:

(ii) For each $q\in[1,p]$, there exist positive finite constants c_q and ρ_q such that (4) holds.

(iii) For any measurable function $h:\mathcal{X}\mapsto\mathbb{R}$ with $\|h\|_{f_p}<\infty$, (5) and (6) hold.

(iv) For any measurable function $h:\mathcal{X}\mapsto\mathbb{R}$ with $\|h^q\|_{f_p}<\infty$ for some $q>2$, there exist non-negative finite constants σ_h and γ_h such that (7) and (8) hold as $n\to\infty$ $\mathbb{P}_x$-weakly in $\mathcal{D}[0,1]$ for all $x\in\mathcal{X}$.

We note that X is called *ergodic* if it has a stationary distribution π and the convergence (3) holds, whereas called *f-exponentially ergodic* if

$$\|\mathbb{P}_x(X(t)\in\cdot)-\pi(\cdot)\|_f\leq c(x)e^{-\rho t},\quad t\geq 0,\ x\in\mathcal{X},$$

for some functions $f:\mathcal{X}\mapsto[1,\infty)$, $c:\mathcal{X}\mapsto\mathbb{R}_+$ and some positive finite constant ρ. Clearly, X is f_p-exponentially ergodic under the assumptions of Theorem 1(ii) or Theorem 2(ii).

The key condition imposed here to establish positive Harris recurrence of AJDs is that $\beta+\mathbb{E}(Z)\kappa^{\intercal}$ is a stable matrix. If we adopt the convention that $0\times\infty=0$, when $\kappa=\mathbf{0}$ this condition is reduced to that β is a stable matrix regardless of the finiteness of $\mathbb{E}\|Z\|$. The condition that β is a stable matrix is typically assumed in the literature, including Sato and Yamazato (1984), Glasserman and Kim (2010), and Jena et al. (2012), in order that the process be mean reverting and have a stationary distribution. However, the first of the three articles works on a special Lévy-driven SDE, whereas the other two study ADs, so none of them involves state-dependent jump intensities as AJDs do. It can be shown that the stability of $\beta+\mathbb{E}(Z)\kappa^{\intercal}$ implies that of β. Thus, our condition is stronger and we call it the *strong mean reversion condition.*

Note that $\mathbb{E}(Z)$ is the mean jump size and that κ mostly determines the magnitude of the jump intensity when the AJD takes on large values. To some extent, $\mathbb{E}(Z)\kappa^\intercal$ captures the impact of the jumps. Thus, by imposing the stability of $\beta + \mathbb{E}(Z)\kappa^\intercal$, we essentially assume that mean reversion is a dominating factor, more significant than the jumps, in the dynamics of the process. On the other hand, this condition is technically mild. Indeed, we show in Section 3.3 that it cannot be relaxed in general if positive Harris recurrence of an AJD is desired.

3 Stochastic Stability

In this section, we apply the Lyapunov approach to address the stochastic stability of X. A key step in this approach is to judiciously construct suitable Lyapunov functions that induce suitable Lyapunov inequalities; see Meyn and Tweedie (1993c) for an extensive treatment of this approach. Nevertheless, we do not directly use the results there because their theory uses a definition of domain that insists on functions inducing martingales, whereas we work with local martingales.

Consider a twice-differentiable function $g : \mathcal{X} \mapsto \mathbb{R}$. By virtue of Itô's formula,

$$\begin{aligned} g(X(t)) = g(X(0)) + \int_0^t \left[\nabla g(X(s\text{-}))^\intercal \mu(X(s\text{-})) + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 g(X(s\text{-}))}{\partial x_i \partial x_j} (\sigma(X(s\text{-}))\sigma(X(s\text{-}))^\intercal)_{ij} \right] \mathrm{d}s \\ + \int_0^t \nabla g(X(s\text{-}))^\intercal \sigma(X(s\text{-}))\, \mathrm{d}W(s) + \int_0^t \int_{\mathcal{X}} (g(X(s\text{-}) + z) - g(X(s\text{-}))) N(\mathrm{d}s, \mathrm{d}z). \end{aligned} \tag{9}$$

By defining operators $\mathscr{G}$, $\mathscr{L}$, and $\mathscr{A}$ on twice-differentiable appropriately integrable functions g via

$$\begin{aligned} \mathscr{G}g(x) &:= \nabla g(x) \cdot (b + \beta x) + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 g(x)}{\partial x_i \partial x_j}\left(a_{i,j} + \sum_{k=1}^d \alpha_{k,ij} x_k \right), \\ \mathscr{L}g(x) &:= (\lambda + \kappa^\intercal x) \int_{\mathcal{X}} (g(x+z) - g(x))\, \nu(\mathrm{d}z), \\ \mathscr{A}g(x) &:= \mathscr{G}g(x) + \mathscr{L}g(x), \end{aligned} \tag{10}$$

we may rewrite (9) as

$$\begin{aligned} g(X(t)) &= g(X(0)) + \int_0^t \mathscr{A}g(X(s\text{-}))\, \mathrm{d}s + S_1(t) + S_2(t), \\ S_1(t) &:= \int_0^t \nabla g(X(s\text{-}))^\intercal \sigma(X(s\text{-}))\, \mathrm{d}W(s), \\ S_2(t) &:= \int_0^t \int_{\mathcal{X}} (g(X(s\text{-}) + z) - g(X(s\text{-}))) \tilde{N}(\mathrm{d}s, \mathrm{d}z), \end{aligned} \tag{11}$$

where $\tilde{N}(\mathrm{d}s, \mathrm{d}z) = N(\mathrm{d}s, \mathrm{d}z) - \Lambda(X(s\text{-}))\mathrm{d}s\nu(\mathrm{d}z)$ is the compensated random measure of $N(\mathrm{d}s, \mathrm{d}z)$.

We introduce some notation to facilitate the construction of Lyapunov inequalities. First, for a $d \times d$ matrix $H \succ 0$, define $\|v\|_H := \sqrt{v^\intercal H v}$. Then, $\|\cdot\|_H$ is a *vector norm* on $\mathbb{R}^d$ and it is easy to show that

$$\underline{\delta}\|v\|^2 \leq \|v\|_H^2 \leq \bar{\delta}\|v\|^2, \quad v \in \mathbb{R}^d, \tag{12}$$

where $(\delta_i : i = 1, \ldots, d)$ are the eigenvalues of H, $\underline{\delta} = \min\{\delta_i : i = 1, \ldots, d\}$ and $\bar{\delta} = \max\{\delta_i : i = 1, \ldots, d\}$. We can then define the following *induced* matrix norms (see Horn and Johnson (2012, p.340)). For a matrix $A \in \mathbb{R}^{d\times d}$, define

$$|||A||| \coloneqq \sup\left\{\frac{\|Av\|}{\|v\|} : \mathbf{0} \neq v \in \mathbb{R}^d\right\},$$

and

$$|||A|||_H \coloneqq \sup\left\{\frac{\|Av\|_H}{\|v\|_H} : \mathbf{0} \neq v \in \mathbb{R}^d\right\}.$$

3.1 Positive Harris Recurrence and Ergodicity

For each $\Delta > 0$, let $X^\Delta \coloneqq (X(n\Delta) : n = 0, 1, \ldots)$ denote the Δ-skeleton of X.

Proposition 1. *Under Assumptions 1–3, X^Δ is φ-irreducible for any $\Delta > 0$, where φ is the Lebesgue measure on $\mathcal{X}$.*

The proof of Proposition 1 relies on the following result, which is of interest in its own right. It reduces irreducibility of a jump-diffusion process to that of the associated diffusion process.

Lemma 1. *Suppose that X satisfies SDE (1)[1]. Let $\tilde{X} = (\tilde{X}(t) : t \geq 0)$ satisfy the following SDE*

$$\begin{aligned} \mathrm{d}\tilde{X}(t) &= \mu(\tilde{X}(t))\,\mathrm{d}t + \sigma(\tilde{X}(t))\,\mathrm{d}W(t), \\ \tilde{X}(0) &= x \in \mathcal{X}, \end{aligned} \tag{13}$$

where W is the d-dimensional Wiener process in (1). If $\tilde{X}^\Delta$ (resp., $\tilde{X}$) is φ-irreducible, then X^Δ (resp., X) is φ-irreducible.

Proof. Consider a measurable $K \subseteq \mathcal{X}$ and let τ denote the first jump time of X. Then $\mathbb{P}_x(X(t) = \tilde{X}(t)) = 1$ for $t < \tau^*$. It follows that for any $t > 0$,

$$\begin{aligned} \mathbb{P}_x(X(t) \in K, \tau^* > t) &= \mathbb{E}_x\left[\mathbb{E}\left(\mathbb{I}(\tilde{X}(t) \in K, \tau^* > t) | X(s), 0 \leq s \leq t\right)\right] \\ &= \mathbb{E}_x\left[\mathbb{I}(\tilde{X}(t) \in K)\,\mathbb{P}\left(\tau^* > t | \tilde{X}(s), 0 \leq s \leq t\right)\right] \\ &= \mathbb{E}_x\left[\mathbb{I}(\tilde{X}(t) \in K) e^{-\int_0^t \Lambda(\tilde{X}(s))\,\mathrm{d}s}\right]. \end{aligned}$$

Hence, $\mathbb{P}_x(X(t) \in K, \tau^* > t) = 0$ if and only if $\mathbb{P}_x(\tilde{X}(t) \in K) = 0$ for any $t > 0$. It is then clear that the φ-irreducibility of $\tilde{X}^\Delta$ (resp., $\tilde{X}$) implies that of X^Δ (resp., X). □

Proof of Proposition 1. The key in the proof is to convert the AJD by a linear transformation used in Filipović and Mayerhofer (2009) into a canonical representation in which the matrices involved

[1] Here, we do not restrict its coefficients μ, σ, Λ to follow the affine form (2).

are of special form. Specifically, note that if X satisfies SDE (1) with coefficients (2), then for any nonsingular matrix $A \in \mathbb{R}^{d\times d}$, the linear transformation $Y = AX$ satisfies

$$\begin{aligned} \mathrm{d}Y(t) &= (Ab + A\beta A^{-1}Y(t))\,\mathrm{d}t + A\sigma\left(A^{-1}Y(t)\right)\,\mathrm{d}W(t) + \int_{\mathbb{R}^d} AzN(\mathrm{d}t,\mathrm{d}z), \\ Y(0) &= Ax, \end{aligned} \tag{14}$$

where $N(\mathrm{d}t,\mathrm{d}z)$ has the compensator measure $\Lambda(A^{-1}Y(t\text{-}))\mathrm{d}t\nu(\mathrm{d}z)$. So the drift, diffusion matrix, and intensity of SDE (14) are

$$Ab + A\beta A^{-1}y, \quad A\sigma\left(A^{-1}y\right)\sigma\left(A^{-1}y\right)^{\intercal}A^{\intercal}, \quad \text{and} \quad \lambda + \kappa^{\intercal}A^{-1}y$$

respectively, which are all affine in y. Consequently, the existence and uniqueness of a strong solution to (1) is invariant with respect to nonsingular linear transformations.

Since $\alpha_{i,ii} > 0$ for all $i = 1,\ldots,m$, it follows from Lemma 7.1 of Filipović and Mayerhofer (2009) that there exists a nonsingular matrix $A \in \mathbb{R}^{d\times d}$ that maps $\mathbb{R}^m_+ \times \mathbb{R}^{d-m}$ to itself and renders the transformed diffusion matrix in the following block-diagonal form

$$A\sigma\left(A^{-1}y\right)\sigma\left(A^{-1}y\right)^{\intercal}A^{\intercal} = \begin{pmatrix} \mathrm{diag}(\alpha_{1,11}y_1,\ldots,\alpha_{m,mm}y_m) & \mathbf{0} \\ \mathbf{0} & h + \sum_{i=1}^m y_i\eta_i \end{pmatrix}$$

for some $(d-m)\times(d-m)$ matrices $h \succeq 0$ and $\eta_i \succeq 0$, $i = 1,\ldots,m$. In particular, A is of the form

$$A = \begin{pmatrix} I_m & \mathbf{0} \\ D & I_{d-m} \end{pmatrix},$$

for some $(d-m)\times m$ matrix D, where I_m and I_{d-m} are identity matrices. Moreover, it is straightforward to verify that Ab, $A\beta A^{-1}$, and $\kappa^{\intercal}A^{-1}$ satisfy both Assumption 2 and Assumption 3 in lieu of b, β, and κ. Hence, we can assume without loss of generality that the diffusion matrix of (1) has the form

$$\sigma(x)\sigma(x)^{\intercal} = \begin{pmatrix} \mathrm{diag}(\alpha_{1,11}x_1,\ldots,\alpha_{m,mm}x_m) & \mathbf{0} \\ \mathbf{0} & a_{\mathcal{J}\mathcal{J}} + \sum_{i=1}^m x_i\alpha_{i,\mathcal{J}\mathcal{J}} \end{pmatrix}. \tag{15}$$

Hence, $\tilde{X}_{\mathcal{I}}(t)$ satisfies

$$\begin{aligned} \mathrm{d}\tilde{X}_{\mathcal{I}}(t) &= (b_{\mathcal{I}} + \beta_{\mathcal{I}\mathcal{I}}\tilde{X}_{\mathcal{I}}(t))\,\mathrm{d}t + \mathrm{diag}(\sqrt{\alpha_{1,11}x_1},\ldots,\sqrt{\alpha_{m,mm}x_m})\,\mathrm{d}W_{\mathcal{I}}(t), \\ \tilde{X}_{\mathcal{I}}(0) &= x_{\mathcal{I}} \in \mathbb{R}^m_+. \end{aligned}$$

With the assumption that $2b_i > \alpha_{i,ii}$, $i = 1,\ldots,m$, we can directly verify the conditions of the theorem on p.388 of Duffie and Kan (1996) to conclude that $\mathbf{0} \in \mathbb{R}^m_+$ is not attainable in finite time, i.e. $\tilde{X}_i(t) > 0$ for all $t > 0$ and $i = 1,\ldots,m$, if $\tilde{X}_i(0) > 0$, $i = 1,\ldots,m$.

We now consider a bijective transformation $\tilde{Y} := f(\tilde{X})$, where $f : \mathcal{X} \mapsto \mathcal{X}$ is defined as follows:

$f_i(x) = 2\sqrt{x_i}$ for $i = 1, \ldots, m$ and $f_i(x) = x_i$ for $x = m+1, \ldots, d$. Then,

$$\frac{\partial f_i(x)}{\partial x_j} = \begin{cases} x_i^{-1/2}, & \text{if } i = j, i = 1, \ldots, m, \\ 1, & \text{if } i = j, i = m+1, \ldots, d, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\frac{\partial^2 f_i(x)}{\partial x_k \partial x_l} = \begin{cases} -\frac{1}{2} x_i^{-3/2}, & \text{if } i = k = l, i = 1, \ldots, m, \\ 0, & \text{otherwise.} \end{cases}$$

It follows that, by Itô's formula,

$$\mathrm{d} f_i(\tilde{X}(t)) = \zeta_i(\tilde{X}(t))\,\mathrm{d}t + \nabla f_i(\tilde{X}(t))^\intercal \sigma(\tilde{X}(t))\,\mathrm{d}W(t),$$

for $i = 1, \ldots, d$, where

$$\zeta_i(x) = \frac{\partial f_i(x)}{\partial x_i} \mu_i(x) + \frac{1}{2} \frac{\partial^2 f_i(x)}{\partial x_i^2} (\sigma(x)\sigma(x)^\intercal)_{ii}.$$

Note that we have shown that $x_i > 0$, $i = 1, \ldots, m$ for $x \in \mathcal{X}$, so the function $\zeta(x)$ is well-defined for all $x \in \mathcal{X}$. Let f^{-1} denote the inverse mapping of f, i.e. $f_i^{-1}(y) = y_i^2$ for $i = 1, \ldots, m$ and $f_i^{-1}(y) = y_i$, for $i = m+1, \ldots, d$. Then,

$$\mathrm{d}\tilde{Y}(t) = \zeta(f^{-1}(\tilde{Y}(t)))\,\mathrm{d}t + \nabla f(f^{-1}(\tilde{Y}(t)))\sigma(f^{-1}(\tilde{Y}(t)))\,\mathrm{d}W(t), \tag{16}$$

where $\nabla f := (\frac{\partial f_i}{\partial x_j})_{1 \le i,j \le d}$ is the Jacobian matrix of f. A straightforward calculation reveals that the diffusion matrix of (16) is

$$\nabla f(f^{-1}(y))\sigma(f^{-1}(y))\sigma(f^{-1}(y))^\intercal \nabla f(f^{-1}(y))^\intercal = \begin{pmatrix} \mathrm{diag}(\alpha_{1,11} \ldots, \alpha_{m,mm}) & \mathbf{0} \\ \mathbf{0} & a_{\mathcal{J}\mathcal{J}} + \sum_{i=1}^m y_i^2 \alpha_{i,\mathcal{J}\mathcal{J}} \end{pmatrix}.$$

Hence, in light of the assumption that $\alpha_{i,ii} > 0$, $i = 1, \ldots, m$ and $a_{\mathcal{J}\mathcal{J}} \succ 0$, the diffusion matrix of (16) is *uniformly elliptic*. It is well known that such diffusion processes admit a positive probability density; see, e.g., Theorem 3.3.4 of Davies (1989). Since the mapping f is bijective, we conclude that $\tilde{X}$ also admits a positive density; in particular,

$$\mathbb{P}_x(\tilde{X}(t) \in K) = \int_K p(t, x, y)\,\mathrm{d}y,$$

for all $t > 0$, $x \in \mathcal{X}$ and measurable $K \subseteq \mathcal{X}$, where $p(t, x, y) > 0$ is the density. Therefore, letting φ be the Lebesgue measure restricted to $\mathcal{X}$, $\mathbb{P}_x(\tilde{X}(t) \in K) > 0$ whenever $\varphi(K) > 0$. In particular, we have $\mathbb{P}_x(\tilde{X}(\Delta) \in K) > 0$ whenever $\varphi(K) > 0$. So $\tilde{X}^\Delta$ is φ-irreducible, which completes the proof in light of Lemma 1. $\square$

Proposition 2. *Under Assumptions 1 and 2, X is a regular affine process.*

Proof. For $T > 0$ and any purely imaginary vector $u \in i\mathbb{R}^d$, define $M(t) \coloneqq e^{\phi(T-t,u)+\psi(T-t,u)^\intercal X(t)}$, where $\phi : \mathbb{R}_+ \times i\mathbb{R}^d \mapsto \mathbb{C}$ and $\psi(t,u) : \mathbb{R}_+ \times i\mathbb{R}^d \mapsto \mathbb{C}^d$ are functions that are differentiable with respect to t. Applying Itô's formula,

$$
\begin{aligned}
&M(t)\\
&= M(0) + \int_0^t M(s\text{-})\psi(T-s,u)^\intercal \sigma(X(s))\,\mathrm{d}W(s) + \int_0^t M(s\text{-}) \int_{\mathcal{X}} \left(e^{\psi(T-s,u)^\intercal z} - 1\right) N(\mathrm{d}s, \mathrm{d}z)\\
&+ \int_0^t M(s\text{-}) \left[-\partial_t \phi(T-s,u) - \partial_t \psi(T-s,u)^\intercal X(s) + \psi(T-s,u)^\intercal \mu(X(s\text{-}))\right] \mathrm{d}s\\
&+ \frac{1}{2} \int_0^t M(s\text{-})\psi(T-s,u)^\intercal \sigma(X(s\text{-}))\sigma(X(s\text{-}))^\intercal \psi(T-s,u)\,\mathrm{d}s\\
&= M(0) + \int_0^t M(s\text{-})\psi(T-s,u)^\intercal \sigma(X(s))\,\mathrm{d}W(s) + \int_0^t M(s\text{-}) \int_{\mathcal{X}} \left(e^{\psi(T-s,u)^\intercal z} - 1\right) \tilde{N}(\mathrm{d}s, \mathrm{d}z)\\
&+ \int_0^t M(s\text{-}) \left[-\partial_t \phi(T-s,u) + \psi(T-s,u)^\intercal b + \frac{1}{2}\psi(T-s,u)^\intercal a \psi(T-s,u)\right] \mathrm{d}s\\
&+ \int_0^t M(s\text{-}) \Bigg[- \partial_t \psi(T-s,u)^\intercal X(s\text{-}) + \psi(T-s,u)^\intercal \beta X(s\text{-}) + \frac{1}{2}\sum_{i=1}^d \psi(s,u)^\intercal \alpha_i \psi(s,u) X_i(s\text{-}) \Bigg] \mathrm{d}s\\
&+ \int_0^t M(s\text{-})(\lambda + \kappa^\intercal X(s\text{-})) \int_{\mathcal{X}} \left(e^{\psi(T-s,u)^\intercal z} - 1\right) \nu(\mathrm{d}z)\mathrm{d}s.
\end{aligned}
$$

Hence, if ϕ and ψ satisfy the following generalized Riccati equations

$$
\begin{aligned}
\partial_t \phi(t,u) &= \psi(t,u)^\intercal b + \frac{1}{2}\psi(t,u)^\intercal a \psi(t,u) + \lambda \int_{\mathcal{X}} \left(e^{\psi(t,u)^\intercal z} - 1\right) \nu(\mathrm{d}z),\\
\partial_t \psi_i(t,u) &= \psi(t,u)^\intercal \beta_i + \frac{1}{2}\sum_{i=1}^d \psi(t,u)^\intercal \alpha_i \psi(t,u) + \kappa_i \int_{\mathcal{X}} \left(e^{\psi(t,u)^\intercal z} - 1\right) \nu(\mathrm{d}z), \quad i = 1, \ldots, d,
\end{aligned}
$$

with $\phi(0,u) = 0$ and $\psi(0,u) = u$, where β_i is the i-th column of β, then

$$
M(t) = M(0) + \int_0^t M(s\text{-})\psi(T-s,u)^\intercal \sigma(X(s))\,\mathrm{d}W(s) + \int_0^t M(s\text{-}) \int_{\mathcal{X}} \left(e^{\psi(T-s,u)^\intercal z} - 1\right) \tilde{N}(\mathrm{d}s, \mathrm{d}z). \tag{17}
$$

It follows from Proposition 6.1 and Proposition 6.4 of Duffie et al. (2003) that under Assumption 2, the preceding generalized Riccati equations have a unique solution $(\phi(\cdot,u), \psi(\cdot,u)) : \mathbb{R}_+ \mapsto \mathbb{C}_- \times \mathbb{C}_-^m \times i\mathbb{R}^{d-m}$ for all $u \in \mathbb{C}_-^m \times i\mathbb{R}^{d-m}$, where $\mathbb{C}_-^m = \{z \in \mathbb{C}^m | \mathrm{Re}(z) \in \mathbb{R}_-^m\}$. Hence,

$$
\phi(t,u) + \psi(t,u)^\intercal x \in \mathbb{C}_-, \quad x \in \mathcal{X}, \tag{18}
$$

under Assumption 1. Further, Proposition 7.4 of Duffie et al. (2003) asserts that

$$
\begin{aligned}
\phi(t+s,u) &= \phi(t,u) + \phi(s, \psi(t,u))\\
\psi(t+s,u) &= \psi(t, \psi(s,u))
\end{aligned} \tag{19}
$$

for all $t, s \in \mathbb{R}_+$ and $u \in \mathbb{C}^m_- \times i\mathbb{R}^{d-m}$.

In light of (17) and (18), $(M(t) : 0 \leq t \leq T)$ is a local martingale with $|M(t)| \leq 1$ for all t, thereby a martingale. So

$$\mathbb{E}_x[e^{u^\intercal X(T)}] = \mathbb{E}_x[M(T)] = \mathbb{E}_x[M(0)] = e^{\phi(T,u)+\psi(T,u)^\intercal x}, \tag{20}$$

namely the characteristic function $\mathbb{E}_x[e^{u^\intercal X(t)}]$ is exponential-affine in x. In addition, it is easy to verify via (19) and (20) the Chapman–Kolmogorov equation

$$\mathbb{P}_x(X(t+s) \in \cdot) = \int_{\mathcal{X}} \mathbb{P}_x(X(t) \in \mathrm{d}y)\, \mathbb{P}_y(X(s) \in \cdot),$$

implying that X is a time-homogeneous Markov process, thereby an affine process by (20).

At last, $\mathbb{E}_x[e^{u^\intercal X(t)}]$ is clearly continuous in t by (20), indicating that X is stochastically continuous. Then, Theorem 5.1 of Keller-Ressel et al. (2011) asserts that a stochastically continuous affine process on state space $\mathbb{R}^m_+ \times \mathbb{R}^{d-m}$ is regular. □

Proof of Theorem 1(i). We first show that X is Harris recurrent. Theorem 1.1 of Meyn and Tweedie (1993a) asserts that X is Harris recurrent if (i) X is a *Borel right process* (Getoor 1975, p.55), and (ii) there exists a *petite* set K for X, such that $\mathbb{P}_x(\tau_K < \infty) = 1$ for all $x \in \mathcal{X}$, where $\tau_K = \inf\{t \geq 0 : X(t) \in K\}$.

For condition (i), we note that X is a Feller process by Proposition 8.2 of Duffie et al. (2003) and Proposition 2. The Feller property of X trivially implies that X is a Borel right process.

For condition (ii), fix an arbitrary $\Delta > 0$ and note that X^Δ is a Feller chain since X is a Feller process. By Theorem 3.4 of Meyn and Tweedie (1992), the Feller property of X^Δ and Proposition 1 immediately imply that all compact sets are petite for X^Δ, thereby petite for X. In the sequel, we will show that there exists a compact set K such that $\mathbb{P}_x(\tau_K < \infty) = 1$ for all $x \in \mathcal{X}$. To that send, we first establish the following Lyapunov inequality

$$\mathscr{A} g(x) \leq -c_1 + c_2\, \mathbb{I}_K(x), \quad x \in \mathcal{X}, \tag{21}$$

for some compact set K and some positive finite constants c_1 and c_2, where $g(x) = \log(1 + \|x\|^2_H)$ for some $d \times d$ matrix $H \succ 0$.

Since β is a stable matrix, there exists a $d \times d$ matrix $H \succ 0$ for which $-(H\beta + \beta^\intercal H) \succ 0$; see Berman and Plemmons (1994, Theorem 2.3(G), p.134). It is straightforward to calculate the gradient and Hessian of $g(x)$ as follows

$$\nabla g(x) = \frac{2Hx}{1 + \|x\|^2_H} \quad \text{and} \quad \nabla^2 g(x) = \frac{2(1 + \|x\|^2_H)H - 4Hxx^\intercal H}{(1 + \|x\|^2_H)^2}.$$

Hence,

$$\mathscr{G} g(x) = \frac{2}{1 + \|x\|^2_H} \left[x^\intercal H(b + \beta x) + \frac{1}{2} \sum_{i,j=1}^d \left(a_{ij} + \sum_{k=1}^d \alpha_{k,ij} x_k \right) \left(H - \frac{2Hxx^\intercal H}{1 + \|x\|^2_H} \right)_{ij} \right]. \tag{22}$$

We note that for any $i, j = 1, \ldots, d$,

$$|(Hx x^{\intercal} H)_{ij}| = |(Hx)_i (Hx)_j| \leq \|Hx\|^2 \leq |||H|||^2 \|x\|^2 \leq \underline{\delta}^{-1} |||H|||^2 \|x\|_H^2,$$

where the last inequality follows from (12). Hence,

$$\frac{|(Hxx^{\intercal}H)_{ij}|}{1 + \|x\|_H^2} = O(1), \tag{23}$$

as $\|x\|_H \to \infty$. Therefore, we can rewrite (22) as

$$\mathscr{G} g(x) = \frac{2x^{\intercal} H \beta x + O(\|x\|_H)}{1 + \|x\|_H^2} = \frac{2x^{\intercal} H \beta x}{\|x\|_H (1 + \|x\|_H)} + o(1),$$

as $\|x\|_H \to \infty$. Moreover, by virtue of (12) and the fact that $-(H\beta + \beta^{\intercal} H) \succ 0$,

$$-2x^{\intercal} H \beta x = -x^{\intercal} (H\beta + \beta^{\intercal} H) x \geq \underline{\gamma} \|x\|^2 \geq \underline{\gamma} \bar{\delta}^{-1} \|x\|_H^2,$$

where $\underline{\gamma} > 0$ is the smallest eigenvalue of $-(H\beta + \beta^{\intercal} H)$. Therefore,

$$\limsup_{\|x\|_H \to \infty} \mathscr{G} g(x) = \limsup_{\|x\|_H \to \infty} \frac{2x^{\intercal} H \beta x}{\|x\|_H (1 + \|x\|_H)} \leq -\underline{\gamma} \bar{\delta}^{-1}. \tag{24}$$

On the other hand, it is easy to see that $1 + (\|x\|_H + \|z\|_H)^2 \leq 2(1 + \|x\|_H^2)(1 + \|z\|_H^2)$ for all $x, z \in \mathbb{R}^d$. Thus,

$$\log \left(\frac{1 + \|x + z\|_H^2}{1 + \|x\|_H^2} \right) \leq \log \left(\frac{1 + (\|x\|_H + \|z\|_H)^2}{1 + \|x\|_H^2} \right) \leq \log(2(1 + \|z\|_H^2)).$$

It is easy to see that $\log(2(1 + \|z\|_H^2))$ is integrable on $\mathcal{X}$ since $\mathbb{E} \log(1 + \|Z\|_H) < \infty$ if and only if $\mathbb{E} \log(1 + \|Z\|) < \infty$ in light of (12). Then, we apply the reverse Fatou's lemma to obtain

$$\limsup_{\|x\|_H \to \infty} \int_{\mathcal{X}} \log \left(\frac{1 + \|x + z\|_H^2}{1 + \|x\|_H^2} \right) \nu(\mathrm{d}z) \leq \int_{\mathcal{X}} \limsup_{\|x\|_H \to \infty} \log \left(\frac{1 + \|x + z\|_H^2}{1 + \|x\|_H^2} \right) \nu(\mathrm{d}z) = 0.$$

With $\kappa = \mathbf{0}$,

$$\limsup_{\|x\|_H \to \infty} \mathscr{L} g(x) = \limsup_{\|x\|_H \to \infty} \lambda \int_{\mathcal{X}} \log \left(\frac{1 + (\|x\|_H + \|z\|_H)^2}{1 + \|x\|_H^2} \right) \nu(\mathrm{d}z) \leq 0. \tag{25}$$

We then conclude from (24) and (25) that there exists $k > 0$ for which

$$\mathscr{A} g(x) = \mathscr{G} g(x) + \mathscr{L} g(x) \leq -\frac{1}{2} \underline{\gamma} \bar{\delta}^{-1},$$

for all $x \in \mathcal{X}$ with $\|x\|_H > k$. Then, it is easy to check that the inequality (21) holds by setting

$K = \{x \in \mathcal{X} : \|x\|_H \le k\}$, $c_1 = \gamma\bar{\delta}^{-1}/2$, and $c_2 = \max\{1, \sup_{x\in K}(\mathscr{A} g(x) + c_1)\}$.

We are now ready to show $\mathbb{P}_x(\tau_K < \infty) = 1$ for all $x \in \mathcal{X}$. Define $T_n = \inf\{t \ge 0 : |X(t)| > n\}$. It follows from (11) and (21) that

$$g(X(t \wedge T_n)) \le g(X(0)) + \int_0^{t\wedge T_n} [-c_1 + c_2\, \mathbb{I}_K(X(s\text{-}))]\, \mathrm{d}s + S_1(t \wedge T_n) + S_2(t \wedge T_n), \quad n \ge 1. \tag{26}$$

Noting that $|X(t\text{-})| \le n$ is bounded for $t \in [0, T_n)$, $(S_i(t \wedge T_n) : t \ge 0)$ is a martingale, $i = 1, 2$. Then by the optional sampling theorem (see, e.g., Karatzas and Shreve (1991, p.19))

$$\mathbb{E}_x[g(X(t \wedge \tau_K \wedge T_n))] \le g(x) - c_1\, \mathbb{E}_x(t \wedge \tau_K \wedge T_n), \quad x \in \mathcal{X} \setminus K,\ n \ge 1.$$

Therefore,

$$c_1\, \mathbb{E}_x(t \wedge \tau_K \wedge T_n) \le g(x), \quad x \in \mathcal{X} \setminus K,\ n \ge 1,$$

since $g(x) \ge 0$ for all $x \in \mathcal{X}$. Note that X is non-explosive, so $T_n \to \infty$ as $n \to \infty$ $\mathbb{P}_x$-a.s. for all $x \in \mathcal{X}$. Therefore, by sending $n \to \infty$ and then sending $t \to \infty$, we conclude from the monotone convergence theorem that $c_1\, \mathbb{E}_x(\tau_K) \le g(x)$ for $x \in \mathcal{X} \setminus K$. Hence, $\mathbb{P}_x(\tau_K < \infty) = 1$ for all $x \in \mathcal{X}$. Consequently, X is Harris recurrent by Theorem 1.1 of Meyn and Tweedie (1993a).

Theorem 1.2 of Meyn and Tweedie (1993a) states that given the Harris recurrence, X is positive Harris recurrent if $\sup_{x\in K} \mathbb{E}_x(\tau_K(\Delta)) < \infty$. We now show this is indeed the case. For any $\Delta > 0$, let $\tau_K(\Delta) \coloneqq \Delta + \Theta^{\Delta} \circ \tau_K$ be the first hitting time on K after Δ, where Θ^{Δ} is the *shift operator*; see Sharpe (1988, p.8). Then,

$$\mathbb{E}_x(\tau_K(\Delta) - \Delta) = \int_{\mathcal{X}} \mathbb{P}_x(X(\Delta) \in \mathrm{d}y)\, \mathbb{E}_y(\tau_K) \le \int_{\mathcal{X}} c_1^{-1} g(y)\, \mathbb{P}_x(X(\Delta) \in \mathrm{d}y) = c_1^{-1}\, \mathbb{E}_x\, g(X(\Delta)), \tag{27}$$

for all $x \in \mathcal{X}$. In addition, it follows from (26) that

$$\mathbb{E}_x\, g(X(\Delta \wedge T_n)) \le g(x) + (c_2 - c_1)\, \mathbb{E}_x(\Delta \wedge T_n), \quad x \in \mathcal{X},\ n \ge 1.$$

Then, by Fatou's lemma and the monotone convergence theorem,

$$\mathbb{E}_x\, g(X(\Delta)) \le \liminf_{n\to\infty} \mathbb{E}_x\, g(X(\Delta \wedge T_n)) \le g(x) + (c_2 - c_1)\Delta, \quad x \in \mathcal{X}. \tag{28}$$

Combining (27) and (28) yields that

$$\mathbb{E}_x(\tau_K(\Delta)) \le c_1^{-1}(g(x) + d), \quad x \in \mathcal{X}.$$

Hence, $\sup_{x\in K} \mathbb{E}_x(\tau_K(\Delta)) < \infty$, which implies that X is positive Harris recurrent by Theorem 1.2 of Meyn and Tweedie (1993a).

Finally, Theorem 6.1 of Meyn and Tweedie (1993b) asserts that if X^{Δ} is φ-irreducible, which is true by Proposition 1, then a positive Harris recurrent process is ergodic, i.e. (3) holds. $\square$

Proof of Theorem 2(i). Following the proof Theorem 1(i), it suffices to show the Lyapunov inequality (21) holds under the assumptions of Theorem 2(i). In fact, we prove the following stronger result

$$\mathscr{A} g(x) \leq -c_1 g(x) + c_2 \mathbb{I}_K(x), \quad x \in \mathcal{X}, \tag{29}$$

for some compact set K and some positive finite constants c_1 and c_2, where $g(x) = (1 + \|x\|_H^2)^{p/2}$ for some $d \times d$ matrix $H \succ 0$ and some constant $p \geq 1$.

Since $\mathbb{E}\,\|Z\| < \infty$, there exists $p \geq 1$ for which $\mathbb{E}\,\|Z\|^p < \infty$. Since $\beta + \mathbb{E}(Z)\kappa^\intercal$ is stable, there exists a matrix $H \succ 0$ such that

$$-\left[H(\beta + \mathbb{E}(Z)\kappa^\intercal) + (\beta + \mathbb{E}(Z)\kappa^\intercal)^\intercal H\right] \succ 0. \tag{30}$$

It is straightforward to calculate the gradient and Hessian of $g(x)$ as follows

$$\nabla g(x) = \frac{pg(x)}{1 + \|x\|_H^2} Hx \quad \text{and} \quad \nabla^2 g(x) = \frac{pg(x)}{1 + \|x\|_H^2}\left[H + \frac{(p-2)Hxx^\intercal H}{1 + \|x\|_H^2}\right].$$

It then follows from (23) that as $\|x\|_H \to \infty$,

$$\begin{aligned}
\mathscr{G} g(x) &= \frac{pg(x)}{1 + \|x\|_H^2}\left[x^\intercal H(b + \beta x) + \frac{1}{2}\sum_{i,j=1}^{d}(a_{i,j} + \sum_{k=1}^{d}\alpha_{k,ij}x_k)\left(H + \frac{(p-2)Hxx^\intercal H}{1 + \|x\|_H^2}\right)_{i,j}\right] \\
&= pg(x)\left(\frac{x^\intercal H\beta x}{\|x\|_H^2} + o(1)\right).
\end{aligned} \tag{31}$$

To analyze the asymptotic behavior of $\mathscr{L} g(x)$, we apply the mean value theorem, namely

$$g(x+z) - g(x) = \nabla g(\xi)^\intercal z = p(1 + \|\xi\|_H^2)^{p/2-1}\xi^\intercal Hz,$$

where $\xi = x + uz$ for some $u \in (0,1)$. Note that $\|\xi\|_H$ lies between $\|x\|_H$ and $\|x+z\|_H$ and $\xi^\intercal Hz\kappa^\intercal x$ lies between $x^\intercal Hz\kappa^\intercal x$ and $(x+z)^\intercal Hz\kappa^\intercal x$. It then follows that

$$\frac{\kappa^\intercal x(g(x+z) - g(x))}{g(x)} = p \cdot \frac{(1 + \|\xi\|_H^2)^{p/2-1}}{(1 + \|x\|_H^2)^{p/2}} \cdot \xi^\intercal Hz\kappa^\intercal x \sim p \cdot \frac{x^\intercal Hz\kappa^\intercal x}{\|x\|_H^2} \tag{32}$$

as $\|x\|_H \to \infty$ for all $z \in \mathbb{R}^d$.[2] Moreover,

$$\begin{aligned}
|g(x+z) - g(x)| &= p(1 + \|\xi\|_H^2)^{p/2-1}|z^\intercal H\xi| \\
&\leq p(1 + \|\xi\|_H^2)^{p/2-1}\|z\|\|H\xi\| \\
&\leq p\underline{\delta}^{-1}(1 + \|\xi\|_H^2)^{p/2-1}\|z\|_H\|H\xi\|_H \\
&\leq p\underline{\delta}^{-1}(1 + \|\xi\|_H^2)^{p/2-1}\|z\|_H|||H|||_H\|\xi\|_H
\end{aligned}$$

[2] Here, we use the notation that $f(x) \sim g(x)$ if $\lim_{\|x\|_H \to \infty} \frac{f(x)}{g(x)} = 1$.

$$\leq p\underline{\delta}^{-1}(1+\|\xi\|_H^2)^{p/2-1/2}\|z\|_H\|\!|H|\!\|_H, \tag{33}$$

where the second inequality follows from (12). So

$$\begin{aligned}\left|\frac{\kappa^\intercal x(g(x+z)-g(x))}{g(x)}\right| &\leq \frac{|\kappa||x|p\underline{\delta}^{-1}(1+\|\xi\|_H^2)^{p/2-1/2}\|z\|_H\|\!|H|\!\|_H}{(1+\|x\|_H^2)^{p/2}} \\ &\leq p\underline{\delta}^{-2}\|\!|H|\!\|_H\|\kappa\|_H\|z\|_H\frac{(1+\|\xi\|_H^2)^{p/2-1/2}}{(1+\|x\|_H^2)^{p/2-1/2}},\end{aligned}$$

where the second inequality follows from (12). Note that

$$1+\|\xi\|_H^2 = 1+\|x+uz\|_H^2 \leq 2(1+\|x\|_H^2)(1+\|uz\|_H^2) \leq 2(1+\|x\|_H^2)(1+\|z\|_H^2),$$

so

$$\int_{\mathcal{X}}\left|\frac{\kappa^\intercal x(g(x+z)-g(x))}{g(x)}\right|\nu(\mathrm{d}z) \leq 2^{p/2-1/2}p\underline{\delta}^{-2}\|\!|H|\!\|_H\|\kappa\|_H\int_{\mathcal{X}}(1+\|z\|_H^2)^{p/2}\,\nu(\mathrm{d}z) < \infty. \tag{34}$$

By (32) and (34), the dominated convergence theorem dictates that

$$\kappa^\intercal x\int_{\mathcal{X}}(g(x+z)-g(x))\,\nu(\mathrm{d}z) \sim pg(x)\cdot\int\frac{x^\intercal Hz\kappa^\intercal x}{\|x\|_H^2}\,\nu(\mathrm{d}z) = pg(x)\cdot\frac{x^\intercal H\,\mathbb{E}(Z)\kappa^\intercal x}{\|x\|_H^2},$$

and thus

$$\mathscr{L}g(x) = (\lambda+\kappa^\intercal x)\int_{\mathcal{X}}(g(x+z)-g(x))\,\nu(\mathrm{d}z) \sim pg(x)\frac{x^\intercal H\,\mathbb{E}(Z)\kappa^\intercal x}{\|x\|_H^2}, \tag{35}$$

as $\|x\|_H \to \infty$. Combining (31) and (35),

$$\mathscr{A}g(x) = \mathscr{G}g(x)+\mathscr{L}g(x) = pg(x)\left(\frac{x^\intercal H(\beta+\mathbb{E}(Z)\kappa^\intercal)x}{\|x\|_H^2}+o(1)\right)$$

as $\|x\|_H \to \infty$. By (30), the definition of the matrix H,

$$-x^\intercal H(\beta+\mathbb{E}(Z)\kappa^\intercal)x = -\frac{1}{2}x^\intercal\left[H(\beta+\mathbb{E}(Z)\kappa^\intercal)+(\beta+\mathbb{E}(Z)\kappa^\intercal)^\intercal H\right]x \geq \frac{1}{2}\underline{\gamma}\|x\|^2 \geq \frac{1}{2}\underline{\gamma}\bar{\delta}^{-1}\|x\|_H^2,$$

where $\underline{\gamma} > 0$ is the smallest eigenvalue of $-\left[H(\beta+\mathbb{E}(Z)\kappa^\intercal)+(\beta+\mathbb{E}(Z)\kappa^\intercal)^\intercal H\right]$. Hence, there exists $k > 0$ for which

$$\mathscr{A}g(x) \leq -\frac{1}{4}p\underline{\gamma}\bar{\delta}^{-1}g(x)$$

for all $x \in \mathcal{X}$ with $\|x\|_H > k$. Therefore, the inequality (29) holds by setting $K = \{x \in \mathcal{X} : \|x\|_H \leq k\}$, $c_1 = p\underline{\gamma}\bar{\delta}^{-1}/4$, and $c_2 = \max\{1, \sup_{x\in K}(\mathscr{A}g(x)+c_1g(x))\}$. □

3.2 Exponential Ergodicity

Proof of Theorem 1(ii). Note that if $\mathbb{E}\,\|Z\|^p < \infty$ for some $p > 0$, then $\mathbb{E}\,\|Z\|^q < \infty$ for all $q \in (0, p]$. We assume that $p \in (0, 1)$, because the case that $p \geq 1$ is covered by Theorem 2(ii), which will be proved later.

Since β is stable, there exists a matrix $H \succ 0$ such that $-(H\beta + \beta^\intercal H) \succ 0$. We show that $g_q(x) = (1 + \|x\|_H^2)^{q/2}$ satisfies the inequality (29) for some compact set K and some positive finite constants c_1, c_2. Note that

$$g_q(x+z) - g_q(x) \leq (1 + \|x\|_H^2 + \|z\|_H^2)^{q/2} - (1 + \|x\|_H^2)^{q/2} = \frac{q}{2}\xi^{q/2-1}\|z\|_H^q,$$

where the equality follows from the mean value theorem and $\xi \in (1 + \|x\|_H^2, 1 + \|x\|_H^2 + \|z\|_H^2)$. Since $\xi > 1$ and $p \in (0, 1)$, we have $g_q(x+z) - g_q(x) \leq \frac{q}{2}\|z\|_H^q$. Likewise, it can be shown that $g_q(x) - g_q(x+z) \leq \frac{q}{2}\|z\|_H^q$. Hence, $|g_q(x+z) - g_q(x)| \leq \frac{q}{2}\|z\|_H^q$ and

$$\left|\int_{\mathcal{X}} g_q(x+z) - g_q(x)\,\nu(\mathrm{d}z)\right| \leq \int_{\mathcal{X}} |g_q(x+z) - g_q(x)|\,\nu(\mathrm{d}z) \leq \frac{q}{2}\,\mathbb{E}\,\|Z\|_H^q < \infty,$$

It follows that with $\kappa = \mathbf{0}$,

$$\mathscr{L} g_q(x) = \lambda \int_{\mathcal{X}} (g_q(x+z) - g_q(x))\,\nu(\mathrm{d}z) = O(1),$$

as $\|x\|_H \to \infty$. Moreover, applying (31) to $g_q(x)$,

$$\mathscr{G} g_q(x) = q g_q(x)\left(\frac{x^\intercal H\beta x}{\|x\|_H^2} + o(1)\right).$$

Hence,

$$\mathscr{A} g_q(x) = \mathscr{G} g_q(x) + \mathscr{L} g_q(x) = q g_q(x)\left(\frac{x^\intercal H\beta x}{\|x\|_H^2} + o(1)\right),$$

as $\|x\|_H \to \infty$. By the definition of the matrix H,

$$-x^\intercal H\beta x = -\frac{1}{2}x^\intercal(H\beta + \beta^\intercal H)x \geq \frac{1}{2}\underline{\gamma}\|x\|^2 \geq \frac{1}{2}\underline{\gamma}\bar{\delta}^{-1}\|x\|_H^2,$$

where $\underline{\gamma} > 0$ is the smallest eigenvalue of $-(H\beta + \beta^\intercal H)$. Hence, there exists $k > 0$ for which

$$\mathscr{A} g(x) \leq -\frac{1}{4}p\underline{\gamma}\bar{\delta}^{-1}g(x)$$

for all $x \in \mathcal{X}$ with $\|x\|_H > k$. Therefore,

$$\mathscr{A} g_q(x) \leq -c_1 g_q(x) + c_2\,\mathbb{I}_K(x), \quad x \in \mathcal{X}, \tag{36}$$

where $K = \{x \in \mathcal{X} : \|x\|_H \le k\}$, $c_1 = p\underline{\gamma}\bar{\delta}^{-1}/4$, and $c_2 = \max\{1, \sup_{x\in K}(\mathscr{A} g(x) + c_1 g(x))\}$.

We apply Itô's formula to $e^{c_1 t} g_q(X(t))$. In particular, by (11),

$$\begin{aligned} e^{c_1 t} g_q(X(t)) = {} & g_q(X(0)) + \int_0^t e^{c_1 s}[c_1 g_q(X(s\text{-})) + \mathscr{A} g_q(X(s\text{-}))]\mathrm{d}s \\ & + \int_0^t e^{c_1 s} \nabla g_q(X(s\text{-}))^\intercal \sigma(X(s))\, \mathrm{d}W(s) \\ & + \int_0^t e^{c_1 s} \int_{\mathcal{X}} (g_q(X(s\text{-}) + z) - g_q(X(s\text{-})))\tilde{N}(\mathrm{d}s, \mathrm{d}z). \end{aligned}$$

Clearly, the two stochastic integrals above are both martingales up to time T_n, where $T_n = \{t \ge 0 : |X(t)| > n\}$. It follows from (36) and the optional sampling theorem that

$$e^{c_1 t}\, \mathbb{E}_x\, g_q(X(t\wedge T_n)) \le g_q(x) + \mathbb{E}_x \int_0^{t\wedge T_n} e^{c_1 s} \cdot c_2\, \mathbb{I}_K(X(s))\, \mathrm{d}s \le g_q(x) + c_2 c_1^{-1}\, \mathbb{E}_x\, e^{t\wedge T_n}.$$

We now apply Fatou's lemma and the monotone convergence theorem to conclude that

$$e^{c_1 t}\, \mathbb{E}_x\, g_q(X(t)) \le g_q(x) + c_2 c_1^{-1} \cdot \liminf_{n\to\infty} \mathbb{E}_x\, e^{t\wedge T_n} = g_q(x) + c_2 c_1^{-1} e^{c_1 t}. \tag{37}$$

Then we can adopt the argument used in the proof of Theorem 6.1 of Meyn and Tweedie (1993c) to conclude that because of (37), there exist positive finite constants d_q and ρ_q such that

$$\|\mathbb{P}_x(X(t) \in \cdot) - \pi(\cdot)\|_{g_q+1} \le d_q(g_q(x) + 1)e^{-\rho_q t}, \quad t \ge 0,\ x \in \mathcal{X}.$$

Moreover, thanks to (12), there exist positive finite constants d_1 and d_2 such that

$$d_1 \le \left| \frac{f_q(x)}{g_q(x) + 1} \right| \le d_2, \quad x \in \mathcal{X}.$$

Hence,

$$\|\mathbb{P}_x(X(t) \in \cdot) - \pi(\cdot)\|_{f_q} \le c_q f_q(x) e^{-\rho_q t}, \quad t \ge 0,\ x \in \mathcal{X},$$

where $c_q = d_q d_2/d_1$. □

Proof of Theorem 2(ii). Following the proof of Theorem 1(ii), it suffices to show that (36) holds under the present assumptions. Note that $\mathbb{E}\,\|Z\|^q < \infty$ for all $q \in [1, p]$ since $\mathbb{E}\,\|Z\|^p < \infty$. Hence, we can apply the Lyapunov inequality (29) to $g_q(x)$, which results in (36). □

3.3 Remarks on the Strong Mean-Reversion Condition

The key condition that we impose to establish positive Harris recurrence of X is the strong mean-reversion condition, i.e, $\beta + \mathbb{E}(Z)\kappa^\intercal$ is a stable matrix. Indeed, this condition cannot be relaxed in general as illustrated by the following example.

Proposition 3. *Suppose that $d = 1$, $m = 1$, and Assumptions 1 - 3 hold. If $\mathbb{E}|Z| < \infty$ and $\beta + \mathbb{E}(Z)\kappa > 0$, then X is transient.*

Proof. The proof also relies on Lyapunov inequalities; see Theorem 3.3 of Stramer and Tweedie (1994). Specifically, it suffices to show that there exists a nontrivial bounded function g and a closed set K such that

$$\mathscr{A} g(x) \geq 0, \quad x \in \mathcal{X} \setminus K, \tag{38}$$

and

$$\sup_{x\in K} g(x) < g(x_0), \quad x_0 \in \mathcal{X} \setminus K. \tag{39}$$

Let $g(x) = 1 - e^{-\epsilon x}$ for some $\epsilon > 0$. Obviously, g is bounded for $x \in \mathcal{X} = \mathbb{R}_+$. Then,

$$\begin{aligned}
\mathscr{A} g(x) &= (b + \beta x)g'(x) + \frac{1}{2}(a + \alpha x)g''(x) + (\lambda + \kappa x)\int_{\mathbb{R}_+}(g(x + z) - g(x))\,\nu(\mathrm{d}z) \\
&= e^{-\epsilon x}\left[\epsilon(b + \beta x) - \frac{\epsilon^2}{2}(a + \alpha x) + (\lambda + \kappa x)\int_{\mathbb{R}_+}(1 - e^{-\epsilon z})\,\nu(\mathrm{d}z)\right] \\
&= e^{-\epsilon x}\left[\left(\epsilon\beta - \frac{\epsilon^2}{2}\alpha + \kappa(1 - \mathbb{E}\,e^{-\epsilon Z})\right)x + \epsilon b - \frac{\epsilon^2}{2}a + \lambda(1 - \mathbb{E}\,e^{-\epsilon Z})\right].
\end{aligned}$$

Let $h(\epsilon)$ be the coefficient of x in the brackets above, i.e.,

$$h(\epsilon) := \epsilon\beta - \frac{\epsilon^2}{2}\alpha + \kappa(1 - \mathbb{E}\,e^{-\epsilon Z}).$$

Clearly, $h(0) = 0$ and $h'(0) = \beta + \kappa\,\mathbb{E}(Z) > 0$, yielding that $h(\epsilon) > 0$ for some $\epsilon > 0$. Fixing this ϵ, we see that $\mathscr{A} g(x) \sim e^{-\epsilon x}h(\epsilon)x$ as $x \to \infty$. Hence, there exists $k > 0$ such that $\mathscr{A} g(x) > 0$ for $x \in \mathcal{X} \setminus K$, where $K := [0, k]$, proving (38). Moreover, (39) is true since $g(x)$ is increasing in x. $\square$

The "boundary" case, i.e. $\beta + \mathbb{E}(Z)\kappa^\intercal = 0$, is more complicated as the behavior of the process may depend on other parameters. We leave its analysis for future research.

4 Limit Theorems

In this section, we prove SLLNs and FCLTs for additive functionals of X of the form $\int_0^t h(X(s))\,\mathrm{d}s$ or $\sum_{i=1}^n h(X(i\Delta))$ for some function h. Limit theorems for both discrete-time and continuous-time Markov processes have been extensively studied in the past; see, e.g., Glynn and Meyn (1996), Kontoyiannis and Meyn (2003), Meyn and Tweedie (2009, chap.17), and references therein. In particular, positive Harris recurrence is "almost" sufficient for an LLN to hold. Conditions for FCLTs, on the other hand, often include exponential ergodicity, or Lyapunov inequalities of the form similar to (29).

Nevertheless, existing FCLTs for discrete-time Markov processes are not applicable to the skeleton chain X^Δ because they typically require one to establish a "discrete-time" version of Lyapunov inequality of the form $\mathbb{E}_x[g(X(\Delta))] \leq cg(x)$ for some constant $c < 1$, some function $g \geq 1$, and all x

off a compact set. This is challenging given the fact that the transition distribution $\mathbb{P}_x(X(\Delta) \in \cdot)$ is not known explicitly. Our approach to establish (8) is to first consider the scenario in which $X(0)$ follows the stationary distribution. We then apply an FCLT for stationary sequences, i.e., Theorem 3.1 of Ethier and Kurtz (1986, p.351), whose conditions can be verified as a consequence of exponential ergodicity (4). To generalize the FCLT to an arbitrary initial state we apply the shift coupling property of positive Harris recurrent Markov processes; see Aldous and Thorisson (1993) for the application of this technique in a discrete-time setting.

The asymptotic variances, σ_h^2 in (7) and γ_h^2 in (8), can be expressed in terms of the solution to a *Poisson equation*; see, e.g., Glynn and Meyn (1996). But it typically has no closed form in terms of the parameters $(a, \alpha_1 \dots, \alpha_d, b, \beta, \lambda, \kappa, \nu)$ of SDE (1). However, when h is the (vector-valued) identity function, we are indeed able to analytically derive both the asymptotic mean and asymptotic covariance matrix that appear in the corresponding FCLT (see Corollary 1), thanks to the tractable affine structure.

4.1 Strong Law of Large Numbers

Proof of Theorem 1(iii) and Theorem 2(iii). We have established positive Harris recurrence and ergodicity of X in Section 3.1 under the assumptions of Theorem 1(iii) or Theorem 2(iii). So $\pi(|h|) < \infty$ for any measurable function $h : \mathcal{X} \mapsto \mathbb{R}$ with $\|h\|_{f_p} < \infty$. The SLLN (5) then follows from Theorem 2 of Sigman (1990).

For the skeleton chain X^Δ, note that the stationary distribution π of X is necessarily invariant for X^Δ. In addition, X^Δ is φ-irreducible by Proposition 1, so X^Δ is positive Harris recurrent. Hence, the SLLN (6) follows from Theorem 17.1.7 of Meyn and Tweedie (2009, p.427). $\square$

4.2 Functional Central Limit Theorem

Proof of Theorem 1(iv) and Theorem 2(iv). Fix $q > 2$ and an arbitrary measurable function $h : \mathcal{X} \mapsto \mathbb{R}$ with $\|h^q\|_{f_p} < \infty$.

We have shown in Section 3.2 that there exists a matrix $H \succ 0$, a compact set K and positive finite constants c_1, c_2 such that

$$\mathscr{A} g(x) \le -c_1 g(x) + c_2 \mathbb{I}_K(x), \quad x \in \mathcal{X},$$

where $g(x) = (1 + \|x\|_H^2)^{p/2}$. Thanks to (12), $\|h\|_{f_p} < \infty$ if and only if $\|h\|_g < \infty$. Moreover, we have shown in Section 3.1 that K is a petite set for X. It then follows immediately from Theorem 4.4 of Glynn and Meyn (1996) that (7) holds as $n \to \infty$ $\mathbb{P}_x$-weakly in $\mathcal{D}[0, 1]$ for all $x \in \mathcal{X}$.

We now show that (8) holds $\mathbb{P}_\pi$-weakly in $\mathcal{D}[0, 1]$, where π is the stationary distribution of X. This can be done by applying an FCLT for stationary sequences to $\{\bar{h}(X(n\Delta) : n = 0, 1, \dots\}$, which is a mean-zero stationary sequence if $X(0) \sim \pi$, where $\bar{h}(x) \coloneqq h(x) - \pi(h)$.

Specifically, let $\mathscr{F}_k$ and $\mathscr{F}^k$ denote the σ-algebras generated by $(X(n\Delta) : n \le k)$ and $(X(n\Delta) : n \ge k)$, respectively. Let $\varphi_1(l) \coloneqq \sup_{\Gamma \in \mathscr{F}^{k+l}} \mathbb{E}_\pi |\mathbb{P}(\Gamma | \mathscr{F}_k) - \mathbb{P}(\Gamma)|$ denote the *measure of mixing*

(Ethier and Kurtz 1986, p.346) of $\mathscr{F}_k$ and $\mathscr{F}^{k+l}$ associated with the L^1-norm. Then, by Theorem 3.1 and Remark 3.2(b) of Ethier and Kurtz (1986, p.351), it suffices to verify that for some $\epsilon > 0$,

$$\mathbb{E}_\pi\left[\left|\bar{h}(X(n\Delta))\right|^{2+\epsilon}\right] < \infty \quad \text{and} \quad \sum_{l=0}^{\infty}[\varphi_1(l)]^{\epsilon/(2+\epsilon)} < \infty. \tag{40}$$

Let $\epsilon = q - 2 > 0$. Then, $\mathbb{E}_\pi\left[\left|\bar{h}(X(n\Delta))\right|^{2+\epsilon}\right] = \pi(\bar{h}^q) \leq \left\|\bar{h}^q\right\|_{f_p}\pi(f_p) < \infty$, verifying the first condition in (40). To verify the second, notice that by the Markov property, for any $\Gamma \in \mathscr{F}^{k+l}$ there exists a function w_Γ with $|w_\Gamma(\cdot)| \leq 1$ such that $\mathbb{P}[\Gamma|\mathscr{F}_{k+l}] = w_\Gamma(X((k+l)\Delta))$. If $X(0) \sim \pi$, then for any $\Gamma \in \mathscr{F}^{k+l}$,

$$\begin{aligned}
|\mathbb{P}(\Gamma|\mathscr{F}_k) - \mathbb{P}(\Gamma)| &= |\mathbb{E}[w_\Gamma(X((k+l)\Delta))|\mathscr{F}_k] - \mathbb{E}[w_\Gamma(X((k+l)\Delta))]| \\
&= \left|\int_{\mathcal{X}} w_\Gamma(y)\, \mathbb{P}_{X(k\Delta)}(X(l\Delta) \in \mathrm{d}y) - \int_{\mathcal{X}}\int_{\mathcal{X}} w_\Gamma(y)\, \mathbb{P}_x(X((k+l)\Delta) \in \mathrm{d}y)\pi(\mathrm{d}x)\right| \\
&\leq \left\|\mathbb{P}_{X(k\Delta)}(X(l\Delta) \in \cdot) - \pi(\cdot)\right\|,
\end{aligned}$$

where $\|\cdot\|$ is the total variation norm, where the inequality follows from Definition 3 and the fact that $|w_\Gamma(\cdot)| \leq 1$. It follows that

$$\varphi_1(l) \leq \mathbb{E}_\pi \left\|\mathbb{P}_{X(k\Delta)}(X(l\Delta) \in \cdot) - \pi(\cdot)\right\| \leq c_p e^{-\rho_p l\Delta}\, \mathbb{E}_\pi[f(X(k\Delta))] = c_p\pi(f_p)e^{-\rho_p l\Delta},$$

where the second inequality holds because of Theorem 1(ii) and Theorem 2(ii). This immediately implies $\sum_{l=0}^{\infty}[\varphi_1(l)]^{\epsilon/(2+\epsilon)} < \infty$. Therefore, we conclude that (8) holds $\mathbb{P}_\pi$-weakly in $\mathcal{D}[0,1]$.

We now prove that (8) indeed holds as $n \to \infty$ $\mathbb{P}_x$-weakly in $\mathcal{D}[0,1]$ for all $x \in \mathcal{X}$. To that end, we first show that

$$\mathbb{P}_x\left(\lim_{n\to\infty}\sup_{0\leq t\leq 1}|Y_n(t) - Y_{n,l}(t)| = 0\right) = 1, \quad x \in \mathcal{X}, \tag{41}$$

for any positive integer l, where $Y_n(t) := n^{-1/2}\sum_{i=1}^{\lfloor nt\rfloor}\bar{h}(X(i\Delta))$ and $Y_{n,l}(t) := n^{-1/2}\sum_{i=l+1}^{\lfloor nt\rfloor+l}\bar{h}(X(i\Delta))$. Notice that for all sufficiently large n,

$$\begin{aligned}
\sup_{0\leq t\leq 1}|Y_n(t) - Y_{n,l}(t)| &= \frac{1}{n}\left|\sum_{i=1}^{n}\bar{h}(X(i\Delta)) - \sum_{i=l+1}^{n+l}\bar{h}(X(i\Delta))\right|^2 \\
&= \frac{1}{n}\left|\sum_{i=1}^{l}\bar{h}(X(i\Delta)) - \sum_{i=n+1}^{n+l}\bar{h}(X(i\Delta))\right|^2 \\
&\leq \frac{1}{n}\sum_{i=1}^{l}\bar{h}^2(X(i\Delta)) + \frac{1}{n}\sum_{i=n+1}^{n+l}\bar{h}^2(X(i\Delta)) \to 0, \quad \mathbb{P}_x\text{ –a.s.},
\end{aligned}$$

as $n \to \infty$ for all $x \in \mathcal{X}$, because $\frac{1}{n}\sum_{i=1}^{n}\bar{h}^2(X(i\Delta)) \to \pi(\bar{h}^2) < \infty$, $\mathbb{P}_x$ –a.s., as $n \to \infty$ for all $x \in \mathcal{X}$, thanks to Theorem 1(iii) and Theorem 2(iii). This completes the proof of (41).

Let ϕ be a bounded continuous functional ϕ on $\mathcal{D}[0,1]$. Then (41) implies that for any positive integer l, $|\mathbb{E}_x[\phi(Y_n)] - \mathbb{E}_x[\phi(Y_{n,l})]| \to 0$ as $n \to \infty$ for all $x \in \mathcal{X}$. This limit can be rewritten as

$$\lim_{n\to\infty} \left| \mathbb{E}_x[\phi(Y_n)] - \int_{\mathcal{X}} \mathbb{P}_x(X(l\Delta) \in \mathrm{d}y)\, \mathbb{E}_y[\phi(Y_{n,l})] \right| = 0, \quad x \in \mathcal{X}. \tag{42}$$

On the other hand, note that

$$\left| \int_{\mathcal{X}} \mathbb{P}_x(X(l\Delta) \in \mathrm{d}y)\, \mathbb{E}_y[\phi(Y_{n,l})] - \mathbb{E}_\pi[\phi(Y_n)] \right| \le \|\mathbb{P}_x(X(l\Delta) \in \cdot) - \pi(\cdot)\| \cdot \sup_{g\in\mathcal{D}[0,1]} |\phi(g)|.$$

Since $\|\mathbb{P}_x(X(l\Delta) \in \cdot) - \pi(\cdot)\| \to 0$ as $l \to \infty$ by Theorem 1(i) and Theorem 2(i), for any $\delta > 0$ we can choose l so large that

$$\left| \int_{\mathcal{X}} \mathbb{P}_x(X(l\Delta) \in \mathrm{d}y)\, \mathbb{E}_y[\phi(Y_{n,l})] - \mathbb{E}_\pi[\phi(Y_n)] \right| \le \delta. \tag{43}$$

It then follows from (42) and (43) that $\limsup_{n\to\infty} |\mathbb{E}_x[\phi(Y_n)] - \mathbb{E}_\pi[\phi(Y_n)]| \le \delta$. Since (8) holds $\mathbb{P}_\pi$-weakly in $\mathcal{D}[0,1]$, we must have $\lim_{n\to\infty} |\mathbb{E}_\pi[\phi(Y_n)] - \mathbb{E}_\pi[\phi(W)]| = 0$, and thus

$$\limsup_{n\to\infty} |\mathbb{E}_x[\phi(Y_n)] - \mathbb{E}_\pi[\phi(W)]| \le \delta.$$

Sending $\delta \to 0$ yields that (8) holds $\mathbb{P}_x$-weakly in $\mathcal{D}[0,1]$ for all $x \in \mathcal{X}$. $\square$

4.3 A Special Case

Thanks to the affine structure, the asymptotic mean and the asymptotic variance can be derived analytically when h is the identity function, i.e. $h(x) = x$. Note that with h being $\mathbb{R}^d$-valued, the corresponding SLLN and FCLT are multivariate. The calculation follows closely to the approach used in Zhang et al. (2015) so we omit the details.

Corollary 1. *Suppose that Assumptions 1–3 hold. If $\mathbb{E}\,\|Z\| < \infty$, then*

$$\mathbb{P}_x\left(\lim_{t\to\infty} \frac{1}{t} \int_0^t h(X(s))\,\mathrm{d}s = v \right) = 1, \quad x \in \mathcal{X},$$

where $v = -(\beta + \mathbb{E}(Z)\kappa^\intercal)^{-1}(b + \lambda\,\mathbb{E}(Z))$. Further, if $\mathbb{E}\,\|Z\|^{2+\epsilon} < \infty$ for some $\epsilon > 0$, then

$$n^{1/2}\left(\frac{1}{n} \int_0^{n\cdot} X(s)\,\mathrm{d}s - v \right) \Rightarrow \Sigma^{1/2} W(\cdot),$$

as $n \to \infty$ $\mathbb{P}_x$-weakly in $\mathcal{D}_{\mathbb{R}^d}[0,1]$ for all $x \in \mathcal{X}$, where

$$\Sigma = A(a + \lambda\,\mathbb{E}(ZZ^\intercal))A^\intercal + \sum_{i=1}^m v_i A(\alpha_i + \kappa_i\,\mathbb{E}(ZZ^\intercal))A^\intercal,$$

$\mathcal{D}_{\mathbb{R}^d}[0,1]$ *denote the space of right continuous functions* $x : [0,\infty) \mapsto \mathbb{R}^d$ *with left limits, endowed with the Skorokhod topology and* W *is a* d*-dimensional Wiener process.*

Acknowledgments

The research is partially supported by the Hong Kong Research Grants Council under General Research Fund Project No. 624112.

References

Y. Aït-Sahalia. Estimating continuous-time models using discretely sampled data. In R. Blundell, P. Torsten, and W. K. Newey, editors, *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress*, chapter 9. Cambridge University Press, 2007.

D. J. Aldous and H. Thorisson. Shift-coupling. *Stoch. Proc. Appl.*, 44:1–14, 1993.

L. B. G. Andersen and V. V. Piterbarg. Moment explosions in stochastic volatility models. *Financ. Stoch.*, 11:29–50, 2007.

M. Barczy, L. Döring, Z. Li, and G. Pap. Stationarity and ergodicity for an affine two-factor models. *Adv. Appl. Probab.*, 46(3):878–898, 2014.

O. E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Statist. Soc. B*, 63(2):167–241, 2001.

D. S. Bates. Post-'87 crash fears in the S&P 500 futures option market. *J. Econometrics*, 94(1-2):181–238, 2000.

D. S. Bates. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Rev. Financ. Stud.*, 9(1):69–107, 1996.

D. S. Bates. Maximum likelihood estimation of latent affine processes. *Rev. Financ. Stud.*, 19(3):909–965, 2006.

A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences.* SIAM, Philadelphia, 1994.

P. Cheridito, D. Filipović, and R. L. Kimmel. Market price of risk specification for affine models: Theory and evidence. *J. Financ. Econ.*, 83:123–170, 2007.

P. Collin-Dufresne, R. S. Goldstein, and C. S. Jones. Identification of maximal affine term structure models. *J. Finance*, 63(2):743–759, 2008.

J. C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53 (2):385–407, 1985.

Q. Dai and K. J. Singleton. Specification analysis of affine term structure models. *J. Finance*, 55:1943–1978, 2000.

E. B. Davies. *Heat Kernels and Spectral Theory*, volume 92 of *Cambridge Tracts in Mathematics.* Cambridge University Press, 1989.

D. Duffie and P. W. Glynn. Estimation of continuous-time markov processes sampled at random time intervals. *Econometrica*, 72(6):1773–1808, 2004.

D. Duffie and R. Kan. A yield-factor model of interest rates. *Math. Finance*, 6:379–406, 1996.

D. Duffie, J. Pan, and K. J. Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.

D. Duffie, D. Filipović, and W. Schachermayer. Affine processes and applications in finance. *Ann. Appl. Probab.*, 13(3):984–1053, 2003.

S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, Inc., 1986.

D. Filipović and E. Mayerhofer. Affine diffusion processes: Theory and applications. In H. Albrecher, W. Runggaldier, and W. Schachermayer, editors, *Radon Ser. Comput. Appl. Math.*, volume 8, pages 1–40. 2009.

D. Filipović, E. Mayerhofer, and P. Schneider. Density approximations for multivariate affine jump-diffusion processes. *J. Econometrics*, 176:93–111, 2013.

Z. Fu and Z. Li. Stochastic equations of non-negative processes with jumps. *Stoch. Anal. Appl.*, 120(3): 306–330, 2010.

R. K. Getoor. *Markov Processes: Ray Processes and Right Processes*. Lecture Notes in Mathematics. Springer-Verlag Berlin Heidelberg, 1975.

P. Glasserman and K.-K. Kim. Moment explosions and stationary distributions in affine diffusion models. *Math. Finance*, 20(1):1–33, 2010.

P. W. Glynn and S. P. Meyn. A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.*, 24(2): 916–931, 1996.

L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50: 1029–1054, 1982.

L. P. Hansen and J. A. Scheinkman. Back to the future: generating moment implications for continuous-time Markov processes. *Econometrica*, 63(4):767–804, 1995.

S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.*, 6:327–343, 1993.

R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2012.

R. P. Jena, K.-K. Kim, and H. Xing. Long-term and blow-up behaviors of exponential moments in multi-dimensional affine diffusions. *Stoch. Proc. Appl.*, 122:2961–2993, 2012.

P. Jin, B. Rüdiger, and C. Trabelsi. Positive Harris recurrence and exponential ergodicity of the basic affine jump-diffusion. *Stoch. Anal. Appl.*, 34(1):75–95, 2016.

P. Jin, J. Kremer, and B. Rüdiger. Exponential ergodicity of an affine two-factor model based an the α-root process. *Adv. Appl. Probab.*, 49:1144–1169, 2017.

I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 2 edition, 1991.

M. Keller-Ressel. Moment explosions and long-term behavior of affine stochastic volatility model. *Math. Finance*, 21(1):73–98, 2011.

M. Keller-Ressel, W. Schachermayer, and J. Teichmann. Affine processes are regular. *Probab. Theor. Relat. Field.*, 151:591–611, 2011.

I. Kontoyiannis and S. P. Meyn. Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.*, 13(1):304–362, 2003.

R. W. Lee. The moment formula for implied volatlity at extreme strikes. *Math. Finance*, 14(3):469–480, 2004.

H. Masuda. On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli*, 10:97–120, 2004.

S. P. Meyn and R. L. Tweedie. Stability of Markovian processes I: criteria for discrete-time chains. *Adv. Appl. Probab.*, 24:542–574, 1992.

S. P. Meyn and R. L. Tweedie. Generalized resolvents and Harris recurrence of Markov processes. *Contemporary Mathematics*, 149:227–250, 1993a.

S. P. Meyn and R. L. Tweedie. Stability for Markovian processes II: continuous-time processes and sampled chains. *Adv. Appl. Probab.*, 25:487–517, 1993b.

S. P. Meyn and R. L. Tweedie. Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Adv. Appl. Probab.*, 25:518–548, 1993c.

S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, 2nd edition, 2009.

P. E. Protter. *Stochastic Integration and Differential Equations*. Springer, 2 edition, 2003.

K.-i. Sato and M. Yamazato. Operator-self-decomposable distributions as limit distributions of processes of Ornstein–Uhlenbeck type. *Stoch. Proc. Appl.*, 17:73–100, 1984.

M. Sharpe. *General Theory of Markov Processes*. Academic Press, 1988.

K. Sigman. One-dependent regenerative processes and queues in continuous time. *Math. Oper. Res.*, 15(1): 175–189, 1990.

K. J. Singleton. Estimation of affine asset pricing models using the empirical characteristic function. *J. Econometrics*, 102(1):111–141, 2001.

O. Stramer and R. L. Tweedie. Stability and instability of continuous time Markov processes. In F. P. Kelly, editor, *Probability, Statistics and Optimization: A Tribute to Peter Whittle*, pages 173–184. John Wiley & Sons, Inc., 1994.

O. Vasicek. An equilibrium characterization of the term structure. *J. Financ. Econ.*, 5:177–188, 1977.

X. Zhang, J. Blanchet, K. Giesecke, and P. W. Glynn. Affine point processes: Approximation and efficient simulation. *Math. Oper. Res.*, 40(4):797–819, 2015.