Incremental Nonlinear Stability Analysis for Stochastic Systems Perturbed by Lévy Noise

SooJean Han and Soon-Jo Chung, *Senior Member, IEEE*

Abstract— **We extend incremental stability conditions for nonlinear stochastic systems from Gaussian white noise to general Lévy noise. Our main contribution in this paper is to use contraction theory to show that different trajectories of a system perturbed by different sample paths of Lévy noise exponentially converge towards each other in the mean-squared sense, achieving stability within finite time. Our paper takes a step forward in the development of an analytical framework for understanding stochastic systems with non-Gaussian noise particularly in robotic applications, which should be of great interest because it would allow for controller design through a principled approach rather than complete reliance on model-free techniques. Numerical illustrations are provided with the Dubins car and a scalar LQR system subject to constant-variation Lévy noise.**

Index Terms— **Nonlinear systems, Poisson processes, Random processes, Robust stability, Stability analysis, Stability criteria, Stochastic processes, Stochastic systems, Uncertain systems**

I. INTRODUCTION

MANY model-based designs for controllers and observers of real-world stochastic systems aim for robustness against additive white Gaussian noise (AWGN). Theory developed around AWGN is the foundation of the field of information theory [1]. In applications that are more relevant to robotics, such as spacecraft navigation [2], vision-based localization/mapping [3], and motion-planning [4], AWGN is often used to model measurement noise and process noise. This is possibly attributed to the versatility of the Gaussian random process, e.g., an affine combination of Gaussian random variables is still Gaussian-distributed, as well as its widespread presence in the real world, a phenomenon which is theoretically formalized by the Central Limit Theorem (CLT). Such desirable properties make AWGN very appealing to use when considering deterministic systems in a stochastic environment.

However, a major flaw with models which only consider pure white noise is the lack of generality. White noise is typically small in magnitude, continuous in the sense that large drifts occur steadily over time, and affects a system persistently for a measurable duration of time. For instance, the

SooJean Han is with the Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, USA. Email: soojean@caltech.edu.

Soon-Jo Chung is with the Graduate Aerospace Laboratory, California Institute of Technology, Pasadena, CA, 91125, USA. Email: sjchung@caltech.edu.
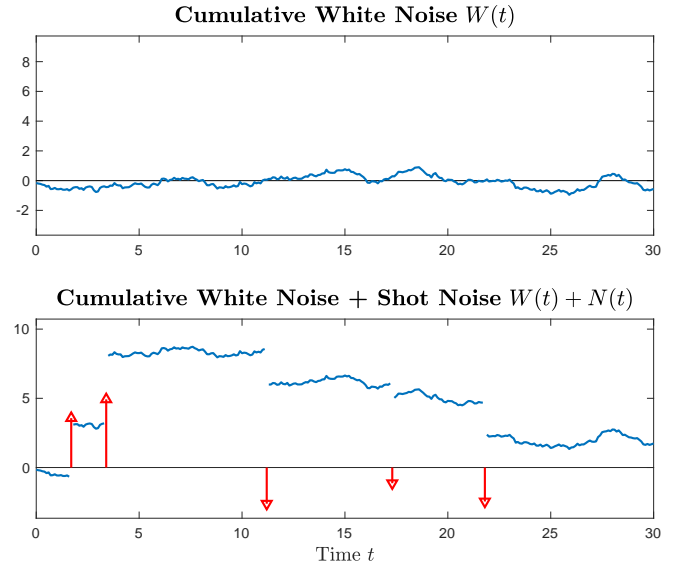
Fig. 1: In the top, a pure white noise process. In the bottom, a white noise process superimposed atop shot noise impulses. As a result of the shot noise, the effect over time is more severe (larger drift) than pure white noise.

white Gaussian noise process has *unbounded variation*, yet the variance is kept proportional to the approximate incremental timestep, which decreases to zero. So there is a very small probability that consecutive random variables will have a large discrepancy between them [5]. Thus, it is unable to characterize a broad class of sudden impulsive perturbations which are often referred to as *shot noise* [6] or *jump noise*. These sudden discrete jumps have distributions which are better modeled as a Poisson process, and are also often referred to as *Poisson noise*; we will be using all three terms interchangeably throughout this paper. Furthermore, there is an even broader class of general *Lévy noise* processes which, loosely-speaking, can be decomposed into a continuous-part white noise process and a discontinuous-part Poisson noise process through the well-known *Lévy-Khintchine Decomposition Theorem* (Section III-A, Theorem 5). Intuitive visualizations of a pure white noise process, and a white noise process superimposed atop shot noise impulses are shown in Fig. 1.

Such types of noise arise in real-world scenarios almost just as commonly as the pure white noise does, like the large fluctuations in stock prices in financial economics settings [7], as well as signal-processing neuronal spikes arising from brain

activity in neuroscientific applications [8]. In the case of robotics, Lévy noise commonly arises due to abrupt contact with the environment: one might consider a flying mechanism experiencing a sudden gust of wind, a collision against a tree branch, or a rough manual push by a human. Extremely noisy readings from proprioceptive sensors are also often a source of shot noise. However, despite being such a prevalent phenomenon, Lévy noise effects, or even simply pure shot noise effects, have rarely ever been included for consideration in robotic and cyberphysical systems to the authors' knowledge.

In this paper, we address a subset of the finite-time stochastic stability questions posed by [9] for nonlinear systems which are perturbed by an affine combination of Gaussian white noise and Poisson shot noise, also known collectively as Lévy noise. In particular, for a given bounded set of initial conditions, can trajectories of the system, arising from different sample paths of the noise, be bounded within some region? To address this question, the primary tool we use in our analysis is contraction theory, which is used to prove incremental stability of the system. Incremental stability is considered because it is stronger than the traditional Lyapunov sense by allowing for global exponential convergence with respect to time-varying trajectories. The bound we derive is achieved in the mean-squared sense and bounded for any given fixed interval of time, implying finite-time stability.

A. Related Work

Traditional characterizations of deterministic system stability arise from the construction of Lyapunov functions [10] as well as the well-known direct and indirect Lyapunov methods, which can be found in any standard control theory textbook (e.g. [11], [12]). In contrast, we will primarily focus on utilizing contraction theory to identify a type of stability known as *incremental stability*, which generalizes the Lyapunov approach by comparing the behavior of multiple solution trajectories with respect to each other or any desired time-varying trajectory instead of an equilibrium point or a limit cycle. Furthermore, systems with solution trajectories which satisfy the incremental stability property have guaranteed global exponential convergence towards the desired trajectory. This particular notion of stability has found application in numerous settings such as cooperative control over multi-agent swarm systems [13] and phase synchronization in directed networks [14], [15]. Further connections between Lyapunov and incremental stability can be found in [16]. There has been an extensive amount of work done on deciding incremental stability for deterministic nonlinear systems [17], [18].

Before we continue forth with our discussion on stability analysis methods for stochastic systems, we remind ourselves the definition of a Lévy process:

Definition 1. A process $L(t)$ is said to be a *Lévy process* if all paths of L are right-continuous and left-limit (rcll), $\mathbb{P}(L(0) = 0) = 1$, and L has stationary and independent increments.

Lévy processes are known to satisfy the *strong Markov property* (informally speaking, this means it satisfies the memorylessness property of Markov processes conditioned on

any nonnegative stopping time). The work of Kushner [9] summarized various problems of interest when considering stability of right-continuous strong Markov processes such as Lévy processes, and subsequently laid out the foundations of Lyapunov-based stochastic stability theory. Control Lyapunov functions are used for disturbance attenuation of processes subject to unknown-covariance Gaussian white noise in [19], and stochastic Lyapunov-like techniques are used to provide sufficient conditions on the existence of smooth stabilizing feedback laws for controlling white-noise-perturbed nonlinear systems [20] and systems which are affine in control input [21].

Incremental stability criteria for nonlinear systems has recently been an area of active research, perhaps due to its generalization of Lyapunov stability to time-varying trajectories as well as its guaranteed globally exponential convergence. In particular, conditions have been developed for nonlinear stochastic systems with Gaussian white noise; the work of [22] takes the first step doing so. Stability is considered with respect to both constant and time-dependent metrics, where a metric is, informally defined, a coordinate transformation of the original system dynamics which allows for easier stability analysis. In addition, [23] extends the theory in [22] to more general state-dependent metrics, which is useful in the construction of nonlinear observers or controllers. But, as mentioned before, there are many real-world systems that suffer stochastic perturbations which are non-Gaussian.

To this end, there have also been literature on characterizing stochastic stability for systems perturbed by non-Gaussian noise. [24] develops a stochastic Lyapunov-like characterization of stability for stochastic discrete-time systems, while [25] uses Lyapunov exponents for stochastic stability criteria on systems driven by Lévy noise. Asymptotic stability of systems driven by Lévy noise is developed in [26] and an exponential stability result is derived in [27]. However, to the authors' knowledge, using incremental stability to prove convergence between two time-varying trajectories perturbed by non-Gaussian noise has not been done prior to this paper, which derives a condition to achieve finite-time incremental stability for nonlinear stochastic systems for two classes of non-Gaussian noise: shot noise and Lévy noise.

Lyapunov-like approaches for hybrid systems and jump-Markov systems are also relevant in the sense that stabilization techniques are focused on handling jump-discontinuities; [28] derives a sufficient Lie-algebraic condition for a family of linear systems to be able to achieve asymptotic stability and [29] makes extensions to LaSalle's Invariance Principle. Additionally, Lyapunov criteria for the control of jump-Markov processes has been studied in [30], and contraction theoretic characterizations of incremental stability for hybrid systems was studied in [31]. There is a poignant distinction between systems perturbed by a general class of jump-discontinuous noise processes (e.g., Lévy) versus hybrid systems or jump-Markov systems which switch between multiple modes; namely, jumps arise solely from the noise process, independently of the open-loop dynamics, whereas hybrid or jump-Markov systems often model switches (i.e., jumps) as an inherent property in the open-loop dynamics. However, the

two types of systems can still be related to one another through the conditions imposed on the *dwell time*. In essence, it can be seen as a form of finite-time stability criterion which ensures that the system has sufficient time to achieve asymptotic or exponential convergence to a desired equilibrium in between consecutive switching phases. One example is in [32], where it is shown that for both linear and nonlinear systems , the dwell time should be "slow-on-average" and that input-to-state induced norms are bounded uniformly between slow-on-the-average switches. The usage of dwell-time criterion in attaining exponential stability has been shown to be effective for robotic applications, in particular walking locomotion and flapping flight [33] as well as autonomous vehicle steering [34].

There have been numerous extensive work devoted to the stochastic control of systems with AWGN perturbations, including the classical Linear Quadratic Gaussian (LQG) model [35], [36], the path integral approach [37], convex optimization-based approaches [38], [39], as well as a number of reinforcement learning-based approaches [40], [41]. There have also been some work devoted to the stochastic control of systems perturbed by non-Gaussian noise. One common branch of approaches is based on the Hamilton-Jacobi-Bellman (HJB) equation, which has been proven useful in deterministic settings with deterministic disturbances [42], [43]. See [44] for stochastic control of systems under Lévy noise perturbations with financial applications. Yet a well-known limitation of HJB approaches in many scenarios is the large amount of computation time and the inability to analytically solve for the value function from the partial differential equation (PDE). To this end, numerical algorithms have been considered, for example, fuzzy interpolation [45] and several machine-learning-based methods [46], [47]. However, this introduces a new class of well-known limitations such as massive amounts of required offline training, which is not suitable for settings where we have insufficient training data or training time. To this end, we are primarily interested in the development of a model-based analytical approach, which can be used as a baseline controller or observer. This consequently necessitates a theoretical framework in which stability analysis of various classes of stochastic systems can be carried out, which is the focus of this very paper.

B. Paper Organization

Section II discusses the relevant background needed to understand the Stochastic Contraction Theorems we present. We begin with a brief review of the fundamental deterministic contraction theorem, then move onto an extensive description of the Poisson random measure and the Poisson integral; the purpose of this subsection is to clarify exactly the definitions that we will be using, as these definitions tend to vary by literature. We conclude this section by a uniqueness and existence result for stochastic differential equations (SDEs) perturbed by shot noise, which we then use for our analysis in Section III. In Section III, we present our main results. First, the Stochastic Contraction Theorem for shot noise SDEs; in particular, we prove exponential and mean-square convergence

towards a bounded error ball within a specified interval of time via incremental stability analysis. Second, the Stochastic Contraction Theorem for Lévy noise SDEs, which is extended from the shot noise SDE case via the Lévy-Khintchine Decomposition Theorem. In Section IV, we numerically demonstrate that trajectories can still remain bounded under Lévy noise and illustrate that a simple linear quadratic regulator (LQR) scheme can still reasonably control linear systems affected by shot noise. Finally, we conclude the paper and provide a few directions for future work in Section V.

II. BACKGROUND AND PRELIMINARIES

A. Deterministic Contraction Theorem

Definition 2. The deterministic, noiseless system $\dot{\mathbf{x}} = f(t, \mathbf{x})$ is said to be *incrementally (globally exponentially) stable* if there exist constants $\kappa, \beta > 0$ such that

$$\|\mathbf{x}(t) - \mathbf{y}(t)\| \leq \kappa \|\mathbf{x}_0 - \mathbf{y}_0\| e^{-\beta t} \tag{1}$$

for all $t \geq 0$ and for all solution trajectories $\mathbf{x}(t)$ and $\mathbf{y}(t)$ of the system with respective initial conditions $\mathbf{x}_0$ and $\mathbf{y}_0$. Assume $\mathbf{x}_0 \neq \mathbf{y}_0$, otherwise the two trajectories are exactly the same for all t and (1) is trivially satisfied with equality.

For $\mathbf{x} \in \mathbb{R}^n$, we denote $\delta \mathbf{x} := (\delta x_1, \cdots, \delta x_n)$ to be the *infinitesimal displacement length* over a fixed interval of time Δt. Its rate of change can be approximated by the dynamics $\dot{\delta \mathbf{x}} = (\partial f / \partial \mathbf{x}) \delta \mathbf{x}$.

A main tool that will be used throughout is the *Comparison Lemma* [11], which can be simply stated as follows. Suppose we have an initial-value problem of the form $\{\dot{u} = g(u, t), u(0) = u_0\}$ and corresponding solution $u(t)$. Then if we were to consider an analogous problem $\{\dot{v} \leq g(v, t), v(0) \leq u_0\}$, the solution $v(t)$ satisfies $v(t) \leq u(t)$ for all $t \geq 0$.

Theorem 1 (Basic Contraction Theorem)**.** If there exists a uniformly positive definite matrix $S(t, \mathbf{x}) := \Theta(t, \mathbf{x})^T \Theta(t, \mathbf{x})$ for Θ being some smooth invertible square matrix that represents the differential coordinate transformation of $\delta \mathbf{z} := \Theta \delta \mathbf{x}$, and if the following condition is satisfied:

$$\left(\frac{\partial f}{\partial \mathbf{x}}\right)^T S(t, \mathbf{x}) + S(t, \mathbf{x}) \left(\frac{\partial f}{\partial \mathbf{x}}\right) + \dot{S}(t, \mathbf{x}) \leq -2\alpha S \tag{2}$$

for some $\alpha > 0$ and $\forall t \geq 0$, then all system trajectories converge globally exponentially to a single trajectory with a convergence rate equal to α. Such a system is called *contracting* with respect to the Riemannian metric associated with $S(t, \mathbf{x})$.

A more comprehensive study of established deterministic contraction results can be found in [13], [22], [48]–[50].

B. The Poisson Integral

We will establish a few definitions and properties before delving into existence and uniqueness theorems for SDEs perturbed by shot noise (Section II-C), and the contraction theorems that arise (Section III).

Definition 3. Let E be a Borel-measurable set such that $0 \in \overline{E}$, where $\overline{E}$ denotes the closure of E. We will refer to E as the *jump space*. Consider a random measure $\overline{N}([0,T] \times E)$ on set E until time $T > 0$ with *intensity measure* Leb $\times \nu$, where Leb denotes the standard Lebesgue measure. We will denote the corresponding *intensity (parameter)* as $\lambda := \nu(E)$. It is important to notice the difference between Leb (the measure in time) and ν (the probability measure describing the distribution of the jump "size"). Then $\overline{N}$ is called a *Poisson random measure* if the following are satisfied:

1) if $E_1, \cdots, E_n$ are pairwise disjoint elements of E, then $\overline{N}([0,T] \times E_1), \cdots, \overline{N}([0,T] \times E_n)$ are independent.
2) for each $E_i \subseteq E$, $\overline{N}([0,T] \times E_i)$ is distributed as a Poisson random variable with intensity parameter $\lambda_i := \nu(E_i)$.

Every Poisson random measure $\overline{N}([0,T] \times E)$ has an associated *standard Poisson process* (counting process) $N(t)$, which counts the number of jumps in E that have occurred in the time interval $[0,t]$ for $t \leq T$. We use the subscript 1 because it is a special case of the Poisson random measure with jump space $E = \{1\}$. Both the random measure and its counting process have the same intensity λ. Denote the time of the i-th arrival with random variable T_i. The mean is given by $\mathbb{E}[N(t)] = \lambda t$ and the variance is $\text{Var}(N(t)) = \lambda t$. The *compensated Poisson process* $N(t) - \lambda t$ is a martingale with respect to $\mathcal{F}_t := \sigma(N(s) : s \in [0,t))$. This results from the fact that $\mathbb{E}[N(t) - \lambda t] = 0$, i.e., has centered increments.

Example 1. A Poisson random measure with heights taking more than one value in E can also be expressed as a *compound Poisson process* $Y(t)$; it is essentially a standard Poisson process where the size of the ith jump is a random variable ξ_i not necessarily equal to 1. We write it as the cumulative sum $Y(t) = \sum_{i=1}^{N(t)} \xi_i$. Again, we denote arrival times by T_i and its intensity by λ.

To make the illustration more concrete, consider a scalar process which models the arrivals of vehicles in a particular parking space and the variables ξ_k which represent the type of the vehicle: $\xi_k = i$ if it is type i, where $i \in E := \{1, \cdots, K\}$. Then we can partition the space into K disjoint sets $E_i = \{i\}, i = 1, \cdots, K$. For each i, we associate a counting process $N_{E_i}(t)$ with intensity λ_i. It counts the number of arrivals of type i there are (i.e. number of jumps in set E_i) in the time interval $[0,t]$. Now suppose an entering vehicle is type i with probability p_i, where $\sum_{i=1}^{K} p_i = 1$. Then the intensity of each split process is related to the intensity λ of the original arrival process via $\lambda_i = p_i \lambda$.

Let $g : [0,T] \times E \rightarrow \mathbb{R}$ be a Borel-measurable function and $\overline{N}$ be a Poisson random measure on $[0,T] \times E$ with intensity measure Leb $\times \nu$. We define the *Poisson integral* of g as follows:

$$I_g := \int_{[0,T] \times E} g(t,y) \overline{N}(dt, dy) \qquad (3)$$

In contrast to the previous Example 1, we use lowercase y to denote the jump sizes instead of ξ.

A well-known interpretation of this integral (see [51] for a formula with time-invariant g) is the following summation

form:

$$I_g = \sum_{0 < t \leq T} g(t, \Delta\overline{N}(t)) \mathbb{1}_E \{\Delta\overline{N}(t)\} \qquad (4)$$

where $\Delta\overline{N}(t) = \overline{N}(t) - \overline{N}(t-)$, and the indicator $\mathbb{1}_E \{\Delta\overline{N}(t)\}$ is equal to 1 if a jump occurred at time t and 0 otherwise. The subscript E in the indicator is to show that a jump is counted only if it satisfies a membership property of E. For example, if $E = (-\infty, -1] \cup [1, \infty)$, then any $\Delta\overline{N}(t)$ such that $|\Delta\overline{N}(t)| < 1$ is not considered a jump.

Overall, the definition of the integral has quite an intuitive interpretation: if we think of the noise process as a sequence of impulses where the ith impulse arrives at time T_i, then integrating a function g with respect to it over an interval of time $[0,T]$ would only pick out the values of g at $T_i \in [0,T]$.

Example 2. We continue from Example 1. Let $g(t,y) = g(y)$ be the time-independent cost of parking in the parking site, which is a function of the vehicle type in the following way: $g(y) = 10y$. Partition the space according to Example 1. Then evaluating I_g over a fixed time period $[0,T]$ is equivalent to determining the total cost of all vehicles which have entered and parked in this time. Suppose that for a specific instance, the arrival process of vehicle types (i.e. the jump heights of the process $Y(t)$) during this time interval is $\{2, 2, 3, 2, 3, 1, 2\}$ so that there have been $N(T) = 7$ total arrivals. We get

$$I_g = \sum_{i=1}^{7} \sum_{k=1}^{3} g(T_i, \Delta Y(T_i)) \mathbb{1}_{E_k} \{\Delta Y(T_i)\}$$
$$= g(2) + g(2) + g(3) + g(2) + g(3) + g(1) + g(2) = 150$$

Another needed result for our analysis is the following: the simplest version of a well-known formula called *Campbell's Formula*. The version in [52] presents the formula for time-invariant functions g; here, we extend the result more generally to functions g that also depend on time.

Lemma 1 (Campbell's Formula). Let $g : [0, \infty) \times E \rightarrow \mathbb{R}$ be a Borel-measurable function and $\overline{N}([0,T] \times E)$ denote the Poisson random measure with intensity $\lambda := \nu(E)$ over the jump space E. Then

$$\mathbb{E}[I_g] = \int_{[0,T] \times E} \mathbb{E}[g(t,y)] dt \nu(dy) \qquad (5)$$

In order for the integral (3) to be well-defined, we need:

$$\int_{[0,T] \times E} |g(t,y)| dt \nu(dy) < \infty \text{ a.s.} \qquad (6)$$

where we write "a.s." for "almost-surely". Campbell's formula and other uses of this integral are stated while assuming this property is satisfied.

Proof: The compensated integral is given by replacing $\overline{N}(dt, dy)$ with a compensated Poisson process:

$$\int_{[0,T] \times E} g(t,y)(\overline{N}(dt, dy) - dt\nu(dy))$$

Because the compensated Poisson is a martingale, this integral has mean zero. Splitting this apart yields our desired result:

$$\mathbb{E}\left[\int_{[0,T] \times E} g(t,y)(\overline{N}(dt, dy) - dt\nu(dy))\right] = 0$$

which implies

$$\mathbb{E}\left[\int_{[0,T]\times E} g(t,y)\overline{N}(dt,dy)\right] = \mathbb{E}\left[\int_{[0,T]\times E} g(t,y)dt\nu(dy)\right]$$

$$= \int_{[0,T]\times E} \mathbb{E}[g(t,y)]dt\nu(dy)$$

where the last equality results from Fubini's theorem. $\qquad\square$

Remark 1. A special case of the formula in Lemma 1 is when the function g is deterministic and independent of time.

$$\mathbb{E}\left[\int_{[0,T]\times E} g(y)\overline{N}(dt,dy)\right] = T\int_E g(y)\nu(dy) \qquad (7)$$

Essentially, because the function is independent of time, the integral over t becomes just a constant multiplication.

Finally, we discuss one additional property of the Poisson integral [51], [53], the proof of which we defer to Appendix I. We will be using this in the proof of the existence and uniqueness theorem presented in the following subsection.

Lemma 2. If the following properties hold:

$$\int_{[0,T]\times E} |g(t,y)|dt\nu(dy) < \infty \qquad (8a)$$

$$\int_{[0,T]\times E} g^2(t,y)dt\nu(dy) < \infty \qquad (8b)$$

then

$$\mathbb{E}[I_g^2] = \int_{[0,T]\times E} \mathbb{E}[g(t,y)^2]dt\nu(dy)$$

$$+ \left(\int_{[0,T]\times E} \mathbb{E}[g(t,y)]dt\nu(dy)\right)^2 \qquad (9)$$

Both white noise and Poisson jump noise separately are Lévy processes (again, we emphasize that we will be using the terms "shot", "jump", and "Poisson" noise interchangeably). As such, a linear combination of white and Poisson noise, such as in Fig. 1, is also a Lévy process. More generally, the Lévy-Khintchine Decomposition Formula, which is presented in Theorem 5, states that any Lévy noise process can be expressed as a combination of white and shot noise. Because maintaining incremental stability against white noise has been studied in the past literature, we will focus primarily on a treatment for Poisson noise throughout the remainder of this paper, then discuss the Lévy noise combination at the end.

C. Existence and Uniqueness of Solutions

We consider systems that can be expressed as SDEs of the following form:

$$d\mathbf{x}(t) = f(t,\mathbf{x})dt + \sigma(t,\mathbf{x})dW(t) + \xi(t,\mathbf{x})dN(t) \qquad (10)$$

where $f : \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^n$ is the unperturbed part of the system, $\sigma : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^{n\times d}$ is the variation of the white noise, $W : \mathbb{R}^+ \to \mathbb{R}^d$ is a d-dimensional Brownian motion process, $\xi : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^{n\times\ell}$ is the size of the jumps that occur, and $N(t)$ is the ℓ-dimensional standard

Poisson random measure with intensity $\lambda := \nu(\{1\}^\ell)$. Hence, $N(t) := \overline{N}([0,t]\times\{1\}^\ell)$.

Note that when $\xi(t,\mathbf{x}) \equiv 0$, we recover the standard white or Gaussian noise SDE:

$$d\mathbf{x}(t) = f(t,\mathbf{x})dt + \sigma(t,\mathbf{x})dW(t) \qquad (11)$$

The conditions for existence and uniqueness of solutions are the standard ones (see [54]): the functions must be Lipschitz with respect to the time argument and have bounded growth with respect to the state argument.

A similar result can be derived in the case of shot noise perturbations. For the moment, we consider SDEs with $\sigma(t,\mathbf{x}) \equiv 0$.

$$d\mathbf{x}(t) = f(t,\mathbf{x})dt + \xi(t,\mathbf{x})dN(t) \qquad (12)$$

First, we present the well-known Gronwall inequality, a standard result of which can be found in any classical control-theoretic textbook (e.g. [11], [12]).

Lemma 3 (Gronwall inequality)**.** Let $I = [t_1,t_2] \subset \mathbb{R}$ and $\phi,\psi,\rho : I \to \mathbb{R}^+$ be continuous, nonnegative functions.

If the following inequality holds true:

$$\phi(t) \le \psi(t) + \int_{t_1}^t \rho(s)\phi(s)ds \ \ \forall\, t \in [t_1,t_2] \qquad (13)$$

Then it follows that

$$\phi(t) \le \psi(t) + \int_{t_1}^t \psi(s)\rho(s)e^{\int_s^t \rho(\tau)d\tau}ds \qquad (14)$$

Theorem 2 (Existence and Uniqueness for SDE (12))**.** For fixed $T > 0$, let $f : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ and $\xi : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times\ell}$ be measurable functions satisfying the following conditions

1) Lipschitz: $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, t \in [0,T]$,

$$\|f(t,\mathbf{x}) - f(t,\mathbf{y})\| + \|\xi(t,\mathbf{x}) - \xi(t,\mathbf{y})\|_F$$
$$\le K\|\mathbf{x} - \mathbf{y}\| \qquad (15)$$

2) Bounded growth: $\forall \mathbf{x} \in \mathbb{R}^n, t \in [0,T]$

$$\|f(t,\mathbf{x})\|^2 + \|\xi(t,\mathbf{x})\|_F^2 \le C(1 + \|\mathbf{x}\|^2) \qquad (16)$$

for positive constants C and K where the norm on ξ is the Frobenius norm and the norms on the vector-valued functions are any vector norm. Further, let $\mathbf{x}_0 \in \mathbb{R}^n$ have $\mathbb{E}[\|\mathbf{x}_0\|] < \infty$ and be independent of the noise processes. Then the SDE (12) with initial condition $\mathbf{x}(0) = \mathbf{x}_0$ has a unique solution $\mathbf{x}(t)$ adapted to the filtration $\mathcal{F}_t$ generated by $\mathbf{x}_0$ and $N(s)$, where $s \le t$ and

$$\mathbb{E}\left[\int_0^T \|\mathbf{x}(t)\|^2\, dt\right] < \infty$$

Remark 2. Following the construction of (3), we take the definition of the Poisson integral used in the Lipschitz and bounded growth conditions to be

$$\int_0^t \xi(s,\mathbf{x})dN(s) = \sum_{i=1}^{N_1(t)} \xi(T_i,\mathbf{x}(T_i)) \qquad (17)$$

since the size of every jump of the standard Poisson process is 1. We have essentially isolated the size of each jump as a multiplicative factor $\xi(t, \mathbf{x})$. From here on out in our analysis, we will be using this multiplicative form instead of the general Poisson random measure $N(dt, dy)$. Thus, we can express the Poisson integral as:

$$
\begin{aligned}
\mathbb{E}\left[\int_0^t \xi(s, \mathbf{x}) dN(s)\right] &= \mathbb{E}\left[\int_{[0,t]\times\{1\}^\ell} \xi(s, \mathbf{x}) N(ds, dy)\right] \\
&= \int_{[0,t]\times\{1\}^\ell} \mathbb{E}\left[\xi(s, \mathbf{x})\right] ds\nu(dy) \\
&= \lambda \int_0^t \mathbb{E}\left[\xi(s, \mathbf{x})\right] ds \qquad (18)
\end{aligned}
$$

Remark 3. Theorem 2 was presented in [44] for Lévy noise, but without a proof. We present a complete proof specifically for the shot noise case here; existence and uniqueness criteria for Lévy noise simply involves summing the Lipschitz and bounded growth terms together. The Lévy noise existence and uniqueness criterion was also presented in [55] with proof for when $\mathbb{E}[\mathbf{x}_0] < \infty$ and when $\mathbb{E}[\mathbf{x}_0] = \infty$, but our proof was based on extending the white noise conditions from [54], and developed independently of this work. Again, we only focus on the pure shot noise SDE. Additionally, with our definition of the Poisson integral in Section II-B, we can make simplifications using Lemma 2. Lastly, instead of considering the Poisson integral with respect to the general Poisson random measure, we are specializing to the case of the standard Poisson process, with further simplifies the proof in comparison to that presented in [55]. There have also been previous work done on describing such conditions for solutions to SDEs of the form (12) while imposing different, non-Lipschitz conditions on f and ξ. For instance, [56] relaxes the Lipschitz conditions by instead assuming that f and ξ are bounded above by any concave function of the normed difference in trajectories $\|\mathbf{x} - \mathbf{y}\|$. Alternatively, [57] presents a result for conditions where f is upper-bounded in norm by a constant and the bound on ξ depends on the size of the jump (which is not easily applicable to our case because we are only considering standard Poisson process noise, i.e. the jump size is always one). We choose to work with simple Lipschitz conditions because it is easier to relate to the well-known white noise version.

Proof: First we construct an approximate sequence using Picard iterations, recursively defined as

$$
\begin{aligned}
\mathbf{z}^{(n)}(t) = \mathbf{z}_0^{(n)} &+ \int_0^t f(s, \mathbf{z}^{(n-1)}(s)) ds \\
&+ \int_0^t \xi(s, \mathbf{z}^{(n-1)}(s)) dN(s) \qquad (19)
\end{aligned}
$$

where $n \in \mathbb{N}$.

Taking the difference between two trajectories $\mathbf{z}^{(n)}(t)$ and $\mathbf{z}^{(m)}(t)$ results in

$$
\begin{aligned}
\mathbf{z}^{(n,m)}(t) = \mathbf{z}_0^{(n,m)} &+ \int_0^t f^{(n-1,m-1)}(s) ds \\
&+ \int_0^t \xi^{(n-1,m-1)}(s) dN(s) \qquad (20)
\end{aligned}
$$

with $n, m \in \mathbb{N}$ and the notation

$$
\begin{aligned}
\mathbf{z}^{(n,m)}(t) &:= \mathbf{z}^{(n)}(t) - \mathbf{z}^{(m)}(t), \quad \mathbf{z}^{(n,m)}(0) := \mathbf{z}^{(n,m)}(0) \\
f^{(n,m)}(t) &:= f(t, \mathbf{z}^{(n)}(t)) - f(t, \mathbf{z}^{(m)}(t)) \\
\xi^{(n,m)}(t) &:= \xi(t, \mathbf{z}^{(n)}(t)) - \xi(t, \mathbf{z}^{(m)}(t))
\end{aligned}
$$

Taking the mean-squared difference, and applying the triangle and Cauchy-Schwarz inequalities leads to

$$
\begin{aligned}
\mathbb{E}\left[\left\|\mathbf{z}^{(n,m)}(t)\right\|^2\right] &\le \mathbb{E}\left[\left(\left\|\mathbf{z}_0^{(n,m)}\right\| + \int_0^t \left\|f^{(n-1,m-1)}(s)\right\| ds \right.\right. \\
&\qquad \left.\left. + \int_0^t \left\|\xi^{(n-1,m-1)}(s)\right\|_F dN(s)\right)^2\right] \\
&\le 3\mathbb{E}\left[\left\|\mathbf{z}_0^{(n,m)}\right\|^2\right] + 3\mathbb{E}\left[\int_0^t ds \int_0^t \left\|f^{(n-1,m-1)}(s)\right\|^2 ds\right] \\
&\quad + 3\mathbb{E}\left[\left(\int_0^t \left\|\xi^{(n-1,m-1)}(s)\right\|_F dN(s)\right)^2\right] \qquad (21)
\end{aligned}
$$

Note that the Lipschitz bound (15) can be squared on both sides:

$$
\begin{aligned}
\|f(t, \mathbf{x}) - f(t, \mathbf{y})\|^2 + \|\xi(s, \mathbf{x}) - \xi(s, \mathbf{y})\|_F^2 + \\
2\|f(t, \mathbf{x}) - f(t, \mathbf{y})\| \|\xi(s, \mathbf{x}) - \xi(s, \mathbf{y})\|_F \le K^2 \|\mathbf{x} - \mathbf{y}\|^2 \qquad (22)
\end{aligned}
$$

Because norms are nonnegative and the integral of nonnegative functions (whether it is standard ds or Poisson $dN(s)$) is also nonnegative, the bound also holds for each individual term in the left-hand side sum. Using (22) on the second expectation term yields:

$$
\begin{aligned}
\mathbb{E}\left[\int_0^t ds \int_0^t \left\|f^{(n-1,m-1)}(s)\right\|^2 ds\right] \\
\le K^2 t \int_0^t \mathbb{E}\left[\left\|\mathbf{z}^{(n-1,m-1)}(s)\right\|^2\right] ds \qquad (23)
\end{aligned}
$$

and for the final term, we can apply Lemma 2.

$$
\begin{aligned}
\mathbb{E}&\left[\left(\int_0^t \left\|\xi^{(n-1,m-1)}(s)\right\|_F dN(s)\right)^2\right] \\
&= \lambda \int_0^t \mathbb{E}\left[\left\|\xi^{(n-1,m-1)}(s)\right\|_F^2\right] ds \\
&\quad + \mathbb{E}\left[\left(\lambda \int_0^t \left\|\xi^{(n-1,m-1)}(s)\right\|_F ds\right)^2\right] \qquad (24)
\end{aligned}
$$

using the fact that $\lambda := \int_{\{1\}^\ell} \nu(dy)$, as in Remark 2. By Cauchy-Schwarz inequality and the squared Lipschitz bound (22):

$$
\begin{aligned}
(24) &\le K^2 \lambda \int_0^t \mathbb{E}\left[\left\|\mathbf{z}^{(n-1,m-1)}(s)\right\|^2\right] ds \\
&\quad + \lambda^2 \mathbb{E}\left[\int_0^t ds \int_0^t \left\|\xi^{(n-1,m-1)}(s)\right\|_F^2 ds\right] \\
&\le K^2 \lambda \int_0^t \mathbb{E}\left[\left\|\mathbf{z}^{(n-1,m-1)}(s)\right\|^2\right] ds \\
&\quad + K^2 \lambda^2 t \int_0^t \mathbb{E}\left[\left\|\mathbf{z}^{(n-1,m-1)}(s)\right\|^2\right] ds \qquad (25)
\end{aligned}
$$

Finally, note that $\mathbf{z}_0^{(n,m)} = 0$ because both trajectories $\mathbf{z}^{(n)}$ and $\mathbf{z}^{(m)}$ begin with the same initial conditions. In combination, we get:

$$\mathbb{E}\left[\left\|\mathbf{z}^{(n)}(t) - \mathbf{z}^{(m)}(t)\right\|^2\right] \leq 3K^2(t + \lambda + \lambda^2 t)$$
$$\times \int_0^t \mathbb{E}\left[\left\|\mathbf{z}^{(n-1)}(s) - \mathbf{z}^{(m-1)}(s)\right\|^2\right] ds \quad (26)$$

Choose $n = k+1, m = k$ for $k > 0$. By induction, we get:

$$\mathbb{E}\left[\left\|\mathbf{z}^{(k+1)}(t) - \mathbf{z}^{(k)}(t)\right\|^2\right] \leq \frac{c^k t^{k+1}}{(k+1)!} \quad \forall k \geq 0, t \in [0, T] \quad (27)$$

where $c := 3K^2(T + \lambda + \lambda^2 T)$. From there, it is straightforward to show that $\left\{\mathbf{z}^{(k)}(t)\right\}$ is a Cauchy sequence which converges to a limit since $\mathbf{z} \in \mathbb{R}^n$.

To show that the solution is unique, consider two solution trajectories $\mathbf{x}(t, \omega)$ and $\mathbf{y}(t, \omega)$ of (12) with respective initial conditions $\mathbf{x}_0$ and $\mathbf{y}_0$ where ω is a specific sample path of the noise process N. We can apply the same calculations as before on the mean-squared error difference between $\mathbf{x}$ and $\mathbf{y}$ to get

$$\mathbb{E}\left[\|\mathbf{x}(t) - \mathbf{y}(t)\|^2\right] \leq 3\mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{y}_0\|^2\right]$$
$$+ c\mathbb{E}\left[\int_0^t \|\mathbf{x}(s) - \mathbf{y}(s)\|^2 ds\right] \quad (28)$$

By Gronwall's inequality Lemma 3, (28) becomes

$$\mathbb{E}\left[\|\mathbf{x}(t) - \mathbf{y}(t)\|^2\right] \leq 3\mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{y}_0\|^2\right] e^{ct} \quad (29)$$

Now we set the two initial conditions $\mathbf{x}_0$ and $\mathbf{y}_0$ equal to each other. This implies that $c_1 = 0$ and so $h(t) = 0$ for all $t \geq 0$. Thus,

$$\mathbb{P}\left(\|\mathbf{x} - \mathbf{y}\| = 0\right) = 1 \quad \text{for all } t \geq 0$$

This holds for all sample paths of N. Thus, the solution is indeed unique for all $t \in [0, T]$. The proof is complete. $\square$

Remark 4. We have shown convergence of two trajectories in the mean-squared sense: in expectation, the trajectories will converge toward each other. It is weaker than the almost-sure sense of convergence, meaning we do not guarantee trajectory convergence for every noise process sample paths ω. For a more comprehensive treatment of this topic, see Chapter 2 of [58], [59], and [60], which additionally develops a CLT-like theorem for semimartingales (informally defined, SDEs which involve both a continuous martingale part such as white noise, and a purely discontinuous part such as shot noise).

III. STOCHASTIC CONTRACTION THEOREMS

For deterministic systems, we have established contraction as a concept of convergence between different solution trajectories of the system starting from different initial conditions. In the stochastic setting, the difference between trajectories will also arise from using different noise processes. Specifically,

we define two trajectories $\mathbf{x}(t)$ and $\mathbf{y}(t)$ as solutions to (11) driven by different Lévy noise processes:

$$d\mathbf{x} = f(t, \mathbf{x})dt + \sigma_1(t, \mathbf{x})dW_1(t) + \xi(t, \mathbf{x})dN(t) \quad (30a)$$
$$d\mathbf{y} = f(t, \mathbf{y})dt + \sigma_2(t, \mathbf{y})dW_2(t) \quad (30b)$$

where the pairs of functions (f, σ_1) and (f, σ_2) each satisfy the usual conditions for existence and uniqueness of solutions to white noise SDEs [54], f and ξ satisfy the conditions in Theorem 2, and we assume there exists constants $\eta_1, \eta_2, \eta > 0$ such that

$$\|\sigma_1(t, \mathbf{x})\|_F \leq \eta_1, \quad \|\sigma_2(t, \mathbf{y})\|_F \leq \eta_2, \quad \|\xi(t, \mathbf{x})\|_F \leq \eta \quad (31)$$

In Section III-A, we will first review the contraction criterion for pure white noise systems from [23], then subsequently develop a criterion for pure shot noise systems and explain the asymmetric expression with respect to the shot noise terms. Then in Section III-B, the criteria are combined via the Lévy-Khintchine formula to demonstrate contraction between trajectories with general Lévy noise perturbations.

A. Shot Noise Case

Assumption 1 (Bounded Metric). We will assume that the metric $S(t, \mathbf{x})$ is bounded in both arguments $\mathbf{x}$ and t from above and below, and that its first and second derivatives with respect to the $\mathbf{x}$ argument are also bounded from above. We thus define the following constants

$$\underline{s} = \inf_{t, \mathbf{x}} \lambda_{\min}(S(t, \mathbf{x})), \quad \overline{s} = \sup_{t, \mathbf{x}} \lambda_{\max}(S(t, \mathbf{x})) \quad (32)$$
$$s' = \sup_{t, \mathbf{x}, i, j} \|(\partial_x S(t, \mathbf{x}))_{i,j}\|, \quad s'' = \sup_{t, \mathbf{x}, i, j} \|(\partial_x^2 S(t, \mathbf{x}))_{i,j}\|$$

We can represent the infinitesimal differential length $\delta\mathbf{z}$ as a path integral and reparameterize using measure $\mu \in [0, 1]$:

$$\mathbf{y}(t) - \mathbf{x}(t) = \int_\mathbf{x}^\mathbf{y} \delta\mathbf{z} = \int_0^1 \left(\frac{\partial\mathbf{z}}{\partial\mu}\right) d\mu \quad (33)$$

Note that by Cauchy-Schwarz and the triangle inequality for integrals, we can bound $\underline{s}\|\mathbf{y} - \mathbf{x}\|^2 \leq V(t, \mathbf{z}, \delta\mathbf{z})$, where V is our Lyapunov-like function

$$V(t, \mathbf{z}, \delta\mathbf{z}) = \int_0^1 \left(\frac{\partial\mathbf{z}}{\partial\mu}\right)^T S(t, \mathbf{z}(\mu, t)) \left(\frac{\partial\mathbf{z}}{\partial\mu}\right) d\mu \quad (34)$$

Note that there is a dependence of V on $\delta\mathbf{z}$ because we can alternatively (and informally) express $V(\mathbf{z}, \delta\mathbf{z}, t) = \int_\mathbf{x}^\mathbf{y} \delta\mathbf{z}^T S(t, \mathbf{z})\delta\mathbf{z}$.

Using the above, we construct a *virtual system* in terms of $\mathbf{z}(t) \in \mathbb{R}^n$ such that its particular solutions are $\mathbf{x}(t)$ and $\mathbf{y}(t)$. If this virtual system is contracting in $S(t, \mathbf{z})$, then $\mathbf{x}$ and $\mathbf{y}$ converge towards each other globally and exponentially fast.

For the pure white noise SDE, we compare two trajectories of (30) when $\xi \equiv 0$ and define the measure μ such that:

$$\mathbf{z}(\mu = 0, t) = \mathbf{x}(t), \mathbf{z}(\mu = 1, t) = \mathbf{y}(t) \quad (35)$$
$$\sigma_{\mu=0}(t, \mathbf{z}) = \sigma_1(t, \mathbf{x}), \sigma_{\mu=1}(t, \mathbf{z}) = \sigma_2(t, \mathbf{y})$$
$$W_{\mu=0}(t, \mathbf{z}) = W_1(t), W_{\mu=1}(t, \mathbf{z}) = W_2(t)$$

e.g., $\mathbf{z}(\mu, t) := \mu\mathbf{x}(t) + (1 - \mu)\mathbf{y}(t)$. Using this, we can rewrite (11) as the virtual system $d\mathbf{z}(\mu, t) = f(t, \mathbf{z}(\mu, t))dt +$

$\sigma_\mu(t, \mathbf{z}(\mu, t))dW_\mu(t)$ and the virtual dynamics as $d\delta\mathbf{z} = F\delta\mathbf{z}dt + \delta\sigma_\mu dW_\mu$ where $\delta\sigma_\mu = \left[\frac{\partial\sigma_{\mu,1}}{\partial\mathbf{z}}\delta\mathbf{z}, \cdots, \frac{\partial\sigma_{\mu,d}}{\partial\mathbf{z}}\delta\mathbf{z} \right]$ and $\sigma_\mu := [\sigma_{\mu,1}, \cdots, \sigma_{\mu,d}]$, where $\sigma_{\mu,i}$ is the ith column of σ_μ.

As in the proof for existence and uniqueness, stochastic contraction results for both white and shot noise SDEs will be examined in the mean-square sense.

Theorem 3 (Stochastic Contraction Theorem for White Noise [23])**.** For the SDE system described in (11), suppose the following two conditions are satisfied:

1) the unperturbed system $\dot{\mathbf{x}} = f(t, \mathbf{x})$ satisfies (2) with metric $S(t, \mathbf{x})$ and rate α.
2) there exists $\eta_1, \eta_2 > 0$ such that the first two inequalities of (31) are satisfied.

Further assume that the initial conditions adhere to some probability distribution $p(\mathbf{z}_0) = p(\mathbf{x}_0, \mathbf{y}_0)$. Then (11) is stochastically contracting if

$$\mathbb{E}_{\mathbf{z}_0}[V(t, \mathbf{z}, \delta\mathbf{z})] \leq V(0, \mathbf{z}_0, \delta\mathbf{z}_0)e^{-\beta_w t} + \frac{\kappa_w}{\beta_w}(1 - e^{-\beta_w t}) \tag{36}$$

where $V(t, \mathbf{z}, \delta\mathbf{z})$ is defined in (34) and

$$\beta_w = 2\alpha - \frac{1}{2\underline{s}}(\eta_1^2 + \eta_2^2)\left(\frac{s''}{2} + s' \right) \tag{37a}$$

$$\kappa_w = \frac{1}{2}(\overline{s} + s')(\eta_1^2 + \eta_2^2) \tag{37b}$$

We refer to [23] for proof. Because V can be arbitrarily defined, it is better to express the inequality to in terms of the given trajectories $\mathbf{x}$ and $\mathbf{y}$. Dividing across by $\underline{s}$ and lower-bounding the left-hand side of the inequality according to the construction of (34) yields:

$$\mathbb{E}_{(\mathbf{x}_0, \mathbf{y}_0)}[\|\mathbf{y} - \mathbf{x}\|^2] \leq \frac{1}{\underline{s}}\|\mathbf{y}_0 - \mathbf{x}_0\|^2 e^{-\beta_w t} + \frac{\kappa_w}{\underline{s}\beta_w}\left(1 - e^{-\beta_w t}\right) \tag{38}$$

For the pure shot noise SDE, we compare two trajectories of (30) when $\sigma_1, \sigma_2 \equiv 0$. In contrast to (35), we define the parameter $\mu \in [0, 1]$ such that

$$\mathbf{z}(\mu = 0, t) = \mathbf{x}(t), \mathbf{z}(\mu = 1, t) = \mathbf{y}(t) \tag{39}$$
$$\xi_{\mu=0}(t, \mathbf{z}) = \xi(t, \mathbf{x}), \xi_{\mu=1}(t, \mathbf{z}) = 0$$
$$N_{\mu=0}(t, \mathbf{z}) = N(t), N_{\mu=1}(t, \mathbf{z}) = 0$$

with the corresponding virtual system $d\mathbf{z}(\mu, t) = f(t, \mathbf{z}(\mu, t))dt + \xi_\mu(t, \mathbf{z}(\mu, t))dN_\mu(t)$ and virtual dynamics $d\delta\mathbf{z} = F\delta\mathbf{z}dt + \delta\xi_\mu dN_\mu$.

A fundamental difference for stochastic contraction with respect to shot noise arises in contrast to the white noise case because of the large, discontinuously-occurring nature of the noise process. Thus, we provide guarantees of stability within finite-time [9]: within a fixed interval of time, the difference between the two trajectories will remain bounded by a time-varying function. The following useful fact allows us to do so.

Remark 5. For any fixed $t > 0$, $\mathbb{P}(N(t) = \infty) = 0$. That is, it is with probability zero that the number of jumps that have occurred by time t is infinite. This allows us to further impose

the assumption that the sum of all the sizes of the jumps that have occurred until time t is finite.

We also introduce the Itô formula with jumps, which we will use in our proof of contraction for shot-noise SDEs.

Lemma 4. For functions $F \in \mathcal{C}^{(1,2)}$,

$$\begin{aligned}
F(t, \mathbf{x}(t)) = {} & F(0, \mathbf{x}_0) + \int_0^t \partial_t F(s, \mathbf{x}(s))ds \\
& + \sum_{i=1}^n \int_0^t \partial_{x_i} F(s, \mathbf{x}(s))d\mathbf{x}_i(s) \\
& + \frac{1}{2}\sum_{i,j=1}^n \int_0^t \partial_{x_i}\partial_{x_j}F(s, \mathbf{x}(s))d[\mathbf{x}_i, \mathbf{x}_j]^c(s) \\
& + \int_0^t \left(F(s, \mathbf{x}(s)) - F(s, \mathbf{x}(s-)) \right)dN(s) \tag{40}
\end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^n$.

A comprehensive treatment of this standard formula can be found in many papers (see e.g., [52]). The specific version of the formula stated is also used in [26], [61]. The term $d[\mathbf{x}_i, \mathbf{x}_j]^c(s)$ in (40) is the continuous part of the quadratic variation between two stochastic processes $\mathbf{x}_i$ and $\mathbf{x}_j$, which is defined in the following remark.

Remark 6 (Quadratic Variation)**.** Consider two generic scalar SDEs of the form (10):

$$dx_i = f(t, x_i)dt + \sigma(t, x_i)dW(t) + \xi(t, x_i)dN(t) \tag{41a}$$

for $i = 1, 2$. Then the *quadratic variation* term $d\langle x_1, x_2 \rangle(t)$ is computed to be $\sigma(t, x_1)\sigma(t, x_2)dt + \xi(t, x_1)\xi(t, x_2)dN(t)$ since $dW(t) \cdot dW(t) = dt$ and $dN(t) \cdot dN(t) = dN(t)$ while the dot products between all other terms vanish ($dt \cdot dt = dt \cdot dW(t) = dt \cdot dN(t) = dW(t) \cdot dN(t) = 0$). It is comprised of two parts: the continuous part $d[x_1, x_2]^c(t) = dt$ and the purely discontinuous part $d[x_1, x_2]^d(t) = dN(t)$. For further information about this notation, one may refer to [54], [61] and references therein.

Theorem 4 (Stochastic Contraction Theorem for Shot Noise)**.** For the SDE system described in (12), suppose the following two conditions are satisfied:

1) the unperturbed system satisfies (2).
2) there exists $\eta > 0$ such that the last inequality of (31) is satisfied.

Further assume that the initial conditions adhere to some probability distribution $p(\mathbf{z}_0) = p(\mathbf{x}_0, \mathbf{y}_0)$. Then (12) is stochastically contracting if

$$\mathbb{E}_{(\mathbf{x}_0, \mathbf{y}_0)}[\|\mathbf{y} - \mathbf{x}\|^2] \leq \frac{1}{\underline{s}}\|\mathbf{y}_0 - \mathbf{x}_0\|^2 e^{-\beta_s t} + \frac{\kappa_s(t)}{\underline{s}} \tag{42}$$

where

$$\beta_s := 2\alpha - \lambda\eta\left(\frac{s'}{\underline{s}} + \frac{\overline{s}}{\underline{s}} \right) \tag{43a}$$

$$\kappa_s(t) := \lambda\int_0^t \left(\overline{s}\eta + c(s) \right)e^{-\beta_s(s-t)}ds \tag{43b}$$

and λ is the intensity of the Poisson process $N(t)$ and $c(t)$ is the continuous and bounded function defined in (49).

Remark 7. By Remark 5, the number of jumps by time t is finite with probability 1, so we are taking a maximum over a finite number of values. Hence, the supremum in the bound of Theorem 4 is justified.

Proof: By applying Itô's formula (40) to the Lyapunov-like function (34), we obtain

$$V(t, \mathbf{z}, \delta\mathbf{z}) = V(0, \mathbf{z}_0, \delta\mathbf{z}_0) + \int_0^t \partial_t V(s, \mathbf{z}, \delta\mathbf{z})ds +$$

$$\int_0^t \sum_{i=1}^n \left[\partial_{z_i} V(s, \mathbf{z}, \delta\mathbf{z}) f_i(s, \mathbf{z}) + \partial_{\delta z_i} V(s, \mathbf{z}, \delta\mathbf{z}) \left(\frac{\partial f}{\partial \mathbf{z}} \delta\mathbf{z} \right)_i \right]$$

$$+ \int_0^t \sum_{i=1}^n \left[\partial_{z_i} V(s, \mathbf{z}, \delta\mathbf{z}) \xi_{\mu, i}(s, \mathbf{z}) \right.$$

$$\left. + \partial_{\delta z_i} V(s, \mathbf{z}, \delta\mathbf{z}) \delta\xi_{\mu, i} \right] dN_\mu(s)$$

$$+ \int_0^t (V(s, \mathbf{z}, \delta\mathbf{z}) - V(s, \mathbf{z}(s-), \delta\mathbf{z}(s-))) dN_\mu(s) \quad (44)$$

where the subscript of i in $\xi_{\mu, i}$ (and other similar notation) denotes the ith component of the respective vector. Note that these are dimension $1 \times d$, and the Poisson dN_μ is dimension $d \times 1$, so the overall product is a scalar, as expected.

The first three terms are derived directly from deterministic contraction of an unperturbed system $\dot{\mathbf{x}} = f(t, \mathbf{x})$. For the fourth and fifth terms, we will derive bounds with the expected value included so that we may directly apply Campbell's formula Lemma 1.

By Campbell's formula, the bound on the fourth sum term is written as

$$\mathbb{E}_{\mathbf{z}_0} \left[\int_0^t \sum_{i=1}^n (\partial_{z_i} V(s, \mathbf{z}, \delta\mathbf{z}) \xi_{\mu, i}(\mathbf{z}, s) \right.$$

$$\left. + \partial_{\delta z_i} V(s, \mathbf{z}, \delta\mathbf{z}) \delta\xi_{\mu, i}(\mathbf{z}, s)) dN_\mu(s) \right]$$

$$= \lambda \sum_{i=1}^n \left(\int_0^t \mathbb{E}_{\mathbf{z}_0} \left[\partial_{z_i} V(s, \mathbf{z}, \delta\mathbf{z}) \xi_{\mu, i}(\mathbf{z}, s) \right] ds \right.$$

$$\left. + \int_0^t \mathbb{E}_{\mathbf{z}_0} \left[\partial_{\delta z_i} V(s, \mathbf{z}, \delta\mathbf{z}) \delta\xi_{\mu, i}(\mathbf{z}, s) \right] ds \right) \quad (45)$$

where λ can be taken outside of the integral by Remark 2.

Applying submultiplicativity, we get:

$$(45) \leq \lambda \sum_{i=1}^n \left(\int_0^t \mathbb{E}_{\mathbf{z}_0} \left[\|\partial_{z_i} V(s, \mathbf{z}, \delta\mathbf{z})\| \|\xi(s, z)_i\| \right] ds \right.$$

$$\left. + \int_0^t \mathbb{E}_{\mathbf{z}_0} \left[\|\partial_{\delta z_i} V(s, \mathbf{z}, \delta\mathbf{z})\| \|\delta\xi(s, z)_i\| \right] ds \right) \quad (46)$$

Note that

$$\|\partial_{z_i} V(s, \mathbf{z}, \delta\mathbf{z})\| \leq \sup_{t, \mathbf{z}} \|\partial_{\mathbf{z}_i} S(t, \mathbf{z})\| \|\delta\mathbf{z}^T \delta\mathbf{z}\| \leq \frac{s'}{\underline{s}} V(s, \mathbf{z}, \delta\mathbf{z})$$

$$\|\partial_{\delta z_i} V(s, \mathbf{z}, \delta\mathbf{z})\| \leq 2 \sup_{t, \mathbf{z}} \|S(t, \mathbf{z})\| |\delta\mathbf{z}_i| = 2\overline{s}|\delta\mathbf{z}_i|$$

Hence, (46) is bounded above by:

$$(46) \leq \lambda \left(\frac{s'}{\underline{s}} \right) \int_0^t V(s, \mathbf{z}, \delta\mathbf{z}) \sum_{i=1}^n \|\xi(s, \mathbf{z})_i\| ds$$

$$+ 2\lambda\overline{s} \int_0^t \sum_{i=1}^n |\delta\mathbf{z}_i| \|\delta\xi(s, \mathbf{z})_i\| ds$$

$$\leq \lambda\eta \left(\frac{s'}{\underline{s}} \right) \int_0^t V(s, \mathbf{z}, \delta\mathbf{z}) ds$$

$$+ 2\lambda\overline{s}\eta \int_0^t \|\delta\mathbf{z}\| \text{ by Cauchy-Schwarz}$$

$$\leq \lambda\eta \left(\frac{s'}{\underline{s}} \right) \int_0^t V(s, \mathbf{z}, \delta\mathbf{z}) ds$$

$$+ \lambda\overline{s}\eta \left(t + \frac{1}{\underline{s}} \int_0^t V(s, \mathbf{z}, \delta\mathbf{z}) ds \right) \quad (47)$$

Again by Campbell's formula, the bound on the fifth term is written as:

$$\mathbb{E}_{\mathbf{z}_0} \left[\int_0^t (V(s, \mathbf{z}, \delta\mathbf{z}) - V(s-, \mathbf{z}, \delta\mathbf{z})) dN(s) \right]$$

$$= \lambda \int_0^t \mathbb{E}_{\mathbf{z}_0} \left[V(s, \mathbf{z}, \delta\mathbf{z}) - V(s-, \mathbf{z}, \delta\mathbf{z}) \right] ds \quad (48)$$

Assumption 2. There exists a continuous and bounded function c such that

$$\mathbb{E}_{\mathbf{z}_0} \left[V(t, \mathbf{z}(t), \delta\mathbf{z}(t)) - V(t-, \mathbf{z}(t-), \delta\mathbf{z}(t-)) \right] \leq c(t) \quad (49)$$

Using (49), (48) is bounded by:

$$(48) \leq \lambda \int_0^t c(s) ds =: \lambda K(t)$$

In combination, the expected value of (44) is bounded above as:

$$\mathbb{E}_{\mathbf{z}_0} \left[V(t, \mathbf{z}, \delta\mathbf{z}) \right] \leq V(0, \mathbf{z}_0, \delta\mathbf{z}_0) - 2\alpha \int_0^t \mathbb{E}_{\mathbf{z}_0} [V(s, \mathbf{z}, \delta\mathbf{z})] ds$$

$$+ \lambda\eta \left(\frac{s'}{\underline{s}} + \frac{\overline{s}}{\underline{s}} \right) \int_0^t \mathbb{E}_{\mathbf{z}_0} [V(s, \mathbf{z}, \delta\mathbf{z})] ds + \lambda\overline{s}\eta t + \lambda K(t) \quad (50)$$

where α is the contraction rate of the unperturbed system.

Defining β_s as in (43a), (50) becomes:

$$\mathbb{E}_{\mathbf{z}_0} \left[V(t, \mathbf{z}, \delta\mathbf{z}) \right] \leq V(0, \mathbf{z}_0, \delta\mathbf{z}_0) - \beta_s \int_0^t \mathbb{E}_{\mathbf{z}_0} \left[V(s, \mathbf{z}, \delta\mathbf{z}) \right] ds$$

$$+ \lambda\eta\overline{s}t + \lambda K(t) \quad (51)$$

We can rewrite integral inequality (51) as a differential inequality:

$$d\mathbb{E}_{\mathbf{z}_0} \left[V(t, \mathbf{z}, \delta\mathbf{z}) \right] \leq -\beta_s \mathbb{E}_{\mathbf{z}_0} \left[V(t, \mathbf{z}, \delta\mathbf{z}) \right] + (\lambda\overline{s}\eta + \lambda c(t))$$

Solving the differential equation and using the Comparison lemma to turn equality to an inequality, we obtain:

$$\mathbb{E}_{\mathbf{z}_0} \left[V(t, \mathbf{z}, \delta\mathbf{z}) \right] \leq V(0, \mathbf{z}_0, \delta\mathbf{z}_0) e^{-\beta_s t}$$

$$+ \lambda \int_0^t (\overline{s}\eta + c(s)) e^{-\beta_s(s-t)} ds$$

Defining $\kappa_s(t)$ as in (43b) and expressing $V(t, \mathbf{z}, \delta\mathbf{z})$ in terms of the mean-squared difference between $\mathbf{x}$ and $\mathbf{y}$ yields our desired bound (42).

$\square$

Remark 8. As the intensity λ of the Poisson noise process increases (i.e. the jumps arrive more frequently), then note that the steady-state error $\kappa_s(t)$ increases as well. One way to minimize this effect is to increase the deterministic contraction rate α accordingly, because within each time interval (T_j, T_{j+1}) there are no additional jumps (it can be thought of restarting the process with initial condition at the value of $\mathbf{z}$ immediately after the jump affects the system). Since there is no perturbation in the time interval (T_j, T_{j+1}), the trajectory $\mathbf{y}$ behaves as a deterministic system with initial condition $\mathbf{y}(T_j)$. In this interval of time, it becomes equivalent to comparing two deterministic trajectories against each other. Increasing α also increases the stochastic contraction rate β and the trajectories converge as quickly as possible towards each other in between jumps. This will decrease the supremum term in (43b).

B. Lévy Noise Case

The contraction results Theorem 3 and Theorem 4 further directs us into a treatment of stochastic systems that are perturbed by a broader class of Lévy noise processes. This is due to the following theorem, stated specifically for scalar systems for the sake of simplicity. For further reading, one is referred to [53], [62].

Definition 4. Let L be a Lévy process and $\varphi : \mathbb{R} \to \mathbb{C}$ be its *characteristic function*, defined as $\varphi(\theta) = \mathbb{E}\left[e^{i\theta L(t)}\right]$ for any time t. Then there exists a unique continuous function $\rho : \mathbb{R} \to \mathbb{C}$ such that $e^{\rho(\theta)} = \varphi(\theta)$. This function $\rho(\theta) = \log(\varphi(\theta))$ is referred to as the *characteristic exponent* of $L(t)$.

Theorem 5 (Lévy-Khintchine Decomposition Formula)**.** Let L be a Lévy process with characteristic exponent Ψ. Then there exist (unique) $a \in \mathbb{R}, \sigma \geq 0$, and a measure ν which satisfies $\int_{\mathbb{R}} x^2 \nu(dx) < \infty$ such that

$$\Psi(\theta) = ia\theta - \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}} \left(e^{i\theta x} - 1\right)\nu(dx) \\ - \int_{\mathbb{R}} i\theta x \mathbb{1}_{[-1,1]}(x)\nu(dx) \quad (52)$$

Conversely, given any triplet (a, σ, ν), there exists a Lévy process L with characteristic exponent given by (52).

Note that a, which is often called the *center* of L, captures the deterministic drift component, *Gaussian coefficient* σ captures the variance of the Brownian motion component, and the *Lévy measure* ν captures the size and intensity of the "large" jumps of L. In essence, Theorem 5 tells us that any Lévy process (general noise) can be split into three simpler Lévy processes: a time shift of a white noise process, a time shift of a Poisson process with "large" jumps, and an additional Poisson process of jumps small enough to be compensated by a deterministic drift to turn it into a martingale.

Motivated by Theorem 5, a contraction result can be easily obtained for the combined system (10). Assume the same metric assumptions as in Assumption 1. The two trajectories we will compare against each other are given in (30). As in the individual white and shot noise SDE cases, the infinitesimal differential length $\delta\mathbf{z}$ can be represented as a path integral and reparameterized using measure parameter $\mu \in [0, 1]$ such that:

$$\mathbf{z}(\mu = 0, t) = \mathbf{x}(t), \mathbf{z}(\mu = 1, t) = \mathbf{y}(t) \quad (53)$$
$$\sigma_{\mu=0}(t, \mathbf{z}) = \sigma_1(t, \mathbf{x}), \sigma_{\mu=1}(t, \mathbf{z}) = \sigma_2(t, \mathbf{y})$$
$$\xi_{\mu=0}(t, \mathbf{z}) = \xi(t, \mathbf{x}), \xi_{\mu=1}(t, \mathbf{z}) = 0$$
$$W_{\mu=0}(t, \mathbf{z}) = W_1(t), W_{\mu=1}(t, \mathbf{z}) = W_2(t)$$
$$N_{\mu=0}(t, \mathbf{z}) = N(t), N_{\mu=1}(t, \mathbf{z}) = 0$$

Thus, we can rewrite (10) as the virtual system $d\mathbf{z}(\mu, t) = f(t, \mathbf{z}(\mu, t))dt + \sigma_\mu(t, \mathbf{z}(\mu, t))dW_\mu(t) + \xi_\mu(t, \mathbf{z}(\mu, t))dN_\mu(t)$ and the virtual dynamics as follows $d\delta\mathbf{z} = F\delta\mathbf{z}dt + \delta\sigma_\mu dW_\mu + \delta\xi_\mu dN_\mu$.

Corollary 1 (Stochastic Contraction Theorem for General Lévy Noise)**.** For the SDE system described in (10), suppose the following two conditions are satisfied:
   1) the unperturbed system satisfies (2).
   2) there exists $\eta, \eta_1, \eta_2 > 0$ such that (30) holds.

Further assume that the initial conditions adhere to some probability distribution $p(\mathbf{z}_0) = p(\mathbf{x}_0, \mathbf{y}_0)$. Then (10) is stochastically contracting if (42) holds with β_s and $\kappa_s(t)$ replaced by

$$\beta_\ell := \beta_w + \beta_s - 2\alpha \quad (54a)$$
$$\kappa_\ell(t) := \kappa_w + \kappa_s(t) \quad (54b)$$

where β_w and κ_w are defined in Theorem 3, β_s and $\kappa_s(t)$ are defined in Theorem 4, and α is the contraction rate of the unperturbed system.

The additional 2α in (54a) arises to prevent double-counting of the contraction rate for the deterministic part. Aside from this term, note that the values β_ℓ and $\kappa_\ell(t)$ for the combined SDE are a direct summation of the corresponding values for (11) and (12) separately. This is due to the similar differential inequality form that $\mathbb{E}_{\mathbf{z}_0}[V(t, \mathbf{z}, \delta\mathbf{z})]$ has in terms of $V(0, \mathbf{z}_0, \delta\mathbf{z}_0)$ and $\int_0^t \mathbb{E}_{\mathbf{z}_0}[V(s, \mathbf{z}, \delta\mathbf{z})]ds$ prior to multiplying by the integrating factor and applying the Comparison lemma. Additionally, simply setting $\sigma_2 \equiv 0$ in (30) allows us to make a comparison between trajectories of a system perturbed by Lévy noise and the unperturbed system; this corresponds to setting $\eta_2 = 0$ in the bound in Corollary 1.

IV. NUMERICAL EXAMPLES

A. Dubins' Car

Consider the dynamics of the 5-dimensional Dubins' car model from [42]:

$$\dot{\xi} = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \\ \dot{v} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} v\cos(\theta) + w_1 + +n_1 \\ v\sin(\theta) + w_2 + n_2 \\ \omega + w_3 + n_3 \\ a \\ \alpha \end{bmatrix} \quad (55)$$

where $w(t) = (w_1, w_2, w_3)^T$ is a white noise vector, $n(t) = (n_1, n_2, n_3)^T$ is a Poisson noise vector, and the control inputs

are given by translational acceleration a and angular acceleration α.

White noise vectors are generated from a Gaussian distribution with mean $\mathbf{0}$ and covariance $3I$, while shot noise vectors are generated from a Poisson process with rate $\lambda = \frac{1}{10}$ and heights:

$$
n(t) = \begin{cases} \begin{bmatrix} \text{Unif}[400, 600] \\ \text{Unif}[200, 400] \\ \text{Unif}[-200, 200] \end{bmatrix} & \text{with prob } 0.5 \\ \begin{bmatrix} \text{Unif}[-600, -400] \\ \text{Unif}[-400, -200] \\ \text{Unif}[-200, 200] \end{bmatrix} & \text{with prob } 0.5 \end{cases}
$$

We simulate 50 different Monte-Carlo trajectories of this system under the influence of different noise processes generated from the distributions described. Each trajectory is simulated until the time of its first jump, which varies by trajectory, starting from the initial conditions $\xi_0 = \left(5, 0, \frac{\pi}{2}, 0, 0\right)^T$. For the first one second, we add in constant linear acceleration $a = 10$ and constant angular acceleration $\alpha = \frac{\pi}{4}$, then set both control inputs to zero afterwards. When the system is not perturbed by noise, this traces out a circular trajectory of approximate radius of 12.7602 and approximate center $C = (-7.7569, -0.0199)$.

Under various sample paths of Lévy noise processes described above, the system trajectory is able to remain reasonably bounded within the norm ball of the unperturbed trajectory, prescribed by Corollary 1 with $\sigma_2 \equiv 0$ and Lyapunov-like function $V(t, \mathbf{z}, \delta\mathbf{z}) = \delta\mathbf{z}^T \delta\mathbf{z}$ (i.e., identity contraction metric $S(t, \mathbf{x}) = I$). We illustrate this in Fig. 2, where we partition the largest bounded set into three regions. The fraction of trajectories that has its maximum deviation in each region for this particular experiment is given by $\left(\frac{7}{50}, \frac{36}{50}, \frac{13}{50}\right)$ lying in the smallest, medium, and largest bounded sets, respectively.

B. 1D Compound Poisson Control

Consider the scalar linear system

$$
\dot{x}(t) = ax(t) + u(t) + w(t) + n(t) \tag{56}
$$

with $a = 2$ so that it is unstable in open-loop. This system tries to track the nominal reference trajectory (in dashed blue) while being perturbed by a white noise process $w(t)$ and a shot noise process $n(t)$.

We design a baseline controller to track a sinusoid trajectory $x_r(t) = \sin(t)$:

$$
\begin{aligned} u_r(t) &= \dot{x}_r(t) - ax_r(t) \\ u(t) &= u_r(t) - k(x(t) - x_r(t)) \end{aligned} \tag{57}
$$

and gain k is chosen $a + 0.5$ so that $x(t)$ exponentially converges to the sinusoid with a rate of -0.5.

The white noise process is generated zero-mean with variance 3. The shot noise process is Poisson with interarrival times distributed Exponential with mean λ, and heights are distributed according to the following law

$$
\xi_k \sim \begin{cases} \text{Unif}[100, 400] & \text{with prob } 0.5 \\ \text{Unif}[-400, -100] & \text{with prob } 0.5 \end{cases}
$$

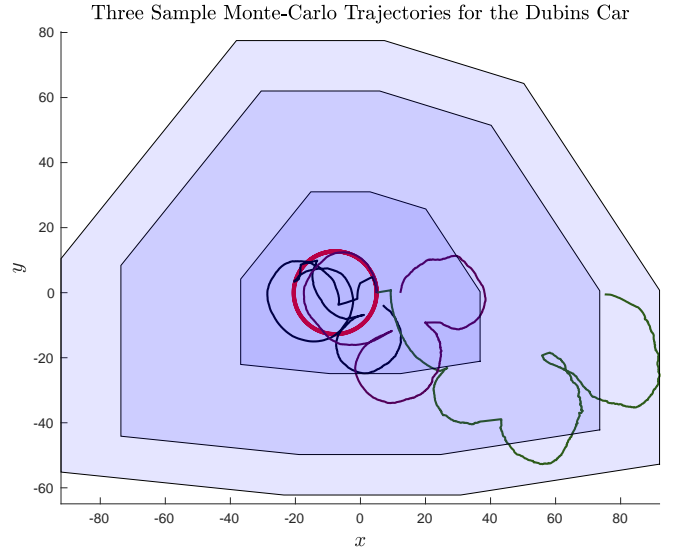Three Sample Monte-Carlo Trajectories for the Dubins Car

Fig. 2: Three sample open-loop white and shot noise perturbed trajectories corresponding to each of three disjoint partitions of the maximum norm ball (lightest shade of blue). The black trajectory lies fully within the smallest set, the dark red lies fully in the medium-sized set, and the green trajectory remains in the largest bounded set. The brighter red circle represents open-loop trajectory of the unperturbed system.

We modify the control law so that the gain is increased according to a Linear Quadratic Regulator (LQR) law with state cost and control cost equal to 1 so as to encourage the system to converge quickly back to the reference trajectory if it has been suddenly perturbed by a jump. We impose two criteria for when this change in the control law should be applied:

1) the size of the jump must be larger than a predetermined height. For our purposes, we set this threshold to 1.
2) the jump perturbs the system further away from the reference trajectory. There is no need to force an increase in the gain if the jump is beneficial.

Once the system trajectory converges within ε distance of the reference, we revert the gain back to $a + 0.5$. Furthermore, there may be a slight time delay in when we observe the jump noise, so the control law is not applied immediately after the jump occurs. We choose $\varepsilon = 0.01$ and a time delay of 0.5 seconds. Three simulated trajectories, with $\lambda = 3, 7, 11$, under the setting described is shown in Fig. 3. Performing 100 simulations and perturbing the system for 100 different combinations of Lévy noise yields the same behavior of converging to the sinusoidal reference trajectory as much as possible.

Remark 9. Note that in contrast to the criteria derived in Section III-A, which essentially bounded the maximum deviation possible in open-loop, contraction for closed-loop would refer to the system's ability to globally exponentially converge to within some bounded distance away from the nominal trajectory before it is hit by another jump. Hence, the bound will be dependent upon the time that elapses in
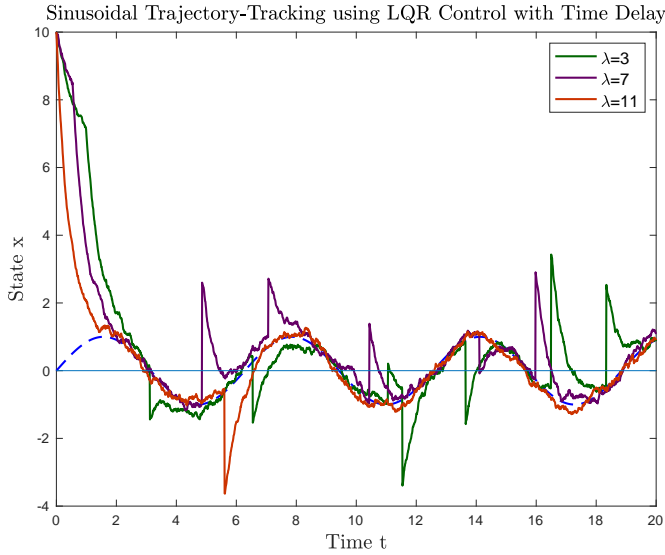
Fig. 3: Three system responses of an open-loop unstable linear system when influenced by white and shot noise, and controlled by a simple PD control scheme. The parameter λ of the shot noise process is varied while the mean and variance of the white noise is kept constant.

between jumps to ensure sufficient convergence of the system trajectory to the reference one.

V. CONCLUSION

To summarize, we have performed stability analysis for (10), stochastic systems perturbed by the broad class of Lévy noise. We leveraged techniques from contraction theory to prove incremental stability, which offers a more general characterization of stability than Lyapunov methods do because of its guarantee of globally exponential convergence with respect to time-varying trajectories rather than stationary equilibrium points or periodic limit cycles. We derived a finite-time-stability result and show that trajectories perturbed by different sample paths of noise converge exponentially to a time-varying bound in the mean-squared sense. An additional tool that was developed to facilitate the derivation was the existence and uniqueness theorem for (12), as well as rigorous definitions of the Poisson random measure and the Poisson integral. Our primary focus on pure shot noise SDEs as opposed to the general Lévy noise is justified by the Lévy-Khintchine Decomposition Formula, which allows us to represent Lévy noise as an affine combination of white and shot noise. Finally, we illustrated by simulation of a 5D Dubins car that system trajectories with shot noise are still reasonably bounded according to the error bound we derived through contraction theory analysis. We additionally illustrated that the usual LQR control law suffices for reference-tracking in a 1D linear system perturbed by constant-variation Lévy noise.

One direction of future investigation is a methodological, principled approach to controller and filter designs for higher-dimensional nonlinear systems perturbed by Lévy noise. We

may then be able to use this as an improved baseline controller upon which the conventional machine learning-based approaches mentions in Section I-A can improve upon, rather than learning from a pure white noise controller or from scratch. Possible demonstrations on real-life robotic dynamical systems such as a quadrotor is also of high interest to the authors.

APPENDIX I
PROOF OF LEMMA 2 AND
ADDITIONAL PROPERTIES OF THE POISSON INTEGRAL

In this section, we prove Lemma 2. We provide a more complete version than what was given in [53]. Consider the simple function $g(t,y) = c_1 \mathbb{1}(y \in E_1) + \cdots + c_n \mathbb{1}(y \in E_n)$. We can decompose the integral into a sum

$$
\begin{aligned}
I_g &= \int_{[0,T] \times E} \sum_{k=1}^{n} c_k \mathbb{1}(y \in E_k) N(dt, dy) \\
&= \sum_{k=1}^{n} \int_{[0,T] \times E_k} c_k N(dt, dy)
\end{aligned}
\tag{58}
$$

Then expanding out the left side of the above identity:

$$
\begin{aligned}
\mathbb{E}[I_g^2] &= \mathbb{E}\left[\left(\sum_{k=1}^{n} c_k \int_{[0,T] \times E_k} N(dt, dy) \right)^2 \right] \\
&= \sum_{k=1}^{n} c_k^2 \mathbb{E}[N_k^2(t)] + \sum_{k \neq j} c_k c_j \mathbb{E}[N_k(t) N_j(t)]
\end{aligned}
\tag{59}
$$

where $N_k(t)$ is the standard Poisson process defined on Borel set E_k, with intensity $\lambda_k := \lambda \cdot \nu(E_k)/\nu(E)$. First note that since $N_k(t)$ is Poisson distributed, $\mathbb{E}[N_k^2(t)] = \mathrm{Var}(N_k(t)) + \mathbb{E}[N_k(t)]^2 = \lambda_k t + \lambda^2 t^2$. Next, since $E_j \cap E_k = \varnothing$, we have that $N_j(t)$ and $N_k(t)$ are independent random variables. Thus

$$
\begin{aligned}
0 = \mathrm{Cov}(N_j(t), N_k(t)) &= \mathbb{E}[N_j(t)N_k(t)] - \mathbb{E}[N_j(t)]\mathbb{E}[N_k(t)] \\
&= \mathbb{E}[N_j(t)N_k(t)] - \lambda_j \lambda_k t^2
\end{aligned}
$$

Substituting these values into (59):

$$
\begin{aligned}
&\sum_{k=1}^{n} c_k^2 (\lambda_k t + \lambda_k^2 t^2) + \sum_{k \neq j} c_k c_j \lambda_k \lambda_j t^2 \\
&= \sum_{k=1}^{n} c_k^2 \lambda_k t + \sum_{k=1}^{n} c_k^2 \lambda_k^2 t^2 + \sum_{k \neq j} c_k c_j \lambda_k \lambda_j t^2 \\
&= \sum_{k=1}^{n} \int_{[0,T] \times E_k} c_k^2 \, dt\nu(dy) \\
&\quad + \left(\sum_{k=1}^{n} \int_{[0,T] \times E_k} c_k \, dt\nu(dy) \right)^2 \\
&= \int_{[0,T] \times E} \mathbb{E}[g(t,y)^2] \, dt\nu(dy) \\
&\quad + \left(\int_{[0,T] \times E} \mathbb{E}[g(t,y)] \, dt\nu(dy) \right)^2
\end{aligned}
$$

Use the Monotone Convergence Theorem to apply the same argument as above to general functions g by approximating g with step-functions g_n. This concludes our proof.

For further intuition of the computation of Poisson integrals, consider the following example.

Example 3. If I_g is a.s. absolutely convergent, then

$$\mathbb{E}\left[e^{i\beta I_g}\right] = \exp\left(-\int_{[0,T]\times E}\left(1 - e^{i\beta g(t,y)}\right)dt\nu(dy)\right) \quad (60)$$

where $\beta \in \mathbb{R}$. This formula is often called the *exponential formula*.

Again, we will consider the simple function $g(t,y) := c_1\mathbb{1}(y \in E_1) + \cdots + c_n\mathbb{1}(y \in E_n)$ where $n \in \mathbb{N}$, $c_k \in (0,T)$, and $E_1, \cdots, E_n$ are disjoint Borel sets.

Substituting the decomposition (58) into the left side of (60),

$$\mathbb{E}\left[e^{i\beta I_g}\right] = \mathbb{E}\left[e^{i\beta\sum_{k=1}^{n}\int_{[0,T]\times E_k}c_k N(dt,dy)}\right] \quad (61)$$

$$= \mathbb{E}\left[\prod_{k=1}^{n}e^{i\beta\int_{[0,T]\times E_k}c_k N(dt,dy)}\right] = \prod_{k=1}^{n}\mathbb{E}\left[e^{i\beta c_k N_k(t)}\right]$$

where the last equality follows from disjointness of the Borel sets E_k, hence independence between the terms in the product. Now we can use the characteristic function of a (scaled) Poisson

$$\mathbb{E}\left[e^{i\beta c_j N_j(t)}\right] = \sum_{k=0}^{\infty}e^{i\beta c_j k}e^{-\lambda_j t}\frac{(\lambda_j t)^k}{k!} = e^{\lambda_j t(e^{i\beta c_j}-1)}$$

to further simplify (61)

$$\prod_{k=1}^{n}e^{\lambda_k t(e^{i\beta c_k}-1)} = \exp\left(\sum_{k=1}^{n}\lambda_k t\left(e^{i\beta c_k}-1\right)\right)$$

$$= \exp\left(-\sum_{k=1}^{n}\int_{[0,T]\times E_k}\left(1 - e^{i\beta c_k}\right)N(dt,dy)\right) \quad (62)$$

Now to make the extension from simple functions to general functions, we can construct a sequence of simple functions g_n with limit $g_n \uparrow g$ as $n \to \infty$. Again, we apply the Monotone Convergence Theorem to get that the integrals with respect to the Poisson measure N converge as well. Using this argument, we can substitute in the general function g in place of the c_k in (62). Further using $E = \cup_k E_k$, with E_k disjoint sets, yields our desired form:

$$\mathbb{E}\left[e^{i\beta I_g}\right] = \exp\left(-\sum_{k=1}^{n}\int_{[0,T]\times E_k}\left(1 - e^{i\beta g(t,y)}N(dt,dy)\right)\right)$$

$$= \exp\left(-\int_{[0,T]\times E}\left(1 - e^{i\beta g(t,y)}N(dt,dy)\right)\right)$$

APPENDIX II
BOUNDS ON THE STOCHASTIC TERMS

Lemma 5. Consider the function

$$V(t,\mathbf{z},\delta\mathbf{z}) = \int_0^1\left(\frac{\partial\mathbf{z}}{\partial\mu}\right)^T S(t,\mathbf{z}(\mu,t))\left(\frac{\partial\mathbf{z}}{\partial\mu}\right)d\mu$$

with respect to metric S, which satisfies Assumption 1. Then the following hold true, where constants are as in (31) and $\varepsilon > 0$ arbitrary:

$$\sum_{i,j=1}^{n}\frac{\partial^2 V}{\partial\delta\mathbf{z}_i\partial\delta\mathbf{z}_j}(t,\mathbf{z},\delta\mathbf{z})d\langle\delta\mathbf{z}_i,\delta\mathbf{z}_j\rangle \le \overline{s}(\eta_1^2 + \eta_2^2) \quad (63a)$$

$$\sum_{i,j=1}^{n}\frac{\partial^2 V}{\partial\mathbf{z}_j\partial\delta\mathbf{z}_i}(t,\mathbf{z},\delta\mathbf{z})d\langle\mathbf{z}_i,\delta\mathbf{z}_j\rangle$$

$$\le \frac{1}{2}s'(\eta_1^2 + \eta_2^2)\left(\int_0^1\left\|\frac{\partial\mathbf{z}}{\partial\mu}\right\|^2 d\mu + 1\right) \quad (63b)$$

$$\sum_{i,j=1}^{n}\frac{\partial^2 V}{\partial\mathbf{z}_i\partial\mathbf{z}_j}(t,\mathbf{z},\delta\mathbf{z})d\langle\mathbf{z}_i,\mathbf{z}_j\rangle$$

$$\le \frac{1}{2}s''(\eta_1^2 + \eta_2^2)\int_0^1\left\|\frac{\partial\mathbf{z}}{\partial\mu}\right\|^2 d\mu \quad (63c)$$

Proof:
First, we compute the quadratic variation terms. We have

$$d\mathbf{z}_i = f(t,\mathbf{z}_i)dt + \sum_{k=1}^{d}\sigma_{\mu,ik}dW_{\mu,k}(t)$$

$$d\delta\mathbf{z}_i = (F\delta\mathbf{z})_i dt + \sum_{k=1}^{d}\delta\sigma_{\mu,ik}dW_{\mu,k}(t)$$

Then we obtain

$$d\langle\delta\mathbf{z}_i,\delta\mathbf{z}_j\rangle = \sum_{k=1}^{d}\delta\sigma_{\mu,ik}\delta\sigma_{\mu,jk}$$

$$d\langle\mathbf{z}_i,\delta\mathbf{z}_j\rangle = \sum_{k=1}^{d}\sigma_{\mu,ik}\delta\sigma_{\mu,jk}$$

$$d\langle\mathbf{z}_i,\mathbf{z}_j\rangle = \sum_{k=1}^{d}\sigma_{\mu,ik}\sigma_{\mu,jk}$$

a. *Proof of (63a):* From matrix multiplication, and because $S(t,\mathbf{z})$ is independent of $\delta\mathbf{z}$:

$$\delta\mathbf{z}^T S(t,\mathbf{z})\delta\mathbf{z} = \sum_{k=1}^{n}\sum_{l=1}^{n}\delta\mathbf{z}_k\delta\mathbf{z}_l S_{kl}(t,\mathbf{z})$$

$$\frac{\partial^2 V}{\partial\delta\mathbf{z}_i\partial\delta\mathbf{z}_j}(t,\mathbf{z},\delta\mathbf{z}) = 2\int_0^1 S_{ij}(t,\mathbf{z}(\mu,t))d\mu$$

Substituting into the left side of (63a), we get:

$$\sum_{i,j=1}^{n}\frac{\partial^2 V}{\partial\delta\mathbf{z}_i\partial\delta\mathbf{z}_j}(t,\mathbf{z},\delta\mathbf{z})d\langle\delta\mathbf{z}_i,\delta\mathbf{z}_j\rangle$$

$$= 2\int_0^1\sum_{i,j=1}^{n}S_{ij}(t,\mathbf{z})\sum_{k=1}^{d}\left(\frac{\partial\sigma_\mu}{\partial\mu}\right)_{ik}\left(\frac{\partial\sigma_\mu}{\partial\mu}\right)_{jk}d\mu$$

$$\le 2\overline{s}\sum_{i,j=1}^{n}\sum_{k=1}^{d}\int_0^1\left(\frac{\partial\sigma_\mu}{\partial\mu}\right)_{ik}\left(\frac{\partial\sigma_\mu}{\partial\mu}\right)_{jk}d\mu \quad (64)$$

We have the following identity for any square matrix A and any pair $i,j = 1,\cdots,n$ such that $i \ne j$

$$\frac{1}{2}\text{tr}(A^T A) = \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{d}a_{ik}^2 \ge \sum_{k=1}^{d}a_{ik}a_{jk}$$

which is easily seen by completing the squares. In the 2D case ($n = 2, d = 2$), we would get:

$$A^T A = \begin{bmatrix} a_{11}^2 + a_{12}^2 & * \\ * & a_{21}^2 + a_{22}^2 \end{bmatrix}$$

$$\implies \operatorname{tr}(A^T A) = \sum_{i,k=1}^{2} a_{ij}^2 \geq 2(a_{11}a_{21} + a_{21}a_{22})$$

$$\implies \frac{1}{2}\operatorname{tr}(A^T A) \geq a_{11}a_{21} + a_{21}a_{22}$$

and when $i = j$, the same bound holds because the sum is simply the ith partial sum of the trace.

This allows us to bound (64) by splitting up into terms with $i = j$ and terms with $i \neq j$ and bounding both parts by a trace:

$$2\overline{s} \sum_{i,j=1}^{n} \sum_{k=1}^{d} \int_0^1 \left(\frac{\partial \sigma_\mu}{\partial \mu}\right)_{ik} \left(\frac{\partial \sigma_\mu}{\partial \mu}\right)_{jk} d\mu \leq 2\overline{s}\operatorname{tr}\left(\sigma_\mu^T \sigma_\mu(t, \mathbf{x})\right)$$

$$\leq \overline{s}\left[\operatorname{tr}\left(\sigma_1^T \sigma_1(t, \mathbf{x})\right) + \operatorname{tr}\left(\sigma_2^T \sigma_2(t, \mathbf{y})\right)\right]$$

$$\leq \overline{s}(\eta_1^2 + \eta_2^2)$$

b. *Proof of (63b)*: First, we can compute the matrix derivative as follows.

$$\frac{\partial^2 V}{\partial \mathbf{z}_i \partial \delta \mathbf{z}_j}(t, \mathbf{z}, \delta \mathbf{z}) = 2 \int_0^1 \frac{\partial}{\partial \mathbf{z}_i} S_{j,i}(t, \mathbf{z})(\delta \mathbf{z})_i d\mu$$

$$\forall 1 \leq i, j \leq n$$

This gives:

$$2 \int_0^1 \sum_{i,j=1}^{n} \frac{\partial}{\partial \mathbf{z}_i} S_{j,i}(t, \mathbf{z})(\delta \mathbf{z})_i \sum_{k=1}^{d} \sigma_{\mu,ik} \left(\frac{\partial \sigma_\mu}{\partial \mu}\right)_{jk} d\mu$$

$$\leq 2s' \int_0^1 \sum_{i=1}^{n} \left(\frac{\partial \mathbf{z}}{\partial \mu}\right)_i \sum_{k=1}^{d} \sigma_{\mu,ik} \delta \sigma_{\mu,jk} d\mu$$

$$\leq s'(\eta_1^2 + \eta_2^2) \int_0^1 \sum_{i=1}^{n} \left(\frac{\partial \mathbf{z}}{\partial \mu}\right)_i d\mu$$

$$\leq s'(\eta_1^2 + \eta_2^2) \int_0^1 \left\|\frac{\partial \mathbf{z}}{\partial \mu}\right\| d\mu \tag{65}$$

where the second-to-last inequality follows from the same trace bound used in (63a).

Note that for any $a, b > 0$, $2ab \leq a^2 + b^2$ and so by Cauchy-Schwarz

$$\int_0^1 \left\|\frac{\partial \mathbf{z}}{\partial \mu}\right\| d\mu \leq \frac{1}{2} \left(\int_0^1 \left\|\frac{\partial \mathbf{z}}{\partial \mu}\right\|^2 d\mu + 1\right) \tag{66}$$

Combining (65) together with (66) yields our final bound:

$$\frac{1}{2}s'(\eta_1^2 + \eta_2^2) \left(\int_0^1 \left\|\frac{\partial \mathbf{z}}{\partial \mu}\right\|^2 d\mu + 1\right)$$

c. *Proof of (63c)*: Again, start by computing the matrix derivative

$$\frac{\partial^2 V}{\partial \mathbf{z}_i \partial \mathbf{z}_j}(t, \mathbf{z}, \delta \mathbf{z}) = \int_0^1 \sum_{k,l=1}^{n} \frac{\partial^2}{\partial \mathbf{z}_i \partial \mathbf{z}_j} S_{kl}(t, \mathbf{z}) \delta \mathbf{z}_k \delta \mathbf{z}_l d\mu$$

Now we can bound using essentially the same technique as in (63b).

$$\sum_{i,j=1}^{n} \frac{\partial^2 V}{\partial \mathbf{z}_i \partial \mathbf{z}_j}(t, \mathbf{z}, \delta \mathbf{z}) d\langle \mathbf{z}_i, \mathbf{z}_j \rangle$$

$$= \sum_{i,j=1}^{n} \sum_{k,l=1}^{n} \int_0^1 \frac{\partial^2}{\partial \mathbf{z}_i \partial \mathbf{z}_j} S_{kl}(t, \mathbf{z})$$

$$\times \left(\frac{\partial \mathbf{z}_k}{\partial \mu}\right)^T \cdot \sum_{m=1}^{d} \sigma_{\mu,im} \sigma_{\mu,jm} \cdot \left(\frac{\partial \mathbf{z}_l}{\partial \mu}\right) d\mu$$

$$\leq \frac{1}{2}s''(\eta_1^2 + \eta_2^2) \sum_{k,l=1}^{n} \int_0^1 \left(\frac{\partial \mathbf{z}_k}{\partial \mu}\right)^T \left(\frac{\partial \mathbf{z}_l}{\partial \mu}\right) d\mu$$

$$\leq \frac{1}{2}s''(\eta_1^2 + \eta_2^2) \int_0^1 \left\|\frac{\partial \mathbf{z}}{\partial \mu}\right\|^2 d\mu$$

$\square$

REFERENCES

[1] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, Jan 1949.

[2] V. Capuano, A. Harvard, Y. Lin, and S.-J. Chung, "DGNSS-vision integration for robust and accurate relative spacecraft navigation," in *Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019)*, Sep 2019, pp. 2923–2939.

[3] J. Yang, A. Dani, S.-J. Chung, and S. Hutchinson, "Inertial-aided vision-based localization and mapping in a riverine environment with reflection measurements," in *AIAA Guidance, Navigation, and Control (GNC) Conference*, Aug 2013.

[4] M. Kalakrishnan, S. Chitta, E. Theodorou, P. Pastor, and S. Schaal, "STOMP: Stochastic trajectory optimization for motion planning," in *2011 IEEE International Conference on Robotics and Automation*, Aug 2011, pp. 4569–4574.

[5] M. H. A. Davis, "Construction of Brownian motion," 2011. [Online]. Available: http://wwwf.imperial.ac.uk/~mdavis/course_material/sp1/BROWNIAN_MOTION.PDF

[6] F. Baccelli and B. Blaszczyszyn, *Stochastic Geometry and Wireless Networks, Part I: Theory*. Hanover, MA, USA: Now Publishers Inc., 2009.

[7] R. Cont and P. Tankov, *Financial Modelling With Jump Processes*. London, UK: CRC Press UK, 2004.

[8] A. Patel and B. Kosko, "Stochastic resonance in continuous and spiking neuron models with Lévy noise," *IEEE Transactions on Neural Networks*, vol. 19, pp. 1993–2008, Jan 2009.

[9] H. J. Kushner, *Stochastic Stability and Control*. Academic Press, 1967.

[10] E. Sontag, "A Lyapunov-like characterization of asymptotic controllability," *Siam Journal on Control and Optimization*, vol. 21, no. 3, pp. 462–471, May 1983.

[11] H. K. Khalil, *Nonlinear Systems; 3rd ed.* Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[12] J. Slotine and W. Li, *Applied Nonlinear Control*. Eaglewood Cliffs, NJ, USA: Prentice Hall, 1991.

[13] S.-J. Chung and J. E. Slotine, "Cooperative robot control and concurrent synchronization of Lagrangian systems," *IEEE Transactions on Robotics*, vol. 25, no. 3, pp. 686–700, Jun 2009.

[14] J.-J. Slotine, W. Wang, and K. Rifai, "Contraction analysis of synchronization in networks of nonlinearly coupled oscillators," in *Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems*, Jan 2004, pp. 5–8.

[15] S.-J. Chung, S. Bandyopadhyay, I. Chang, and F. Y. Hadaegh, "Phase synchronization control of complex networks of Lagrangian systems on adaptive digraphs," *Automatica*, vol. 49, no. 5, pp. 1148–1161, May 2013.

[16] D. Angeli, "A Lyapunov approach to incremental stability properties," *IEEE Transactions on Automatic Control*, vol. 47, no. 3, pp. 410–421, Mar 2002.

[17] W. Lohmiller and J.-J. E. Slotine, "Nonlinear process control using contraction theory," *American Institute of Chemical Engineers Journal*, no. 3, pp. 588–596, Mar 2000.

[18] ——, "Contraction analysis of nonlinear distributed systems," *International Journal of Control*, no. 9, pp. 678–688, 2005.

[19] Hua Deng, M. Krstic, and R. J. Williams, "Stabilization of stochastic nonlinear systems driven by noise of unknown covariance," *IEEE Transactions on Automatic Control*, vol. 46, no. 8, pp. 1237–1253, Aug 2001.

[20] P. Florchinger, "Lyapunov-like techniques for stochastic stability," in *Proceedings of 1994 33rd IEEE Conference on Decision and Control*, Dec 1994, pp. 1145–1150.

[21] ——, "Feedback stabilization of affine in the control stochastic differential systems by the control lyapunov function method," in *Proceedings of 35th IEEE Conference on Decision and Control*, vol. 1, Dec 1996, pp. 1047–1052.

[22] Q.-C. Pham, N. Tabareau, and J.-J. Slotine, "A contraction theory approach to stochastic incremental stability," *IEEE Transactions on Automatic Control*, pp. 816–820, 2009.

[23] A. P. Dani, S.-J. Chung, and S. Hutchinson, "Observer design for stochastic nonlinear systems via contraction-based incremental stability," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 700–714, Mar 2015.

[24] Y. Qin, M. Cao, and B. D. O. Anderson, "Lyapunov criterion for stochastic systems and its applications in distributed computation," *IEEE Transactions on Automatic Control*, Apr 2019. early access.

[25] Y. Xu, X.-Y. Wang, H.-Q. Zhang, and W. Xu, "Stochastic stability for nonlinear systems driven by Lévy noise," *Nonlinear Dynamics*, vol. 68, pp. 7–15, Apr 2012.

[26] D. Applebaum and M. Siakalli, "Asymptotic stability of stochastic differential equations driven by Lévy noise," *Journal of Applied Probability*, vol. 46, pp. 1116–1129, Dec 2009.

[27] X. Mao, "Exponential stability for stochastic differential equations with respect to semimartingales," *Stochastic Processes and their Applications*, vol. 35, no. 2, pp. 247–277, Aug 1990.

[28] D. Liberzon, J. P. Hespanha, and A. Morse, "Stability of switched systems: a lie-algebraic condition," *Systems & Control Letters*, vol. 37, no. 3, pp. 117 – 122, Jul 1999.

[29] J. P. Hespanha, "Uniform stability of switched linear systems: extensions of lasalle's invariance principle," *IEEE Transactions on Automatic Control*, vol. 49, no. 4, pp. 470–482, Apr 2004.

[30] A. Mesquita and J. Hespanha, "Construction of Lyapunov functions for piecewise-deterministic Markov processes," in *Proceedings of the IEEE Conference on Decision and Control*, Dec 2010, pp. 2408–2413.

[31] Q. C. Pham, "Analysis of discrete and hybrid stochastic systems by nonlinear contraction theory," in *International Conference on Control, Automation, Robotics and Vision*, Jan 2009, pp. 1054 – 1059.

[32] J. P. Hespanha and A. S. Morse, "Stability of switched systems with average dwell-time," in *Proceedings of the 38th IEEE Conference on Decision and Control*, vol. 3, Dec 1999, pp. 2655–2660.

[33] M. Dorothy and S.-J. Chung, "Switched systems with multiple invariant sets," *Systems & Control Letters*, vol. 96, pp. 103–109, Oct 2016.

[34] A. R. Mesquita and J. P. Hespanha, "Jump control of probability densities with applications to autonomous vehicle motion," *IEEE Transactions on Automatic Control*, vol. 57, no. 10, pp. 2588–2598, Oct 2012.

[35] D. S. Bernstein and W. M. Haddad, "LQG control with an h_∞ performance bound: a Riccati equation approach," in *1988 American Control Conference*, Jun 1988, pp. 796–802.

[36] V. Gupta, B. Hassibi, and R. M. Murray, "Optimal LQG control across packet-dropping links," *Systems & Control Letters*, vol. 56, no. 6, pp. 439 – 446, Jun 2007.

[37] E. Theodorou, J. Buchli, and S. Schaal, "A generalized path integral control approach to reinforcement learning," *Journal of Machine Learning Research*, vol. 11, p. 31373181, Mar 2010.

[38] K. Rawlik, M. Toussaint, and S. Vijayakumar, "On stochastic optimal control and reinforcement learning by approximate inference," in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, Aug 2013, pp. 3052–3056.

[39] Y. Nakka and S.-J. Chung, "Trajectory optimization for chance-constrained nonlinear stochastic systems," in *58th IEEE Conference on Decision and Control*, Sep 2019, to appear.

[40] R. Munos and P. Bourgine, "Reinforcement learning for continuous stochastic control problems," in *Advances in Neural Information Processing Systems 10*. MIT Press, 1998, pp. 1029–1035.

[41] M. P. Deisenroth, D. Fox, and C. E. Rasmussen, "Gaussian processes for data-efficient learning in robotics and control," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 2, pp. 408–423, Feb 2015.

[42] S. L. Herbert, M. Chen, S. Han, S. Bansal, J. F. Fisac, and C. J. Tomlin, "FaSTrack: A modular framework for fast and guaranteed safe motion planning," in *IEEE 56th Annual Conference on Decision and Control*, Dec 2017, pp. 1517–1522.

[43] S. Bansal, M. Chen, J. Fisac, and C. Tomlin, "Safe sequential path planning of multi-vehicle systems under presence of disturbances and imperfect information," in *American Control Conference*, May 2017.

[44] B. Øksendal and A. Sulem-Bialobroda, *Applied Stochastic Control of Jump Diffusions*, ser. Universitext. Berlin, Germany: Springer-Verlag, 2007.

[45] W.-H. Chen and B. Chen, "Robust stabilization design for nonlinear stochastic system with poisson noise via fuzzy interpolation method," *Fuzzy Sets and Systems*, vol. 217, p. 4161, 04 2013.

[46] J. Han and E. Weinan, "Deep learning approximation for stochastic control problems," *ArXiv*, Nov 2016. [Online]. Available: https://arxiv.org/pdf/1611.07422.pdf

[47] G. Shi, W. Hönig, Y. Yue, and S.-J. Chung, "Neural-swarm: Decentralized close-proximity multirotor control using learned interactions," in *International Conference on Robotics and Automation*, 2020, to appear.

[48] W. Lohmiller and J.-J. E. Slotine, "On contraction analysis for non-linear systems," *Automatica*, vol. 34, no. 6, pp. 683–696, Jun 1998.

[49] S.-J. Chung and J.-E. Slotine, "On synchronization of coupled Hopf-Kuramoto oscillators with phase delays," in *49th IEEE Conference on Decision and Control (CDC)*, Dec 2010, pp. 3181–3187.

[50] J. Jouffroy and J. . E. Slotine, "Methodological remarks on contraction theory," in *IEEE Conference on Decision and Control (CDC)*, Dec 2004, pp. 2537–2543.

[51] P. Protter, *Stochastic Integration and Differential Equation*, 2nd ed. Berlin, Heidelberg: Springer-Verlag, 1992.

[52] M. Jeanblanc. (2007, Apr) Jump processes. [Online]. Available: https://www.maths.univ-evry.fr/pages_perso/jeanblanc/conferences/Cimpa-sauts.pdf

[53] A. Watson. (2016, Sep) Introduction to Lévy processes. [Online]. Available: https://personalpages.manchester.ac.uk/staff/alex.watson/lectures/intro-levy.pdf

[54] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, ser. Universitext. Berlin, Germany: Springer Berlin, 2010.

[55] D. Applebaum, *Lévy Processes and Stochastic Calculus*, 2nd ed., ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

[56] J. Liu, "On the existence and uniqueness of solutions to stochastic differential equations of mixed Brownian and Poissonian sheet type," *Stochastic Processes and their Applications*, vol. 94, no. 2, pp. 339–354, Aug 2001.

[57] Y. Kasahara and K. Yamada, "Stability theorem for stochastic differential equations with jumps," *Stochastic Processes and their Applications*, vol. 38, no. 1, pp. 13 – 32, 1991.

[58] B. Hajek, "Notes for ECE534: An exploration of random processes for engineers," 2006. [Online]. Available: http://www.ifp.uiuc.edu/~hajek/Papers/randomprocJuly06.pdf

[59] J. Jacod and A. Shiryaev, *Limit Theorems for Stochastic Processes*. Berlin, Germany: Springer Berlin, 2003.

[60] M. Podolskij and M. Vetter, "Understanding limit theorems for semi-martingales: a short survey," *Special Issue: Statistical Inference for Lévy Processes with Applications to Finance*, vol. 6, no. 3, pp. 329–351, Aug 2010.

[61] H. Kunita, "Ito's stochastic calculus: Its surprising power for applications," *Stochastic Processes and their Applications*, vol. 120, pp. 622–652, 05 2010.

[62] R. F. Bass. (2009, Apr) SDEs with jumps; Notes for Barrett lectures. [Online]. Available: http://www.math.utk.edu/~jxiong/barrett/Barrett_Bass.pdf
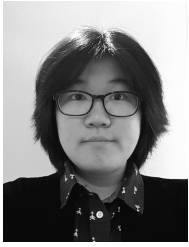
SooJean Han (BS'16) received her B.S. degree in Electrical Engineering and Computer Science, and Applied Mathematics at the University of California, Berkeley (UC Berkeley) in 2016. In 2017, she was a research assistant in the Hybrid Systems Lab at UC Berkeley, and a research assistant in the Center for Autonomous Systems and Technology (CAST) Lab at California Institute of Technology (Caltech). She is currently a Ph.D. candidate in Control and Dynamical Systems at Caltech. Her research interests include stochastic nonlinear systems, decentralized control and estimation over multi-agent swarms, decision-making networked systems, and machine learning. She is a recipient of the Caltech Special EAS Fellowship and the National Science Foundation Graduate Research Fellowship (NSF GRFP).

Soon-Jo Chung (M06–SM12) received the B.S. degree (summa cum laude) in aerospace engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 1998, and the S.M. degree in aeronautics and astronautics and the Sc.D. degree in estimation and control from Massachusetts Institute of Technology, Cambridge, MA, USA, in 2002 and 2007, respectively.

He is currently Bren Professor of Aerospace and a Jet Propulsion Laboratory Research Scientist in the California Institute of Technology, Pasadena, CA, USA. He was with the faculty of the University of Illinois at Urbana-Champaign (UIUC) during 2009–2016. His research interests include spacecraft and aerial swarms and autonomous aerospace systems, and in particular, on the theory and application of complex nonlinear dynamics, control, estimation, guidance, and navigation of autonomous space and air vehicles.

Dr. Chung was the recipient of the UIUC Engineering Deans Award for Excellence in Research, the Beckman Faculty Fellowship of the UIUC Center for Advanced Study, the U.S. Air Force Office of Scientific Research Young Investigator Award, the National Science Foundation Faculty Early Career Development Award, and three Best Conference Paper Awards from the IEEE and the American Institute of Aeronautics and Astronautics. He is an Associate Editor of IEEE TRANSACTIONS ON ROBOTICS, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and Journal of Guidance, Control, and Dynamics.