Stochastic Processes for Control Engineers

SooJean Han*

December 27, 2020

Abstract

This manuscript provides a treatment of concepts from the theory of stochastic processes with a focus on applications to control theory. We look into stochastic systems which can be represented as stochastic differential equations (SDEs) which are perturbed by three specific types of noise processes: Gaussian white, Poisson shot, and Lévy. Although many controller and observer design approaches already exist in literature for white noise, there isn't as much extensive treatment for other two types, despite being equally prevalent in stochastic systems. The practical utilities include the design of stochastic controllers and observers which do not immediately resort to model-free techniques to synthesize from scratch simply because the noise process is non-Gaussian.

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

Contents

Chapter 1

Poisson Point Processes and Poisson Random Measure

Definition 1 (Poisson Random Measure)**.** Let $E \subseteq \mathbb{N}^\ell$. We define random measure $N([0,t] \times E)$ on *jump space* E until some time $t > 0$ with *intensity measure* $\text{Leb} \times \nu$, where Leb denotes the standard Lebesgue measure (the measure in time) and ν is the probability measure on the jump space E (describing the distribution of the jumps). We denote the *intensity (parameter)* for the Poisson process corresponding to the intensity measure ν as λ. One can think of λ as the average number of arrivals over time. N is called a *Poisson random measure* if the following are satisfied:

1. if $E_1, \cdots, E_n$ are pairwise disjoint subsets of E, then $N([0,t]\times E_1), \cdots, N([0,t]\times E_n)$ are independent.
2. for each $E_i \subseteq E$, the random measure $N([0,t]\times E_i)$ has a corresponding Poisson process with intensity parameter $\lambda_i := (\nu(E_i)/\nu(E))\lambda$.

We denote the time of the ith arrival with random variable T_i. $\square$

As a visual example to the definition above, we will take $E := \mathbb{R}^2$. Color the points either black with probability p or white with probability $q := 1 - p$. For any subset $A \subseteq E$, let $N(A)$ denote the number of points which belong in A. Further, let $N_b(A)$ be the number of black points among those points in A, $N_w(A)$ be the number of white points in A. Then $N(A)$ is distributed as a Poisson random variable with mean measure $\nu(A)$. Additionally, the conditional distribution of $(N_b(A)|N(A)) \sim \text{Binom}(N(A), p)$. A 2D example illustration of this notion is shown in Fig. 1.1.

This concept can be generalized to more than two colors by the following theorem.

Theorem 1 (Coloring Theorem)**.** Suppose N is a Poisson point process on E with mean measure μ. Color the points randomly with one of K colors such that the ith color is chosen with probability p_i and $\sum_{k=1}^{K} p_k = 1$. Denote N_i to be the subset of the total points in N which are colored with color i. Then the N_i are independent Poisson processes with mean measure $\mu_i := p_i\mu$.

Colors of points can also be denoted by more complex items such as a function ξ on position x: for each point $x \in N$, associate a random variable $\xi(X) \in \mathbb{R}^\ell$ to be its *mark*. For instance, in the coloring setup above, $\xi(X)$ is just the color; in fact, the mark is actually independent of the point X in this case, since the color is randomly assigned with some probability.

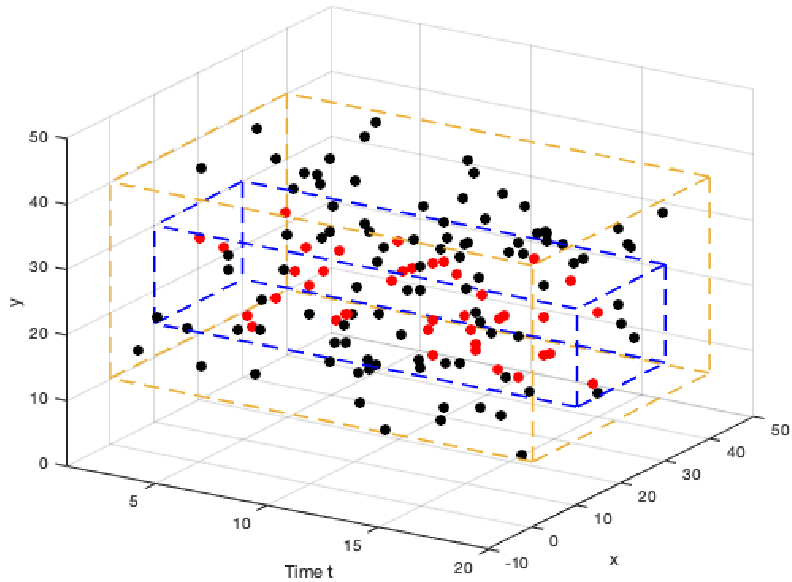

Figure 1.1: The locations of the points of the form $(x, y) \in \mathbb{R}^2 =: E$ for all time $t \leq 20$. Points which are in E are marked in red. In this figure, $E := \{0 \leq x \leq 40, 10 \leq y \leq 40\}$ and $E := \{10 \leq x \leq 30, 15 \leq y \leq 30\}$.

Theorem 2 (Marking Theorem)**.** Definition of N^*? N^* as constructed above is a Poisson point process on $E \times \mathbb{R}^\ell$ with mean measure μ^* given by

$$\mu^*(E^*) = \int \int_{(x,m) \in E^*} p(dm|x)\mu(dx)$$

for all subsets $E^* \subseteq E \times \mathbb{R}^\ell$. $p(m|x)$ represents the conditional distribution of the marks in N^* given the position of the points x generated from N.

Ultimately, this construction leads to a formal definition of the Poisson random measure. Throughout the manuscript, we use a version that closely follows Theorem 2.3.5 in [1] and Definition 3.1 in [2].

Chapter 2

Stochastic Integration

2.1 With Respect to Brownian Motion Processes

Stochastic integration of a function g with respect to a Brownian motion process can be defined in different ways depending on the class that g belongs to. We consider two cases here.

Denote $\mathcal{L}^2(I;\mathbb{R})$ to be the space of deterministic, square-integrable functions mapping $I := [a,b] \subseteq \mathbb{R}$ to $\mathbb{R}$. Further denote $\mathcal{C}_0^1(I;\mathbb{R}) := \left\{h \in C^1(I;\mathbb{R}) \mid h(a) = h(b) = 0\right\}$ to be the space of continuously-differentiable functions such that the endpoints take value 0.

Definition 2 (Paley-Wiener-Zygmund, [3])**.** For $g \in \mathcal{L}^2(I;\mathbb{R})$, we define the *Paley-Wiener-Zygmund (PWZ) integral* as:

$$\int_0^T g(s)\, dW(s) = -\lim_{n\to\infty} \int_0^T g_n'(s) W(s)\, ds$$

where the limit is defined in the L^2 sense, $g_n \in C_0^1$, and $\lim_{n\to\infty} \|g - g_n\|_{L^2(I;\mathbb{R})}$. □

Remark 1. The definition of the PWZ Integral is inspired by the concept of integration-by-parts. Informally, suppose that $W(\cdot)$ is differentiable. Then:

$$\int_0^T g(s) \frac{dW(s)}{ds}\, ds = g(s)W(s)\Big|_0^T - \int_0^T g'(s)W(s)\, ds = -\int_0^T g'(s)W(s)\, ds$$

where the last equality comes from the compact support conditions imposed on g. □

Theorem 3. The PWZ Integral satisfies the following properties:

$$\mathbb{E}\left[\int_0^T g(s)\, dW(s)\right] = 0 \tag{2.1}$$

$$\mathbb{E}\left[\left(\int_0^T g(s)\, dW(s)\right)^2\right] = \int_0^T g(s)^2\, ds \tag{2.2}$$

Now denote $\mathbb{L}^2(I;\mathbb{R})$ to be the space of stochastic functions which are mean-square integrable.

Definition 3. Define the *Itô Stochastic Integral* by:

$$\int_0^T g(s)\, dW(s) \approx \sum_{k=0}^{m-1} g_k(\underbrace{W(t_{k+1}) - W(t_k)}_{=:\Delta W_k})$$

where $\{g_k\}_{k=0}^{m-1}$ in $\mathbb{R}$ is a sequence of real-valued random variables for which the function g can be approximated by: $g(t) \approx g_k, t_k < t < t_{k+1}$. $\square$

Theorem 4. The Itô Integral satisfies properties analogous to those of the PWZ Integral described above:

$$\mathbb{E}\left[\int_0^T g(s)\, dW(s)\right] = 0 \tag{2.3}$$

$$\mathbb{E}\left[\left(\int_0^T g(s)\, dW(s)\right)^2\right] = \int_0^T \mathbb{E}[g(s)^2]\, ds \tag{2.4}$$

The second property is known as the "Itô Isometry".

Remark 2. More general step function approximations to g can be defined as follows:

$$\int_0^T g(s)\, dW(s) \approx \sum_{k=0}^{m-1} g_{k,\lambda} \Delta W_k$$

where $g_{k,\lambda} := g(\lambda t_k + (1-\lambda)t_{k+1})$ and $\lambda \in [0,1]$. When $\lambda = 1$, this is simply the Itô Integral again, and $g_{k,\lambda}$ is independent of ΔW_k. When $\lambda = \frac{1}{2}$, this is the *Stratonovich Integral* and $g_{k,\lambda}$ is not independent of ΔW_k. $\square$

We can derive a relationship between the Itô Integral and the Stratonovich Integral as follows. Let the interval $[0,T)$ be partitioned into n intervals $0 = t_0 \leq t_1 \leq t_2 \leq \cdots t_n = T$ and let $t_{j^*} = \frac{1}{2}(t_{j+1} + t_j)$. For simplicity, we will assume that the integral is equally partitioned with stepsize $\Delta t > 0$. Further denote $W_k = W(t_k)$ for any $k \in [0,T)$.

$$\int_0^T W(t) \circ dW(t) = \lim_{\Delta t \to 0} \sum_{j=0}^{n-1} W_{j^*}(W_{j+1} - W_j)$$

$$\begin{aligned}
\sum_{j=0}^{n-1} W_{j^*}(W_{j+1} - W_j) &= \sum_{j=0}^{n-1} W_{j*}\left[W_{j+1} - W_{j^*} + W_{j^*} - W_j\right] \\
&= \sum_{j=0}^{n-1} W_{j^*}(W_{j+1} - W_{j^*}) + \sum_{j=0}^{n-1} W_{j^*}(W_{j^*} - W_j) \\
&= \sum_{j=0}^{n-1} W_{j^*}(W_{j+1} - W_{j^*}) + \sum_{j=0}^{n-1} W_j(W_{j^*} - W_j) + \sum_{j=0}^{n-1} W_{j^*}(W_{j^*} - W_j) - \sum_{j=0}^{n-1} W_j(W_{j^*} - W_j) \\
&= \sum_{j=0}^{n-1} \left[W_{j^*}(W_{j+1} - W_{j^*}) + W_j(W_{j^*} - W_j)\right] + \sum_{j=0}^{n-1} (W_{j^*} - W_j)^2
\end{aligned}$$

Note that the first sum term simply becomes the Itô Integral when $\Delta t \to 0$. The partition is chosen so that the intervals are half as long (i.e, there are twice as many grid points than the one chosen above).

Now:

$$\mathbb{E}\left[\int_0^T W(t)\circ dW(t)\right] = \mathbb{E}\left[\int_0^T W(t)\, dW(t)\right] + \lim_{\Delta t\to 0}\mathbb{E}\left[\sum_{j=1}^{n}(W_{j^*}-W_j)^2\right] = \mathbb{E}\left[\int_0^T W(t)\, dW(t)\right] + \frac{T}{2}$$

For further details about the above integrals, the interested reader is referred to e.g. [4] and references therein.

Proposition 1. Define

$$S_n := \sum_{k=0}^{n-1}(W(t_{k+1}^n) - W(t_k^n))^2$$

Then $\mathbb{E}[S_n] = T$ and $\lim_{n\to\infty} S_n = T$ a.s. This is known as the Quadratic Variation property of the Brownian motion.

Proof. We prove the result in expectation, not almost surely. Then we have $S_n = \sum_{k=0}^{n-1}(t_{k+1}^n - t_k^n)(\xi_k^n)^2$, where $\xi_k^n \sim \mathcal{N}(0,1)$ i.i.d with respect to k and a fixed n.

Thus:

$$\mathbb{E}[S_n] = \sum_{k=0}^{n-1}\delta = n\delta = T$$

and we are done. ■

Proposition 2. Define

$$R_n := \sum_{k=0}^{n-1} W\underbrace{(\lambda t_k^n + (1-\lambda)t_{k+1}^n)}_{=:\tau_k^n}(W(t_{k+1}^n) - W(t_k^n))$$

Then

$$\lim_{n\to\infty}\mathbb{E}[R_n] = (1-\lambda)T \text{ and } \lim_{n\to\infty} R_n = \frac{W(T)^2}{2} + \left(\lambda - \frac{1}{2}\right)T \text{ a.s.}$$

Proof. We shall prove the result in expectation, not almost surely.

$$\begin{aligned}
\mathbb{E}[R_n] = \sum_{k=0}^{n-1}\mathbb{E}[W(\tau_k^n)W(t_{k+1}^n)] - \mathbb{E}[W(\tau_k^n)W(t_k^n)] &= \sum_{k=0}^{n-1}\tau_k^n - t_k^n \\
&= \sum_{k=0}^{n-1}\left(\lambda t_k^n + (1-\lambda)t_{k+1}^n - t_k^n\right) \\
&= (1-\lambda)\sum_{k=0}^{n-1}(t_{k+1}^n - t_k^n) = (1-\lambda)T
\end{aligned}$$

which is the desired result.

To prove the second expression, let us first denote $W_{k+1} = W(t_{k+1}), W_k = W(t_k)$, and $W_{k^*} = W(t_k^*)$ where $t_{k^*} = t_k + \lambda(t_{k+1} - t_k)$. Then we follow the same trick as the above:

$$\begin{aligned}
R_n = \sum_{k=0}^{n-1} W_{k^*}(W_{k+1} - W_k) &= \sum_{k=0}^{n-1} W_{k^*}(W_{k+1} - W_{k^*}) + \sum_{k=0}^{n-1} W_{k^*}(W_{k^*} - W_k) \\
&= \sum_{k=0}^{n-1} \left[W_{k^*}(W_{k+1} - W_{k^*}) + W_k(W_{k^*} - W_k)\right] + \sum_{k=0}^{n-1}(W_{k^*} - W_k)^2
\end{aligned}$$

Again, the first term simply becomes the Itô Integral when $\Delta t \to 0$, with twice as many (uneven if $\lambda \neq 1/2$) partitions as chosen initially. Furthermore, $\sum_{k=0}^{n-1}(W_{k^*} - W_k)^2 \to \lambda T$ as $n \to \infty$.

We can rewrite the Itô Integral in the following way:

$$\begin{aligned}
W^2(T) = \sum_{k=0}^{n-1} \left[W_{k+1}^2 - W_k^2\right] &= \sum_{k=0}^{n-1} \left[(W_{k+1} - W_k)^2 + 2W_k W_{k+1} - 2W_k^2\right] \\
&= \sum_{k=0}^{n-1}(W_{k+1} - W_k)^2 + 2\sum_{k=0}^{n-1} W_k(W_{k+1} - W_k)
\end{aligned}$$

and as $n \to \infty$, $\sum_{k=0}^{n-1}(W_{k+1} - W_k)^2 \to T$ and $\sum_{k=0}^{n-1} W_k(W_{k+1} - W_k)$ becomes just the Itô Integral of $W(t)$.

In combination:

$$\begin{aligned}
\frac{1}{2}W^2(T) &= \frac{T}{2} + \int_0^T W(t)\, dW(t) \\
\implies \lim_{n\to\infty} R_n &= \int_0^T W(t)\, dW(t) + \lambda T = \frac{1}{2}W^2(T) - \frac{T}{2} + \lambda T
\end{aligned}$$

and we have our desired result. ■

Proposition 3. Let us write $M(t) = \int_0^t g(s)\, dW(s)$ the Itô Stochastic Integral, and $\mathcal{F}(s) = \sigma(W(\tau), 0 \leq \tau \leq s)$. Then $M(t)$ satisfies the Martingale property.

We can see this as follows:

$$\begin{aligned}
\mathbb{E}\left[M(t)|\mathcal{F}_s\right] = \mathbb{E}\left[\int_0^t g(\tau)\, dW(\tau)\middle|\mathcal{F}_s\right] &= \mathbb{E}\left[\int_0^s g(\tau)\, dW(\tau)\middle|\mathcal{F}_s\right] + \mathbb{E}\left[\int_s^t g(\tau)\, dW(\tau)\middle|\mathcal{F}_s\right] \\
&= \int_0^s g(\tau)\, dW(\tau) + \int_s^t \mathbb{E}\left[\sum_{k=0}^{n-1} g(\tau_k)(W_{k+1} - W_k)\middle|\mathcal{F}_s\right] \text{ by definition} \\
&= \int_0^s g(\tau)\, dW(\tau) + \int_s^t \sum_{k=0}^{n-1} \mathbb{E}\left[g(\tau_k)\right] \underbrace{\mathbb{E}\left[(W_{k+1} - W_k)|\mathcal{F}_s\right]}_{=0 \text{ by independent increments}} \\
&= \int_0^s g(\tau)\, dW(\tau) = M(s)
\end{aligned}$$

Indeed, $\mathbb{E}[M(t) \mid \mathcal{F}(s)] = M(s)$.

2.2 With Respect to Poisson Processes

Definition 4 (Poisson Integral)**.** Let $\xi : [0,\infty) \times E \to \mathbb{R}$ be a predictable, bounded Borel-measurable function and N be a Poisson random measure on $[0,t] \times E$ with intensity measure Leb $\times \nu$. We define the *Poisson integral* of ξ as follows:

$$I_\xi := \int_{[0,t]\times E} \xi(s,y) N(ds,dy) \tag{2.5}$$

$\square$

If the interpretation of the mark in Theorem 2 is such that it is exactly a function of the point x, e.g., a type of cost such as battery charge for electric vehicles, then **Campbell's formula** is a useful notion which allows us to determine the distribution of the sum $\sum_{X\in N} f(X)$. It even covers that case where f is not a deterministic function of X. The formula for time-invariant functions $\xi(y)$ is often presented in the literature, e.g., Section 3.2 of [5] and Proposition 2.7 of [6].

Theorem 5 (Campbell's Theorem)**.** Let N be a Poisson point process on E with mean measure μ, and suppose $f : E \to \mathbb{R}$ is a measurable function. Then the sum $S := \sum_{X\in N} f(X)$ is absolutely convergent with probability iff

$$\int_E \min(|f(x)|, 1)\mu(dx) < \infty$$

Furthermore,

$$\mathbb{E}[e^{\theta S}] = e^{\int_E \left(e^{\theta f(x)}-1\right)\mu(dx)}$$

for $\theta \in \mathbb{C}$ such that the integral on the right side of the equation converges, and:

$$\mathbb{E}\left[\sum_{X\in N} f(X)\right] = \int_E f(x)\mu(dx), \quad \text{Var}\left(\sum_{X\in N} f(X)\right) = \int_E f(x)^2\mu(dx)$$

The following example illustrates a more careful setup of a scenario where Campbell's formula is useful.

Example 1 (Simplified Model of the Universe)**.** Consider a very simple classical model of the universe. Denote galaxies as point masses scattered over $E := \mathbb{R}^3$ according to a Poisson process with position-dependent intensity $\lambda(\mathbf{x})$.

The marks $\xi(\mathbf{x})$ of this process correspond to the galaxy masses with density $\rho(m|\mathbf{x})$ for each position $\mathbf{x} \in E$ and infinitesimal mass element $m > 0$. Thus, the augmented Poisson point process N^* is defined on $\mathbb{R}^3 \times (0,\infty)$ with intensity $\lambda^*(\mathbf{x}, m) = \lambda(\mathbf{x})\rho(m|\mathbf{x})$.

The function $f : \mathbb{R}^3 \times (0,\infty) \to \mathbb{R}^3$ that we apply to each point is precisely the gravitational field of the galaxy:

$$f(\mathbf{x}, \xi(\mathbf{x})) = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix} \quad \text{where} \quad F_j := \frac{g\xi(\mathbf{x})x_j}{\|\mathbf{x}\|_2^{\frac{3}{2}}}$$

where g is the gravitational constant.

Then the gravitational field at the origin is given by $\sum_{\mathbf{x}\in N} f(\mathbf{x}, \xi(\mathbf{x}))$. This can be computed via Campbell's formula.

Using the characteristic function

$$\mathbb{E}\left[e^{i\sum_{j=1}^{3} t_j F_j}\right] = \exp\left(\int_0^\infty \int_{\mathbb{R}^3} (e^{igm\langle t,\psi(\mathbf{x})\rangle} - 1)\lambda(\mathbf{x})\rho(m|\mathbf{x})d\mathbf{x}dm\right) \tag{2.6}$$

where ψ is the vector function with components

$$\psi_j(x) = \|\mathbf{x}\|_2^{-3} x_j$$

(2.6) is valid as long as the following condition holds:

$$\int_0^\infty \int_{\mathbb{R}^3} m|\psi(\mathbf{x})|\lambda(\mathbf{x})\rho(m|\mathbf{x})d\mathbf{x}dm < \infty \tag{2.7}$$

Denoting the expected mass of a galaxy at $\mathbf{x}$ by

$$\overline{\xi}(\mathbf{x}) = \int m\rho(m|\mathbf{x})dm$$

(2.7) turns into

$$\int_{\mathbb{R}^3} \frac{\overline{\xi}(\mathbf{x})\lambda(\mathbf{x})}{\|\mathbf{x}\|_2^2} d\mathbf{x} < \infty \tag{2.8}$$

If (2.8) holds, the characteristic function (2.6) can help us determine the mean and covariance of the forces:

$$\mathbb{E}\left[F_j\right] = \int_0^\infty \int_{\mathbb{R}^3} gm\psi_j(\mathbf{x})\lambda(\mathbf{x})\rho(m|\mathbf{x})d\mathbf{x}dm = g\int_{\mathbb{R}^3} \overline{\xi}(\mathbf{x})\psi_j(\mathbf{x})\lambda(\mathbf{x})d\mathbf{x}$$

and

$$\mathrm{Cov}(F_i, F_j) = \int\int (gm\psi_i(\mathbf{x}))(gm\psi_j(\mathbf{x}))\lambda(\mathbf{x})\rho(m|\mathbf{x})d\mathbf{x}dm = g^2\int_{\mathbb{R}^3} \overline{\xi}_2(\mathbf{x})\psi_i(\mathbf{x})\psi_j(\mathbf{x})\lambda(\mathbf{x})d\mathbf{x}$$

where

$$\overline{\xi}_2(\mathbf{x}) = \int_0^\infty m^2\rho(m|\mathbf{x})dm$$

If the universe is uniform (i.e, $\overline{\xi}(x)$ and $\lambda(x)$ are nonzero constants) then (2.8) does not hold and the sum in the definition of each component F_j is not absolutely convergent. $\square$

Lemma 1. If the following properties hold:

$$\int_{[0,T]\times E} |g(t,y)|dt\nu(dy) < \infty, \quad \int_{[0,T]\times E} g^2(t,y)dt\nu(dy) < \infty \tag{2.9}$$

then

$$\mathbb{E}[I_g^2] = \int_{[0,T]\times E} \mathbb{E}[g(t,y)^2]dt\nu(dy) + \left(\int_{[0,T]\times E} \mathbb{E}[g(t,y)]dt\nu(dy)\right)^2 \tag{2.10}$$

Chapter 3

Stochastic Differential Equations

3.1 Basic Setup

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

$$d\mathbf{x}(t) = f(t, \mathbf{x})dt + \sigma(t, \mathbf{x})dW(t) + \xi(t, \mathbf{x})dN(t) \tag{3.1}$$

where

- $f : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$ is a deterministic function in $\mathcal{C}^{(1,2)}$, i.e., f is once-differentiable in time and twice-differentiable in state.
- $\sigma : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^{n\times d}$ is the variation of the white noise which belongs in $\mathcal{C}^{(1,1)}$, while $W : \mathbb{R}^+ \to \mathbb{R}^d$ is a d-dimensional Brownian motion process. Collectively, the $\sigma(t, \mathbf{x})dW(t)$ denotes the additive *white noise* of the system.
- $\xi : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^{n\times \ell}$, which also belongs in $\mathcal{C}^{(1,1)}$ describe the jumps that occur, and $N(t)$ is the ℓ-dimensional standard Poisson process with intensity λ. Collectively, the $\xi(t, \mathbf{x})dN(t)$ denotes the additive *shot noise* of the system.

Note that if $\sigma(t, \mathbf{x}) \equiv 0$, we have the following *shot-noise SDE*

$$d\mathbf{x}(t) = f(t, \mathbf{x})dt + \xi(t, \mathbf{x})dN(t) \tag{3.2}$$

while if $\xi(t, \mathbf{x}) \equiv 0$, we recover the *white-noise SDE*:

$$d\mathbf{x}(t) = f(t, \mathbf{x})dt + \sigma(t, \mathbf{x})dW(t) \tag{3.3}$$

We focus mostly on the shot noise case, before combining it with white noise to discuss the Lévy noise case.

3.2 Additive White Noise

We will use the integral defined in Section 2.1 to consider two important stochastic differential equations (SDEs) and their solutions.

Example 2 (Ornstein-Uhlenbeck Process)**.** The scalar Ornstein-Uhlenbeck (OU) process is written as follows:

$$dX(t) = -\mu X(t)dt + \sigma dW(t)$$

where $\mu, \sigma > 0$ are constants. Initial condition is $X(0) = X_0$.

Solving using integrating factors:

$$d\left(X(t)e^{\mu t}\right) = \sigma dW(t)e^{\mu t} \Longrightarrow X(t) = X_0 e^{-\mu t} + \sigma \int_0^t e^{-\mu(t-s)} dW(s)$$

$\square$

Example 3 (Geometric Brownian Motion)**.** The scalar geometric Brownian motion SDE is written as:

$$dX(t) = rX(t)dt + \alpha X(t)dW(t)$$

where $r, \alpha > 0$ are constants. Initial condition is $X(0) = X_0$.

First, divide across by $X(t)$ and then integrate both sides:

$$d(\ln X(t)) = rdt + \alpha dW(t) \Longrightarrow \ln X(t) = \ln X(0) + rt + \alpha \int_0^t dW(s)$$
$$\Longrightarrow X(t) = X_0 e^{rt+\alpha(W(t)-W(0))} = X_0 e^{rt+\alpha W(t)}$$

$\square$

3.3 Additive Shot Noise

The lemma below is a simple extension of the formula to functions which are both time-varying and dependent on some underlying SDE dynamics $\xi(t, \mathbf{x}(t), y)$.

Lemma 2 (Campbell's Formula)**.** Let $\xi : [0, \infty) \times \mathbb{R}^n \times E \to \mathbb{R}$ be a predictable, locally-bounded Borel-measurable function and $N([0, t] \times E)$ denote the Poisson random measure with intensity λ over the jump space E. If ξ also satisfies the integrability condition $\int_{[0,t]\times E} |\xi(s, \mathbf{x}(s), y)| ds\nu(dy) < \infty$ almost-surely, then

$$\mathbb{E}[I_\xi] = \int_{[0,t]\times E} \mathbb{E}[\xi(s, \mathbf{x}(s), y)] ds\nu(dy) \tag{3.4}$$

Remark 3. In Chapter 5, we make use of Lemma 2 when the Poisson process is standard, i.e., it is a simple counting random measure. Following the construction of (2.5):

$$\int_0^t \xi(s, \mathbf{x}(s)) dN(s) = \sum_{0<t\leq T} \xi(t, \Delta\mathbf{x}(t)) \mathbb{1}\{\Delta N(t) \neq 0\} = \sum_{i=1}^{N(t)} \xi(T_i, \Delta\mathbf{x}(T_i)) \tag{3.5}$$

where $\Delta\mathbf{x}(t) = \mathbf{x}(t) - \mathbf{x}(t-)$, $\Delta N(t) = N(t) - N(t-)$, and the indicator in the second line of the equation determines whether or not a jump occurred at time t. We denote this special case by $N(t)$ without the argument y, since we have essentially isolated the jump as the multiplicative factor $\xi(T_i, \mathbf{x}(T_i))$ for each i; all we need to represent is the time of each arrival and the cumulative number of arrivals until time t. This representation has 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 ξ with respect to it over an interval of time $[0, t]$ would only pick out the values of ξ at $T_i \in [0, t]$. $\square$

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

$$F(t,\mathbf{x}(t)) = F(0,\mathbf{x}_0) + \int_0^t \partial_t F(s,\mathbf{x}(s-))ds + \sum_{i=1}^n \int_0^t \partial_{x_i} F(s,\mathbf{x}(s-))d\mathbf{x}_i^c(s)$$
$$+ \frac{1}{2}\sum_{i,j=1}^n \int_0^t \partial_{x_i}\partial_{x_j} F(s,\mathbf{x}(s-))d[\mathbf{x}_i,\mathbf{x}_j]^c(s) + \sum_{s\le t}\left(F(s,\mathbf{x}(s)) - F(s-,\mathbf{x}(s-))\right) \tag{3.6}$$

where $\mathbf{x}\in\mathbb{R}^n$ and $F\in\mathcal{C}^{(1,2)}$.

A comprehensive treatment of the scalar version of this standard formula can be found in many references (see e.g., Theorem 32 of [7] or Theorem 3.7 of [8]). The specific version of the formula stated is also used in [9, 10]. The term $d[\mathbf{x}_i,\mathbf{x}_j]^c(s)$ in (3.6) 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 4 (Quadratic Variation)**.** Consider two generic scalar SDEs of the form (3.1):

$$dx_i = f(t,x_i)dt + \sigma(t,x_i)dW(t) + \xi(t,x_i)dN(t) \tag{3.7a}$$

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 $d[x_1,x_2]^c(t) = dt$ □

3.4 Existence and Uniqueness of Solutions

The conditions for existence and uniqueness of solutions for (3.3) are the standard ones (see [11]): 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 sys First, we present the well-known Gronwall inequality, a standard result of which can be found in any classical control-theoretic textbook (e.g. [12, 13]).

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:

$$\phi(t) \le \psi(t) + \int_{t_1}^t \rho(s)\phi(s)ds \;\; \forall\, t\in[t_1,t_2] \tag{3.8}$$

Then it follows that

$$\phi(t) \le \psi(t) + \int_{t_1}^t \psi(s)\rho(s)e^{\int_s^t \rho(\tau)d\tau}ds \tag{3.9}$$

Theorem 6 (Existence and Uniqueness for SDE (3.2))**.** 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: $\forall \mathbf{x},\mathbf{y}\in\mathbb{R}^n, t\in[0,T]$,

$$\|f(t,\mathbf{x}) - f(t,\mathbf{y})\| + \|\xi(t,\mathbf{x}) - \xi(t,\mathbf{y})\|_F \le K\|\mathbf{x}-\mathbf{y}\| \tag{3.10}$$

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

$$\|f(t, \mathbf{x})\|^2 + \|\xi(t, \mathbf{x})\|_F^2 \leq C(1 + \|\mathbf{x}\|^2) \tag{3.11}$$

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 (3.2) 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 $s \leq t$ and

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

Remark 5. Theorem 6 was presented in [14] 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 [1] 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 [11], and developed independently of this work. Again, we only focus on the pure shot noise SDE. 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 [1]. There have also been previous work done on describing such conditions for solutions to SDEs of the form (3.2) while imposing different, non-Lipschitz conditions on f and ξ. For instance, [15] 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, [16] 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

$$\mathbf{z}^{(n)}(t) = \mathbf{z}_0^{(n)} + \int_0^t f(s, \mathbf{z}^{(n-1)}(s))ds + \int_0^t \xi(s, \mathbf{z}^{(n-1)}(s))dN(s) \tag{3.12}$$

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

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

$$\mathbf{z}^{(n,m)}(t) = \mathbf{z}_0^{(n,m)} + \int_0^t f^{(n-1,m-1)}(s)ds + \int_0^t \xi^{(n-1,m-1)}(s)dN(s) \tag{3.13}$$

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

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

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

$$\begin{aligned}
&\mathbb{E}\left[\left\|\mathbf{z}^{(n,m)}(t)\right\|^2\right] \leq \mathbb{E}\left[\left(\left\|\mathbf{z}_0^{(n,m)}\right\| + \int_0^t \left\|f^{(n-1,m-1)}(s)\right\| ds + \int_0^t \left\|\xi^{(n-1,m-1)}(s)\right\|_F dN(s)\right)^2\right] \\
&\leq 3\mathbb{E}\left[\left\|\mathbf{z}_0^{(n,m)}\right\|^2\right] + 3\mathbb{E}\left[\int_0^t ds \int_0^t \left\|f^{(n-1,m-1)}(s)\right\|^2 ds\right] + 3\mathbb{E}\left[\left(\int_0^t \left\|\xi^{(n-1,m-1)}(s)\right\|_F dN(s)\right)^2\right]
\end{aligned} \tag{3.14}$$

Note that the Lipschitz bound (3.10) can be squared on both sides:

$$\|f(t,\mathbf{x}) - f(t,\mathbf{y})\|^2 + \|\xi(s,\mathbf{x}) - \xi(s,\mathbf{y})\|_F^2 + 2\|f(t,\mathbf{x}) - f(t,\mathbf{y})\| \|\xi(s,\mathbf{x}) - \xi(s,\mathbf{y})\|_F \le K^2 \|\mathbf{x} - \mathbf{y}\|^2 \tag{3.15}$$

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 (3.15) on the second expectation term yields:

$$\mathbb{E}\left[\int_0^t ds \int_0^t \left\|f^{(n-1,m-1)}(s)\right\|^2 ds\right] \le K^2 t \int_0^t \mathbb{E}\left[\left\|\mathbf{z}^{(n-1,m-1)}(s)\right\|^2\right] ds \tag{3.16}$$

and for the final term, we can apply Lemma 1.

$$\begin{aligned}\mathbb{E}&\left[\left(\int_0^t \left\|\xi^{(n-1,m-1)}(s)\right\|_F dN(s)\right)^2\right] \\ &= \lambda \int_0^t \mathbb{E}\left[\left\|\xi^{(n-1,m-1)}(s)\right\|_F^2\right] ds + \mathbb{E}\left[\left(\lambda \int_0^t \left\|\xi^{(n-1,m-1)}(s)\right\|_F ds\right)^2\right]\end{aligned} \tag{3.17}$$

using the fact that $\lambda := \int_{\{1\}^\ell} \nu(dy)$, as in Remark 3. By Cauchy-Schwarz inequality and the squared Lipschitz bound (3.15):

$$\begin{aligned}(3.17) &\le K^2 \lambda \int_0^t \mathbb{E}\left[\left\|\mathbf{z}^{(n-1,m-1)}(s)\right\|^2\right] ds + \lambda^2 \mathbb{E}\left[\int_0^t ds \int_0^t \left\|\xi^{(n-1,m-1)}(s)\right\|_F^2 ds\right] \\ &\le K^2 \lambda \int_0^t \mathbb{E}\left[\left\|\mathbf{z}^{(n-1,m-1)}(s)\right\|^2\right] ds + K^2 \lambda^2 t \int_0^t \mathbb{E}\left[\left\|\mathbf{z}^{(n-1,m-1)}(s)\right\|^2\right] ds\end{aligned} \tag{3.18}$$

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:

$$\mathbb{E}\left[\left\|\mathbf{z}^{(n)}(t) - \mathbf{z}^{(m)}(t)\right\|^2\right] \le 3K^2(t + \lambda + \lambda^2 t) \int_0^t \mathbb{E}\left[\left\|\mathbf{z}^{(n-1)}(s) - \mathbf{z}^{(m-1)}(s)\right\|^2\right] ds \tag{3.19}$$

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

$$\mathbb{E}\left[\left\|\mathbf{z}^{(k+1)}(t) - \mathbf{z}^{(k)}(t)\right\|^2\right] \le \frac{c^k t^{k+1}}{(k+1)!} \quad \forall k \ge 0, t \in [0,T] \tag{3.20}$$

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 (3.2) 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

$$\mathbb{E}\left[\|\mathbf{x}(t) - \mathbf{y}(t)\|^2\right] \le 3\mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{y}_0\|^2\right] + c\mathbb{E}\left[\int_0^t \|\mathbf{x}(s) - \mathbf{y}(s)\|^2 ds\right] \tag{3.21}$$

By Gronwall's inequality Lemma 3, (3.21) becomes

$$\mathbb{E}\left[\|\mathbf{x}(t) - \mathbf{y}(t)\|^2\right] \le 3\mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{y}_0\|^2\right] e^{ct} \tag{3.22}$$

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. ∎

Remark 6. 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 [17], [18], and [19], 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). □

Chapter 4

Change of Measure Formulas

4.1 Change of Measure for the Brownian Motion Process

A motivation for why a change of measure would be useful in the context of SDEs can be provided by looking at the simple scalar Gaussian random variable.

Example 4 (Scalar Gaussian Distribution)**.** Consider the random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. The probability density can be computed as follows:

$$
\begin{aligned}
\mathbb{P}(\alpha \leq X \leq \beta) &= \int_\alpha^\beta \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_\alpha^\beta \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x^2-2\mu x+\mu^2)} dx \\
&= \int_\alpha^\beta \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2}}_{\mathbb{P}(Y\in[\alpha,\beta]) \text{ where } \mathbb{E}[Y]=0} \cdot e^{-\frac{1}{2\sigma^2}(-2\mu x+\mu^2)} dx
\end{aligned}
$$

Even for general distributions not necessarily Gaussian, we can define

$$Z(x) := e^{-\frac{1}{2\sigma^2}(-2\mu x+\mu^2)}$$

to be the "change of measure" term, more formally known as the **Radon-Nikodym derivative**.

One can think of the probability measure P such that $dP(x) = \mathbb{P}(Y = x)dx$, and Q such that $dQ(x) = Z(x)dP(x)$, or

$$Z(x) = \frac{dQ(x)}{dP(x)} \tag{4.1}$$

The interpretation goes as follows: with respect to measure P, X is normally distributed with mean μ, variance σ^2, and with respect to measure Q, X is normally distributed with mean 0 (same variance σ^2). $\square$

In general, there is a property that P and Q need to satisfy.

Definition 5 (Absolute Continuity)**.** A measure Q is said to be **absolutely continuous** with respect to another measure P (also denoted mathematically by $Q << P$) if the null set of P is also under the null set of Q. That is, for any event set E, if $P(E) = 0$ then $Q(E) = 0$. $\square$

A change of measure can be induced from P to Q via the transformation $Z(x)$. Its interpretation is similar to that of the fundamental theorem of calculus:

$$\forall \text{ Borel set } E, \quad Q[E] = \int_E dQ(x) = \int_E Z(x)dP(x)$$

We can use the following computed expected values as a definition for the change of measure:

$$\mathbb{E}_P[Z(x)] = \int Z(x)dP(x) = \int dQ(x) = \mathbb{E}_Q[1]$$

and for general functions $h \in \mathcal{L}^1$, $\mathbb{E}_P[Z(x)h(x)] = \mathbb{E}_Q[h(x)]$.

Example 5 (Gaussian Random Process)**.** Now we extend the notion introduced in Example 4 to random processes instead of the random variable.

Let $X_0, X_1, \cdots, X_n$ be a collection of random variables at times $t_0 < t_1 < \cdots < t_n$, with $X_0 = 0$, $\Delta t := t_{i+1} - t_i$ and $\Delta X_i := X_i - X_{i-1}$ for all i. This implies $X_i = \sum_{j=0}^{i} \Delta X_j$.

Suppose $\Delta X_i \sim \mathcal{N}(\mu_i \Delta t, \Delta t)$ independent across i. This implies $X_i \sim \mathcal{N}\left(\sum\limits_{j=1}^{i} \mu_j \Delta t, t_i\right)$, but the X_i's are not independent of each other. The conditional distribution of X_i given $X_{i-1} = x_{i-1}$ is normal with mean $x_{i-1} + \mu_i \Delta t$ and variance Δt.

Denote the event

$$E := \{\alpha_1 \leq X_1 \leq \beta_1, \alpha_2 \leq X_2 \leq \beta_2, \cdots, \alpha_n \leq X_n \leq \beta_n\}$$

Then the probability distribution associated with this event is given by a multidimensional Gaussian:

$$\mathbb{P}(E) = \frac{1}{(2\pi\Delta t)^{\frac{n}{2}}} \int_{\alpha_1}^{\beta_1} \cdots \int_{\alpha_n}^{\beta_n} e^{-\frac{1}{2\Delta t}\sum\limits_{i=1}^{n}(x_i - x_{i-1} - \mu_i \Delta t)^2} dx_1 \cdots dx_n \tag{4.2}$$

We will apply the same concept invoked by the scalar case Example 4 to determine the Radon-Nikodym derivative for interval shifts in this random process setting. By expanding the square in each term, the sum in the exponent becomes:

$$\sum_{i=1}^{n}(x_i - x_{i-1} - \mu_i \Delta t)^2 = \sum_{i=1}^{n} \Delta x_i^2 - 2\sum_{i=1}^{n} \Delta x_i \mu_i \Delta t + \sum_{i=1}^{n} \mu_i^2 \Delta t^2$$

Substituting this back into (4.2) yields:

$$\mathbb{P}(E) = \underbrace{\frac{1}{(2\pi\Delta t)^{\frac{n}{2}}} \int_{\alpha_1}^{\beta_1} \cdots \int_{\alpha_n}^{\beta_n} e^{-\frac{1}{2\Delta t}\sum\limits_{i=1}^{n}\Delta x_i^2}} \cdot e^{\sum\limits_{i=1}^{n}\Delta x_i \mu_i - \frac{1}{2}\sum\limits_{i=1}^{n}\mu_i^2 \Delta t}$$

The underbraced term is the probability of event E occurring when there each increment ΔX_i is normally distributed with mean 0 (and variance Δt). Analogously, the Radon-Nikodym derivative is given by:

$$Z_n = e^{\sum\limits_{i=1}^{n} \mu_i \Delta x_i - \frac{1}{2} \sum\limits_{i=1}^{n} \mu_i^2 \Delta t}$$

□

The multidimensional Gaussian distribution described in Example 5 can be thought of as a finite-dimensional, discrete stochastic process. This is easy to extend to the usual definition of a stochastic random process, namely one that is infinite-dimensional (continuous with time). We summarize the analogy as follows, then use invoke **Lévy's characterization of the Brownian motion** to show that the transformed white noise process is indeed a Brownian motion under the new measure Q.

The two analogies are below:

1. Discrete-Time Case: We begin with $X_n = \sum_{i=1}^{n} \Delta X_i \sim \mathcal{N}[0, t_n]$ under measure P. The shift term is given by

$$Y_n = \sum_{i=1}^{n} \mu_i \Delta t$$

which corresponds to the Radon-Nikodym derivative

$$Z_n = \frac{dQ}{dP} = e^{\sum\limits_{i=1}^{n} \mu_i \Delta x_i - \frac{1}{2} \sum\limits_{i=1}^{n} \mu_i^2 \Delta t}$$

This causes the shift to $X_n - \sum_{i=1}^{n} \mu_i \Delta t \sim \mathcal{N}(0, t_n)$ under new measure Q.

2. Continuous-Time Case: We begin with $W(t)$, the standard Brownian motion process, under measure P. The shift term is given by

$$Y(t) = \int_0^t \mu(s) ds$$

which corresponds to the Radon-Nikodym derivative

$$Z(t) = \frac{dQ}{dP} = e^{\int_0^t \mu(s) dW(s) - \frac{1}{2} \int_0^t \mu^2(s) ds} \tag{4.3}$$

This causes the shift to $\tilde{W}(t) := W(t) - \int_0^t \mu(s) ds$ under new measure Q. Furthermore, $Z(t)$ for this Brownian motion case is commonly referred to as the **Doléans-Dade exponential**. A further treatment of the continuous-time case can be found in [11].

How does this fit within the context of Brownian motion SDEs? Consider the following example.

Example 6 (Drift-to-Drift Transformation)**.** Consider the following example transformations of the Brownian motion process. Let $\sigma(t) \not\equiv 0$ for all t.

1. Suppose we want to transform a drift-perturbed Brownian motion process to a process without drift.

$$dX(t) = \mu(t) dt + \sigma(t) dW(t) = \sigma(t) \left(\frac{\mu}{\sigma(t)} dt + dW(t) \right)$$

We want to find a $\tilde{W}(t)$ such that $dX(t) = \sigma(t)d\tilde{W}(t)$.

Choose the shift term $Y(t) = -(\mu(t)/\sigma(t))t$ so that $\tilde{W}(t) = W(t) + (\mu(t)/\sigma(t))t$. Then:

$$\tilde{Z}(t) = \frac{dQ(t)}{dP(t)} = e^{-\frac{\mu(t)}{\sigma(t)}W(t) - \frac{1}{2}\left(\frac{\mu(t)}{\sigma(t)}\right)^2 t}$$

2. Now suppose we want to transform a zero-drift Brownian motion process to one with nonzero drift.

$$dX(t) = \sigma(t)\tilde{W}(t) = \nu(t)dt + (-\nu(t)dt + \sigma(t)\tilde{W}(t)) = \nu(t)dt + \sigma(t)\left(-\frac{\nu(t)}{\sigma(t)}dt + \tilde{W}(t)\right)$$

We want to find $W'(t)$ such that $dX(t) = \nu(t)dt + \sigma(t)dW'(t)$.

Choose the shift term $\tilde{Y}(t) = (\nu(t)/\sigma(t))t$ so that $W'(t) = \tilde{W}(t) - (\nu/\sigma)t$. Then:

$$Z'(t) = \frac{dR(t)}{dQ(t)} = e^{\frac{\nu(t)}{\sigma(t)}\tilde{W}(t) - \frac{1}{2}\left(\frac{\nu(t)}{\sigma(t)}\right)^2 t}$$

Overall, the change of measure from drift $\mu(t)$ to drift $\nu(t)$ is performed using the Radon-Nikodym derivative

$$\frac{dR(t)}{dP(t)} = \frac{\cancel{dQ(t)}}{dP(t)} \cdot \frac{dR(t)}{\cancel{dQ(t)}} = \tilde{Z}(t)Z'(t)$$

with

$$dW'(t) = dW(t) - \frac{\nu(t) - \mu(t)}{\sigma(t)}dt$$

□

Now, some formalities. We want to show that $\tilde{W}(t) := W(t) - \mu(t)$ is a Brownian motion under the measure Q. First, we need the following proposition.

Proposition 4. The Doléans-Dade exponential $Z(t)$ (4.3) is a martingale.

Proof. We will take the derivative and show that it is a white-noise stochastic process without drift. Define

$$X(t) = -\int_0^t \mu(s)dW(s), \quad Y(t) = -\frac{1}{2}\int_0^t \mu^2(s)ds$$

so that $Z(t) = e^{X(t)+Y(t)}$.

Applying the Itô formula to $Z(t)$ yields:

$$\begin{aligned}
dZ(t) &= Z(t)dX(t) + Z(t)dY(t) + \frac{1}{2}d[X,X](t)Z(t) + \frac{1}{2}Z(t)d[Y,Y](t) + Z(t)d[X,Y](t) \\
&= -Z(t)\mu(t)dW(t) - \frac{1}{2}Z(t)\mu^2(t)dt + \frac{1}{2}\mu^2(t)dt + 0 + 0 \\
&= -Z(t)\mu(t)dW(t)
\end{aligned}$$

Since $dZ(t)$ is a multiple of $dW(t)$ with no additional drift term, it follows that $Z(t)$ is a local martingale. ■

In addition to the above proposition, we will invoke **Lévy's Characterization of the Brownian motion** to show that $\tilde{W}(t)$ is a martingale. The characterization consists of three parts:

1. $\mathbb{P}(\tilde{W}(0) = 0) = 1$
2. $\tilde{W}(t)$ is a martingale with continuous sample paths
3. $\tilde{W}^2(t) - t$ is a martingale.

For simplicity of notation, we will verify for the simple linear case where $\mu(t) = \mu t$. The first property is obvious to see upon substituting $t = 0$. For the second and third properties:

2. Want to show $\mathbb{E}_Q[\tilde{W}(t)|\mathcal{F}_s] = \tilde{W}(s)$. From the left side, we can apply Bayes' theorem to get

$$\mathbb{E}_Q[\tilde{W}(t)|\mathcal{F}_s] = \frac{\mathbb{E}_P[Z(t)\tilde{W}(t)|\mathcal{F}_s]}{\mathbb{E}_P[Z(t)|\mathcal{F}_s]} \tag{4.4}$$

   In the denominator, we have $\mathbb{E}_P[Z(t)|\mathcal{F}_s] = Z(s)$ because $Z(t)$ is a martingale. It is only left to show that the numerator $\mathbb{E}_P[Z(t)\tilde{W}(t)|\mathcal{F}_s]$ is a martingale.

   By the Itô product rule

$$\begin{aligned} d(Z\tilde{W}) &= Zd\tilde{W} + \tilde{W}dZ + dZd\tilde{W} \\ &= Z(dW - \mu dt) + (W - \mu t)\mu ZdW + \mu ZdW(dW - \mu dt) \\ &= ZdW - \mu Zdt + \mu Z(W - \mu t)dW + \mu Zdt - \underbrace{\mu^2 ZdW \cdot dt}_{=0} \\ &= (Z + \mu Z(W - \mu t))dW \end{aligned}$$

   Note that the expression is purely in terms of dW(t), i.e, there is no drift. The expected value of this stochastic integral is therefore equal to 0. We get:

$$(4.4) = \frac{Z(s)\tilde{W}_s}{Z(s)} = \tilde{W}(s)$$

3. The proof follows in much the same way: apply Bayes' theorem, then use Itô's product rule to show the numerator is a martingale.

$$\mathbb{E}_Q[\tilde{W}^2(t) - t|\mathcal{F}_s] = \frac{\mathbb{E}_P[Z(t)(\tilde{W}^2(t) - t)|\mathcal{F}_s]}{\mathbb{E}_P[Z(t)|\mathcal{F}_s]} \tag{4.5}$$

   Since the denominator is a martingale, we just need to look at the numerator. First, recall that:

$$d(\tilde{W}^2 - t) = (2\tilde{W}d\tilde{W} + dt) - dt = 2(W - \mu t)dW - 2\mu(W - \mu t)dt$$

   Then:

$$\begin{aligned} d(Z(\tilde{W}^2 - t)) &= dZ(\tilde{W}^2 - t) + Z(2(W - \mu t)dW - 2\mu(W - \mu t)dt) \\ &\qquad + dZ(2(W - \mu t)dW - 2\mu(W - \mu t)dt) \\ &= (\mu Z(\tilde{W}^2 - t) + 2Z(W - \mu t))dW \end{aligned}$$

   which is again purely stochastic. Thus:

$$(4.5) = \frac{Z(s)(\tilde{W_s}^2 - s)}{Z(s)} = \tilde{W_s}^2 - s$$

Recall that the solution to the equation $dZ(t) = Z(t)dt, Z(0) = 1$ is given by the exponential $Z(t) = e^t$. More generally, for the deterministic function $X(t)$, the solution to the system $dZ(t) = Z(t)dX(t), Z(0) = 1$ is given by $Z(t) = e^{(X(t)-X(0))}$. The Doléans-Dade exponential is the analogous result when $X(t)$ is stochastic, and we've seen that for the Brownian motion case $dZ(t) = \mu Z(t)dW(t), Z(0) = 1$, (4.3) is the solution. We now present the formula of the solution to $dZ(t) = Z(t)dX(t), Z(0) = 1$ when $X(t)$ is a semimartingale.

Proposition 5. The Doléans-Dade exponential for semimartingale X is given by:

$$Z(t) = e^{X(t)-\frac{1}{2}[X,X](t)} \prod_{0<s\leq t} (1+\Delta X(s))e^{-\Delta X(s)+\frac{1}{2}\Delta X^2(s)} \tag{4.6}$$

Example 7. Before we prove the proposition, we apply the formula (4.6) for the simple constant linear semimartinagle $X(t) = at + bW(t) + cN(t)$.

$$Z(t) = e^{at+bW(t)+cN(t)-\frac{1}{2}b^2t-\frac{1}{2}\sum_{s\leq t} c^2\mathbb{1}\{N(s)=1\}} \cdot \prod_{s\leq t} \left(1+c\mathbb{1}\{N(s)=1\}\right) e^{(-c+\frac{1}{2}c^2)\mathbb{1}\{N(s)=1\}}$$

Note that the exponential term $e^{\frac{1}{2}\sum_{s\leq t} c^2\mathbb{1}\{N(s)=1\}}$ cancels out, and we get:

$$Z(t) = e^{at+bW(t)+cN(t)-\frac{1}{2}b^2t} \cdot \prod_{s\leq t} \left(1+c\mathbb{1}\{N(s)=1\}\right) e^{-c\mathbb{1}\{N(s)=1\}}$$

□

This simplification occurs for the general formula (4.6) too:

$$Z(t) = e^{X(t)-\frac{1}{2}[X,X]^c(t)} \prod_{0<s\leq t} (1+\Delta X(s))e^{-\Delta X(s)}$$

Proof of Proposition 5. The system $\Delta Z_n = Z_{n-1}\Delta X_n, Z_0 = 1$, with $\Delta Z_n := Z_n - Z_{n-1}$ can be recursively substituted backwards until the base case. We get:

$$Z_n = (1+\Delta X_n)Z_{n-1} = \prod_{i=1}^{n}(1+\Delta X_i)$$

This is easily extendable to the continuous-time case:

$$Z(t) = (1+\Delta X(t))Z(t-) \text{ for all } t > 0$$

Apply Itô's formula for semimartingales:

$$d\left(\ln(Z(t))\right) = \frac{dZ(t)}{Z(t)} - \frac{1}{2Z^2(t)}d[Z,Z]^c(t) + \sum 0 < s \leq t \left[\ln(Z(s)) - \ln(Z(s-)) - \frac{\Delta Z(s)}{Z(s-)}\right] \tag{4.7}$$

where $\Delta Z(s) := Z(s) - Z(s-)$.

We have $d[Z,Z]^c(t) = Z^2 d[X,X]^c(t)$ and

$$\ln(Z(s)) - \ln(Z(s-)) = \ln\left(\frac{Z(s-)(1+\Delta X(s))}{Z(s-)}\right) = \ln(1+\Delta X(s))$$
$$\frac{\Delta Z(s)}{Z(s-)} = \frac{Z(s-)\Delta X(s)}{Z(s-)} = \Delta X(s)$$

Substituting, we get:

$$\begin{aligned}(4.7) &= dX(t) - \frac{1}{2}d[X,X]^c(t) + \sum 0 < s \le t\,[\ln(1+\Delta X(s)) - \Delta X(s)] \\ \implies Z(t) &= e^{X(t)-\frac{1}{2}[X,X]^c(t)} \prod_{0<s\le t} (1+\Delta X(s))e^{-\Delta X(s)}\end{aligned}$$

which is exactly our desired formula. ∎

4.2 Change of Measure for the Poisson Process

We present the change of probability formula for the standard Poisson process. It is easy to verify that the Radon-Nikodym derivative is expressed as

$$L_t = e^{t(\lambda-\tilde{\lambda})}\left(\frac{\tilde{\lambda}}{\lambda}\right)^{N(t)}$$

by writing out the probability distributions:

$$\tilde{\mathbb{P}}(N(t)=k) = e^{-t(\tilde{\lambda}-\lambda)}\left(\frac{\tilde{\lambda}}{\lambda}\right)^k \mathbb{P}(N(t)=k) = e^{-t(\tilde{\lambda}-\lambda)}\left(\frac{\tilde{\lambda}}{\lambda}\right)^k \cdot e^{-\lambda t}\frac{(\lambda t)^k}{k!} = e^{-\tilde{\lambda}}\frac{\tilde{\lambda}^k t^k}{k!}$$

Thus, $N(t)$ is also Poisson under the measure $\tilde{\mathbb{P}}$, just with intensity $\tilde{\lambda}$ as opposed to λ.

Now consider the compound Poisson process Y_t, written as follows:

$$Y_t = \sum_{i=1}^{N(t)} \xi_i$$

where $N(t)$ is the standard Poisson process of intensity λ, and the jump sizes ξ_i are iid random variables. Suppose this is the construction of the compound Poisson process under probability measure P, with $\lambda = \nu(\mathbb{R}) > 0$.

If is of great use to compute the moment-generating function of the increment $Y(T) - Y(t)$. Given that $\{N(T) = n\}$, the jump sizes of $Y(t)$ on $[0,T]$ are independent random variables which are distributed on $\mathbb{R}$ according to $\nu(dx)$. Based on this, we get that for any $t \in [0,T]$, we have

$$\mathbb{E}\left[e^{\alpha(Y_T-Y_t)}\right] = e^{\left(\lambda(T-t)\int_{-\infty}^{\infty}(e^{\alpha y}-1)\nu(dy)\right)}, \quad \alpha \in \mathbb{R}$$

This can be seen through the following argument. Since $N(t)$ has a Poisson distribution with parameter $t > 0$ and is independent of $\{\xi_k\}$, we get

$$\begin{aligned}\mathbb{E}\left[e^{\alpha(Y_T-Y_t)}\right] &= \mathbb{E}\left[e^{\left(\alpha \sum_{k=N(t)+1}^{N(T)} \xi_k\right)}\right] \\ &= \mathbb{E}\left[e^{\left(\alpha \sum_{k=1}^{N(T)-N(t)} \xi_k\right)}\right] \quad \text{since } \xi_k \text{ iid} \\ &= \sum_{n=0}^{\infty} \mathbb{E}\left[e^{\left(\alpha \sum_{k=1}^{n} \xi_k\right)} \middle| N(T)-N(t)=n\right] \mathbb{P}(N(T)-N(t)=n) \qquad (4.8)\end{aligned}$$

where the last equality follows by definition of expectation. From the fact that the Poisson process has stationary increments, (4.8) becomes:

$$\begin{aligned}
& e^{-\lambda(T-t)} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}(T-t)^n \mathbb{E}\left[e^{\left(\alpha \sum_{k=1}^{N(T)-N(t)} \xi_k\right)} \right] \\
&= e^{-\lambda(T-t)} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}(T-t)^n \prod_{k=1}^{n} \mathbb{E}\left[e^{\alpha \xi_k}\right] \quad \text{since } \xi_k \text{ iid} \\
&= e^{-\lambda(T-t)} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}(T-t)^n \mathbb{E}\left[e^{\alpha \xi_k}\right]^n \quad \text{again } \xi_k \text{ iid} \\
&= e^{-\lambda(T-t)} e^{\lambda(T-t)\mathbb{E}\left[e^{\alpha\xi_k}\right]} \quad \text{by Taylor expansion} \\
&= e^{\lambda(T-t)\left(\mathbb{E}\left[e^{\alpha\xi_k}\right]-1\right)} \\
&= e^{\lambda(T-t)\int_{-\infty}^{\infty} e^{\alpha y}\nu(dy) - \lambda(T-t)\int_{-\infty}^{\infty}\nu(dy)} \\
&= e^{\lambda(T-t)\int_{-\infty}^{\infty}(e^{\alpha y}\nu(dy)-1)\nu(dy)}
\end{aligned}$$

where the second-to-last equality comes from the fact that the probability distribution $\nu(dy)$ of ξ satisfies

$$\mathbb{E}[e^{\alpha\xi}] = \int_{-\infty}^{\infty} e^{\alpha y}\nu(dy) \quad \text{and} \quad \int_{-\infty}^{\infty} \nu(dy) = 1$$

Now we can use the moment generating function as an alternative way (aside from Wald's Equality) to compute the expected value of $Y(t)$:

$$\mathbb{E}[Y(t)] = \frac{\partial}{\partial\alpha}\mathbb{E}\left[e^{\alpha Y(t)}\right]\Big|_{\alpha=0} = \lambda t \int_{-\infty}^{\infty} y\nu(dy) = \mathbb{E}[N(t)]\mathbb{E}[\xi] = \lambda t \mathbb{E}[\xi]$$

Theorem 7. Consider an alternative probability measure Q under which Y_t is a compound Poisson process with intensity $\tilde{\lambda} = \tilde{\nu}(\mathbb{R}) > 0$. Denote $d\nu = \lambda\nu(dx)$ and $d\tilde{\nu} = \tilde{\lambda}\tilde{\nu}(dx)$.

Define the factor

$$L_t = e^{t(\lambda-\tilde{\lambda})} \left(\frac{\tilde{\lambda}}{\lambda}\right)^{\sum_{k=1}^{N(t)} \xi_k}$$

Then the change of probability measure is given by $\frac{dQ}{dP}\Big|_{\mathcal{F}_t} = L_t$ for a given filtration $\mathcal{F}_t$.

Before we prove this theorem, we must first show that L_t is a martingale. Write the integral of some function $f \in \mathcal{L}^1$ with respect to the standard Poisson process $N(t)$ in a way similar to before. Similar to the notion of the Poisson random measure in Definition 1, we specifically define N as

$$N(\omega, [0,t] \times E) = N(dt, dx) = \sum_{k=1}^{N(t)(\omega)} \mathbb{1}[Y_k(\omega) \in E]$$

from which

$$\int_0^t \int_{\mathbb{R}} f(x)N(dt,dx) = \int_{\mathbb{R}} f(x)N(t)(dx) = \int_0^t \int_{\mathbb{R}} f(x) \sum_{k=1}^{N(t)} f(Y_k)$$

Lemma 4. The process

$$\int_0^t \int_{\mathbb{R}} f(x)(N(ds,dx) - \nu(dx)ds)$$

is a martingale. This follows easily from the fact that $N(ds,dx) - \nu(dx)ds$ is a martingale.

This further follows that

$$\exp\left(\sum_{k=1}^{N(t)} f(Y_k) - t\int_{-\infty}^{\infty}(e^{f(x)}-1)\nu(dx)\right)$$
$$= \exp\left(\int_0^t \int_{\mathbb{R}} f(x)N(ds.dx) - t\int_{-\infty}^{\infty}(e^{f(x)}-1)\nu(dx)\right)$$

is a martingale. Now to see that L_t is a martingale, one can simply substitute the function $f = \ln\left(\frac{d\tilde{\nu}}{d\nu}\right)$ into the above:

$$\exp\left(\sum_{k=1}^{N(t)} \ln\left(\frac{d\tilde{\nu}}{d\nu}\right)(\Delta X_s) - t\int_{-\infty}^{\infty}(e^{\ln\left(\frac{d\tilde{\nu}}{d\nu}\right)(x)}-1)\nu(dx)\right)$$
$$= \exp\left(\sum_{k=1}^{N(t)} \ln\left(\frac{d\tilde{\nu}}{d\nu}\right)(\Delta X_s) - \underbrace{t\int_{-\infty}^{\infty}\left(\frac{\tilde{\lambda}\tilde{\nu}(dx)}{\lambda\nu(dx)}\right)}_{\frac{t}{\lambda}\int_{-\infty}^{\infty}\tilde{\lambda}\tilde{\nu}(dx)-\lambda\nu(dx)}\right)$$

Now we will prove Theorem 7:

Proof. We begin by writing the characteristic function:

$$\mathbb{E}_Q(e^{iuY_t}) = \mathbb{E}_P\left[e^{iuY_t}\cdot e^{t(\lambda-\tilde{\lambda})}\left(\frac{\tilde{\lambda}}{\lambda}\right)^{\sum_{k=1}^{N(t)}\xi_k}\right] = e^{t(\lambda-\tilde{\lambda})}\mathbb{E}_P\left[e^{iu\sum_{k=1}^{N(t)}\xi_k}\left(\frac{\tilde{\lambda}}{\lambda}\right)^{\sum_{k=1}^{N(t)}\xi_k}\right]$$
$$= e^{t(\lambda-\tilde{\lambda})}\mathbb{E}_P\left[\left(e^{iu\xi_1}\left(\frac{\tilde{\lambda}}{\lambda}\right)^{\xi_1}\right)^{N(t)}\right] \tag{4.9}$$

where the last equality follows from the distribution of ξ_k. We can use the expected value of $N(t)$ to rewrite (4.9) as:

$$\sum_{n=0}^{\infty} e^{t(\lambda-\tilde{\lambda})}\mathbb{E}_P\left[\left(e^{iu\xi_1}\left(\frac{\tilde{\lambda}}{\lambda}\right)^{\xi_1}\right)\right]^n\cdot e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$
$$= e^{-\tilde{\lambda}t}\frac{(\lambda t)^n}{n!}\sum_{n=0}^{\infty}\mathbb{E}_P\left[\left(e^{iu}\frac{\tilde{\lambda}}{\lambda}\right)^{\xi_1}\right]^n$$
$$= e^{-\tilde{\lambda}t}\sum_{n=0}^{\infty}\frac{1}{n!}\left(\lambda t\mathbb{E}_P\left[\left(e^{iu}\frac{\tilde{\lambda}}{\lambda}\right)^{\xi_1}\right]\right)^n$$
$$= e^{-\tilde{\lambda}t}\cdot e^{\lambda t\mathbb{E}_P\left[\left(e^{iu}\frac{\tilde{\lambda}}{\lambda}\right)^{\xi_1}\right]} \quad \text{by Taylor expansion}$$

We have

$$
\begin{aligned}
&\mathbb{E}_P\left[\left(e^{iu}\cdot\frac{\tilde{\lambda}}{\lambda}\right)^{\xi_1}\right] = \mathbb{E}_P\left[\left(e^{iu+\ln\frac{\tilde{\lambda}}{\lambda}}\right)^{\xi_1}\right] \\
&= \int_{\mathbb{R}} \left(e^{iu+\ln\frac{\tilde{\lambda}}{\lambda}}\right)^{x} \nu(dx) \ \text{ take expectation over } \xi_1 \\
&= \int_{\mathbb{R}} e^{iux} e^{\ln\frac{\tilde{\lambda}}{\lambda}x} \nu(dx)
\end{aligned} \tag{4.10}
$$

Note that

$$
e^{\ln\left(\frac{\tilde{\lambda}}{\lambda}x\right)} = \frac{d\tilde{\nu}}{d\nu}(x) \implies (4.10) = \int_{\mathbb{R}} e^{iux}\cdot\frac{1}{\lambda}d\tilde{\nu}(x)
$$

Substituting this back into (4.9) yields:

$$
\mathbb{E}_Q[e^{iuY_t}] = e^{-\tilde{\lambda}t}e^{\lambda t\cdot\frac{1}{\lambda}\int_{\mathbb{R}}e^{iux}d\tilde{\nu}(x)} = e^{t\int_{\mathbb{R}}e^{iux}d\tilde{\nu}(x)-\tilde{\lambda}} = e^{t\int_{\mathbb{R}}(e^{iux}-1)d\tilde{\nu}(x)}
$$

Characteristic function for the Poisson is:

$$
\mathbb{E}[e^{iuN(t)}] = \sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)}{n!}\cdot e^{iun} = e^{-\lambda t}\sum_{n=0}^{\infty}\frac{1}{n!}(e^{iu}\lambda t)^n = e^{-\lambda t}e^{e^{iu}\lambda t} = e^{\lambda t(e^{iu}-1)}
$$

So indeed, under Q, it is a compound Poisson process with intensity $\tilde{\lambda}$. ■

Chapter 5

Stochastic Contraction Theorems

In this chapter, we present the two main results of this paper. In Sec. 5.3, we state Theorem 9, a stochastic contraction criterion for shot noise system (3.2), then prove it using the path integral Lyapunov-like function form (5.5). We also draw attention to a tradeoff between the intensity λ of the shot noise and the size of the steady-state error ball, which allows our derived stability criteria to be interpreted as a condition on the class shot noise process with mean interarrival time large enough for the system to converge within the bounded error ball of each other between consecutive jumps. In Sec. 5.4, the stochastic contraction theorem for Lévy noise systems is presented and proven via combination of the white and the shot noise cases. However, in order to take advantage of the affine representation of Lévy noise prescribed by the Lévy-Khintchine Decomposition Theorem, we need a derivation of the white noise case which is different from what was presented in the previous literature [20, 21]. Namely, the path integral Lyapunov-like function (5.5) is used instead of (5.7), and a noise-perturbed trajectory is compared against an unperturbed one. Appropriate modifications to the white noise case are made before presenting the Lévy noise case in combination. Before we begin, we provide a brief summary of past related work.

Lyapunov-Sense Stability: Traditional characterizations of deterministic and stochastic system stability are defined in the Lyapunov sense [22, 23]. For deterministic systems, we have the well-known direct and indirect Lyapunov methods, which can be found in any standard control theory textbook (e.g. [12, 13]). For white-noise stochastic systems, sufficient stability conditions can be derived for nonlinear systems [24] or systems which are affine in control input [25]. Although Gaussian white noise is the most prevalent model for stochastic dynamics, there have also been works on characterizing stability for stochastic systems perturbed by non-Gaussian noise beyond the seminal work of [23], which laid out the foundations of Lyapunov-based stochastic stability theory. Asymptotic stability of systems driven by Lévy noise is developed in [10] and exponential stability is studied in [26].

Incremental Stability: Formally defined in Definition 6, incremental stability generalizes Lyapunov-sense stability be considering the convergence of solution trajectories towards some desired trajectory rather than an equilibrium point or a limit cycle. Systems that satisfy the incremental stability property have guaranteed global exponential convergence towards the desired trajectory. Applications of incremental stability arise in numerous settings such as cooperative control over multi-agent swarm systems [27] and phase synchronization in directed networks [28, 29]. For deterministic systems, there has been an extensive amount of work characterizing incremental stability for nonlinear dynamics [30, 31]. The work of [20] took the first step in extending incremental stability to Gaussian white noise stochastic systems, and [21] extended this theory to more general state-dependent metrics, which is useful in the construction of nonlinear observers or controllers. However, to the authors' knowledge, incremental stability for dynamics perturbed by non-Gaussian noise processes has not been considered; this paper intends to take the first step in doing so for

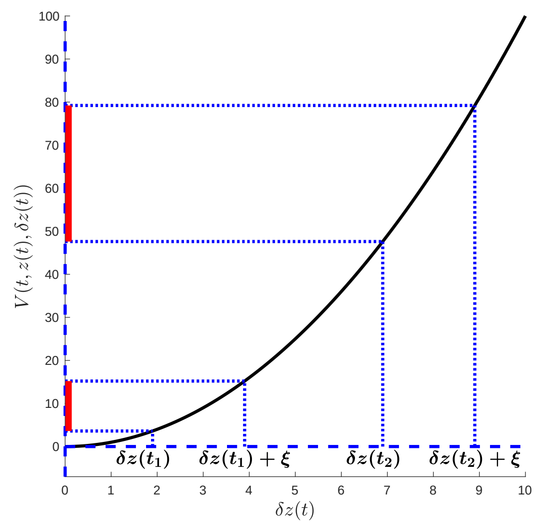

Figure 5.1: Scalar example of the norm-squared Lyapunov function (5.7) violating (5.23). Although the height of the jumps, indicated on the x-axis, are η at both times t_1 and t_2, the difference of $V(t, z(t), \delta z(t))$ at time t_2 is larger than the difference at time t_1, as shown by the red lines on the y-axis. Thus, there is no way to bound the squared difference.

shot and Lévy noise systems.

Hybrid Systems: We emphasize a poignant distinction between the literature of hybrid or jump-Markov systems versus our considered setting of systems perturbed by a general class of jump-discontinuous noise processes. In shot or Lévy noise systems, the jumps arise solely from the noise process, independently of the open-loop dynamics, unlike hybrid systems where the switches (i.e., jumps) arise as an inherent property in the open-loop dynamics. Despite this important distinction, the two settings can still be closely related to one another through two concepts. First, stability analysis techniques are primarily focused on handling the jump-discontinuities of the system. Literature towards this direction of research for hybrid systems include Lyapunov-sense conditions for asymptotic stability [32, 33] and characterizations of incremental stability was studied in [34]. Second, *dwell time* can be related to the interarrival time by viewing it as a form of stability criterion which ensures that the system has sufficient time to converge to within a bounded error ball of a desired state in between consecutive switching/jumping phases. One notable example which utilizes dwell-time criterion for nonlinear systems is in [35], where it is shown that input-to-state induced norms should be bounded uniformly between switches. In terms of applications, dwell-time criterion for attaining exponential stability has been shown to be effective for robotic systems, in particular walking locomotion and flapping flight [36] as well as autonomous vehicle steering [37].

5.1 Previous Results on Notions of Stability

Definition 6 (Incremental Stability)**.** For any nonlinear function $f \in \mathcal{C}^{(1,2)}$, the deterministic, noiseless system $\dot{\mathbf{x}} = f(t, \mathbf{x})$ is said to be *incrementally (globally exponentially) stable* if there exist constants $\kappa, \alpha > 0$ such that

$$\|\mathbf{x}(t) - \mathbf{y}(t)\|_m \leq \kappa \|\mathbf{x}_0 - \mathbf{y}_0\|_m e^{-\alpha t} \tag{5.1}$$

for any norm $m \geq 1$, all $t \geq 0$, and all solution trajectories $\mathbf{x}(t)$ and $\mathbf{y}(t)$ of the system with respective initial conditions $\mathbf{x}_0$ and $\mathbf{y}_0$. We assume $\mathbf{x}_0 \neq \mathbf{y}_0$, otherwise the two trajectories are exactly the same for all t and (5.1) is trivially satisfied with equality. $\square$

We denote $\delta \mathbf{z}$ to be the infinitesimal displacement length between the two trajectories, and represent it as a path integral by parametrizing it using a measure parameter $\mu \in [0,1]$:

$$\mathbf{y}(t) - \mathbf{x}(t) = \int_{\mathbf{x}}^{\mathbf{y}} \delta \mathbf{z} = \int_0^1 \left(\frac{\partial \mathbf{z}}{\partial \mu} \right) d\mu \Longrightarrow \| \mathbf{y}(t) - \mathbf{x}(t) \| \leq \int_{\mathbf{x}}^{\mathbf{y}} \| \delta \mathbf{z} \| = \int_0^1 \left\| \frac{\partial \mathbf{z}}{\partial \mu} \right\| d\mu \tag{5.2}$$

Its rate of change can be approximated by the dynamics $\delta \dot{\mathbf{z}} = (\partial_{\mathbf{x}} f) \delta \mathbf{z}$, where $\partial_{\mathbf{x}} f$ denotes the derivative of $f(t, \mathbf{x})$ being taken with respect to the second argument. Oftentimes, performing a differential coordinate transform from $\delta \mathbf{z}$ to $\Theta \delta \mathbf{z}$, where $\Theta \in \mathbb{R}^{n \times n}$, makes it easier to verify (5.1).

The transformed virtual dynamics become

$$(d/dt) \delta \mathbf{z} = F \delta \mathbf{z}, \quad F := (\dot{\Theta} + \Theta(\partial_{\mathbf{x}} f)) \Theta^{-1} \tag{5.3}$$

In many cases, $\Theta := \Theta(t, \mathbf{x})$ is dependent on time and state and we can assume the existence of bounds:

$$\underline{h} = \inf_{t, \mathbf{x}} \lambda_{\min}(\Theta), \;\; \overline{h} = \sup_{t, \mathbf{x}} \lambda_{\max}(\Theta), \;\; h' = \sup_{t, \mathbf{x}, i} \| (\partial_x \Theta)_i \|, h'' = \sup_{t, \mathbf{x}, i, j} \left\| (\partial_x^2 \Theta)_{ij} \right\| \tag{5.4}$$

Definition 7 (Basic Contraction)**.** An equivalent way to say that a system is incrementally stable is to say that it is *contracting* with respect to a differential coordinate transform $\Theta(t, \mathbf{x})$ and convergence rate $\alpha > 0$. For instance, the inequality (5.1) implies that the system $\dot{\mathbf{x}}(t) = f(t, \mathbf{x})$ is *contracting* under $\Theta(t, \mathbf{x}) = I$ with rate α. But instead of using Definition 6 directly, it is oftentimes easier to prove that a system is contracting by analyzing the time-derivative of a choice of Lyapunov-like function, and invoking the Comparison Lemma (see Sec. 9.3 of [12]). For deterministic systems, one common choice is the modified path length:

$$V(t, \mathbf{z}, \delta \mathbf{z}) = \int_0^1 \| \Theta(t, \mathbf{z}(\mu, t)) \delta \mathbf{z} \| \, d\mu \tag{5.5}$$

If there exists a $\Theta(t, \mathbf{x})$ and $\alpha > 0$ such that the following condition is satisfied:

$$\dot{\Theta}(t, \mathbf{x}) + \Theta(t, \mathbf{x}) \left(\frac{\partial f}{\partial \mathbf{x}} \right) \leq -\alpha \Theta(t, \mathbf{x}) \tag{5.6}$$

then the system is contracting. Another common choice is the norm-squared Lyapunov-like function:

$$V(t, \mathbf{z}, \delta \mathbf{z}) = \int_0^1 \delta \mathbf{z}^T S(t, \mathbf{z}(\mu, t)) \delta \mathbf{z} d\mu \tag{5.7}$$

where $S(t, \mathbf{x}) := \Theta^T \Theta(t, \mathbf{x})$. If there exists such a $S(t, \mathbf{x})$ and $\alpha > 0$ so that the following condition is satisfied:

$$\left(\frac{\partial f}{\partial \mathbf{x}} \right)^T S(t, \mathbf{x}) + S(t, \mathbf{x}) \left(\frac{\partial f}{\partial \mathbf{x}} \right) + \dot{S}(t, \mathbf{x}) \leq -2\alpha S(t, \mathbf{x}) \tag{5.8}$$

then the system is contracting. □

Remark 7. The two conditions (5.6) and (5.8) are exactly the same: both imply that all system trajectories converge globally exponentially to a single trajectory with a convergence rate equal to α. Yet, there is a very specific reason why we chose to distinguish between the two forms. Standard contraction results for deterministic systems, e.g. Definition 2 of [38], present contraction in terms of only (5.8), derived from using the Lyapunov-like function (5.7). As shown in Theorem 3 of [20] and Lemma 2 of [21], (5.7) can also be used to prove contraction in white noise systems (3.3). However, for shot noise SDEs (3.2), such a Lyapunov-like function cannot be used to prove exponential convergence due to (5.23): Fig. 5.1 intuitively shows how the square of differences can be unbounded in the simple 1D case with identity metric. Therefore, our results Theorem 9 and Corollary 1 uses the path length form (5.5) instead. □

Incremental stability for deterministic systems has been established as a concept of convergence between different solution trajectories starting from different initial conditions [38, 39]. In the stochastic setting, the difference between trajectories also arises from using different noise processes. Stochastic contraction for white noise systems (3.3) is defined in Definition 2 of [20]. We define a similar notion of stochastic contraction for the shot noise system (3.2) and the Lévy noise SDE (3.1).

Definition 8 (Stochastically Contracting)**.** The system (3.1) is said to be *stochastically contracting* if:

1. the unperturbed system $\dot{\mathbf{x}} = f(t, \mathbf{x})$ is contracting with some differential coordinate transform $\Theta(t, \mathbf{x})$ and convergence rate α.

2. there exist constants $\gamma, \eta > 0$ such that $\sup_{t,\mathbf{x}} \|\sigma(t, \mathbf{x})\|_F \leq \gamma$ and $\sup_{t,\mathbf{x}} \|\xi(t, \mathbf{x})\|_F \leq \eta$.

Separately, (3.3) is stochastically contracting under the same two conditions above without η and ξ in condition 2, while (3.2) removes mention of γ and σ. □

5.2 White Noise Case

First, we briefly review the version of the white noise contraction theorem from [20, 21], where the Lyapunov-like function (5.7) is used. This requires the following assumption.

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

$$\begin{aligned} \underline{s} &= \inf_{t,\mathbf{x}} \lambda_{\min}(S(t, \mathbf{x})), \quad \overline{s} = \sup_{t,\mathbf{x}} \lambda_{\max}(S(t, \mathbf{x})) \\ s' &= \sup_{t,\mathbf{x},i,j} \|(\partial_x S(t, \mathbf{x}))_{i,j}\|, \quad s'' = \sup_{t,\mathbf{x},i,j} \left\|(\partial_x^2 S(t, \mathbf{x}))_{i,j}\right\| \end{aligned} \tag{5.9}$$

□

We have previously 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 stem from using different noise processes. Specifically, we define two trajectories $\mathbf{x}(t)$ and $\mathbf{y}(t)$ as solutions to (3.3) driven by completely different noise processes with different variances:

$$d\mathbf{x} = f(t, \mathbf{x})dt + \sigma_1(t, \mathbf{x})dW_1(t) \tag{5.10a}$$
$$d\mathbf{y} = f(t, \mathbf{y})dt + \sigma_2(t, \mathbf{y})dW_2(t) \tag{5.10b}$$

and for both of these systems stochastically contracting, we will denote η_1 and η_2 such that $\|\sigma_1(t, \mathbf{x})\|_F \leq \eta_1$ and $\|\sigma_2(t, \mathbf{y})\|_F \leq \eta_2$. Then, the following holds:

$$\mathrm{tr}(\sigma_1^T S \sigma_1(t, \mathbf{x})) \leq \overline{s}\eta_1^2, \quad \mathrm{tr}(\sigma_2^T S \sigma_2(t, \mathbf{y})) \leq \overline{s}\eta_2^2 \tag{5.11}$$

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.

Definition 9 (Virtual System for (3.3))**.** We can represent the infinitesimal differential length $\delta\mathbf{z}$ as a path integral and reparameterize using measure $\mu \in [0,1]$

$$\mathbf{y}(t) - \mathbf{x}(t) = \int_{\mathbf{x}}^{\mathbf{y}} \delta\mathbf{z} = \int_0^1 \left(\frac{\partial \mathbf{z}}{\partial \mu}\right) d\mu \tag{5.12}$$

where $\mu \in [0,1]$ is a measure parameter such that:

$$\begin{gathered} \mathbf{z}(\mu = 0, t) = \mathbf{x}(t),\, \mathbf{z}(\mu = 1, t) = \mathbf{y}(t) \\ \sigma_{\mu=0}(t, \mathbf{z}) = \sigma_1(t, \mathbf{x}),\, \sigma_{\mu=1}(t, \mathbf{z}) = \sigma_2(t, \mathbf{y}) \\ W_{\mu=0}(t, \mathbf{z}) = W_1(t),\, W_{\mu=1}(t, \mathbf{z}) = W_2(t) \end{gathered} \tag{5.13}$$

e.g., $\mathbf{z}(\mu, t) := \mu\mathbf{x}(t) + (1-\mu)\mathbf{y}(t)$. Using this, we can rewrite (3.3) as the virtual system:

$$d\mathbf{z}(\mu, t) = f(t, \mathbf{z}(\mu, t))dt + \sigma_\mu(t, \mathbf{z}(\mu, t))dW_\mu(t) \tag{5.14}$$

This further enables us to describe the virtual dynamics as follows:

$$d\delta\mathbf{z} = F\delta\mathbf{z}dt + \delta\sigma_\mu dW_\mu \tag{5.15}$$

where we denote

$$\delta\sigma_\mu = \begin{bmatrix} \frac{\partial \sigma_{\mu,1}}{\partial \mathbf{z}}\delta\mathbf{z} & \cdots & \frac{\partial \sigma_{\mu,d}}{\partial \mathbf{z}}\delta\mathbf{z} \end{bmatrix}$$

for $\sigma_\mu := \begin{bmatrix} \sigma_{\mu,1} & \cdots & \sigma_{\mu,d} \end{bmatrix}$, where $\sigma_{\mu,i}$ is the ith column of σ_μ. □

Note that by Cauchy-Schwarz and the triangle inequality for integrals, we can bound the path integral as follows:

$$\|\mathbf{y} - \mathbf{x}\|^2 = \left\| \int_{\mathbf{x}}^{\mathbf{y}} \delta\mathbf{z} \right\|^2 \leq \int_{\mathbf{x}}^{\mathbf{y}} \|\delta\mathbf{z}\|^2 = \int_0^1 \left\| \frac{\partial \mathbf{z}}{\partial \mu} \right\|^2 d\mu \tag{5.16}$$

and multiplying across by $\underline{s}$ yields

$$\begin{aligned} \underline{s}\,\|\mathbf{y} - \mathbf{x}\|^2 &\leq \underline{s} \int_0^1 \left\| \frac{\partial \mathbf{z}}{\partial \mu} \right\|^2 d\mu \\ &\leq \int_0^1 \left(\frac{\partial \mathbf{z}}{\partial \mu}\right)^T S(t, \mathbf{z}(\mu, t)) \left(\frac{\partial \mathbf{z}}{\partial \mu}\right) d\mu \end{aligned} \tag{5.17}$$

We will use the right-side expression above as our Lyapunov function:

$$V(t, \mathbf{z}, \delta\mathbf{z}) = \int_0^1 \left(\frac{\partial \mathbf{z}}{\partial \mu}\right)^T S(t, \mathbf{z}(\mu, t)) \left(\frac{\partial \mathbf{z}}{\partial \mu}\right) d\mu \tag{5.18}$$

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

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

Theorem 8 (Stochastic Contraction Theorem for White Noise [21])**.** For the SDE system described in (3.3), suppose the following two conditions are satisfied:

1. the unperturbed system $\dot{\mathbf{x}} = f(t, \mathbf{x})$ is contracting in $S(t, \mathbf{x})$ at a rate α:
2. there exists $\eta > 0$ such that $\|\sigma(t, \mathbf{x})\|_F \leq \eta$ for all $\mathbf{x} \in \mathbb{R}^n$ and $t \geq 0$. This implies that $\mathrm{tr}(\sigma^T S \sigma(t, \mathbf{x})) \leq \overline{s}\eta^2$, and justifies the assumption of (5.11).

Further assume that the initial conditions adhere to some probability distribution $p(\mathbf{z}_0) = p(\mathbf{x}_0, \mathbf{y}_0)$. Then (3.3) is stochastically contracting if

$$\mathbb{E}_{\mathbf{z}_0}[V(t, \mathbf{z}, \delta \mathbf{z})] \leq V(0, \mathbf{z}_0, \delta \mathbf{z}_0) e^{-\beta t} + \frac{\kappa}{\beta}(1 - e^{-\beta t}) \tag{5.19}$$

where

$$\beta = \alpha - \frac{1}{2\underline{s}}(\eta_1^2 + \eta_2^2)\left(\frac{s''}{2} + s'\right) \tag{5.20a}$$

$$\kappa = \frac{1}{2}(\overline{s} + s')(\eta_1^2 + \eta_2^2) \tag{5.20b}$$

Because the Lyapunov-like function V can be arbitrarily defined for stability analysis, we simplify the inequality further to express it in terms of the given trajectories $\mathbf{x}$ and $\mathbf{y}$. Dividing across by $\underline{s}$ yields:

$$\frac{1}{\underline{s}}\mathbb{E}_{\mathbf{z}_0}[V(t, \mathbf{z}, \delta \mathbf{z})] \leq \frac{1}{\underline{s}}\mathbb{E}_{\mathbf{z}_0}[V(0, \mathbf{z}_0, \delta \mathbf{z}_0)] e^{-\beta t} + \frac{\kappa}{\underline{s}\beta}(1 - e^{-\beta t})$$

Finally, we lower-bound the left-hand side of the inequality according to the construction in (5.18). This gives us:

$$\mathbb{E}_{(\mathbf{x}_0, \mathbf{y}_0)}[\|\mathbf{y} - \mathbf{x}\|^2] \leq \frac{1}{\underline{s}}\|\mathbf{y}_0 - \mathbf{x}_0\|^2 e^{-\beta t} + \frac{\kappa}{\underline{s}\beta}\left(1 - e^{-\beta t}\right) \tag{5.21}$$

5.3 Shot Noise Case

Consider two trajectories of a system: $\mathbf{x}(t)$ a solution of (3.2), and $\mathbf{y}(t)$ a solution of the deterministic $\dot{\mathbf{y}} = f(t, \mathbf{y})$. We define the parameter $\mu \in [0, 1]$ such that

$$\begin{aligned} \mathbf{z}(\mu = 0, t) &= \mathbf{x}(t), & \mathbf{z}(\mu = 1, t) &= \mathbf{y}(t) \\ \xi_{\mu=0}(t, \mathbf{z}) &= \xi(t, \mathbf{x}), & \xi_{\mu=1}(t, \mathbf{z}) &= 0 \\ N_{\mu=0}(t, \mathbf{z}) &= N(t), & N_{\mu=1}(t, \mathbf{z}) &= 0 \end{aligned} \tag{5.22}$$

with 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 the virtual dynamics $d\delta\mathbf{z}(\mu, t) = F\delta\mathbf{z}(\mu, t)dt + \delta\xi_\mu(t, \mathbf{z})dN_\mu$ with the generalized Jacobian F from (5.3), where $\delta\xi_\mu = \left[\frac{\partial \xi_{\mu,1}}{\partial \mathbf{z}}\delta\mathbf{z}, \cdots, \frac{\partial \xi_{\mu,\ell}}{\partial \mathbf{z}}\delta\mathbf{z}\right]$ and $\xi_\mu := [\xi_{\mu,1}, \cdots, \xi_{\mu,\ell}]$, where $\xi_{\mu,i}$ is the ith column of ξ_μ.

Theorem 9 (Shot Noise Contraction Theorem)**.** Suppose that (3.2) is stochastically contracting in the sense of Definition 8 under a differential coordinate transform $\Theta(t, \mathbf{x})$ which satisfies (5.4). Further assume that the initial conditions adhere to some probability distribution $p(\mathbf{z}_0) = p(\mathbf{x}_0, \mathbf{y}_0)$, where p is either compactly-supported, or is a distribution with finite second moment. Moreover, for the Lyapunov-like path length function (5.5), suppose there exists a continuously differentiable function $c_{\mathbf{z}_0} : \mathbb{R}^+ \to \mathbb{R}^+$, which is bounded for each fixed initial condition $\mathbf{z}_0$ such that

$$\mathbb{E}_{\mathbf{z}_0}\left[V(t, \mathbf{z}(t), \delta\mathbf{z}(t)) - V(t-, \mathbf{z}(t-), \delta\mathbf{z}(t-))\right] \leq c_{\mathbf{z}_0}(t) \tag{5.23}$$

and such that $\mathbb{E}[c_{\mathbf{z}_0}(t)] := \int c_{\mathbf{z}_0}(t)dp(\mathbf{z}_0) \leq c(t)$ for some continuously differentiable, bounded, deterministic function $c : \mathbb{R}^+ \to \mathbb{R}^+$. With the setup and assumptions above, we have the following inequality:

$$\mathbb{E}[\|\mathbf{y} - \mathbf{x}\|] \leq \frac{1}{\underline{h}}\mathbb{E}[\|\mathbf{y}_0 - \mathbf{x}_0\|]e^{-\beta_s t} + \frac{\kappa_s(t)}{\underline{h}} \tag{5.24}$$

where

$$\beta_s := \alpha \tag{5.25a}$$

$$\kappa_s(t) := \frac{\eta\lambda}{\beta_s}\left(c(t) - c(0)e^{-\beta_s t} - \int_0^t c'(s)e^{-\beta_s(t-s)}ds\right) \tag{5.25b}$$

λ is the intensity of the Poisson process $N(t)$, $\underline{h}$ is defined in (5.4), $c'(t)$ is the derivative of $c(t)$, α is the deterministic contraction rate, and η is the bound on the magnitude of the jumps $\xi(t, \mathbf{x})$ described by Definition 8.

Proof. We apply Itô's formula to our Lyapunov-like function (5.5). The version of the formula for scalar processes is standard, e.g. see Theorem 32 of [7] or Theorem 3.7 of [8], and an extension to multiple dimensions is straightforward. (5.5) becomes

$$\begin{aligned} V(t, \mathbf{z}, \delta\mathbf{z}) = V(0, \mathbf{z}_0, \delta\mathbf{z}_0) + \int_0^t \partial_t V(s, \mathbf{z}, \delta\mathbf{z})ds + \int_0^t \sum_{i=1}^n \left[\partial_{z_i} V(s, \mathbf{z}, \delta\mathbf{z}) f_i(s, \mathbf{z}) + \partial_{\delta z_i} V(s, \mathbf{z}, \delta\mathbf{z}) \left(F\delta\mathbf{z}\right)_i\right] \\ + \sum_{s\leq t} \left(V(s, \mathbf{z}(s), \delta\mathbf{z}(s)) - V(s-, \mathbf{z}(s-), \delta\mathbf{z}(s-))\right) \end{aligned} \tag{5.26}$$

where F is the generalized Jacobian from (5.3), and 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. Note that the terms of Itô's formula which correspond to the continuous part of the process disappear because (3.2) has none, simplifying the expression considerably.

A bound on the first three terms are derived directly from deterministic contraction of an unperturbed system $\dot{\mathbf{x}} = f(t, \mathbf{x})$, and we can use the condition (5.6). A bound on the last term comes from the fourth condition (5.23) and by taking the expected value $\mathbb{E}_{\mathbf{z}_0}$ so that Campbell's formula can be applied. Applying $\mathbb{E}_{\mathbf{z}_0}$ across the entire equation (5.26), and using combining the upper bounds on the deterministic and shot noise terms, we get:

$$\mathbb{E}_{\mathbf{z}_0}\left[V(t, \mathbf{z}, \delta\mathbf{z})\right] - V(0, \mathbf{z}_0, \delta\mathbf{z}_0) \leq -\alpha \int_0^t \mathbb{E}_{\mathbf{z}_0}[V(s, \mathbf{z}, \delta\mathbf{z})]ds + \eta\lambda c_{\mathbf{z}_0}(t) \tag{5.27}$$

where α is the contraction rate of the unperturbed system.

We can then obtain a bound on the solution $\mathbb{E}_{\mathbf{z}_0}[V(t, \mathbf{z}, \delta\mathbf{z})]$ using the Comparison lemma. Integrating (5.27) over the probability distribution $p(\mathbf{z}_0)$ and using the fact that $\underline{h}\mathbb{E}\left[\|\mathbf{y}(t) - \mathbf{x}(t)\|\right] \leq \mathbb{E}\left[\mathbb{E}_{\mathbf{z}_0}[V(t, \mathbf{z}, \delta\mathbf{z})]\right]$ where $\mathbb{E}\left[\|\mathbf{y} - \mathbf{x}\|\right] := \int \|\mathbf{y} - \mathbf{x}\| \, dp(\mathbf{x}_0, \mathbf{y}_0)$ gives us our desired bound (5.24), with β_s as in (5.25a) and $\kappa_s(t)$ as in (5.25b). An additional step using integration-by-parts is taken to explicitly show the dependence on $1/\beta_s$, which intuitively tells us that a slower convergence rate corresponds to a larger bounded error ball. ■

Remark 8. Note that both the shot noise system and the deterministic system have contraction rate α. This is because in between consecutive jumps, the shot noise system behaves exactly as the deterministic system. However, the difference between the two is the nonzero bounded error ball $\kappa_s(t)$ (5.25b). Furthermore, note that $\kappa_s(t) \propto \lambda$, and the interarrival times $T_i - T_{i-1} \sim \text{Exp}(1/\lambda)$; indeed, shorter interarrival times correspond to a larger error ball due to the more rapid accumulation of deviations away from the nominal trajectory. □

5.4 Lévy Noise Case

The investigation of stochastic systems that are perturbed by only white noise $d\mathbf{x}(t) = f(t, \mathbf{x})dW(t)$ and only shot noise $d\mathbf{x}(t) = f(t, \mathbf{x})dN(t)$ will further segway into a treatment of stochastic systems that are perturbed by a broader class of noise using a tool known as the *Lévy-Khintchine decomposition theorem* [2, 8], including combinations such as $d\mathbf{x}(t) = f(t, \mathbf{x})dW(t) + g(t, \mathbf{x})dN(t)$. We begin with a comparison of the definitions for a Brownian motion process and a Poisson process.

1. A real-valued process $\{W(t) : t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a Brownian motion if the following hold:
   - the paths of W are $\mathbb{P}$ almost-surely continuous.
   - $\mathbb{P}(W(0) = 0) = 1$
   - for $0 \leq s \leq t$, $W(t) - W(s)$ is equal in distribution to $W(t - s)$.
   - for $0 \leq s \leq t$, $W(t) - W(s)$ is independent of $W(r)$ for $r \leq s$.
   - for $t > 0$, $W(t)$ is equal in distribution to a normal random variable with variance t.
2. A real-valued process $\{N(t) : t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a Poisson process with intensity $\lambda > 0$ if the following hold:
   - the paths of N are $\mathbb{P}$ right-continuous with left-limits.
   - $\mathbb{P}(N(0) = 0) = 1$
   - for $0 \leq s \leq t$, $N(t) - N(s)$ is equal in distribution to $N(t - s)$.
   - for $0 \leq s \leq t$, $N(t) - N(s)$ is independent of $N(r)$ for $r \leq s$.
   - for $t > 0$, $N(t)$ is equal in distribution to a Poisson random variable with parameter λt.

The similarity between the two definitions motivates the definition of a general Lévy process.

Definition 10. A process $\{L(t) : t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a *Lévy process* if the following hold:

- **Cádlág Paths**: the paths of L are almost-surely $\mathbb{P}$ right-continuous with left-limits. This means that every sample path of L must be right-continuous with left-limits.
- **Zero Initial Condition**: $\mathbb{P}(L(0) = 0) = 1$
- **Stationary Increments**: for $0 \leq s \leq t$, $L(t) - L(s)$ is equal in distribution to $L(t - s)$.
- **Independent Increments**: for $0 \leq s \leq t$, $L(t) - L(s)$ is independent of $L(r)$ for $r \leq s$.

□

We closely follow the version of the formula stated formally in Theorem 1.6 of [2] or Theorem 2.7 of [8].

Theorem 10 (Lévy-Khintchine 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 satisfying $\int_{\mathbb{R}} 1 \wedge x^2 \nu(dx) < \infty$ such that

$$\Psi(\theta) = ia\theta - \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}} \left(e^{i\theta x} - 1\right) \nu(dx) - \int_{\mathbb{R}} i\theta x \mathbb{1}_{[-1,1]}(x)\nu(dx) \tag{5.28}$$

Conversely, given any triplet (a, σ, ν), there exists a Lévy process L with characteristic exponent given by (5.28).

a is called the *center* of L and captures the deterministic drift component, σ is the *Gaussian coefficient* and captures the variance of the Brownian motion component, and the Lévy measure ν captures the size and intensity of the "large" jumps of L.

Proof. We will prove the converse of Theorem 10. First, we briefly derive the characteristic exponents of the Gaussian (for the Brownian motion part) and the Poisson process for a fixed time interval of $[0, 1]$.

For the Brownian motion with drift term a and variation σ^2, the random variable $W(1)$ is normally distributed with mean a and variance σ^2.

$$\mathbb{E}[e^{i\theta W(1)}] = e^{ai\theta - \frac{1}{2}\theta^2\sigma^2}$$

and so the characteristic exponent is $ai\theta - \frac{1}{2}\theta^2\sigma^2$.

For the standard Poisson with intensity λ:

$$\mathbb{E}[e^{i\theta N(1)}] = e^{\lambda\left(e^{i\theta}-1\right)}$$

and so the characteristic exponent is $\lambda(e^{i\theta} - 1)$.

Analogously, for the compound Poisson process

$$Y(t) = \sum_{k=1}^{N(t)} \xi_k$$

with iid sequence of height random variables ξ_k distributed according to measure $\nu(dx)$. We get by the tower property of conditional expectations:

$$\begin{aligned}
\mathbb{E}\left[e^{i\theta Y(1)}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{i\theta\sum\limits_{k=1}^{n}\xi_k}\middle| N(1) = n\right]\right] &= \sum_{n=0}^{\infty}\mathbb{E}\left[e^{i\theta\sum\limits_{k=1}^{n}\xi_k}\right]e^{-\lambda}\frac{\lambda^n}{n!} \\
&= \sum_{n=0}^{\infty}\mathbb{E}\left[e^{i\theta\xi}\right]^n e^{-\lambda}\frac{\lambda^n}{n!} \quad \text{by iid } \xi_k \\
&= e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\left(\int_{\mathbb{R}} e^{i\theta x}\nu(dx)\right)^n \\
&= e^{-\lambda}e^{\lambda\left(\int_{\mathbb{R}} e^{i\theta x}\nu(dx)\right)} \quad \text{by Taylor expansion} \\
&= e^{\lambda\left(\int_{\mathbb{R}} e^{i\theta x}\nu(dx)-1\right)}
\end{aligned}$$

The characteristic exponent is given by $\lambda\int_{\mathbb{R}}(e^{i\theta y} - 1)\nu(dy)$.

If a drift rate of $c \in \mathbb{R}$ is also added to the compound Poisson process:

$$Y_t' = \sum_{k=1}^{N(t)} \xi_k + ct, \quad t \geq 0$$

then we can repeat the whole calculation above, just with the additional c term in the exponent:

$$\begin{aligned}
\mathbb{E}\left[e^{i\theta Y(1)'}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{i\theta\left(\sum_{k=1}^{n}\xi_k + c\right)}\middle| N(1)=n\right]\right] &= \sum_{n=0}^{\infty}\mathbb{E}\left[e^{i\theta\sum_{k=1}^{n}\xi_k + i\theta c}\right]e^{-\lambda}\frac{\lambda^n}{n!} \\
&= \sum_{n=0}^{\infty}\mathbb{E}\left[e^{i\theta\xi}\right]^n e^{i\theta c}e^{-\lambda}\frac{\lambda^n}{n!} \quad \text{by iid } \xi_k \\
&= e^{-\lambda}e^{i\theta c}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\left(\int_{\mathbb{R}} e^{i\theta x}\nu(dx)\right)^n \\
&= e^{-\lambda}e^{i\theta c}e^{\lambda\left(\int_{\mathbb{R}} e^{i\theta x}\nu(dx)\right)} \quad \text{by Taylor expansion} \\
&= e^{\lambda\left(\int_{\mathbb{R}} e^{i\theta x}\nu(dx)-1\right)+i\theta c}
\end{aligned}$$

We typically choose c to be of the form $-\lambda\int_{\mathbb{R}} x\nu(dx)$ so that the compound Poisson process is centered (i.e, compensated). The characteristic exponent then becomes:

$$\lambda\int_{\mathbb{R}}(e^{i\theta y} - 1 - i\theta x)\nu(dy)$$

We further have that the sum of two independent Lévy processes L_1 and L_2 with characteristic exponents Ψ_1 and Ψ_2 is a Lévy process with characteristic exponent $\Psi_1 + \Psi_2$. Hence, we will take the sum of the Brownian motion process and the compound Poisson process.

Let $A_\varepsilon = [-1,1]/(-\varepsilon,\varepsilon)$ be a closed annulus around the origin with inner radius ε (and outer radius 1). We can then decompose Ψ of (5.28) as follows:

$$\begin{aligned}
\Psi(\theta) &= ai\theta - \frac{1}{2}\theta^2\sigma^2 + \lambda\int_{\mathbb{R}}(e^{i\theta y}-1)\nu(dy) \\
&= ai\theta - \frac{1}{2}\theta^2\sigma^2 + \nu(\mathbb{R}/[-1,1])\int_{\mathbb{R}/[-1,1]}\left(e^{i\theta x}-1\right)\frac{\nu(dx)}{\nu(\mathbb{R}/[-1,1])} + \lim_{\varepsilon\downarrow 0}\left[\int_{A_\varepsilon}\left(e^{i\theta x}-1\right)\nu(dx) - i\theta\int_{A_\varepsilon} x\nu(dx)\right]
\end{aligned}$$

which essentially tells us that any Lévy process can be split into three simpler Lévy processes: a Brownian motion, a compound Poisson process of "large" jumps, and a compound Poisson process of jumps small enough to be compensated by a deterministic drift to turn it into a Martingale like the Brownian motion part.

The final steps of the proof require us to show that the final limit gives a characteristic exponent of an infinitely-divisible random variable. This is a nontrivial detail, and requires a bit more background content in order to proceed. We will conclude the proof here for now and come back to it later. ∎

Recall that 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 [2, 40]. Note that both Gaussian white noise and Poisson shot noise separately are Lévy processes. As such, a linear combination of white and Poisson noise is also a Lévy process. In particular, we can invoke the Lévy-Khintchine Decomposition Formula (5.28) to decompose any Lévy noise process into an affine combination of white and shot noise. Analogously, the white noise contraction theorem [21] and Theorem 9 further direct us into a contraction condition for stochastic systems perturbed by Lévy noise.

Consider two trajectories of a system: $\mathbf{x}(t)$ a solution of (3.1), and $\mathbf{y}(t)$ a solution of the deterministic

$d\mathbf{y}(t) = f(t, \mathbf{y})dt$. We define the parameter $\mu \in [0, 1]$ such that (5.22) holds and

$$\begin{aligned} \sigma_{\mu=0}(t, \mathbf{z}) &= \sigma(t, \mathbf{x}), & \sigma_{\mu=1}(t, \mathbf{z}) &= 0 \\ W_{\mu=0}(t, \mathbf{z}) &= W(t), & W_{\mu=1}(t, \mathbf{z}) &= 0 \end{aligned} \tag{5.29}$$

The virtual system for this setting amounts to $d\mathbf{z}(\mu, t) = f(t, \mathbf{z}(\mu, t))dt + \sigma_\mu(t, \mathbf{z}(\mu, t))dW_\mu(t) + \xi_\mu(t, \mathbf{z}(\mu, t))dN_\mu(t)$ and the virtual dynamics $d\delta\mathbf{z} = F\delta\mathbf{z}dt + \delta\sigma_\mu dW_\mu + \delta\xi_\mu dN_\mu$, with again F as in (5.3).

Analogous to the shot noise parameters (5.25), denote β_w, κ_w and β_ℓ, κ_ℓ to be the contraction rate and steady-state error bound for the white noise SDE (3.3) and the Lévy noise SDE, respectively. Although one possible exact form of β_w and κ_w can be found in [20, 21], we do not use those versions for the following two reasons. First, there is a discrepancy between the Lyapunov-like functions used to prove contraction for the white noise case in [20, 21] and the shot noise case; this is detailed in Remark 7. Second, both [20, 21] consider the difference between two noise-perturbed trajectories – $\mathbf{x}(t)$, solution to (3.3) with noise term $\sigma_1(t, \mathbf{x})dW_1(t)$, and $\mathbf{y}(t)$, solution to (3.3) with $\sigma_2(t, \mathbf{x})dW_2(t)$ – instead of one noise-perturbed trajectory against the deterministic trajectory. With these two differences addressed, our new parameters for the white noise case become:

$$\beta_w := \alpha - \frac{h''\gamma^2}{2\underline{h}}, \quad \kappa_w := \frac{h'\gamma^2}{2\beta_w}\left(1 - e^{-\beta_w t}\right) \tag{5.30}$$

We will formally derive how (5.30) was obtained as a part of the proof to our main corollary.

Corollary 1 (Stochastic Contraction Theorem for General Lévy Noise)**.** With the setup described above, suppose (3.1) is stochastically contracting in the sense of Definition 8. Furthermore, we impose the same assumptions as in Theorem 9: the existence of a continuously-differentiable $c(t)$ from the construction of (5.23), metric bounds (5.4), and the distribution on the initial conditions $p(\mathbf{z}_0) = p(\mathbf{x}_0, \mathbf{y}_0)$ with either compact support or finite second moment. Then the following inequality is satisfied:

$$\mathbb{E}[\|\mathbf{y} - \mathbf{x}\|] \le \frac{1}{\underline{h}}\mathbb{E}\left[\|\mathbf{y}_0 - \mathbf{x}_0\|\right] e^{-\beta_\ell t} + \frac{\kappa_\ell(t)}{\underline{h}} \tag{5.31}$$

where

$$\beta_\ell := \alpha - \frac{h''\gamma^2}{2\underline{h}} \tag{5.32a}$$

$$\kappa_\ell(t) := \frac{\eta\lambda}{\beta_\ell}\left(c(t) - c(0)e^{-\beta_\ell t} - \int_0^t c'(s)e^{-\beta_\ell(t-s)}ds\right) + \frac{h'\gamma^2}{2\beta_\ell}\left(1 - e^{-\beta_\ell t}\right) \tag{5.32b}$$

λ is the intensity of the Poisson process $N(t)$, $\underline{h}, h', h''$ are defined in (5.4), $c'(t)$ is the derivative of $c(t)$, α is the deterministic contraction rate, and γ and η are defined in the second condition of Definition 8.

Proof. Applying Itô's formula to (5.7):

$$V(t,\mathbf{z},\delta\mathbf{z}) - V(0,\mathbf{z}_0,\delta\mathbf{z}_0) =$$

$$+\int_0^t \partial_t V(s,\mathbf{z},\delta\mathbf{z})ds + \int_0^t \sum_{i=1}^n \left[\partial_{z_i} V(s,\mathbf{z},\delta\mathbf{z}) f_i(s,\mathbf{z}) + \partial_{\delta z_i} V(s,\mathbf{z},\delta\mathbf{z})\left(F\delta\mathbf{z}\right)_i\right] \tag{5.33a}$$

$$+\int_0^t \sum_{i=1}^n \left[\partial_{z_i} V(s,\mathbf{z},\delta\mathbf{z})\sigma_{\mu,i}(\mathbf{z}) + \partial_{\delta z_i} V(s,\mathbf{z},\delta\mathbf{z})\delta\sigma_{\mu,i}(\mathbf{z})\right] dW_\mu(s) \tag{5.33b}$$

$$+\frac{1}{2}\Bigg[\int_0^t \sum_{i,j=1}^n \frac{\partial^2 V}{\partial \delta z_i \partial \delta z_j}(s,\mathbf{z}(s-),\delta\mathbf{z}(s-))d\left[\delta\mathbf{z}_i,\delta\mathbf{z}_j\right]^c \tag{5.33c}$$

$$+\int_0^t \sum_{i,j=1}^n \frac{\partial^2 V}{\partial z_i \partial \delta z_j}(s,\mathbf{z}(s-),\delta\mathbf{z}(s-))d\left[\mathbf{z}_i,\delta\mathbf{z}_j\right]^c \tag{5.33d}$$

$$+\int_0^t \sum_{i,j=1}^n \frac{\partial^2 V}{\partial z_i \partial z_j}(s,\mathbf{z}(s-),\delta\mathbf{z}(s-))d\left[\mathbf{z}_i,\mathbf{z}_j\right]^c\Bigg] \tag{5.33e}$$

$$+\sum_{s\le t}\left(V(s,\mathbf{z}(s),\delta\mathbf{z}(s)) - V(s-,\mathbf{z}(s-),\delta\mathbf{z}(s-))\right) \tag{5.33f}$$

where F is the generalized Jacobian from (5.3), and the brackets $[\cdot]^c$ denote the continuous-part quadratic variation. As in the proof of Theorem 9, a bound on the first three terms are derived directly from deterministic contraction of an unperturbed system $\dot{\mathbf{x}} = f(t,\mathbf{x})$, and we can use the condition (5.6). With $\mathbb{E}_{\mathbf{z}_0}[\cdot]$ taken, the bound on the last term (5.33f) comes from Campbell's formula and (5.23). When we take expectation over the entire equation, the white noise term between (5.33a) and (5.33c) disappears to 0 due to being a martingale with zero mean.

Simplifying the quadratic variation terms (5.33c) to (5.33e) requires computing the partial derivatives of V. Using matrix multiplication, the submultiplicativity property, and the fact that $\Theta(t,\mathbf{z})$ is independent of $\delta\mathbf{z}$, we obtain the following relationships for each fixed time t:

$$\frac{\partial^2 V}{\partial \delta z_i \partial \delta z_j} d\left[\delta\mathbf{z}_i,\delta\mathbf{z}_j\right]^c = 0, \text{ since } \frac{\partial^2 V}{\partial \delta z_i \partial \delta z_j} = 0$$

$$\frac{\partial^2 V}{\partial z_i \partial \delta z_j} d\left[\mathbf{z}_i,\delta\mathbf{z}_j\right]^c \le \int_0^1 \|\partial_z \Theta(t,\mathbf{z})\| \sum_{k=1}^d \sigma_{\mu,ik}\delta\sigma_{\mu,jk} d\mu \le h'\gamma^2$$

$$\frac{\partial^2 V}{\partial z_i \partial z_j} d\left[\mathbf{z}_i,\mathbf{z}_j\right]^c \le \int_0^1 \left\|\partial_z^2 \Theta(t,\mathbf{z})\right\| \left\|\frac{\partial \mathbf{z}}{\partial \mu}\right\| \sum_{k=1}^d \sigma_{\mu,ik}\sigma_{\mu,jk} d\mu \le \frac{h''\gamma^2}{\underline{h}} V$$

where γ is the white-noise bound described in Definition 8.

Combining the bounds of each individual term in (5.33) and taking the expected value with respect to fixed initial condition $\mathbf{z}_0$ yields the following inequality:

$$\mathbb{E}_{\mathbf{z}_0}\left[V(t,\mathbf{z},\delta\mathbf{z})\right] - V(0,\mathbf{z}_0,\delta\mathbf{z}_0) \le -\left(\alpha - \frac{h''\gamma^2}{2\underline{h}}\right)\int_0^t \mathbb{E}_{\mathbf{z}_0}[V(s,\mathbf{z},\delta\mathbf{z})]ds + \frac{h'\gamma^2}{2} + \eta\lambda c_{\mathbf{z}_0}(t)$$

As in the proof to Theorem 9, we obtain a bound on the solution $\mathbb{E}_{\mathbf{z}_0}[V(t,\mathbf{z},\delta\mathbf{z})]$ using the Comparison lemma. Then we integrate (5.34) over the probability distribution $p(\mathbf{z}_0)$ and use the fact that $\underline{h}\mathbb{E}\left[\|\mathbf{y}(t) - \mathbf{x}(t)\|\right] \le \mathbb{E}\left[\mathbb{E}_{\mathbf{z}_0}[V(t,\mathbf{z},\delta\mathbf{z})]\right]$ to get our desired bound (5.31), with β_ℓ set as in (5.32a) and $\kappa_\ell(t)$ as in (5.32b). ■

Note that with the modified white noise parameters (5.30), β_ℓ and $\kappa_\ell(t)$ for the combined SDE are nearly a direct summation of the parameters (5.30) and (5.25). Specifically, β_ℓ requires an extra α term to be removed so that otherwise the convergence rate due to the deterministic part of the system would be counted twice:

$$\beta_\ell = \beta_s + \beta_w - \alpha = \beta_w \tag{5.34}$$

and $\kappa_\ell(t)$ is an exact summation of the two parts, with contraction rate β_ℓ used in place of β_w or β_s. Again, because $\beta_s = \alpha$, $\beta_\ell = \beta_w$ because in between consecutive jumps, the shot noise system (3.2) behaves exactly like as $\dot{\mathbf{x}} = f(t, \mathbf{x})$.

Bibliography

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

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

[3] A. Zygmund, K. Wiener, and R. Paley, "Notes on random functions." *Mathematische Zeitschrift*, vol. 37, pp. 647–668, 1933.

[4] H. McKean, *Stochastic Integrals*, ser. AMS Chelsea Publishing Series. Academic Press, 1969.

[5] J. F. C. Kingman, *Poisson processes*. NY, USA: Oxford University Press, 1993, vol. 3.

[6] G. Last and M. Penrose, *Lectures on the Poisson process*. NY, USA: Cambridge University Press, 2017. [Online]. Available: https://www.math.kit.edu/stoch/~last/seite/lectures_on_the_poisson_process/media/lastpenrose2017.pdf

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

[8] R. F. Bass. (2014) Jump processes. [Online]. Available: http://bass.math.uconn.edu/jump-processes.pdf

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

[10] 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.

[11] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, ser. Universitext. Berlin, Germany: Springer Berlin, 2010. [Online]. Available: https://books.google.com/books?id=EQZEAAAAQBAJ

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

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

[14] B. Øksendal and A. Sulem-Bialobroda, *Applied Stochastic Control of Jump Diffusions*, ser. Universitext. Berlin, Germany: Springer-Verlag, 2007. [Online]. Available: https://books.google.com/books?id=s4Qi_WKEtkMC

[15] 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.

[16] 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.

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

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

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

[20] 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.

[21] 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.

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

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

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

[25] ——, "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.

[26] X. Mao, "Exponential stability for stochastic differential equations with respect to semimartingales," *Stochastic Processes and their Applications*, vol. 35, no. 2, pp. 267–277, Aug 1990.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

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

[32] 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.

[33] 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.

[34] 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.

[35] 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.

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

[37] 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.

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

[39] Z. Aminzare and E. D. Sontag, "Contraction methods for nonlinear systems: A brief introduction and some open problems," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 3835–3847.

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