Large deviations of affine processes

Matthew Varble

September 20, 2022

Abstract

This is an abstract of the entire dissertation; summarize a history of large deviations and affine processes, then abstractly summarize our large deviations result.

Acknowledgment

This is where I acknowledge how I am useless without others.

Contents

Introduction

This thesis aims to develop a comprehensive picture of large deviations of affine processes. Affine processes have been studied so extensively, that it would not be foolish to admit that their theory is somewhat complete. Though some questions still remain, the thesis of [Cuc11] establishes a comprehensive understanding of affine processes on general convex state spaces. Starting with the basic definition of an affine process as a time-homogeneous Markov process with conditional marginal distributions having a characteristic function with log-affine dependence on the initial state, this paper establishes numerous useful properties of the process. Among these properties include that affine processes have càdlàg modifications (paths have left-limits and are right-continuous), and such modifications are themselves jump-diffusions with respect to their natural filtration. This result allowed [KRM15] to prove the transform formula for the real and complex moments, which establishes existence of the process and its finite moments, so long as a system of generalized Riccati differential equations have some solution (which need not be unique).

Despite this very comprehensive understanding, large deviation asymptotics of families of affine processes do not have a similar general approach. Seeing as the class of affine processes is quite general, surely there are large deviation principles for *certain* families of affine processes. The simplest of which include families of state-independent processes like the Brownian motion or Poisson process, whose large deviations are understood asymptotically $\epsilon \to 0$ as time and space are scaled linearly with respective factors $1/\epsilon$ and ϵ; this is a simple consequence of the *classical* Mogulskii's theorem (see [DZ10]). More general results indeed exist, like [KK14] which proves a large deviation principle for continuous affine processes on the so-called *canonical* state space. We are concerned with large deviation principles of similar asymptotics—those which could be interpreted as regularizing the noise of a dynamical system—but on the level of generality that affine processes are already understood.

Though our original intent was to prove a large deviation principle for the general class of affine processes on convex state spaces, we discovered other remarkable facts along the way. The following two are of key interest. Firstly, the measure-change techniques one often uses in large deviations are typically split into those of a finite-dimensional (*twisting/tilting*) and infinite-dimensional (*exponential martingale*) flavor, but for affine processes one may in fact parameterize the finite-dimensional measure-densities in terms of the infinite-dimensional ones. This was not a fact seen in [KK14], which—as they mentioned—forced them to choose a proof using exponential martingales. We show how their original finite-dimensional approach was sufficient in establishing an integral form for their rate function. Secondly, seeing as no large deviations result has been made on the general level of affine processes, there has been no attempt at understanding the structure of the associated rate functions. In our work of developing a proof for the principle, we saw the techniques mentioned in [DRL04], which ultimately inspired us to develop a calculus associated with our rate functions. These results remarkably have no dependence on the affine assumption, and so if a future large deviation principle is to be proved for asymptotic families of jump-diffusions of the same regularization, this thesis will have already established semi-closed forms for the associated rate function.

All this said, we do still prove a large deviation principle for the general class of affine processes on convex state spaces. While we require usual light-tail assumptions, we feel the generality is comparable to the treatment of affine processes in [Cuc11] and [KRM15]. In fact, we even choose to present affine processes with similar generality up until the point of us needing the assumption. At this juncture, we are able to represent the affine processes as special jump-diffusions, which have representations that are far more familiar to the larger audience that have studied Itô processes. For those that have only seen continuous diffusions, we also offer an Appendix which we feel is a great reference to the calculus associated with jump-diffusions.

It goes without saying that our proof is not without the incredible theory of large deviations that already exists. We in particular discuss it at the level of understanding one has from reading [DZ10], but we for the most part cite [FK06] and its large deviation results on the Skorokhod space. We also use [Puh01] to use a *density argument* in developing an integral form for our rate function.

Now that we have a general understanding of what is presented in this thesis, we remark on how it is organized. Chapter 1 appeals to all the facts we need for affine processes, Chapter 2 proves our large deviation principle, Chapter 3 explores our discovered calculus on the rate functions associated with large deviation principles, and Appendix A helps readers understand the notions and calculus associated with jump-diffusions.

Notation and conventions

Throughout, unless specifically referenced elsewhere, all notions of this text are formally defined and explored in [Kal02] or [JS03]. Most of our notation will coincide with these texts (as well as most other literature), except in regards to some particular conventions. Let us establish some of these here. A stochastic process X with a marginal-index-set I and state space $(\mathbb{X}, \mathscr{X})$ will be indifferently recognized as:

- a collection $X = (X_t)_{t\in I}$ of marginals $X_t : \Omega \to \mathbb{X}$,
- a map $X : \Omega \times I \to \mathbb{X}$,
- or its curried version $X : \Omega \to \mathbb{X}^I$.

With this convention, we find it appropriate to denote filtrations $\mathscr{F} = (\mathscr{F}_t)_{t\geq 0}$ of increasing σ-algebras $\mathscr{F}_t$. Seeing as $\mathscr{F}$ denotes the actual family of σ-algebras, we denote the joined algebra with an infinity subscript, $\mathscr{F}_\infty \coloneqq \bigvee_{t\geq 0} \mathscr{F}_t$. The blackboard notation will generally correspond to a topological space, including those objects we typically introduce in analysis.

- The real $\mathbb{R}$, the complex $\mathbb{C} = \mathbb{R} \oplus \mathrm{i}\mathbb{R}$, the and non-negative $\mathbb{R}_+ = [0, \infty)$ numbers with the usual Euclidean topologies.
- For real normed vector spaces $\mathbb{V}$, $\mathbb{W}$, the space $\mathbb{L}(\mathbb{V}, \mathbb{W})$ of real linear maps $\mathbb{V} \to \mathbb{W}$, equipped with operator norm.
$$|T| \coloneqq \sup_{|v|=1} |Tv|$$

We also concisely denote $\mathbb{L}(\mathbb{V}) \coloneqq \mathbb{L}(\mathbb{V}, \mathbb{V})$.

- For the a separable metric space $\mathbb{X}$ and an interval $I \subseteq \mathbb{R}_+$, the space $\mathbb{D}(I, \mathbb{X})$ of càdlàg functions, equipped with the Skorokhod J1 topology.

- For topological spaces $\mathbb{X}, \mathbb{Y}$, the space $\mathbb{C}(\mathbb{X}, \mathbb{Y})$ of continuous functions, equipped with the supremum norm.

- For finite-dimensional normed vector spaces $\mathbb{V}, \mathbb{W}$ and open $\mathbb{U} \subseteq \mathbb{V}$, the subspace $\mathbb{C}^1(\mathbb{U}, \mathbb{W})$ of functions $f \in \mathbb{C}(\mathbb{U}, \mathbb{W})$ in which there is a derivative map $\mathrm{D}f \in \mathbb{C}(\mathbb{U}, \mathbb{L}(\mathbb{V}, \mathbb{W}))$.

$$\lim_{|v| \to 0} \frac{\left|f(u+v) - f(u) - \mathrm{D}f(u) \cdot v\right|}{|v|} = 0$$

For $f \in \mathbb{C}^1(\mathbb{U}, \mathbb{R})$, we denote $\nabla f \in \mathbb{C}(\mathbb{U}, \mathbb{V})$ the gradient,

$$\left\langle v, \nabla f(u) \right\rangle \coloneqq \mathrm{D}f(u) \cdot v,$$

If there is some canonical ordered basis $(e_1, \ldots, e_{\dim \mathbb{V}})$ of $\mathbb{V}$, denote $\mathrm{D}_i f \in \mathbb{C}(\mathbb{U}, \mathbb{R})$ the i-th partial derivative.

$$\mathrm{D}_i f(u) \coloneqq \mathrm{D}f(u) \cdot e_i, \quad i = 1, \ldots, d$$

- For finite-dimensional normed vector space $\mathbb{V}$ and open $\mathbb{U} \subseteq \mathbb{V}$, the subspace $\mathbb{C}^2(\mathbb{U}, \mathbb{R})$ of $f \in \mathbb{C}^1(\mathbb{U}, \mathbb{R})$ in which we also have $\nabla f \in \mathbb{C}^1(\mathbb{U}, \mathbb{V})$. In such a case, we denote $\mathrm{D}^2 f \in \mathbb{C}(\mathbb{U}, \mathbb{L}(\mathbb{V}))$ the Hessian.

$$\mathrm{D}^2 f(u) \coloneqq \mathrm{D}\big(\nabla f(u)\big)$$

If there is some canonical ordered basis $(e_1, \ldots, e_{\dim \mathbb{V}})$ of $\mathbb{V}$, denote $\mathrm{D}_{ij} f \in \mathbb{C}(\mathbb{U}, \mathbb{R})$ the second-order ij-th partial derivative.

$$\mathrm{D}_{ij} f(u) \coloneqq \left\langle e_i, \mathrm{D}^2 f(u) \cdot e_j \right\rangle, \quad i, j = 1, \ldots, d$$

For any topological space $\mathbb{X}$ and subset $A \subseteq \mathbb{X}$, we denote A° and $\overline{A}$ its interior and closure, respectively. In the case that this topology on $\mathbb{X}$ is induced by some metric d, we denote $B(x, \delta), \overline{B}(x, \delta) \subseteq \mathbb{X}$ the respectively open and closed balls centered at $x \in \mathbb{X}$ with radius $\delta > 0$.

$$B(x, \delta) \coloneqq \Big\{x' \in \mathbb{X} : d(x', x) < \delta\Big\}, \quad \overline{B}(x, \delta) \coloneqq \Big\{x' \in \mathbb{X} : d(x', x) \leq \delta\Big\}$$

When the topology on $\mathbb{X}$ is canonical, we will denote the associated Borel algebra $\mathscr{B}(\mathbb{X})$. Particular examples of this convention are:

- the Borel algebra $\mathscr{B}(\mathbb{V})$ associated to the topology induced from a canonical inner-product $\langle \cdot, \cdot \rangle$ on a vector space $\mathbb{V}$.

- the Borel algebra $\mathscr{B}(\mathbb{X})$ associated to the relative topology of some subset $\mathbb{X}$ of a space $\mathbb{V}$ with itself some canonical topology.

In the case that we are dealing with a finite-dimensional real vector space $\mathbb{V}$ with inner-product $\langle\cdot,\cdot\rangle$, we assume some canonical orthonormal basis $e_1,\ldots,e_{\dim\mathbb{V}}\in\mathbb{V}$ and establish the associated isometric isomorphism $\mathbb{V}\equiv\mathbb{R}^d$.

$$v\in\mathbb{V}\quad\longleftrightarrow\quad\big(v^1,\ldots,v^{\dim\mathbb{V}}\big);\qquad v^i\coloneqq\langle v,e_i\rangle,\quad i=1,\ldots,\dim\mathbb{V}$$

Similarly identify any map $f:\mathbb{A}\to\mathbb{V}$ with component maps $f_1,\ldots,f_d:\mathbb{A}\to\mathbb{R}$.

$$f:\mathbb{A}\to\mathbb{V}\quad\longleftrightarrow\quad\big(f_1,\ldots,f_d\big):\mathbb{A}\to\mathbb{R}^d;\qquad f_i(a)\coloneqq\langle f(a),e_i\rangle$$

Extend the inner-product symmetrically to a bilinear form on $\mathbb{V}\oplus\mathrm{i}\mathbb{V}$,

$$\big\langle v_1+\mathrm{i}w_1,v_2+\mathrm{i}w_2\big\rangle=\big(\langle v_1,v_2\rangle-\langle w_1,w_2\rangle\big)+\mathrm{i}\big(\langle v_1,w_2\rangle+\langle w_1,v_2\rangle\big),$$

and define the trace of an operator $T\in\mathbb{L}(\mathbb{V})$, as follows.

$$\operatorname{tr}(T)=\sum_{i=1}^{d}\langle e_i,Te_i\rangle$$

We adopt that $(\Omega,\Sigma,\mathrm{P})$ is an abstract probability space that—through the process of enlargement via Kolmogorov's extension theorem—we without loss of generality assume it is equipped with identifications of various quantities $X:\Omega\to\mathbb{X}$ into measurable spaces $(\mathbb{X},\mathscr{X})$ associated with distributions μ on $(\mathbb{X},\mathscr{X})$. We typically presume such maps X to be measurable without mention and will otherwise specify this fact explicitly by using the notation $X\in\Sigma/\mathscr{X}$. For each probability measure P on (Ω,Σ), we denote the P-distribution of such X by P_X or pushfoward notation, $X_\#\mathrm{P}$.

$$\mathrm{P}_X\Gamma\coloneqq(X_\#\mathrm{P})(\Gamma)\coloneqq\mathrm{P}(X\in\Gamma)\coloneqq\mathrm{P}\big(X^{-1}\Gamma\big),\quad\Gamma\in\mathscr{X}$$

For intuition, we will also denote integration against this distribution as follows.

$$\int_{\mathbb{X}}\mathrm{P}(X\in\mathrm{d}x)f(x)\coloneqq\int_{\mathbb{X}}\mathrm{P}_X(\mathrm{d}x)f(x)=\int_{\Omega}\mathrm{P}(\mathrm{d}\omega)f\big(X(\omega)\big)=:\mathrm{E}_{\mathrm{P}}f(X)$$

Just as E_{P} denotes the expectation operator of the measure P, we will denote $\mathrm{E}_{\mathrm{P}}(\cdot|\mathscr{G})$ the conditional expectation operator of P associated with a filtration $\mathscr{G}$. Should we choose a target space $(\mathbb{Y},\mathscr{Y})$ and a natural σ-algebra $Y^{-1}\mathscr{Y}$ from some quantity $Y\in\Sigma/\mathscr{Y}$, we denote $\mathrm{E}_{\mathrm{P}}(\cdot|Y=\cdot)$ the factoring of $\mathrm{E}_{\mathrm{P}}(\cdot|Y^{-1}\mathscr{Y})$ through Y.

$$\mathrm{E}_{\mathrm{P}}(X|Y=y)=\mathrm{E}_{\mathrm{P}}\big(X|Y^{-1}\mathscr{Y}\big)\Big|_{Y=y}$$

Also, any quantity $X:\Omega\to\mathbb{X}$ will be identified with the identity map on its codomain, so that we may abusingly use the convenient expectation notation.

$$\mathrm{E}_{\mathrm{P}_X}f(X)\coloneqq\mathrm{E}_{\mathrm{P}_X}f=\int_{\mathbb{X}}f(x)\mathrm{P}_X(\mathrm{d}x)=\int_{\Omega}f\big(X(\omega)\big)\mathrm{P}(\mathrm{d}\omega)=\mathrm{E}_{\mathrm{P}}f(X)$$

This will particularly be useful for when we discuss Markov processes and their associated identities.

Chapter 1

Affine processes

Our first chapter will get us familiar with affine processes. Most of the results will be simple consequences of [Cuc11] and [KRM15], which pretty much comprehensively prove everything there is to know about affine processes on convex state spaces. It is still important for us to have this chapter, for we will need to have a careful understanding of how the spaces of finite moments change with respect to time, and these facts only seem to exist in the case of affine diffusions. Furthermore, the magnificent papers above provide us so much information at the cost of being very formal. Readers which have not found themselves familiar with massive texts like [JS03] may not be able to decipher the results involving semimartingale characteristics in intuitive terms. It is for this reason that we found it beneficial to provide some calculus-heavy proofs that somehow *demystify* these objects. In particular, we translate many of the results of [JS03] to the special setting of *jump-diffusions*—those semimartingales in which the characteristics are differentiable. While certainly less powerful than their general semimartingale brethren, jump-diffusions involve deterministic functions which make calculations far more intuitive. Furthermore, they can always be understood as weak solutions to stochastic differential equations involving a Brownian motion and a Poisson random measure, which gives them a somewhat *generative* flavor. These results are mostly in Appendix A—a choice which was made less because the information is unrelated to affine processes, and more that the extra generality of non-affine differential characteristics comes at little-to-no cost. In any case, we suggest the reader familiarize themselves with Appendix A sooner than later, as we use this material very often throughout the chapters. The remainder of this chapter is organized as follows.

Section 1.1. Defines an affine process and identifies some relevant notions from which we begin our investigation.

Section 1.2. Explores consequences of the affine transform formula on real moments.

Section 1.3. Lifts the notions of the preceding section from marginals to finite-dimensional distributions.

Section 1.4. Explores more on the path-based properties—particularly the jump-diffusion nature—of affine processes.

1.1 Formulation

We start by specifying our affine processes as in [KRM15]. That is to say, we fix a finite-dimensional real vector space $\mathbb{V}$ with inner-product $\langle \cdot, \cdot \rangle$ and select a convex, closed $\mathbb{X} \subseteq \mathbb{V}$ satisfying $0 \in \mathbb{X}$ and $\operatorname{span} \mathbb{X} = \mathbb{V}$. Associate this space with the finite exponentials.

$$\mathcal{U}_{\mathbb{X}} := \left\{ u \in \mathbb{V} \oplus \mathrm{i}\mathbb{V} : \sup_{x \in \mathbb{X}} \exp \langle \Re(u), x \rangle < \infty \right\}$$

We may now define the notion of an affine process on $\mathbb{X}$.

Definition 1.1.1. *For a probability space $(\Omega, \Sigma, \mathrm{P})$ with filtration $\mathscr{F} = (\mathscr{F}_t)_{t \geq 0}$, an affine process X on $\mathbb{X}$ is a stochastically continuous, time-homogeneous $(\mathrm{P}, \mathscr{F})$-Markov process on $\mathbb{X}$ in which the bounded moments have the following log-affine dependence on the initial state.*

$$\begin{aligned} \mathrm{E}_{\mathrm{P}_x} \exp \langle u, X_t \rangle &= \exp \Psi(t, u, x) \\ \Psi(t, u, x) &= \psi_0(t, u) + \big\langle \psi(t, u), x \big\rangle, \end{aligned} \qquad t \geq 0,\ u \in \mathcal{U}_{\mathbb{X}} \tag{1.1.2}$$

Above, we are denoting $(\mathrm{P}_x)_{x \in \mathbb{X}}$ the conditional P-distributions of X factored through the initial state $x \in \mathbb{X}$.

Remark 1.1.3. *(a) See [KRM15, Remark 2.3] for an argument on how our assumptions on $\mathbb{X}$ are at no loss of generality; $\mathbb{X}$ may as well be any nonempty convex set.*

(b) Note how (1.1.2) specifies the characteristic function of each transition kernel of the Markov process X; thus, should an affine process exist for choice of Ψ, only *one will exist, up to distribution.*

(c) See how our notation (ψ_0, ψ) differs from that of [KRM15] and other papers, which typically use (ϕ, ψ). We choose to do this because affine functions prevail throughout our investigation of affine processes, and we saw this an opportunity to have more cohesive notation of all such affine functions.

$$\alpha(x) = a_0 + \sum_{i=1}^{d} x^i a_i$$

(d) If we have a vector space $\mathbb{A}$ and affine map $\alpha : \mathbb{X} \to \mathbb{A}$ determined by $a_0, \ldots, a_d \in \mathbb{A}$ via $\alpha(x) = a_0 + \sum_{i=1}^{d} x^i a_i$, then our linear assumptions $0 \in \mathbb{X}$ and $\operatorname{span} \mathbb{X} = \mathbb{V}$ uniquely determine $a_0, \ldots, a_d \in \mathbb{A}$. In particular, the map Ψ uniquely identifies its parts $\psi_i : \mathbb{R}_+ \times \mathcal{U}_{\mathbb{X}} \to \mathbb{C}$ for $i = 0, \ldots, d$.

In [Cuc11, Theorem 1.2.7], it is shown that, without loss of generality on conditional distributions $(\mathrm{P}_x)_{x \in \mathbb{X}}$, an affine process X can be chosen to have càdlàg paths. Thus, each distribution P_x may be recognized as a measure on the Borel algebra associated with the space $\mathbb{D}([0, \infty), \mathbb{X})$ of càdlàg functions equipped with the Skorokhod topology. By imposing this regularity, the following theorem tells us that an affine process X as in Definition 1.1.1 is a $(\mathrm{P}_x, \mathscr{F})$ jump-diffusion for each $x \in \mathbb{X}$. For relevant definitions and results pertaining to jump-diffusions, we refer the reader to Appendix A.

Theorem 1.1.4. *An affine process X on $\mathbb{X}$ is a $(\mathrm{P}_x, \mathscr{F})$ jump-diffusion in which the differential χ-characteristics $(\beta^\chi, \alpha, \mu)$ are affine maps of the following form.*

$$\beta^\chi(x) \coloneqq b_0^\chi + \sum_{i=1}^d x^i b_i^\chi, \quad \alpha(x) \coloneqq a_0 + \sum_{i=1}^d x^i a_i, \quad \mu(x, \mathrm{d}v) \coloneqq m_0(\mathrm{d}v) + \sum_{i=1}^d x^i m_i(\mathrm{d}v)$$

The associated Lévy-Khintchine map Λ *then also affine,*

$$\begin{aligned}\Lambda(u,x) &= \big\langle u, \beta^\chi(x)\big\rangle + \frac{1}{2}\big\langle u, \alpha(x)u\big\rangle + \int_{\mathbb{V}} \Big(e^{\langle u,v\rangle} - 1 - \big\langle u, \chi(v)\big\rangle\Big)\mu(x,\mathrm{d}v) \\ &= L_0(u) + \sum_{i=1}^d x^i L_i(u) \\ L_i(u) &\coloneqq \big\langle u, b_i^\chi(x)\big\rangle + \frac{1}{2}\big\langle u, a_i(x)u\big\rangle + \int_{\mathbb{V}} \Big(e^{\langle u,v\rangle} - 1 - \big\langle u, \chi(v)\big\rangle\Big)m_i(x,\mathrm{d}v),\end{aligned}$$

and each $u \in \mathrm{i}\mathbb{V}$ induces the following differential equation.

$$\begin{cases} \psi_0(t,u) = L_0\big(\psi(t,u)\big) & t \geq 0 \\ \psi(t,u) = L\big(\psi(t,u)\big) & t \geq 0 \\ \psi_0(0,u) = 0 \\ \psi(0,u) = u \end{cases} \tag{1.1.5}$$

Proof. This is simply a restatement of [Cuc11, Theorem 1.5.4].

Remark 1.1.6. *By Remark 1.1.3(d) and linearity of differentiation, the equation in (1.1.5) is equivalent to the following system of equations.*

$$\forall\ x \in \mathbb{X}, \qquad \begin{cases} \dot{\Psi}(t,u,x) = \Lambda\big(\psi(t,u),x\big) & t \geq 0 \\ \Psi(0,u,x) = \langle u, x\rangle \end{cases} \tag{1.1.7}$$

Henceforth, we fix X a càdlàg affine process with conditional distributions $(\mathrm{P}_x)_{x\in\mathbb{X}}$ on $\mathbb{D}([0,\infty),\mathbb{X})$, induced filtration $\mathscr{F} = (\mathscr{F}_t)_{t\geq 0}$, and moment function Ψ as in Definition 1.1.1. We will use the truncation function $\chi(v) = v1_{|v|\leq 1}$ and fix the differential χ-characteristics $(\beta^\chi, \alpha, \mu)$ and Lévy-Khintchine map Λ as in Theorem 1.1.4.

1.2 Existence of real moments

This section elaborates upon the extension of the transform formula in (1.1.2) and equations (1.1.5) and (1.1.7) to real moments $u \in \mathbb{V}$. Clearly, should any extension exist for some $u \in \mathbb{V}$, the value $\Lambda(u,x) = \dot{\Psi}(0,u,x)$ should be well-defined. Throughout this section, we recall our exploration in Section A.4 of the Lévy-Khintchine map Λ and its essential domain of real moments.

$$\mathcal{D}_\Lambda \coloneqq \Big\{u \in \mathbb{V} : \Lambda(u,x) \text{ is well-defined for all } x \in \mathbb{X}\Big\}$$

These will allow us to establish existence results of $\Psi(\cdot,u,\cdot)$ for real moments $u \in \mathbb{V}$. The following definition will get us started.

Definition 1.2.1. *For each $\tau \geq 0$ and $u \in \mathcal{D}_\Lambda$, we say a function $Q^u : [0,\tau] \times \mathbb{X} \to \mathbb{R}$ satisfies* system(Λ, τ, u) *if the following hold.*

$$\begin{aligned} &\forall\, t \in [0,\tau],\ x \in \mathbb{X}, \quad Q^u(t,x) = q_0^u(t) + \langle q^u(t), x\rangle, \\ &\forall\, x \in \mathbb{X}, \quad \begin{cases} \dot{Q}^u(t,x) = \Lambda\big(q^u(t), x\big), & t \in [0,\tau] \\ Q^u(0,x) = \langle u, x\rangle \end{cases} \end{aligned} \tag{1.2.2}$$

Now define the following sets.

$$\begin{aligned} \mathcal{D}_\Psi(\tau) &\coloneqq \Big\{ u \in \mathcal{D}_\Lambda : \textit{there exists a solution to } \operatorname{system}(\Lambda,\tau,u) \Big\} \\ \mathcal{D}_\Psi &\coloneqq \bigcup_{\tau \geq 0} \Big(\{\tau\} \times \mathcal{D}_\Psi(\tau) \Big) \end{aligned}$$

Theorem 1.2.3. *(a) There exists a map $\Psi : \mathcal{D}_\Psi \times \mathbb{X} \to \mathbb{R}$ of the form*

$$\Psi(t,u,x) = \psi_0(t,u) + \big\langle \psi(t,u), x \big\rangle$$

such that, for each $(\tau, u) \in \mathcal{D}_\Psi$, $\Psi(\cdot, u, \cdot)$ is a solution to system(Λ, τ, u) *dominated by all other such solutions. Moreover, this map satisfies the following for each $(\tau,u) \in \mathcal{D}_\Psi$ and $x \in \mathbb{X}$.*

$$\mathrm{E}_{\mathrm{P}_x} \exp \langle u, X_t \rangle = \exp \Psi(t,u,x), \quad t \in [0,\tau] \tag{1.2.4}$$

(b) If $\tau \geq 0$, $u \in \mathbb{V}$, and $x \in \mathbb{X}^\circ$ are such that $\mathrm{E}_{\mathrm{P}_x} \exp \langle u, X_\tau \rangle < \infty$, then $(\tau, u) \in \mathcal{D}_\Psi$.

Proof. With Remark 1.1.6, this is the same as [KRM15, Theorem 2.14].

Now that we have two characterizations for the space $\mathcal{D}_\Psi$, we seek to understand properties of it and the associated moment map $\Psi : \mathcal{D}_\Psi \times \mathbb{X} \to \mathbb{R}$.

Proposition 1.2.5. *(a) For each $\tau > 0$, $\mathcal{D}_\Psi(\tau)$ is open in $\mathcal{D}_\Lambda^\circ$,*

(b) For each $\tau > 0$ and $u \in \mathcal{D}_\Psi(\tau) \cap \mathcal{D}_\Lambda^\circ$, $\Psi(\cdot, u, \cdot)$ from Theorem 1.2.3 is the unique solution to system(Λ, τ, u).

(c) Ψ is continuously differentiable on $\mathcal{D}_\Psi^\circ \times \mathbb{X}$.

Proof. Fix $\tau > 0$ and $u \in \mathcal{D}_\Psi(\tau) \cap \mathcal{D}_\Lambda^\circ$. Because $u \in \mathcal{D}_\Psi(\tau)$, $\Psi(\cdot, u, \cdot)$ exists on $[0,\tau] \times \mathbb{X}$ as a solution to system(Λ, τ, u). As mentioned in Remark 1.1.6, the function $\psi(\cdot, u)$ associated with $\Psi(\cdot, u, \cdot)$ is a solution to the following equation,

$$\begin{cases} \dot{\psi}(t,u) = f\big(t, \psi(t,u)\big) & t \in [0,\tau] \\ \psi(0,u) = u \end{cases} \tag{1.2.6}$$

where $f : \mathbb{R} \times \mathcal{D}_\Lambda^\circ \to \mathbb{V}$ is defined by $f(t,u) \coloneqq L(u)$. Seeing as f is continuously differentiable on $\mathbb{R} \times \mathcal{D}_\Lambda^\circ$ by Lemma A.4.4, we may use [Wal98, III.13 Theorem X] to ensure some $\epsilon > 0$ such that the band

$$S_\epsilon \coloneqq \Big\{ (t,v) \in [0,\tau] \times \mathbb{V} : |v - \psi(t,u)| \leq \epsilon \Big\}$$

is contained in $\mathbb{R} \times \mathcal{D}_\Lambda^\circ$ and provides us to each $(t_0, v) \in S_\epsilon$ a unique solution $q(\cdot, t_0, v)$ to the following initial value problem,

$$\begin{cases} \dot{q}(t, t_0, v) = f\big(t, q(t, t_0, v)\big) & t \in [t_0, \tau] \\ q(t_0, t_0, v) = v \end{cases}$$

which is continuously differentiable with derivatives $\partial_{t_0} q(t, t_0, v) \in \mathbb{V}$ and $Dq(t, t_0, v) \in \mathbb{L}(\mathbb{V})$ satisfying the following equations.

$$\partial_{t_0} q(t, t_0, v) = -f(t_0, u) + \int_{t_0}^{t} Df\big(s, q(s, t_0, v)\big)\partial_{t_0} q(s, t_0, v)\mathrm{d}s$$

$$Dq(t, t_0, v) = \mathrm{id}_{\mathbb{V}} + \int_{t_0}^{t} Df\big(s, q(s, t_0, v)\big) Dq(s, t_0, v)\mathrm{d}s$$

In particular, for each $v \in B(u, \epsilon)$, we have $|v - \psi(0, u)| = |v - u| < \epsilon$, and so $(0, v) \in S_\epsilon$; this allows us to disregard the middle coordinate and have $q : [0, \tau] \times B(u, \epsilon) \to \mathbb{V}$ such that $q(\cdot, v)$ is the unique solution to

$$\begin{cases} \dot{q}(t, v) = L\big(q(t, v)\big), & t \in [0, \tau] \\ q(0, v) = v \end{cases}$$

and the derivative in the second coordinate $Dq(t, v) \in \mathbb{L}(\mathbb{V})$ satisfies the following equation.

$$Dq(t, v) = \mathrm{id}_{\mathbb{V}} + \int_0^t DL\big(q(s, v)\big) Dq(s, v)\mathrm{d}s$$

From here, we may define $Q : [0, \tau] \times B(u, \epsilon) \times \mathbb{X} \to \mathbb{R}$ as follows.

$$\begin{aligned} Q(t, v, x) &\coloneqq q_0(t, v) + \big\langle q(t, v), x \big\rangle \\ q_0(t, v) &\coloneqq \int_0^t L_0\big(q(s, v)\big)\mathrm{d}s \\ L_0(v) &\coloneqq \Lambda(v, 0) \end{aligned}$$

Because the image of $q(\cdot, v)$ on $[0, \tau]$ remains in $\mathcal{D}_\Lambda^\circ$, on which L is continuously differentiable, q_0 is continuously differentiable with derivatives $\dot{q}_0$ and Dq_0 satisfying the following.

$$\begin{aligned} \dot{q}_0(t, v) &= L_0\big(q(s, v)\big) \\ Dq_0(t, v) &= \int_0^t DL_0\big(q(s, v)\big) Dq(s, v)\mathrm{d}s \end{aligned}$$

By linearity, $Q(\cdot, v, \cdot)$ is a solution to system(Λ, τ, v) and so $v \in \mathcal{D}_\Psi(\tau)$. We now have $B(u, \epsilon) \subseteq \mathcal{D}_\Psi(\tau)$, concluding part (a). Meanwhile, any solution Q^u to system(Λ, τ, u) is such that the associated q^u solves (1.2.6) and so $q^u = q(\cdot, u)$. From here, it is thus the case that $Q^u = Q(\cdot, u, \cdot)$. This means Ψ from Theorem 1.2.3 is such that $\Psi(\cdot, u, \cdot)$ is the unique solution to system(Λ, u, τ), concluding part (b). Lastly, for each $x \in \mathbb{X}$, linearity also shows us that $\Psi(\cdot, \cdot, x)$ is continuously differentiable in a neighborhood of (t, u), with derivative in the second coordinate $D\Psi(\cdot, \cdot, x)$ satisfying the following.

$$D\Psi(t, u, x) = D\psi_0(t, u) + D\psi(t, u) \cdot x$$

$$
\begin{aligned}
&= Dq_0(t,u) + Dq(t,u)\cdot x \\
&= \int_0^t DL_0\big(q(s,u)\big)Dq(s,u)\mathrm{d}s + \left(\mathrm{id}_{\mathbb{V}} + \int_0^t DL\big(q(s,u)\big)Dq(s,u)\mathrm{d}s\right)\cdot x \\
&= x + \int_0^t \Big(DL_0\big(q(s,u)\big)Dq(s,u) + \sum_{i=1}^d x_i DL_i\big(q(s,u)\big)Dq(s,u)\Big)\mathrm{d}s \\
&= x + \int_0^t D\Big(L_0 + \sum_{i=1}^d x_i L_i\Big)\big(q(s,u)\big)Dq(s,u)\mathrm{d}s \\
&= x + \int_0^t D\Lambda\big(q(s,u),x\big)Dq(s,u)\mathrm{d}s \\
&= x + \int_0^t D\Lambda\big(\psi(s,u),x\big)D\psi(s,u)\mathrm{d}s
\end{aligned}
$$

This concludes part (c).

Proposition 1.2.7. *For each compact set $K \subseteq \mathcal{D}_\Lambda^\circ$, there exists $\delta > 0$ such that $K \subseteq \mathcal{D}_\Psi(\delta)$. Moreover, $\Psi(\cdot,u,\cdot)$ from Theorem 1.2.3 is the unique solution to* system(Λ,δ,u) *for each $u \in K$.*

Proof. Firstly, we recognize that by virtue of $K \subseteq \mathcal{D}_\Lambda^\circ$ being compact, we have some $\epsilon > 0$ such that the associated open set

$$K^\epsilon := \Big\{u \in \mathbb{V} : \inf_{u' \in K} |u - u'| < \epsilon\Big\}$$

has closure $\overline{K^\epsilon}$ contained in $\mathcal{D}_\Lambda^\circ$. Note in particular that this provides us with a buffer of radius ϵ around each point in $\mathcal{D}_\Lambda^\circ$.

$$\overline{B}(u,\epsilon) := \Big\{u' \in \mathbb{V} : |u' - u| \le \epsilon\Big\}$$

$$\bigcup_{u \in \mathcal{D}_\Lambda} \overline{B}(u,\epsilon) \subseteq \overline{K^\epsilon} \subseteq \mathcal{D}_\Lambda^\circ$$

With these sets established, we mitigate the task of finding a solution Q^u to system(Λ,δ,u) to that of finding a solution q^u to the related equation.

$$\left\{\begin{array}{ll} \dot{q}^u(t) = L\big(q^u(t)\big) & t \in [0,\delta] \\ q^u(0) = u & \end{array}\right. \tag{1.2.8}$$

For a fixed $u \in \mathcal{D}_\Lambda^\circ$, the existence of some $\delta_u > 0$ and solution q^u to (1.2.8) may be obtained from the usual fixed-point method (see [Wal98, II.6 Theorem III]). Indeed, Remark 1.1.3(d) and Lemma A.4.4 provide us a Lipschitz property for L on $\overline{B}(u,\epsilon)$,

$$\big|L(v) - L(w)\big| \le |v - w| C_{u,\epsilon}, \qquad v, w \in \overline{B}(u,\epsilon)$$

$$C_{u,\epsilon} := \sup_{u' \in \overline{B}(u,\epsilon)} \Big|DL\big(u',x\big)\Big|$$

and so a Banach space $(\mathbb{B}_u, \|\cdot\|_{\mathbb{B}_u})$ defined by

$$\delta_u := 1 \wedge \frac{\epsilon}{\sup_{u' \in \overline{B}(u,\epsilon)} |L(u')|}$$

$$\mathbb{B}_u := \mathbb{C}([0,\delta_u],\mathbb{V})$$
$$\|f\|_{\mathbb{B}_u} := \sup_{t\in[0,\delta_u]} |f(t)|e^{-2C_{u,\epsilon}t}$$

is partially equipped with a map $T : \mathbb{C}([0,\delta_u],K) \to \mathbb{C}([0,\delta_u],\overline{K^\epsilon})$ defined by

$$Tf(t) := u + \int_0^t L\big(f(s)\big)\mathrm{d}s,$$

satisfying a contraction property,

$$\|Tf - Tg\|_{\mathbb{B}_u} \le \frac{1}{2}\|f-g\|_{\mathbb{B}_u},$$

which induces a unique solution $q^u \in \mathbb{C}([0,\delta_u],\overline{K^\epsilon})$ to the associated fixed-point equation, $Tq^u = q^u$. This solution q^u is thus a unique solution to (1.2.8).

From here, we define the following positive δ,

$$\delta := \inf_{u\in K}\delta_u \ge 1 \wedge \inf_{u\in K}\frac{\epsilon}{\sup_{u'\in\overline{B}(u,\epsilon)}|L(u')|} \ge 1\wedge \frac{\epsilon}{\sup_{u'\in\overline{K}^\epsilon}|L(u')|} > 0$$

so that each $u \in K$ has a unique solution q^u to (1.2.8). This induces the following map $Q^u : [0,\delta]\times\mathbb{X}\to\mathbb{R}$ for each $u\in K$.

$$Q^u(t,x) := q_0^u(t) + \big\langle q^u(t), x\big\rangle$$
$$q_0^u(t) := \int_0^t L_0\big(q^u(s)\big)\mathrm{d}s$$

By linearity, Q^u is a solution to system(Λ,δ,u) for each $u\in K$, and so $K\subseteq\mathcal{D}_\Psi(\delta)$. For each $u\in K\subseteq\mathcal{D}_\Psi(\delta)$, a solution $\tilde{Q}^u$ to system(Λ,δ,u) is such that the associated $\tilde{q}^u$ solves (1.2.8) and so $\tilde{q}^u = q^u$. From here, it is thus the case that $\tilde{Q}^u = Q^u$. This means Ψ from Theorem 1.2.3 is such that $\Psi(\cdot,u,\cdot)$ is the unique solution to system(Λ,u,δ) for all $u\in K$.

Proposition 1.2.9. *For any compact subset $K\subseteq\mathcal{D}_\Psi^\circ$, there exists $C_K>0$ such that the following holds for all $(t,u)\in K$.*

$$\big|\Psi(t,u,x)-\Psi(0,u,x)\big| \le C_K\cdot t\cdot\big(1+|x|\big)$$

Proof. Let $K\subseteq\mathcal{D}_\Psi^\circ$ be compact. By Remark 1.1.6 and Proposition 1.2.5(c), we have that the functions ψ_i for $i=0,\dots,d$ are continuously differentiable on $\mathcal{D}_\Psi^\circ$. Thus, we may define the following positive numbers.

$$C_{K,i} := \sup_{(t,u)\in K}|\dot{\psi}_i(t,u)|,\quad i=0,\dots,d$$
$$C_K := \max\Big\{C_{K,0}, C_{K,1}\sqrt{d},\dots,C_{K,d}\sqrt{d}\Big\} < \infty$$

Using the fundamental theorem of calculus and that $\Psi(\cdot,u,\cdot)$ solves system(Λ,τ,u), we produce the following bound for all $(t,u)\in K$.

$$\big|\Psi(t,u,x)-\Psi(0,u,x)\big| = \Big|\psi_0(t,u) + \big\langle\psi(t,u)-u,x\big\rangle\Big|$$

$$
\begin{aligned}
&\leq \big|\psi_0(t,u)\big| + \big|\psi(t,u) - u\big| \cdot |x| \\
&= \left|\int_0^t \dot{\psi}_0(s,u)\mathrm{d}s\right| + \left|\int_0^t \dot{\psi}_i(s,u)\mathrm{d}s\right| \cdot |x| \\
&\leq C_{K,0} \cdot t + \Big(\sum_{i=1}^d C_{K,i}^2\Big)^{1/2} \cdot t \cdot |x| \\
&\leq C_K \cdot t \cdot \big(1 + |x|\big)
\end{aligned}
$$

While we will not need the following result until the next chapter, we find it important to put it here, as a *reality check* with our spaces $\mathcal{D}_\Lambda$ and $\mathcal{D}_\Psi$.

Proposition 1.2.10. *We have $\mathcal{D}_\Lambda = \mathbb{V}$ if and only if each $u \in \mathbb{V}$ is such that there exists $\tau > 0$ in which u is a finite moment of the associated marginal X_τ.*

$$\mathrm{E}_{\mathrm{P}_x} \exp\langle u, X_\tau\rangle < \infty, \quad x \in \mathbb{X}$$

Proof. For sufficiency, first see that any $u \in \mathbb{V}$ is such that $\overline{B}(u,1) \subseteq \mathbb{V} = \mathcal{D}_\Lambda^\circ$ is compact. By Proposition 1.2.7, we then have some $\tau > 0$ so that $u \in \overline{B}(u,1) \subseteq \mathcal{D}_\Psi(\tau)$. Then, Theorem 1.2.3(a) gives us the desired finiteness condition.

$$\mathrm{E}_{\mathrm{P}_x} \exp\langle u, X_\tau\rangle = \exp\Psi(\tau,u,x) < \infty, \quad x \in \mathbb{X}$$

Conversely, if $u \in \mathbb{V}$ satisfies $\mathrm{E}_{\mathrm{P}_x} \exp\langle u, X_\tau\rangle < \infty$ for all $x \in \mathbb{X}$, then it certainly does for some $x \in \mathbb{X}^\circ$, and so Theorem 1.2.3(b) gives us $(\tau,u) \in \mathcal{D}_\Psi$. Surely, this then means that $u \in \mathcal{D}_\Psi(\tau) \subseteq \mathcal{D}_\Lambda$.

1.3 Finite-dimensional distributions

With a good grasp of the finite real moments associated with our affine process X and their correspondence with Ψ, we now leverage these results to the finite-dimensional distributions. In other words, this section serves to lift Theorem 1.2.4 on marginals X_t to one on finite-dimensional distributions $(X_{t_1},\ldots,X_{t_n})$. Let us establish some notation.

For any space $\mathbb{A}$, positive integer $n \in \mathbb{N}$, and $\underline{a} \in \mathbb{A}^n$, adopt the convention of denoting $\underline{a} = (a_1,\ldots,a_n)$ and

$$\underline{a}_{\ell:m} = (a_\ell,\ldots,a_m) \in \mathbb{A}^{m-\ell+1}, \quad 1 \leq \ell \leq m \leq n.$$

For each $n \in \mathbb{N}$ and $\underline{t} \in [0,\infty)^n$, define the projection map $\pi_{\underline{t}} : \mathbb{X}^{[0,\infty)} \to \mathbb{X}^n$ by

$$\pi_{\underline{t}}(\xi) \coloneqq \xi(\underline{t}) \coloneqq \big(\xi(t_1),\ldots,\xi(t_n)\big).$$

Denote $\underline{t} \vdash [0,\infty)$ to mean that $\underline{t}$ is additionally a partition of the following form.

$$0 < t_1 < \cdots < t_n$$

For each such partition $\underline{t} \vdash [0,\infty)$, associate the following notation.

$$
\begin{aligned}
t_0 &\coloneqq 0 \\
\Delta t_k &\coloneqq t_k - t_{k-1}, && 1 \leq k \leq n
\end{aligned}
$$

$$|\underline{t}| \coloneqq n$$

Lastly, for any $A \subseteq [0,\infty)$, denote $\underline{t} \vdash A$ to mean $\underline{t} \vdash [0,\infty)$ and $t_1, \ldots, t_{|\underline{t}|} \in A$. For each $n \in \mathbb{N}$, extend the linear operations of $\mathbb{V}$ to $\mathbb{V}^n$, componentwise. Similarly, extend the definition of our inner-product on $\mathbb{V} \oplus \mathrm{i}\mathbb{V}$ to one on $(\mathbb{V} \oplus \mathrm{i}\mathbb{V})^n$, like so.

$$\langle \underline{u}, \underline{v} \rangle \coloneqq \sum_{k=1}^{n} \langle u_k, v_k \rangle$$

We now clearly specify the extension of Ψ to finite-dimensional projections from the perspective of Theorem 1.2.3 and equation (1.2.4). Note that this specifically *permits* infinite values.

Definition 1.3.1. *To each $\underline{t} \vdash [0,\infty)$, define $\Psi(\underline{t}, \cdot, \cdot) : (\mathbb{V} \oplus \mathrm{i}\mathbb{V})^{|\underline{t}|} \times \mathbb{X} \to (-\infty, \infty]$ as the cumulant generating function of $X_{\underline{t}}$.*

$$\mathrm{E}_{\mathrm{P}_x} \exp \langle \underline{u}, X_{\underline{t}} \rangle \eqqcolon \exp \Psi(\underline{t}, \underline{u}, x)$$

Note that this extends the definition of Ψ in that we may always consider some time $t > 0$ as a partition $t \vdash [0,\infty)$.

Before we investigate real moments, let us establish the easier result on purely complex moments. This will give us intuition for the objects we create in the sequel.

Lemma 1.3.2. *For each $t \geq 0$ and $u \in \mathcal{U}_{\mathbb{X}}$, we have $\psi(t,u) \in \mathcal{U}_{\mathbb{X}}$.*

Proof. This is an immediate result of Definition 1.1.1. For any $x \in \mathbb{X}$, we have the following uniform bound.

$$\begin{aligned}
\exp \Re\big\langle \psi(t,u), x \big\rangle &= \left| e^{\langle \psi(t,u), x \rangle} \right| \\
&= \left| e^{-\psi_0(t,u) + \Psi(t,u,x)} \right| \\
&= \left| e^{-\psi_0(t,u)} \right| \cdot \left| \mathrm{E}_{\mathrm{P}_x} \exp \langle u, X_t \rangle \right| \\
&\leq \left| e^{-\psi_0(t,u)} \right| \cdot \mathrm{E}_{\mathrm{P}_x} \left| \exp \langle u, X_t \rangle \right| \\
&\leq \left| e^{-\psi_0(t,u)} \right| \cdot \sup_{x' \in \mathbb{X}} \exp \Re \langle u, x' \rangle
\end{aligned}$$

Proposition 1.3.3. *For any $\underline{t} \vdash [0,\infty)$, $\underline{u} \in \mathrm{i}\mathbb{V}^{|\underline{t}|}$, and $x \in \mathbb{X}$, we have the following identity, where we denote $n \coloneqq |\underline{t}|$ for brevity.*

$$\begin{aligned}
&\theta_n \coloneqq u_n \\
&\theta_k \coloneqq u_k + \psi(\Delta t_{k+1}, \theta_{k+1}), \quad k = n-1, \ldots, 1 \\
&\Psi(\underline{t}, \underline{u}, x) = \sum_{k=1}^{|\underline{t}|} \psi_0(\Delta t_k, \theta_k) + \big\langle \psi(\Delta t_k, \theta_k), x \big\rangle
\end{aligned}$$

Proof. We start by recognizing that $\underline{u} \in \mathrm{i}\mathbb{V}$ means the following identity.

$$\left| e^{\langle u_k, x \rangle} \right| = \exp \Re \langle u_k, x \rangle = 1, \quad k = 1, \ldots, n$$

In particular, we have $\theta_n = u_n \in \mathcal{U}_{\mathbb{X}}$; we show $\theta_k \in \mathcal{U}_{\mathbb{X}}$ for $k = n-1, \dots, 1$ by induction and Lemma 1.3.2.

$$\begin{aligned}
\sup_{x\in\mathbb{X}} \exp \Re\langle \theta_k, x\rangle &= \sup_{x\in\mathbb{X}} \left| e^{\langle \theta_k, x\rangle} \right| \\
&= \sup_{x\in\mathbb{X}} \left| e^{\langle u_k + \psi(\Delta t_{k+1}, \theta_{k+1}), x\rangle} \right| \\
&= \sup_{x\in\mathbb{X}} \left| e^{\langle u_k, x\rangle} \right| \cdot \left| e^{\langle \psi(\Delta t_{k+1}, \theta_{k+1}), x\rangle} \right| \\
&= \sup_{x\in\mathbb{X}} \exp \Re\langle \psi(\Delta t_{k+1}, \theta_{k+1}), x\rangle < \infty
\end{aligned}$$

Now observe the following identity.

$$\begin{aligned}
&\Psi(\underline{t}, \underline{u}, x) \\
&= \log \mathrm{E}_{\mathrm{P}_x} \exp \langle \underline{u}, X_{\underline{t}} \rangle \\
&= \log \mathrm{E}_{\mathrm{P}_x} \left(\exp \sum_{k=1}^{n-1} \langle u_k, X_{t_k} \rangle \cdot \exp \langle \theta_n, X_{t_n} \rangle \right) \qquad (1.3.4) \\
&= \log \mathrm{E}_{\mathrm{P}_x} \left(\exp \sum_{k=1}^{n-1} \langle u_k, X_{t_k} \rangle \cdot \mathrm{E}_{\mathrm{P}_x} \Big(\exp \langle \theta_n, X_{t_n} \rangle | \mathscr{F}_{t_{n-1}} \Big) \right) \\
&= \log \mathrm{E}_{\mathrm{P}_x} \left(\exp \sum_{k=1}^{n-1} \langle u_k, X_{t_k} \rangle \cdot \exp \Psi(\Delta t_n, \theta_n, X_{t_{n-1}}) \right) \\
&= \psi_0(\Delta t_n, \theta_n) \\
&\quad + \log \mathrm{E}_{\mathrm{P}_x} \left(\exp \sum_{k=1}^{n-1} \langle u_k, X_{t_k} \rangle \cdot \exp \Big(\big\langle \psi(\Delta t_n, \theta_n), X_{t_{n-1}} \big\rangle \Big) \right) \\
&= \psi_0(\Delta t_n, \theta_n) \\
&\quad + \log \mathrm{E}_{\mathrm{P}_x} \left(\exp \sum_{k=1}^{n-2} \langle u_k, X_{t_k} \rangle \cdot \exp \Big(\big\langle u_{n-1} + \psi(\Delta t_n, \theta_n), X_{t_{n-1}} \big\rangle \Big) \right) \\
&= \psi_0(\Delta t_n, \theta_n) \qquad (1.3.5) \\
&\quad + \log \mathrm{E}_{\mathrm{P}_x} \left(\exp \sum_{k=1}^{n-2} \langle \theta_k - \psi(\Delta t_{k+1}, \theta_{k+1}), X_{t_k} \rangle \cdot \exp \Big(\big\langle \theta_{n-1}, X_{t_{n-1}} \big\rangle \Big) \right)
\end{aligned}$$

The final term of (1.3.4) is identical to that of (1.3.5) where we have reduced $k = n$ to $k = n-1$. Repeating these equalities inductively $k = n-1, \dots, 1$ will result in the desired identity.

$$\Psi(\underline{t}, \underline{u}, x) = \sum_{k=2}^{n} \psi_0(\Delta t_k, \theta_k) + \log \mathrm{E}_{\mathrm{P}_x} \exp \langle \theta_1, X_{t_1} \rangle = \sum_{k=1}^{n} \psi_0(\Delta t_k, \theta_k) + \big\langle \psi(\Delta t_1, \theta_1), x \big\rangle$$

As the preceding result shows, the $\mathbb{X}$-affine structure of Ψ allows us to iteratively factor the exponentials in our expectation. The problem with extending this to real moments like in Theorem 1.2.3 is that each of the produced quantities θ_k need not produce an integrable exponential on which we apply the transform formula. The next result is our way of coercing such a property to occur; the map $U_{\underline{t}}$ serves to parameterize those moments $\underline{u} \in \mathbb{V}^{|\underline{t}|}$ which the resulting $\underline{\theta}$ is in $\prod_{k=1}^{|\underline{t}|} \mathcal{D}_\Psi(\Delta t_k)$, since this is precisely the set on which we may perform the calculations between (1.3.4) and (1.3.5). This set turns out to be important in our discussion, so we will reserve it special notation.

$$\mathcal{D}_\Psi(\underline{t}) \coloneqq \prod_{k=1}^{|\underline{t}|} \mathcal{D}_\Psi(\Delta t_k), \quad \underline{t} \vdash [0,\infty)$$

Proposition 1.3.6. *For each $\underline{t} \vdash [0,\infty)$, the following map $U_{\underline{t}}$ is a continuous injection, where we denote $n \coloneqq |\underline{t}|$ for brevity.*

$$U_{\underline{t}} : \mathcal{D}_\Psi(\underline{t}) \to \mathbb{V}^{|\underline{t}|}, \quad U_{\underline{t}}(\underline{\theta}) \coloneqq \Big(\theta_1 - \psi(\Delta t_2, \theta_2), \dots, \theta_{n-1} - \psi(\Delta t_n, \theta_n), \theta_n\Big)$$

Moreover, for each $x \in \mathbb{X}$ and $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$, we have the following (finite) identity.

$$\Psi\big(\underline{t}, U_{\underline{t}}(\underline{\theta}), x\big) = \sum_{k=1}^{|\underline{t}|} \psi_0(\Delta t_k, \theta_k) + \big\langle \psi(\Delta t_1, \theta_1), x \big\rangle \tag{1.3.7}$$

Proof. Fix $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$. By definition, this means that to each $k = 1, \dots, |\underline{t}|$, we have $\theta_k \in \mathcal{D}_\Psi(\Delta t_k)$, and so $\psi(\Delta t_k, \theta_k)$ is well-defined. This ensures that $U_{\underline{t}}$ is well-defined. Now select another point $\underline{\theta}' \in \mathcal{D}_\Psi(\underline{t})$ such that $U_{\underline{t}}(\underline{\theta}) = U_{\underline{t}}(\underline{\theta}')$. The final component of $U_{\underline{t}}$ ensures that $\theta_n = \theta_n'$; by means of induction, we then get $\theta_{k-1} = \theta_{k-1}'$ for $k = n, \dots, 2$, via the equality on the respective component map.

$$\theta_{k-1} - \psi(\Delta t_k, \theta_k) = U_{\underline{t},k-1}(\underline{\theta}) = U_{\underline{t},k-1}(\underline{\theta}') = \theta_{k-1}' - \psi(\Delta t_k, \theta_k')$$

This indicates to us that $U_{\underline{t}}$ is an injection, and continuity comes simply from continuity of each $\psi(\Delta t_k, \cdot)$ via Proposition 1.2.5(c).

It now remains to show the identity in (1.3.7). This reduces down to repeatedly applying iterated expectations; fix $x \in \mathbb{X}$ and observe the following.

$$\begin{aligned}
&\Psi\big(\underline{t}, U_{\underline{t}}(\underline{\theta}), x\big) \\
&= \log \mathrm{E}_{\mathrm{P}_x} \exp \big\langle U_{\underline{t}}(\underline{\theta}), X_{\underline{t}} \big\rangle \\
&= \log \mathrm{E}_{\mathrm{P}_x} \left(\exp \sum_{k=1}^{n-1} \big\langle \theta_k - \psi(\Delta t_{k+1}, \theta_{k+1}), X_{t_k} \big\rangle \cdot \exp \big\langle \theta_n, X_{t_n} \big\rangle \right) \qquad (1.3.8) \\
&= \log \mathrm{E}_{\mathrm{P}_x} \left(\exp \sum_{k=1}^{n-1} \big\langle \theta_k - \psi(\Delta t_{k+1}, \theta_{k+1}), X_{t_k} \big\rangle \cdot \mathrm{E}_{\mathrm{P}_x} \Big(\exp \big\langle \theta_n, X_{t_n} \big\rangle | \mathscr{F}_{t_{n-1}} \Big) \right) \\
&= \log \mathrm{E}_{\mathrm{P}_x} \left(\exp \sum_{k=1}^{n-1} \big\langle \theta_k - \psi(\Delta t_{k+1}, \theta_{k+1}), X_{t_k} \big\rangle \cdot \exp \Psi(\Delta t_n, \theta_n, X_{t_{n-1}}) \right) \\
&= \psi_0(\Delta t_n, \theta_n) \\
&\quad + \log \mathrm{E}_{\mathrm{P}_x} \left(\exp \sum_{k=1}^{n-1} \big\langle \theta_k - \psi(\Delta t_{k+1}, \theta_{k+1}), X_{t_k} \big\rangle \cdot \exp \Big(\big\langle \psi(\Delta t_n, \theta_n), X_{t_{n-1}} \big\rangle \Big) \right) \\
&= \psi_0(\Delta t_n, \theta_n) \qquad (1.3.9) \\
&\quad + \log \mathrm{E}_{\mathrm{P}_x} \left(\exp \sum_{k=1}^{n-2} \big\langle \theta_k - \psi(\Delta t_{k+1}, \theta_{k+1}), X_{t_k} \big\rangle \cdot \exp \Big(\big\langle \theta_{n-1}, X_{t_{n-1}} \big\rangle \Big) \right)
\end{aligned}$$

The final term of (1.3.9) is identical to that of (1.3.8), where we have reduced $k = n$ to $k = n-1$. Repeating these equalities inductively $k = n-1, \dots, 1$ will result in the desired identity.

$$\Psi\big(\underline{t}, U_{\underline{t}}(\underline{\theta}), x\big) = \sum_{k=2}^{n} \psi_0(\Delta t_k, \theta_k) + \log \mathrm{E}_{\mathrm{P}_x} \exp\langle \theta_1, X_{t_1}\rangle = \sum_{k=1}^{n} \psi_0(\Delta t_k, \theta_k) + \big\langle \psi(\Delta t_1, \theta_1), x\big\rangle$$

We now turn to the analogue of Theorem 1.2.3(b), in which P_x-finite moments $\underline{u} \in \mathbb{V}^{|\underline{t}|}$ for initial points $x \in \mathbb{X}^\circ$ are precisely those $\underline{u} \in \mathcal{D}_\Psi(\underline{t})$.

Proposition 1.3.10. *Fix $\underline{t} \vdash [0,\infty)$ and denote $n \coloneqq |\underline{t}|$ for brevity. If $\underline{u} \in \mathbb{V}^{|\underline{t}|}$ is such that $\Psi(\underline{t}, \underline{u}, x) < \infty$ for some $x \in \mathbb{X}^\circ$, then the following recursion holds.*

$$\begin{aligned} &\theta_n \coloneqq u_n \in \mathcal{D}_\Psi(\Delta t_n) \\ &\theta_k \coloneqq u_k + \psi\big(\Delta t_{k+1}, \theta_{k+1}\big) \in \mathcal{D}_\Psi(\Delta t_k), \quad k = n-1, \dots, 1 \end{aligned} \tag{1.3.11}$$

Proof. Consider $\underline{u} \in \mathbb{V}^{|\underline{t}|}$ from which we may not construct the recursion in (1.3.11). In other words, there exists maximal $j \in \{1, \dots, n\}$ in the recursion which fails; i.e. $\theta_k \in \mathcal{D}_\Psi(\Delta t_k)$ for all $k = n, \dots, j+1$ and $\theta_j \notin \mathcal{D}_\Psi(\Delta t_j)$. We now repeat the work as in (1.3.8)-(1.3.9) for a fixed $x \in \mathbb{X}^\circ$ to get the following identity.

$$\begin{aligned}
&\log \mathrm{E}_{\mathrm{P}_x} \exp\langle \underline{u}, X_{\underline{t}}\rangle \\
&= \log \mathrm{E}_{\mathrm{P}_x}\Bigg(\exp\Big(\sum_{k=1}^{n-1} \langle u_k, X_{t_k}\rangle \Big) \cdot \mathrm{E}_{\mathrm{P}_x}\Big(\exp\big\langle u_n, X_{t_n}\big\rangle \big| \mathscr{F}_{t_{n-1}} \Big)\Bigg) \\
&= \psi_0(\Delta t_n, u_n) \\
&\qquad + \log \mathrm{E}_{\mathrm{P}_x}\Bigg(\exp\Big(\sum_{k=1}^{n-2} \langle u_k, X_{t_k}\rangle \Big) \cdot \mathrm{E}_{\mathrm{P}_x}\Big(\exp\big\langle u_{n-1} + \psi(\Delta t_n, u_n), X_{t_{n-1}}\big\rangle \big| \mathscr{F}_{t_{n-2}} \Big)\Bigg) \\
&= \psi_0\big(\Delta t_n, \theta_n\big) \\
&\qquad + \log \mathrm{E}_{\mathrm{P}_x}\Bigg(\exp\Big(\sum_{k=1}^{n-2} \langle u_k, X_{t_k}\rangle \Big) \cdot \mathrm{E}_{\mathrm{P}_x}\Big(\exp\big\langle \theta_{n-1}, X_{t_{n-1}}\big\rangle \big| \mathscr{F}_{t_{n-2}} \Big)\Bigg) \\
&\ \ \vdots \\
&= \sum_{k=j+1}^{n} \psi_0(\Delta t_k, \theta_k) + \log \mathrm{E}_{\mathrm{P}_x}\Bigg(\exp\Big(\sum_{k=1}^{j-1} \langle u_k, X_{t_k}\rangle \Big) \cdot \mathrm{E}_{\mathrm{P}_x}\Big(\exp\big\langle \theta_j, X_{t_j}\big\rangle \big| \mathscr{F}_{t_{j-1}} \Big)\Bigg) \\
&= \sum_{k=j+1}^{n} \psi_0(\Delta t_k, \theta_k) + \log \mathrm{E}_{\mathrm{P}_x}\Bigg(\exp\Big(\sum_{k=1}^{j-1} \langle u_k, X_{t_k}\rangle \Big) \cdot \mathrm{E}_{\mathrm{P}_{X_{t_{j-1}}}} \exp\big\langle \theta_j, X_{\Delta t_j}\big\rangle \Bigg)
\end{aligned}$$

By Theorem 1.2.3, since $\theta_j \notin \mathcal{D}_\Psi(\Delta t_j)$, we have $\mathrm{E}_{\mathrm{P}_{x'}} \exp\langle \theta_j, X_{\Delta t_j}\rangle = \infty$ for all $x' \in \mathbb{X}^\circ$, so the above integrand is infinite on the set $X_{t_{j-1}} \in \mathbb{X}^\circ$. Seeing as $x \in \mathbb{X}^\circ$, this set is P_x non-negligible, and so the quantity is infinite. We conclude that $\underline{u} \notin \mathcal{D}_\Psi(\underline{t})$, which finishes the proof by contrapositive.

Our final result of this section explores more on how finite moments $\underline{u}$ of $X_{\underline{t}}$ relate to those $\underline{\theta}$ of the increments $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$. To see this, we define the following increment cumulant generating function,

$$\varphi(t, \theta, x) \coloneqq \log \mathrm{E}_{\mathrm{P}_x} \exp\langle \theta, X_t - x\rangle = \Psi(t, \theta, x) - \langle \theta, x\rangle$$

Theorem 1.3.12. *Fix $\underline{t} \vdash [0,\infty)$ and $x_0 \in \mathbb{X}^\circ$. The map $U_{\underline{t}}$ is a homeomorphism from $\mathcal{D}_\Psi(\underline{t})$ to the collection of $\underline{u} \in \mathbb{V}^{|\underline{t}|}$ for which $\Psi(\underline{t},\underline{u},x_0) < \infty$. In particular, this means $\underline{u} \in \mathbb{V}^{|\underline{t}|}$ satisfies $\Psi(\underline{t},\underline{u},x_0) < \infty$ if and only if $\underline{u} = U_{\underline{t}}(\underline{\theta})$. Moreover, we have the following identity for all $\underline{x} \in \mathbb{X}^{|\underline{t}|}$.*

$$\langle \underline{u}, \underline{x} \rangle - \Psi(\underline{t},\underline{u},x_0) = \sum_{k=1}^{|\underline{t}|} \Big(\langle \theta_k, x_k - x_{k-1} \rangle - \varphi(\Delta t_k, \theta_k, x_{k-1}) \Big), \quad \underline{u} = U_{\underline{t}}(\underline{\theta})$$

Proof. By Proposition 1.3.6, we have that U is indeed a continuous map from $\mathcal{D}_\Psi(\underline{t})$ into the finite domain of $\Psi(\underline{t},\cdot,x_0)$. Conversely, Proposition 1.3.10 indicates to us that, on the finite domain of $\Psi(\underline{t},\cdot,x_0)$, a recursively-defined map $T_{\underline{t}}$ from (1.3.11) exists. Denoting $n := |\underline{t}|$, we see that this map is continuous by induction and continuity of compositions.

$$T_{\underline{t}}(\underline{u}) = \Big(T_{\underline{t},1}(\underline{u}_{1:n}), \ldots, T_{\underline{t},n}(\underline{u}_{n:n}) \Big), \quad \begin{aligned} T_{\underline{t},n}(\underline{u}_{n:n}) &= u_n \\ T_{\underline{t},k}(\underline{u}_{k:n}) &= u_k + \psi\big(\Delta t_{k+1}, T_{\underline{t}}(\underline{u}_{k+1:n})\big) \end{aligned}$$

Observe that $T_{\underline{t}}$ is the inverse of $U_{\underline{t}}$. To see this, fix $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$ and $\underline{u} := U_{\underline{t}}(\underline{\theta})$. The final coordinate is obvious,

$$T_{\underline{t},n}(\underline{u}_{n:n}) = u_n = U_{\underline{t},n}(\underline{\theta}) = \theta_n,$$

while an inductive hypothesis $T_{\underline{t},k}(\underline{u}_{k:n}) = \theta_k$ gives us the next step.

$$T_{\underline{t},k-1}(\underline{u}_{k-1:n}) = U_{\underline{t},k-1}(\underline{\theta}) + \psi\big(\Delta t_k, T_{\underline{t},k}(\underline{u}_{k:n})\big) = \theta_{k-1} - \psi(\Delta t_k, \theta_k) + \psi(\Delta t_k, \theta_k) = \theta_{k-1}$$

Dual to this, fix $\underline{u} \in \mathbb{V}^{|\underline{t}|}$ for which $\Psi(\underline{t},\underline{u},x_0) < \infty$ and define $\underline{\theta} := T_{\underline{t}}(\underline{u})$. Again, we immediately have

$$U_{\underline{t},n}(\underline{\theta}) = \theta_n = T_{\underline{t},n}(\underline{u}_{n:n}) = u_n,$$

and an inductive hypothesis of $U_{\underline{t},k}(\underline{\theta}) = u_k$ results in the next step.

$$U_{\underline{t},k-1}(\underline{\theta}) = \theta_{k-1} - \psi(\Delta t_k, \theta_k) = T_{\underline{t},k-1}(\underline{u}_{k-1:n}) - \psi\big(\Delta t_k, T_{\underline{t},k}(\underline{u}_{k:n})\big) = u_{k-1}$$

We have now showed that $U_{\underline{t}}$ is a homeomorphism with inverse $T_{\underline{t}}$. It remains to show our identity for a pairing $\underline{u} = U_{\underline{t}}(\underline{\theta})$.

$$\begin{aligned}
&\langle \underline{u}, \underline{x} \rangle - \Psi(\underline{t},\underline{u},x_0) \\
&= \big\langle U_{\underline{t}}(\underline{\theta}), \underline{x} \big\rangle - \Psi\big(\underline{t}, U_{\underline{t}}(\underline{\theta}), x_0\big) \\
&= \sum_{i=1}^{n-1} \big\langle \theta_k - \psi(\Delta t_{k+1}, \theta_{k+1}), x_k \big\rangle + \langle \theta_n, x_n \rangle - \sum_{i=1}^{n} \psi_0(\Delta t_k, \theta_k) - \big\langle \psi(\Delta t_1, \theta_1), x_0 \big\rangle \\
&= \sum_{i=1}^{n} \Big(\langle \theta_k, x_k \rangle - \psi_0(\Delta t_k, \theta_k) - \big\langle \psi(\Delta t_k, \theta_k), x_k \big\rangle \Big) \\
&= \sum_{i=1}^{n} \Big(\langle \theta_k, x_k \rangle - \Psi(\Delta t_k, \theta_k, x_k) \Big) \\
&= \sum_{i=1}^{n} \Big(\langle \theta_k, x_k - x_{k-1} \rangle - \varphi(\Delta t_k, \theta_k, x_k) \Big)
\end{aligned}$$

1.4 Affine jump-diffusions

This section shows how the notions of jump-diffusions explained in Appendix A apply in the affine case. These results are useful because—as we see in Theorem 1.1.4—affine processes are affine jump-diffusions. Throughout this section, we take X to be an affine jump-diffusion with differential χ-characteristics $(\beta^{\chi}, \alpha, \mu)$, but *do not necessarily* assume X is an affine process, in the sense of Definition 1.1.1. Firstly, we prove the uniform-boundedness property for the affine jump kernel μ.

Lemma 1.4.1. *The jump kernel μ satisfies the following uniform-boundedness property. Any function $f \in \mathscr{B}(\mathbb{V})/\mathscr{B}(\mathbb{R})$ that satisfies*

$$\int_{\mathbb{V}} |f(v)|\mu(x, \mathrm{d}v) < \infty$$

for all $x \in \mathbb{X}$ then satisfies the following.

$$x \mapsto \int_{\mathbb{V}} |f(v)|\mu(x, \mathrm{d}v) \text{ bounded on compact sets}$$

Proof. Seeing as $0 \in \mathbb{X}$ and $\operatorname{span} \mathbb{X} = \mathbb{V}$, we can take appropriate linear combinations to get finite integrals for each of the parts $m_0, \dots, m_d$ of μ.

$$F_i \coloneqq \int_{\mathbb{V}} |f(v)| m_i(\mathrm{d}v) < \infty, \quad i = 0, \dots, d$$

From here, the result is a simple effect of our affine property and boundedness of compact sets.

$$\sup_{|x| \le M} \left| \int_{\mathbb{V}} |f(v)|\mu(x, \mathrm{d}v) \right| = \sup_{|x| \le M} \left| F_0 + \sum_{i=1}^{d} x^i F_i \right| \le F_0 + M \sum_{i=1}^{d} F_i < \infty$$

With this result, we can state succinct versions of the results which exist for general jump-diffusions. The first of which can be seen as our *escape hatch* to representing affine processes X as objects subject to a dynamical system $\dot{\xi} = \beta(\xi)$, subject to both continuous and purely-discontinuous martingale noise.

Proposition 1.4.2. *If $0 \in \mathcal{D}_{\Lambda}^{\circ}$, then X is a $(\mathrm{P}_x, \mathscr{F})$ special jump-diffusion for each $x \in \mathbb{X}$. The resulting drift map $\beta : \mathbb{X} \to \mathbb{V}$ in the special semimartingale decomposition,*

$$X_t = x + \beta(X) \bullet \ell_t + X^{\mathrm{c}} + \mathrm{id}_{\mathbb{V}} * \tilde{q}_t^X$$

is also affine, making all the special differential characteristics (β, α, μ) affine.

Proof. By combining Lemma 1.4.1 and Proposition A.4.10, we get that X is special. Now, we perform the algebra to see the affine structure of β.

$$\begin{aligned}
\beta(x) &= \beta^{\chi}(x) + \int_{\mathbb{V}} \big(v - \chi(v)\big)\mu(x, \mathrm{d}v) \\
&= \left(b_0^{\chi} + \sum_{i=1}^{d} x^i b_i^{\chi}\right) + \int_{\mathbb{V}} \big(v - \chi(v)\big)\left(m_0(\mathrm{d}v) + \sum_{i=1}^{d} x^i m_i(\mathrm{d}v)\right) \\
&= \left(b_0^{\chi} + \int_{\mathbb{V}} \big(v - \chi(v)\big)m_0(\mathrm{d}v)\right) + \sum_{i=1}^{d} x^i \left(b_i^{\chi} + \int_{\mathbb{V}} \big(v - \chi(v)\big)m_i(\mathrm{d}v)\right)
\end{aligned}$$

Proposition 1.4.3. *If the jump kernel satisfies $\mu(x, \mathbb{V}) < \infty$ for all $x \in \mathbb{X}$, then X is $(\mathrm{P}_x, \mathscr{F})$ locally countable for all $x \in \mathbb{X}$. In the resulting factorization,*

$$\mu(x, \mathrm{d}v) = \lambda(x)\kappa(x, \mathrm{d}v),$$

the intensity λ is an affine map and the jump distribution κ is a convex mixture of probability distributions $k_0, \ldots, k_d$ whenever $\lambda(x) \neq 0$.

$$\lambda(x) = l_0 + \sum_{i=1}^{d} x^i l_i, \qquad \kappa(x, \mathrm{d}v) = \frac{l_0}{\lambda(x)} k_0(\mathrm{d}v) + \sum_{i=1}^{d} \frac{x^i l_i}{\lambda(x)} k_i(\mathrm{d}v),$$

Proof. By combining Lemmas 1.4.1 and A.3.2, we get the desired local countability. Because $0 \in \mathbb{X}$ and $\operatorname{span} \mathbb{X} = \mathbb{V}$, we are able to take appropriate linear combinations to ensure finiteness of the quantities $l_i \coloneqq m_i(\mathbb{V})$ for each $i = 0, \ldots, d$. This allows us to define our intensity map.

$$\lambda(x) \coloneqq l_0 + \sum_{i=1}^{d} x^i l_i = m_0(\mathbb{V}) + \sum_{i=1}^{d} x^i m_i(\mathbb{V}) = \mu(x, \mathbb{V})$$

Now, just as in Remark A.3.3, each non-zero l_i will produce a probability distribution $k_i(\mathrm{d}v) \coloneqq m_i(\mathrm{d}v)/l_i$; otherwise, simply define $k_i(\mathrm{d}v) \coloneqq \delta_{e_1}$. This way, we have the factoring $m_i(\mathrm{d}v) = l_i k_i(\mathrm{d}v)$ for each $i = 0, \ldots, d$. If $\lambda(x) \neq 0$, we see our other desired identity.

$$\begin{aligned}
\kappa(x, \mathrm{d}v) &\coloneqq \frac{1}{\lambda(x)} \mu(x, \mathrm{d}v) \\
&= \frac{1}{\lambda(x)} \Big(m_0(\mathrm{d}v) + \sum_{i=1}^{d} x^i m_i(\mathrm{d}v) \Big) \\
&= \frac{1}{\lambda(x)} \Big(l_0 k_0(\mathrm{d}v) + \sum_{i=1}^{d} x^i l_i k_i(\mathrm{d}v) \Big) \\
&= \frac{l_0}{\lambda(x)} k_0(\mathrm{d}v) + \sum_{i=1}^{d} \frac{x^i l_i}{\lambda(x)} k_i(\mathrm{d}v)
\end{aligned}$$

Theorem 1.4.4. *If $0 \in \mathcal{D}_\Lambda^\circ$, then any $h \in \mathbb{D}([0, \infty), \mathbb{V})$ of finite variation, compact support, and image contained in $\mathcal{D}_\Lambda^\circ$ makes $\exp(h \bullet X)$ a $(\mathrm{P}_x, \mathscr{F})$ special jump-diffusion and*

$$Z^h \coloneqq \exp\Big(h \bullet X - \Lambda\big(h, X\big) \bullet \ell \Big)$$

a martingale for every $x \in \mathbb{X}$. Moreover, we may define a new measure Q^h by

$$\mathrm{Q}^h(\mathrm{d}\omega) \coloneqq Z^h(\omega) \cdot \mathrm{P}(\mathrm{d}\omega)$$

such that X is a $(\mathrm{Q}^h, \mathscr{F})$ special jump-diffusion with affine $(\mathrm{Q}^h, \mathscr{F})$ special differential characteristics (β^h, α, μ^h).

$$\begin{aligned}
\beta^h(s, x) &\coloneqq \beta(x) + \alpha(x) h(s) + \int_{\mathbb{V}} v \big(e^{\langle h(s), v \rangle} - 1 \big) \mu(x, \mathrm{d}v) \\
\mu^h(s, x, \mathrm{d}v) &\coloneqq e^{\langle h(s), v \rangle} \mu(x, \mathrm{d}v)
\end{aligned}$$

Proof. The quantity $M = \exp\big(h \bullet X - \Lambda(h,X) \bullet \ell\big)$ is a $(\mathrm{P}_x, \mathscr{F})$ local martingale by our hypotheses and Theorem A.4.13. To get the remaining martingale property, we first note that the compact support of h means that there exists $\tau > 0$ such that $h(t) = 0$ for all $t > \tau$. This makes $M = M^\tau$, and so we only need to consider the martingale property on the interval $[0,\tau]$. For this, we use [SV10, Theorem 2.6], which requires the maps

$$(s,x) \mapsto \big\langle h(s), \alpha(x) h(s) \big\rangle, \quad (s,x) \mapsto \int_{\mathbb{V}} \big(e^{\langle h(s),v\rangle} - 1 - \langle h(s), v\rangle\big)\mu(x, \mathrm{d}v)$$

are bounded on compact sets of points (s,x). This comes from the fact that the image of h is contained in some compact subset of $\mathcal{D}_\Lambda^\circ$ and that Λ is uniformly bounded on compact subsets of $\mathcal{D}_\Lambda^\circ \times \mathbb{X}$ by Lemma A.4.4. The $(\mathrm{Q}^h, \mathscr{F})$ special differential characteristics then come from Theorem A.4.16.

Now that we have results on affine jump-diffusions, the question arises if they are affine processes in the sense of our Definition 1.1.1. The answer is *yes*, so long as the associated differential χ-characteristics $(\beta^\chi, \alpha, \mu)$ and subsequent Lévy-Khintchine map have an associated solution to (1.1.7).

Proposition 1.4.5. *Assume that are differential χ-characteristics $(\beta^\chi, \alpha, \mu)$ are such that the Lévy-Khintchine map Λ has a solution $\Psi(\cdot, u, \cdot)$ to the following system for each $u \in \mathcal{U}_{\mathbb{X}}$.*

$$\Psi(t,u,x) = \psi_0(t,u) + \big\langle \psi(t,u), x\big\rangle$$
$$\forall\ x \in \mathbb{X}, \qquad \begin{cases} \dot\Psi(t,u,x) = \Lambda\big(\psi(t,u), x\big) & t \geq 0 \\ \Psi(0,u,x) = \langle u, x\rangle \end{cases}$$

Then we have the remaining necessary condition of Definition 1.1.1.

$$\mathrm{E}_{\mathrm{P}_x} \exp\langle u, X_t\rangle = \exp\Psi(t,u,x), \quad t \geq 0,\ u \in \mathcal{U}_{\mathbb{X}},\ x \in \mathbb{X}$$

Proof. Fix $\tau \geq 0$, $u \in \mathcal{U}_{\mathbb{X}}$, and $x \in \mathbb{X}$ and denote $M_t = \exp\Psi(\tau - t, u, X_t)$. We now apply our jump-diffusion version of Itô's formula for jump-diffusions as in Lemma A.1.7, indifferently treating an element of $\mathbb{V}$ as one of $\mathbb{L}(\mathbb{V}, \mathbb{R})$ in the canonical way.

$$\begin{aligned}
M_t = {} & \exp\Psi(\tau,u,x) - \exp\Psi(\tau-\ell,u,X)\dot\Psi(\tau-\ell,u,X)\bullet\ell_t \\
& + \exp\Psi(\tau-\ell,u,X_-)\psi(\tau-\ell,u)\bullet X_t \\
& + \frac{1}{2}\exp\Psi(\tau-\ell,u,X)\big\langle\psi(\tau-\ell,u),\alpha(X)\psi(\tau-\ell,u)\big\rangle\bullet\ell \\
& + \Big(\exp\Psi(\tau-\ell,u,X) - \exp\Psi(\tau-\ell,u,X_-) - \big\langle\psi(\tau-\ell,u),\mathrm{id}_{\mathbb{V}}\big\rangle\Big) * q^X \\
= {} & \exp\Psi(\tau,u,x) - M\Lambda\big(\psi(\tau-\ell,u),X\big)\bullet\ell_t \\
& + M\bigg(\big\langle\psi(\tau-\ell,u),\beta^\chi(X)\big\rangle + \frac{1}{2}\langle\psi(\tau-\ell,u),\alpha(X)\psi(\tau-\ell,u)\rangle\bigg)\bullet\ell_t \\
& + M_-\psi(\tau-\ell,u)\bullet X^{\mathrm{c}} + M_-\big\langle\psi(\tau-\ell,u),\chi\big\rangle * \tilde q^X_t + M_-\big\langle\psi(\tau-\ell,u),\mathrm{id}_{\mathbb{V}}-\chi\big\rangle * q^X_t \\
& + M_-\Big(\exp\langle\psi(\tau-\ell,u),\mathrm{id}_{\mathbb{V}}\rangle - 1 - \big\langle\psi(\tau-\ell,u),\mathrm{id}_{\mathbb{V}}\big\rangle\Big) * q^X
\end{aligned}$$

Now, we use Lemma 1.3.2 to claim that M is bounded, and so it is special by [JS03, Lemma I.4.24], so we may compensate our jump integrals above.

$$\begin{aligned} M_t = {} & \exp \Psi(\tau, u, x) - M\Lambda\big(\psi(\tau-\ell, u), X\big) \bullet \ell_t \\ & + M\bigg(\big\langle \psi(\tau-\ell, u), \beta^{\mathrm{X}}(X)\big\rangle + \frac{1}{2}\langle \psi(\tau-\ell, u), \alpha(X)\psi(\tau-\ell, u)\rangle\bigg) \bullet \ell_t \\ & + M_- \psi(\tau-\ell, u) \bullet X^{\mathrm{c}} + M_- \big\langle \psi(\tau-\ell, u), \chi\big\rangle * \tilde{q}^X_t \\ & + M_- \Big(\exp \langle \psi(\tau-\ell, u), \mathrm{id}_{\mathbb{V}}\rangle - 1 - \big\langle \psi(\tau-\ell, u), \chi\big\rangle\Big) * \hat{q}^X_t \\ & + M_- \Big(\exp \langle \psi(\tau-\ell, u), \mathrm{id}_{\mathbb{V}}\rangle - 1 - \big\langle \psi(\tau-\ell, u), \chi\big\rangle\Big) * \tilde{q}^X_t \\ = {} & \exp \Psi(\tau, u, x) + M_- \psi(\tau-\ell, u) \bullet X^{\mathrm{c}} + M_- \Big(\exp \langle \psi(\tau-\ell, u), \mathrm{id}_{\mathbb{V}}\rangle - 1\Big) * \tilde{q}^X_t \end{aligned}$$

This shows that M is a local martingale; because it is bounded, it is a martingale, and we have our desired transform.

$$\mathrm{E}_{\mathrm{P}} \exp \langle u, X_\tau\rangle = \mathrm{E}_{\mathrm{P}} \exp \Psi(0, u, X_\tau) = \mathrm{E}_{\mathrm{P}} M_\tau = \exp \Psi(\tau, u, X_\tau)$$

Chapter 2

Large deviations of affine processes

This chapter concerns proving, for a fixed $x \in \mathbb{X}^\circ$, a large deviation principle for a family $(\mathrm{P}_x^\epsilon)_{\epsilon>0}$ of distributions P_x^ϵ of affine processes ϵX^ϵ, where each $(\mathrm{P}_x^\epsilon)_{x\in\mathbb{X}}$ is the family of kernels respectively associated with ϵX^ϵ. The parameterization that establishes these distributions $(\mathrm{P}_x^\epsilon)_{\epsilon>0}$ is stated very simply on the level of parameterizing fixed special differential characteristics (β, α, μ) which already associate with an affine process $X \coloneqq X^1$. To this end, $(\epsilon X^\epsilon)_{\epsilon>0}$ is a family of affine processes in which we are effectively regularizing X as $\epsilon \to 0$, and our large deviation principle explains this regularization.

The principle is established with regarding $(\mathrm{P}_x^\epsilon)_{\epsilon>0}$ as a family of measures P_x^ϵ on the Borel space associated to our Skorokhod topology on $\mathbb{D}([0,\infty), \mathbb{X})$. This means that there exists a lower-semi-continuous function $I_x : \mathbb{D}([0,\infty), \mathbb{X}) \to [0,\infty]$, known as a *rate function*, with the characterizing property that any $\Gamma \in \mathscr{B}(\mathbb{D}([0,\infty), \mathbb{X}))$ is such that the probabilities $\big(\mathrm{P}_x^\epsilon(\epsilon X^\epsilon \in \Gamma)\big)_{\epsilon>0}$ have a first-order exponential asymptotic that corresponds to variational problems on I_x.

$$-\inf_{\xi\in\Gamma^\circ} I_x(\xi) \le \liminf_{\epsilon\to 0} \epsilon \log \mathrm{P}_x^\epsilon\big(\epsilon X^\epsilon \in \Gamma\big) \le \limsup_{\epsilon\to 0} \epsilon \log \mathrm{P}_x^\epsilon\big(\epsilon X^\epsilon \in \Gamma\big) \le -\inf_{\xi\in\overline{\Gamma}} I_x(\xi)$$

Having the full strength of the Borel space on $\mathbb{D}([0,\infty), \mathbb{X})$, our admissible sets Γ can include a variety of tests on ϵX^ϵ. For instance, suitable selection of Γ allows us to derive asymptotics of probabilities associated with stopping times, path-integrals, and finite-dimensional distributions of $(\epsilon X^\epsilon)_{\epsilon>0}$.

A great resource for a systematic approach to the theory of large deviations is [DZ10]. This text offers comprehension on the subject with both intuitive and technically abstract perspectives, along with providing a descriptive history of the subject. While we only use this text to cite Cramér's theorem, its perspective paves the path of our argument. In particular, Cramér is credited with the *measure-change argument* we use throughout, which generally could be explained as defining exponential measure changes,

$$\begin{aligned} Z^{\epsilon,\theta} &\coloneqq \exp\Big(A(X^\epsilon,\theta) - B(X^\epsilon,\theta)\Big) \\ \mathrm{Q}^{\epsilon,\theta}(\mathrm{d}\omega) &\coloneqq Z^{\epsilon,\theta}(\omega) \cdot \mathrm{P}_x^\epsilon(\mathrm{d}\omega), \end{aligned} \tag{2.0.1}$$

where A is some linear form and B is appropriately compensating it to make $\mathrm{E}_{\mathrm{P}_x^\epsilon} Z^{\epsilon,\theta} = 1$. The object θ in the above expression is a parameter suitably chosen to make the measure $\mathrm{Q}^{\epsilon,\theta}(\epsilon X^\epsilon \in \Gamma^c)$ effectively *small enough* to neglect. We will explain this more in the chapter, but we find it important to state here that the nature of our linear form A can be associated with a *finite-dimensional* projection of X^ϵ or be a linear operator on the *infinite-dimensional* space of paths.

The expression (2.0.1) looks familiar to our local martingale in Theorem A.4.13. Such a measure change is of the *infinite-dimensional* nature, and is a commonly used tool in the literature of large deviations which we refer to as the *exponential martingale approach.* Note that the dynamics in Theorem A.4.16 allow for easy calibration our density in (2.0.1). A particular case in which this tool is used is [KK14], where they prove a large deviation principle for continuous affine (jump-)diffusions with special differential characteristics $(\beta, \alpha, 0)$. This approach lends itself to integral expressions which ultimately find their way in the rate function I_x. Having a rate $I_x(\xi)$ involve an integral of ξ is very useful for investigating local properties of the family $(\epsilon X^\epsilon)_{\epsilon>0}$ and will be explained more in depth in the next chapter.

For the *finite-dimensional* approach, we may simply use inner-products on $\mathbb{V}^n$ as our operators A in (2.0.1).

$$\begin{aligned} &\exp\Big(\sum_{k=1}^n \langle u_k, X_{t_k}\rangle - B(t_1,\dots,t_n,u_1,\dots,u_n)\Big), \\ &\mathrm{E}_{\mathrm{P}} \exp \sum_{k=1}^n \langle u_k, X_{t_k}\rangle =: B(t_1,\dots,t_n,u_1,\dots,u_n) \end{aligned} \tag{2.0.2}$$

We refer to this approach as *twisting* or *tilting*, as each parameter u_k decides the distribution of the increment $X_{t_k} - X_{t_{k-1}}$ for calibration of our density. The Dawson-Gärtner theorem [DZ10, Theorem 4.6.1] provides a great abstraction on how to achieve from here a full principle on the paths. This is because a stochastic process $X = (X_t)_{t\geq 0}$—at its weakest—corresponds to a measure on the projective space $(\mathbb{X}^{[0,\infty)}, \bigotimes_{t\in[0,\infty)} \mathscr{B}(\mathbb{X}))$, and the theorem concerns itself with large deviations of abstract projective limit spaces. One issue with this approach is that the resulting rate function I_x does not have an integral form like in the exponential martingale approach. In fact, [KK14] specifically mentioned this issue when proving their principle. Another issue is that the projective limit space corresponds to a product topology on $\mathbb{X}^{[0,\infty)}$, which means we cannot get asymptotics for Γ described above. We need to leverage properties of the Skorokhod topology to then tighten the principle.

We remark in this chapter that, for affine processes, the exponential martingale and twisting approaches are nearly identical. The affine map Ψ which characterizes an affine process provides us with a way of resolving (2.0.2) as a specific selection of parameter θ in the exponential-martingale approach. In this way, our proof is different from [KK14], in that we use this resolution to derive an integral form for our Dawson-Gärtner rate function. A summary of the proof is as simple as:

prove large deviation principles for finite-dimensional distributions, prove exponential tightness, conclude large deviation principle on Skorokhod space

which can be compared to Prokhorov's theorem for weak convergence of measures on a separable metric space,

prove weak convergence of finite-dimensional distributions, prove tightness, conclude weak convergence which was the original intent behind defining the Skorokhod topology (see in [Sko56]). This comparison is not due to us, for researchers are actively studying large deviations theory from a weak convergence perspective. Puhalskii has actively developed such approaches, and we manipulate a proof from [Puh01] to get the integral representation of our rate function. For resources on large deviations on the Skorokhod space, see [FK06], a resource which we use to get some of our results.

Now that we have introduced the key ideas, the remainder of this chapter is organized as follows.

Section 2.1. Describes our parameterization $(\mathrm{P}^{\epsilon}_{x})_{\epsilon>0}$,

Section 2.2. Delineates our assumptions for the result,

Section 2.3. Proves our principle via twisting,

Section 2.4. Resolves the approach of twisting with that of exponential martingales,

Section 2.5. Simplifies our rate function to an integral form.

2.1 Asymptotic family

Fix family $(\epsilon X^{\epsilon})_{\epsilon>0}$ of affine processes ϵX^{ϵ} with associated conditional distributions $(\mathrm{P}^{\epsilon}_{x})_{x\in\mathbb{X}}$ such that the differential χ-characteristics $(\beta^{\chi,\epsilon},\alpha^{\epsilon},\mu^{\epsilon})$ relate through the following parameterization.

$$\beta^{\chi,\epsilon}(x)=\frac{1}{\epsilon}\beta^{\chi,1}(\epsilon x),\quad \alpha^{\epsilon}(x)=\frac{1}{\epsilon}\alpha^{1}(\epsilon x),\quad \mu^{\epsilon}(x,\mathrm{d}v)=\frac{1}{\epsilon}\mu^{1}(\epsilon x,\mathrm{d}v),\quad x\in\mathbb{X} \tag{2.1.1}$$

In effect, for each $x\in\mathbb{X}$, the family $(\mathrm{P}^{\epsilon}_{x})_{\epsilon>0}$ is induced by a *base distribution* $\mathrm{P}_x:=\mathrm{P}^1_x$ associated with *base affine process* $X:=X^1$ and *base differential characteristics* $(\beta^{\chi},\alpha,\mu):=(\beta^{\chi,1},\alpha^1,\mu^1)$. This also implies a similar parameterization for the Lévy-Khintchine maps Λ^{ϵ} associated with $(\beta^{\chi,\epsilon},\alpha^{\epsilon},\mu^{\epsilon})$ in terms of the base map Λ from $(\beta^{\chi},\alpha,\mu)$.

$$\Lambda^{\epsilon}(u,x)=\frac{1}{\epsilon}\Lambda(u,\epsilon x),\quad u\in\mathbb{V},\ x\in\mathbb{X} \tag{2.1.2}$$

Using the notation of Appendix A.4, we see that the set $\mathcal{D}_{\Lambda}(x)$ of finite points of $\Lambda(\cdot,x)$ is identical to that $\mathcal{D}_{\Lambda^{\epsilon}}(\epsilon x)$ of $\Lambda^{\epsilon}(\cdot,\epsilon x)$. So long that $\mathbb{X}$ is a cone—which is to say that $\mathbb{X}$ an additive set, closed under non-negative-scalar multiplication—we have $\mathbb{X}=\epsilon\mathbb{X}$, and so the following sets agree.

$$\mathcal{D}_{\Lambda}=\bigcap_{x\in\mathbb{X}}\mathcal{D}_{\Lambda}(x)=\bigcap_{x\in\mathbb{X}}\mathcal{D}_{\Lambda^{\epsilon}}(\epsilon x)=\bigcap_{x\in\mathbb{X}}\mathcal{D}_{\Lambda^{\epsilon}}(x)=\mathcal{D}_{\Lambda^{\epsilon}}$$

Note that a parameterization like (2.1.1) or (2.1.2) may exist irrespective of the affine property on $(\beta^{\chi},\alpha,\mu)$. However, affine processes are distinct in the existence (from Theorem 1.2.3) of an affine map $\Psi^{\epsilon}:\mathcal{D}_{\Psi^{\epsilon}}\to\mathbb{R}$ respective to X^{ϵ},

$$\Psi^{\epsilon}(t,u,x)=\psi^{\epsilon}_0(t,u)+\big\langle\psi^{\epsilon}(t,u),x\big\rangle,$$

in which $\Psi^{\epsilon}(\cdot,u,\cdot)$ is the minimal solution of system$(\Lambda^{\epsilon},\tau,u)$ for each $(\tau,u)\in\mathcal{D}_{\Psi^{\epsilon}}$,

$$\forall\ x\in\mathbb{X},\quad\begin{cases}\dot{\Psi}^{\epsilon}(t,u,x)=\Lambda^{\epsilon}\big(\psi^{\epsilon}(t,u),x\big), & t\in[0,\tau]\\ \Psi^{\epsilon}(0,u,x)=\langle u,x\rangle\end{cases}$$

and is the cumulant generating function of each marginal.

$$\mathrm{E}_{\mathrm{P}_x^\epsilon} \exp \langle u, X_\tau^\epsilon \rangle = \exp \Psi^\epsilon(\tau, u, x/\epsilon), \quad (\tau, u) \in \mathcal{D}_\Psi,\ x \in \mathbb{X}$$

Above, note that the x/ϵ in the last coordinate is because P_x^ϵ is the distribution associated with ϵX^ϵ, not X^ϵ. The following result shows us that our parameterization in (2.1.1) and (2.1.2) applies these cumulant generating functions, where $\Psi \coloneqq \Psi^1$ and $\mathcal{D}_\Psi \coloneqq \mathcal{D}_\Psi^1$.

Proposition 2.1.3. *Assume $\mathbb{X}$ is a cone satisfying $\operatorname{span} \mathbb{X} = \mathbb{V}$. For each $\epsilon > 0$, we have $\mathcal{D}_\Psi = \mathcal{D}_{\Psi^\epsilon}$ and the following identities.*

$$\Psi^\epsilon(t, u, x) = \frac{1}{\epsilon}\Psi(t, u, \epsilon x), \quad \psi_0^\epsilon(t, u) = \frac{1}{\epsilon}\psi_0(t, u), \quad \psi^\epsilon(t, u) = \psi(t, u),$$

Proof. Start by selecting $(\tau, u) \in \mathcal{D}_\Psi$. This means that $u \in \mathcal{D}_\Psi(\tau)$ and $\Psi(\cdot, u, \cdot)$ is a solution to system(Λ, u, τ). Observe that this implies the following identity for all $x \in \mathbb{X}$.

$$\frac{\partial}{\partial t}\frac{1}{\epsilon}\Psi(t, u, \epsilon x) = \frac{1}{\epsilon}\dot{\Psi}(t, u, \epsilon x) = \frac{1}{\epsilon}\Lambda\big(\psi(t, u), \epsilon x\big) = \Lambda^\epsilon(\psi(t, u), x), \quad t \in [0, \tau]$$
$$\frac{1}{\epsilon}\Psi(0, u, \epsilon x) = \frac{1}{\epsilon}\langle u, \epsilon x\rangle = \langle u, x\rangle$$

This means that $\frac{1}{\epsilon}\Psi(\cdot, u, \epsilon\cdot)$ is a solution to system$(\Lambda^\epsilon, \tau, u)$. By definition, existence of a solution means that $u \in \mathcal{D}_{\Psi^\epsilon}(\tau)$, and so $(\tau, u) \in \mathcal{D}_{\Psi^\epsilon}$. Theorem 1.2.3 then tells us $\Psi^\epsilon(\cdot, u, \cdot)$ exists and is dominated by the other solution.

$$\Psi^\epsilon(t, u, x) \le \frac{1}{\epsilon}\Psi(t, u, \epsilon x), \quad t \in [0, \tau],\ x \in \mathbb{X}$$

On the other hand, if we have $(\tau, u) \in \mathcal{D}_{\Psi^\epsilon}$, then $u \in \mathcal{D}_{\Psi^\epsilon}(\tau)$, and so $\Psi^\epsilon(\cdot, u, \cdot)$ is a solution to system$(\Lambda^\epsilon, \tau, u)$. Now, we have the following identity for all $x \in \mathbb{X}$,

$$\frac{\partial}{\partial t}\epsilon\Psi^\epsilon(t, u, x/\epsilon) = \epsilon\dot{\Psi}^\epsilon(t, u, x/\epsilon) = \epsilon\Lambda^\epsilon\big(\psi^\epsilon(t, u), \epsilon x\big) = \Lambda(\psi(t, u), x), \quad t \in [0, \tau]$$
$$\epsilon\Psi^\epsilon(0, u, x/\epsilon) = \epsilon\langle u, x/\epsilon\rangle = \langle u, x\rangle,$$

and so $\epsilon\Psi^\epsilon(\cdot, u, \cdot)$ is a solution to system(Λ, τ, u). Again, we may conclude from this that $(\tau, u) \in \mathcal{D}_\Psi$ and that $\Psi^\epsilon(\cdot, u, \cdot)$ exists and is dominated by the other solution.

$$\Psi(t, u, x) \le \epsilon\Psi^\epsilon(t, u, x/\epsilon), \quad t \in [0, \tau],\ x \in \mathbb{X}$$

In total, we have now shown that $\mathcal{D}_\Psi = \mathcal{D}_{\Psi^\epsilon}$, and inequalities (17) and (17) indicate to us that the following functions agree.

$$\Psi^\epsilon(t, u, x) = \frac{1}{\epsilon}\Psi(t, u, x), \quad (t, u) \in \mathcal{D}_\Psi,\ x \in \mathbb{X}$$

This means equality of the following affine expressions.

$$\begin{aligned}
\psi_0^\epsilon(t, u) + \big\langle \psi^\epsilon(t, u), x\big\rangle &= \Psi^\epsilon(t, u, x) \\
&= \frac{1}{\epsilon}\Psi(t, u, \epsilon x) \\
&= \frac{1}{\epsilon}\psi_0(t, u) + \frac{1}{\epsilon}\big\langle \psi(t, u), \epsilon x\big\rangle = \frac{1}{\epsilon}\psi_0(t, u) + \big\langle \psi(t, u), x\big\rangle
\end{aligned}$$

Seeing as $\operatorname{span} \mathbb{X} = \mathbb{V}$, we may take appropriate linear combinations to show the remaining identities.

$$\psi_0^\epsilon(t, u) = \frac{1}{\epsilon}\psi_0(t, u), \quad \psi_i^\epsilon(t, u) = \psi_i(t, u), \quad i = 1, \dots, d$$

Remark 2.1.4. *Note that the above proof can be applied to complex moments, since Theorem 1.1.4 and Remark 1.1.6 indicate to us that each $\Psi^\epsilon(\cdot, u, \cdot)$ is a solution to following equation, for each $u \in \mathrm{i}\mathbb{V}$.*

$$\forall\ x \in \mathbb{X}, \qquad \begin{cases} \dot{\Psi}^\epsilon(t,u,x) = \Lambda^\epsilon\big(\psi^\epsilon(t,u),x\big), & t \geq 0 \\ \Psi^\epsilon(0,u,x) = \langle u, x\rangle \end{cases}$$

This parameterization also applies to the liftings $\Psi^\epsilon(\underline{t},\cdot,\cdot)$ of Ψ^ϵ to finite-dimensional projections on partitions $\underline{t} \vdash [0,\infty)$.

$$\mathrm{E}_{\mathrm{P}^\epsilon_x} \exp\langle \underline{u}, \epsilon X^\epsilon_{\underline{t}}\rangle =: \exp \Psi(\underline{t}, \underline{u}, x/\epsilon), \quad \underline{u} \in \mathbb{V}^{|\underline{t}|},\ x \in \mathbb{X}$$

Denoting $\Psi(\underline{t},\cdot,\cdot) \coloneqq \Psi^1(\underline{t},\cdot,\cdot)$, the below result shows just this.

Proposition 2.1.5. *Assume $\mathbb{X}$ is a cone satisfying* $\operatorname{span}\mathbb{X} = \mathbb{V}$. *Fix $\underline{t} \vdash [0,\infty)$, $x_0 \in \mathbb{X}^\circ$, and $\epsilon > 0$ and define $U_{\underline{t}}$ as in Proposition 1.3.6. Each $\underline{u} \in \mathbb{V}^{|\underline{t}|}$ satisfying $\underline{u} = U_{\underline{t}}(\underline{\theta})$ for some $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$ satisfies*

$$\Psi^\epsilon\big(\underline{t}, U_{\underline{t}}(\underline{\theta}), x_0\big) = \frac{1}{\epsilon}\Psi\big(\underline{t}, U_{\underline{t}}(\underline{\theta}), \epsilon x_0\big) < \infty,$$

and if no such $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$ exists, both are infinite.

$$\Psi^\epsilon(\underline{t}, \underline{u}, x_0) = \frac{1}{\epsilon}\Psi(\underline{t}, \underline{u}, \epsilon x_0) = \infty.$$

Proof. We start by recognizing two facts. Firstly, from Proposition 2.1.3, we have an identity of the following sets.

$$\mathcal{D}_{\Psi^\epsilon}(\underline{t}) = \prod_{k=1}^{|\underline{t}|} \mathcal{D}_{\Psi^\epsilon}(\Delta t_k) = \prod_{k=1}^{|\underline{t}|} \mathcal{D}_{\Psi}(\Delta t_k) = \mathcal{D}_\Psi(\underline{t}), \tag{2.1.6}$$

Secondly, Proposition 2.1.3 also shows us that the $U^\epsilon_{\underline{t}}$ associated with X^ϵ is identical to that $U_{\underline{t}}$ of X, as $\psi^\epsilon = \psi$. We now show the desired identity by fixing $\underline{u} \in \mathbb{V}^{|\underline{t}|}$ and considering each case.

First suppose $\underline{u} = U_{\underline{t}}(\underline{\theta})$ for some $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$. The identity of (2.1.6) tells us $\underline{\theta} \in \mathcal{D}_{\Psi^\epsilon}(\underline{t})$ and so Propositions 1.3.6 and 2.1.3 give us the following.

$$\begin{aligned}
\Psi^\epsilon(\underline{t}, \underline{u}, x_0) &= \Psi^\epsilon\big(\underline{t}, U_{\underline{t}}(\underline{\theta}), x_0\big) \\
&= \sum_{k=1}^{|\underline{t}|} \psi^\epsilon_0(\Delta t_k, \theta_k) + \big\langle \psi^\epsilon(\Delta t_1, \theta_1), x_0\big\rangle \\
&= \frac{1}{\epsilon}\left(\sum_{k=1}^{|\underline{t}|} \psi_0(\Delta t_k, \theta_k) + \big\langle \psi(\Delta t_1, \theta_1), \epsilon x_0\big\rangle\right) \\
&= \frac{1}{\epsilon}\Psi\big(\underline{t}, U_{\underline{t}}(\underline{\theta}), \epsilon x_0\big) \\
&= \frac{1}{\epsilon}\Psi(\underline{t}, \underline{u}, \epsilon x_0)
\end{aligned}$$

On the other hand, suppose $\underline{u}$ is not in the image of $\mathcal{D}_\Psi(\underline{t})$ under $U_{\underline{t}}$. Seeing as $x_0 \in \mathbb{X}^\circ$ and $\mathbb{X}$ is a cone, we have $\epsilon x_0 \in \mathbb{X}^\circ$. Applying Theorem 1.3.12, we then have $\Psi(\underline{t}, \underline{u}, \epsilon x_0) = \infty$.

The identity in (2.1.6) also tells us that $\underline{u}$ is not in the image of $\mathcal{D}_\Psi(\underline{t})$ under $U_{\underline{t}}$. Theorem 1.3.12 now tell us $\Psi^\epsilon(\underline{t},\underline{u},x_0)=\infty$. We conclude our final identity.

$$\Psi^\epsilon(\underline{t},\underline{u},x_0)=\frac{1}{\epsilon}\Psi(\underline{t},\underline{u},\epsilon x_0)=\infty$$

Now that we have established parameterizations for just about every object that relates to an affine process, we establish some intuition on the relationship between these distributions $(\mathrm{P}^\epsilon_x)_{\epsilon>0}$. The first of which is immediate from our preceding result, but it only makes sense when we consider the countable sequence $\epsilon_m := 1/m$ for $m\in\mathbb{N}$.

Proposition 2.1.7. *Assume $\mathbb{X}$ is a cone with $\operatorname{span}\mathbb{X}=\mathbb{V}$. For a fixed $x\in\mathbb{X}$, the family $(\mathrm{P}^{\epsilon_m}_x)_{\epsilon>0}$ corresponds to a mean-field regime. That is to say, if we fix a probability space $(\Omega,\Sigma,\mathrm{P})$ equipped with a sequence of independent quantities $(X^{(j)})_{j\in\mathbb{N}}$ each distributing according to P_x, then we may realize each $\epsilon_m X^{\epsilon_m}$ as follows.*

$$\epsilon_m X^{\epsilon_m}=\frac{1}{m}\sum_{j=1}^m X^{(j)},\quad m\in\mathbb{N}$$

Proof. We will prove this by showing that the finite-dimensional distributions agree by identity of their characteristic functions. Fixing $\underline{t}\vdash[0,\infty)$ and $\underline{u}\in\mathbb{V}^{|\underline{t}|}$, we apply Proposition 1.3.3 and Remark 2.1.4.

$$\begin{aligned}
\log \mathrm{E}_{\mathrm{P}}\exp\Big\langle \mathrm{i}\underline{u},\sum_{j=1}^m X^{(j)}_{\underline{t}}\Big\rangle &= \log\Big(\mathrm{E}_{\mathrm{P}_x}\exp\langle \mathrm{i}\underline{u},X_{\underline{t}}\rangle\Big)^m\\
&=\log\Big(\exp\Psi\big(\underline{t},\mathrm{i}\underline{u},x\big)\Big)^m\\
&=m\Psi\big(\underline{t},\mathrm{i}\underline{u},x\big)\\
&=m\bigg(\sum_{k=1}^{|\underline{t}|}\psi_0(\Delta t_k,\theta_k)+\big\langle\psi(\Delta t_1,\theta),x\big\rangle\bigg)\\
&=\sum_{k=1}^{|\underline{t}|}\frac{1}{\epsilon_m}\psi_0(\Delta t_k,\theta_k)+\big\langle\psi(\Delta t_1,\theta),x/\epsilon_m\big\rangle\bigg)\\
&=\sum_{k=1}^{|\underline{t}|}\psi_0^{\epsilon_m}(\Delta t_k,\theta_k)+\big\langle\psi^{\epsilon_m}(\Delta t_1,\theta),x/\epsilon_m\big\rangle\\
&=\Psi^{\epsilon_m}(\underline{t},\underline{u},x/\epsilon_m)\\
&=\log \mathrm{E}_{\mathrm{P}^{\epsilon_m}_x}\exp\langle\underline{u},X_{\underline{t}}\rangle
\end{aligned}$$

We may also intuitively understand the relationship of $(\mathrm{P}^\epsilon_x)_{x\in\mathbb{X}}$ from a dynamical system perspective. In Theorem A.1.14, we see how jump-diffusions X always correspond to a weak solution of some stochastic differential equation driven by standard Brownian motion W and Poisson random measure p.

$$\begin{gathered}
X_t=X_0+\beta^\chi(X)\bullet\ell_t+\sigma(X)\bullet W_t+\chi\circ c(X,\mathrm{id}_\mathbb{V})*\tilde{p}_t+\chi'\circ c(X,\mathrm{id}_\mathbb{V})*p_t\\
\forall\ x\in\mathbb{X},\quad\begin{cases}\mu(x,\Gamma)=\displaystyle\int_\mathbb{V}1_\Gamma\big(c(x,v)\big)\mathrm{d}v, & \Gamma\in\mathscr{B}(\mathbb{V}-\{0\})\\ \alpha(x)=\sigma\sigma^*(x)\end{cases}
\end{gathered}\tag{2.1.8}$$

The following proposition gives perspectives on how the processes ϵX^ϵ may relate through these objects in two different perturbed dynamical systems.

Proposition 2.1.9. *Fix a probability space* $(\Omega, \Sigma, \mathrm{P})$ *equipped with standard Brownian motion* W *on* $\mathbb{V}$ *and Poisson random measure* p *on* $\mathscr{B}(\mathbb{R}_+ \times \mathbb{V})$*. Let* $\sigma : \mathbb{X} \to \mathbb{L}(\mathbb{V})$ *and* $c : \mathbb{X} \times \mathbb{V} \to \mathbb{V}$ *satisfy (2.1.8) for the* χ*-differential characteristics* $(\beta^\chi, \alpha, \mu)$*, as granted by Theorem A.1.14. For each* $x \in \mathbb{X}$*, the family* $(\mathrm{P}_x^\epsilon)_{\epsilon>0}$ *of distributions* P_x^ϵ *may be recognized as each* P_x^ϵ *being a weak solution to the respective* scaled *stochastic dynamical system,*

$$\epsilon X_t^\epsilon = x + \beta^\chi(\epsilon X^\epsilon) \cdot \ell_t + \sqrt{\epsilon}\sigma(X) \cdot W_t + \chi \circ \epsilon c\big(\epsilon X^\epsilon, \sqrt[d]{\epsilon} \cdot \mathrm{id}_{\mathbb{V}}\big) * \tilde{p}_t + \chi' \circ \epsilon c\big(\epsilon X^\epsilon, \sqrt[d]{\epsilon} \cdot \mathrm{id}_{\mathbb{V}}\big) * p_t,$$

or the time-changed *stochastic dynamical system.*

$$\begin{gathered} \epsilon X_t^\epsilon = x + \beta^\chi(\epsilon X^\epsilon) \cdot \ell_t + \epsilon\sigma(X) \cdot W_t^\epsilon + \chi \circ \epsilon c\big(\epsilon X^\epsilon, \mathrm{id}_{\mathbb{V}}\big) * \tilde{p}_t^\epsilon + \chi' \circ \epsilon c\big(\epsilon X^\epsilon, \mathrm{id}_{\mathbb{V}}\big) * p_t^\epsilon, \\ W_t^\epsilon \coloneqq W_{t/\epsilon} \\ p^\epsilon([0,t] \times \Gamma) \coloneqq p([0, t/\epsilon] \times \Gamma) \end{gathered}$$

Proof. By [JS03, III.2.26], it suffices to check if the characteristics of the above dynamical systems match those of ϵX^ϵ. We will instead match those of the scaled X^ϵ; recall its characteristics, as specified by (2.1.1).

$$\big\langle X^{\epsilon,\mathrm{c},i}, X^{\epsilon,\mathrm{c},j}\big\rangle = \frac{1}{\epsilon}\alpha_{i,j}(\epsilon X^\epsilon) \cdot \ell, \quad \hat{q}^{X^\epsilon}(\mathrm{d}s, \mathrm{d}v) = \frac{1}{\epsilon}\mu(\epsilon X^\epsilon, \mathrm{d}v)\mathrm{d}s$$

Let us first address the first system. Note that for any $i, j = 1, \ldots, d$, we use [JS03, Theorem I.4.40(d)] to resolve the predictable quadratic covariation of the continuous local martingale part $\frac{1}{\sqrt{\epsilon}}\sigma(\epsilon X^\epsilon) \cdot W$ of X^ϵ.

$$\begin{aligned} \Big\langle \big(\frac{1}{\sqrt{\epsilon}}\sigma(\epsilon X^\epsilon) \cdot W\big)^i, \big(\frac{1}{\sqrt{\epsilon}}\sigma(\epsilon X^\epsilon) \cdot W\big)^j \Big\rangle &= \Big\langle \sum_{l=1}^d \frac{1}{\sqrt{\epsilon}}\sigma_{i,l}(\epsilon X^\epsilon) \cdot W^l, \sum_{m=1}^d \frac{1}{\sqrt{\epsilon}}\sigma_{j,m}(\epsilon X^\epsilon) \cdot W^m \Big\rangle \\ &= \frac{1}{\epsilon}\sum_{l,m=1}^d \sigma_{i,l}(\epsilon X^\epsilon)\sigma_{j,m}(\epsilon X^\epsilon) \cdot \big\langle W^l, W^m\big\rangle \\ &= \frac{1}{\epsilon}\sum_{l=1}^d \sigma_{i,l}(\epsilon X^\epsilon)\sigma_{j,l}(\epsilon X^\epsilon) \cdot \ell \\ &= \frac{1}{\epsilon}\alpha_{i,j}(\epsilon X^\epsilon) \cdot \ell \\ &= \langle X^{\epsilon,\mathrm{c},i}, X^{\epsilon,\mathrm{c},j}\rangle \end{aligned}$$

Note that the accumulated jump process associated from X^ϵ in these dynamics is the following process.

$$t \mapsto \sum_{0<s\leq t} c\big(\epsilon X_{s-}^\epsilon, \sqrt[d]{\epsilon} \cdot \epsilon\Delta X_s^\epsilon\big)$$

This allows us to see that, for a non-negative predictable process $H : \Omega \times \mathbb{R}_+ \times \mathbb{V} \to \mathbb{R}$, we have the following identities, via changing coordinates.

$$\mathrm{E}_\mathrm{P}\Big(H * q_\infty^{X^\epsilon}\Big) = \mathrm{E}_\mathrm{P}\int_{\mathbb{R}_+ \times \mathbb{V}} H\Big(\cdot, s, c\big(\epsilon X_{s-}^\epsilon, \sqrt[d]{\epsilon} \cdot v\big)\Big) p(\mathrm{d}s, \mathrm{d}v)$$

$$
\begin{aligned}
&= \mathrm{E_P} \int_0^\infty \int_{\mathbb{V}} H\Big(\cdot, s, c\big(\epsilon X^\epsilon_{s-}, \sqrt[d]{\epsilon} \cdot v\big)\Big) \mathrm{d}v \mathrm{d}s \\
&= \mathrm{E_P} \int_0^\infty \int_{\mathbb{V}} H\Big(\cdot, s, c\big(\epsilon X^\epsilon_{s-}, v\big)\Big) \frac{1}{\epsilon} \mathrm{d}v \mathrm{d}s && v \leftarrow \sqrt[d]{\epsilon} \cdot v \\
&= \mathrm{E_P} \int_0^\infty \int_{\mathbb{V}} H(\cdot, s, v) \frac{1}{\epsilon} \mu(\epsilon X^\epsilon_{s-}, \mathrm{d}v) \mathrm{d}s && v \leftarrow c(X_{s-}, v)
\end{aligned}
$$

Now we address the second system using the same calculations; the continuous local martingale term of X^ϵ in this case is $\sigma(\epsilon X^\epsilon) \bullet W$, and we have the following identity.

$$
\begin{aligned}
\Big\langle \big(\sigma(\epsilon X^\epsilon) \bullet W^\epsilon\big)^i, \big(\sigma(\epsilon X^\epsilon) \bullet W^\epsilon\big)^j \Big\rangle &= \Big\langle \sum_{l=1}^d \sigma_{i,l}(\epsilon X^\epsilon) \bullet W^{\epsilon,l}, \sum_{m=1}^d \sigma_{j,m}(\epsilon X^\epsilon) \bullet W^{\epsilon,m} \Big\rangle \\
&= \sum_{l,m=1}^d \sigma_{i,l}(\epsilon X^\epsilon) \sigma_{j,m}(\epsilon X^\epsilon) \bullet \big\langle W^{\epsilon,l}, W^{\epsilon,m} \big\rangle \\
&= \sum_{l=1}^d \sigma_{i,l}(\epsilon X^\epsilon) \sigma_{j,l}(\epsilon X^\epsilon) \bullet \big(\epsilon^{-1} \ell\big) \\
&= \frac{1}{\epsilon} \alpha_{i,j}(\epsilon X^\epsilon) \bullet \ell \\
&= \langle X^{\epsilon,\mathrm{c},i}, X^{\epsilon,\mathrm{c},j} \rangle
\end{aligned}
$$

Meanwhile, our time-change of the Poisson random measure immediately gives us our desired characteristic.

$$
\begin{aligned}
\mathrm{E_P}\Big(H * q^{X^\epsilon}_\infty\Big) &= \mathrm{E_P} \int_{\mathbb{R}_+ \times \mathbb{V}} H\big(\cdot, s, c(\epsilon X^\epsilon_{s-}, v)\big) p^\epsilon(\mathrm{d}s, \mathrm{d}v) \\
&= \mathrm{E_P} \int_0^\infty \int_{\mathbb{V}} H\big(\cdot, s, c(\epsilon X^\epsilon_{s-}, v)\big) \mathrm{d}v \cdot \frac{1}{\epsilon} \mathrm{d}s \\
&= \mathrm{E_P} \int_0^\infty \int_{\mathbb{V}} H(\cdot, s, v) \frac{1}{\epsilon} \mu(\epsilon X^\epsilon_{s-}, \mathrm{d}v) \mathrm{d}s && v \leftarrow c(X_{s-}, v)
\end{aligned}
$$

2.2 Assumptions

We now clearly spell out the assumptions we will need to prove our large deviation principle. The first of which concerns existence of our affine processes.

As mentioned in the previous section, the base parameters $(\beta^\chi, \alpha, \mu)$ parameterize those $(\beta^{\chi,\epsilon}, \alpha^\epsilon, \mu^\epsilon)$ for each $\epsilon > 0$. This means that selecting the base affine process X immediately imposes the laws ϵX^ϵ for all other $\epsilon > 0$, *should they exist.* Generally speaking, there exist functions $(\beta^\chi, \alpha, \mu)$ for which there is no jump-diffusion that makes them differential characteristics. Proposition 2.1.7 indicated that, should distributions $(\mathrm{P}_x)_{x \in \mathbb{X}}$ exist for an affine process X with differential characteristics $(\beta^\chi, \alpha, \mu)$, then we can construct distributions $\mathrm{P}^{\epsilon_m}_x$ for each $m \in \mathbb{N}$ and $x \in \mathbb{X}$ by taking convolutions (recall, $\epsilon_m := 1/m$). However, we find it important to establish our large deviation principle over a continuously defined family $\epsilon > 0$. This now motivates the following assumption, which was already implicitly assumed in the previous section.

Assumption 2.2.1. *The affine parameters $(\beta^{\chi}, \alpha, \mu)$ are chosen so that each $\epsilon \in (0, 1]$, affine processes X^{ϵ} exist which exhibit the parameters $(\beta^{\chi,\epsilon}, \alpha^{\epsilon}, \mu^{\epsilon})$ as in (2.1.1).*

Remark 2.2.2. *(a) If we wanted to show a large deviation principle over only the family $(\epsilon_m X^{\epsilon_m})_{m\in\mathbb{N}}$, then this assumption is unnecessary.*

(b) Selecting $(\beta^{\chi}, \alpha, \mu)$ parameterizes each Λ^{ϵ}, which does specify the systems (1.1.7) and (1.2.2) associated with complex and real moments, respectively. There are already many results which allow one to say that X^{ϵ} exists in this scenario. For instance, if our state space is $\mathbb{X} = \mathbb{R}^m_+ \times \mathbb{R}^n$, then [DFS03, Proposition 7.4] provides us the existence we require, so long as the parameters are chosen to be admissible. *Otherwise, one can specify weaker assumptions on $\mathbb{X}$ at the expense of choosing diffusions ($\mu(\cdot, \mathrm{d}v) = 0$) or pure-jump processes ($\alpha = 0$) on certain factors of the space; see [Cuc11, Section 2.5, Section 3.4].*

Our proof ultimately comes from the parameterizations of our moment functions Ψ^{ϵ} in the previous section. These relied on $\mathbb{X}$ being a cone, so we specifically mention that here.

Assumption 2.2.3. *The space $\mathbb{X}$ is a cone and* $\operatorname{span}\mathbb{X} = \mathbb{V}$.

Remark 2.2.4. *Note that* $\operatorname{span}\mathbb{X}$ *is at no loss of generality, per Remark 1.1.3.*

Our last assumption is of key importance. Large deviations are best understood through moment generating functions, which serve no purpose if they are not finite. For an initial point $x_0 \in \mathbb{X}^{\circ}$, a large deviation principle holds for a family $(\mathrm{P}^{\epsilon}_{x_0})_{\epsilon>0}$ on the projective limit space (i.e. the product topology), so long as $0 \in \mathcal{D}^{\circ}_{\Lambda}$. However, in order to strengthen this result to the Skorokhod topology—a step which is necessary for interesting asymptotics and an integral-form of our rate function—we need the full strength of $\mathcal{D}_{\Lambda} = \mathbb{V}$.

Assumption 2.2.5. *We have $\mathcal{D}_{\Lambda} = \mathbb{V}$; equivalently, by Lemma A.4.2,*

$$\int_{|v|>1} e^{\langle u,v\rangle} \mu(x, \mathrm{d}v) < \infty, \quad x \in \mathbb{X},\ u \in \mathcal{D}_{\Psi}$$

Remark 2.2.6. *(a) It is easy to see in the proofs where we assume $\mathcal{D}_{\Lambda} = \mathbb{V}$ versus simply using $0 \in \mathcal{D}^{\circ}_{\Lambda}$.*

(b) By imposing even the simpler of the two assumptions that $0 \in \mathcal{D}^{\circ}_{\Lambda}$ immediately tells us that all X^{ϵ} are special, and so we proceed the rest of the chapter with a presentation in terms of special differential characteristics (β, α, μ).

With this assumption in mind, we state versions of the propositions at the end of Section 1.2 in our scenario.

Proposition 2.2.7. *(a) For each $\tau > 0$, $\mathcal{D}_{\Psi}(\tau)$ is open; in particular, there exists $\delta > 0$ such that $B(0, \delta) \subseteq \mathcal{D}_{\Psi}(\tau)$.*

(b) For each $\tau > 0$ and $u \in \mathcal{D}_{\Psi}(\tau)$, $\Psi(\cdot, u, \cdot)$ from Theorem 1.2.3 is the unique solution to $\mathrm{system}(\Lambda, \tau, u)$.

(c) *For each $M > 0$, there exists $\delta > 0$ such that $\overline{B}(0, M) \subseteq \mathcal{D}_\Psi(\delta)$.*

(d) *For each $M > 0$, there exist $\delta > 0$ and C_M such that the following holds.*

$$\left|\Psi(t, u, x) - \Psi(0, u, x)\right| \leq C_M \cdot t \cdot \left(1 + |x|\right), \quad t \in [0, \delta],\ u \in \overline{B}(0, M),\ x \in \mathbb{X}$$

Proof. (a) Proposition 1.2.5(a) tells us that $\mathcal{D}_\Psi(\tau)$ is open in $\mathcal{D}_\Lambda^\circ$, which is now $\mathbb{V}$ by our assumption. Also, seeing as $\mathrm{E}_{\mathrm{P}_x} \exp \langle 0, X_\tau \rangle = 1 < \infty$ for all $x \in \mathbb{X}$, $0 \in \mathcal{D}_\Psi(\tau)$. Openness of $\mathcal{D}_\Psi(\tau)$ now grants some $B(0, \delta) \subseteq \mathcal{D}_\Psi(\tau)$.

(b) Proposition 1.2.5(b) gives us uniqueness of $\Psi(\cdot, u, \cdot)$ as a solution to system(Λ, τ, u) for any $u \in \mathcal{D}_\Psi(\tau) \cap \mathcal{D}_\Lambda^\circ$, which is now the same thing as $u \in \mathcal{D}_\Psi(\tau)$.

(c) Seeing as $\mathcal{D}_\Lambda = \mathbb{V}$, we now have $\overline{B}(0, M) \subseteq \mathcal{D}_\Lambda^\circ$, so Proposition 1.2.7 gives us $\delta > 0$ such that $\overline{B}(0, M) \subseteq \mathcal{D}_\Psi(\delta)$.

(d) Fix $M > 0$. By part (c), there exists $\delta > 0$ such that $\overline{B}(0, 2M) \subseteq \mathcal{D}_\Psi(2\delta)$. Now, $[0, \delta] \times \overline{B}(0, M)$ is a compact subset of $\mathcal{D}_\Psi^\circ$, and so Proposition 1.2.9 gives us our desired C_M.

2.3 Dawson-Gärtner

This section proves our large deviation principle from a perspective similar to that of Dawson-Gärtner (see [DZ10, Theorem 4.6.1]), in which we prove the principle for the finite-dimensional projections, so that we get a principle on the projective space associated with $\mathbb{D}([0, \infty), \mathbb{X})$ that we may tighten to the Skorokhod space through exponential tightness. Though we attribute this approach to the names of Dawson and Gärtner, we specifically use results in [FK06] which instead use results which appeal more to weak convergence arguments that are comparable to Prokhorov on an exponential scale.

Theorem 2.3.1. *For each $x \in \mathbb{X}^\circ$ and $\underline{t} \vdash [0, \infty)$, the family $(\pi_{\underline{t}\#} \mathrm{P}_x^\epsilon)_{\epsilon > 0}$ satisfies a large deviation principle on $\mathbb{V}^{|\underline{t}|}$ with good convex rate function $\Psi^*(\underline{t}, \cdot, x)$, the Fenchel-Legendre transform of $\Psi(\underline{t}, \cdot, x)$.*

$$\Psi^*(\underline{t}, \underline{x}, x) \coloneqq \sup_{\underline{u} \in \mathbb{V}^{|\underline{t}|}} \Big(\langle \underline{u}, \underline{x} \rangle - \Psi(\underline{t}, \underline{u}, x) \Big)$$

Proof. We first prove a principle on the discrete family $(\pi_{\underline{t}\#} \mathrm{P}_x^{\epsilon_m})_{m \in \mathbb{N}}$ for $\epsilon_m \coloneqq 1/m$. Note that Proposition 2.1.7 allows us to consider a space $(\Omega, \Sigma, \mathrm{P})$ equipped with an i.i.d. sequence $(X^{(j)})_{j \in \mathbb{N}}$ of elements distributing from P_x and realize each $\epsilon_m X^{\epsilon_m}$.

$$\epsilon_m X^{\epsilon_m} = \frac{1}{m} \sum_{j=1}^{m} X^{(j)}$$

We now use a specific instance of Cramér's theorem [DZ10, Corollary 6.1.6] to conclude that if $\underline{0}$ is an interior point in the finite domain of $\Psi(\underline{t}, \cdot, x)$, then our principle is satisfied with good rate function $\Psi^*(\underline{t}, \cdot, x)$. Note that Proposition 2.2.7(a) tells us that $\underline{0} \in \mathcal{D}_\Psi(\underline{t})$, an open set. Denoting some ball $B(\underline{0}, \delta) \subseteq \mathcal{D}_\Psi(\underline{t})$, Theorem 1.3.12 indicates that $U_{\underline{t}} B(\underline{0}, \delta)$ is an open set containing $\underline{0}$ in the finite domain of $\Psi(\underline{t}, \cdot, x)$.

Now that we have established a large deviation principle for $(\pi_{\underline{t}\#}\mathrm{P}_x^{\epsilon_m})_{m\in\mathbb{N}}$, we seek to establish one for $(\pi_{\underline{t}\#}\mathrm{P}_x^{\epsilon})_{\epsilon>0}$. We start by defining a map $\epsilon \mapsto \tilde{\epsilon}$ which discretizes the nature of $\epsilon > 0$; denoting $[r] \in \mathbb{Z}$ the integer part of $r \in \mathbb{R}$, define $\tilde{\epsilon} := [\epsilon^{-1}]^{-1}$. The following quick inequalities relating ϵ and $\tilde{\epsilon}$,

$$\tilde{\epsilon} - \tilde{\epsilon}^2 < \epsilon \leq \tilde{\epsilon}$$

make it easy to directly show $(\pi_{\underline{t}\#}\mathrm{P}_x^{\tilde{\epsilon}})_{\epsilon>0}$ satisfies a large deviation principle; for each $\Gamma \in \mathscr{B}(\mathbb{V}^{|\underline{t}|})$,

$$\begin{aligned}
-\inf_{\underline{x}\in\Gamma^\circ} \Psi^*(\underline{t}, \underline{x}, x) &\leq \liminf_{m\to\infty} \epsilon_m \log \pi_{\underline{t}\#}\mathrm{P}_x^{\epsilon_m}\Gamma \\
&= \liminf_{\epsilon\to 0} \tilde{\epsilon} \log \pi_{\underline{t}\#}\mathrm{P}_x^{\tilde{\epsilon}}\Gamma \\
&\leq \liminf_{\epsilon\to 0} (\epsilon + \tilde{\epsilon}^2) \log \pi_{\underline{t}\#}\mathrm{P}_x^{\tilde{\epsilon}}\Gamma \\
&= \liminf_{\epsilon\to 0} \epsilon \log \pi_{\underline{t}\#}\mathrm{P}_x^{\tilde{\epsilon}}\Gamma \\
&\leq \limsup_{\epsilon\to 0} \epsilon \log \pi_{\underline{t}\#}\mathrm{P}_x^{\tilde{\epsilon}}\Gamma \\
&\leq \limsup_{\epsilon\to 0} \tilde{\epsilon} \log \pi_{\underline{t}\#}\mathrm{P}_x^{\tilde{\epsilon}}\Gamma \\
&= \limsup_{m\to\infty} \epsilon_m \log \pi_{\underline{t}\#}\mathrm{P}_x^{\epsilon_m}\Gamma \leq -\inf_{\underline{x}\in\overline{\Gamma}} \Psi^*(\underline{t}, \underline{x}, x)
\end{aligned}$$

To obtain a large deviation principle for our family $(\pi_{\underline{t}\#}\mathrm{P}_x^{\epsilon})_{\epsilon>0}$, we show regularity $\epsilon \to \pi_{\underline{t}\#}\mathrm{P}_x^{\epsilon}$ to lift the principle for the discretized family $(\pi_{\underline{t}\#}\mathrm{P}_x^{\tilde{\epsilon}})_{\epsilon>0}$. This notion in the literature is known as an *exponential approximation*, which is explored in [DZ10, Section 4.2.2]. From [DZ10, Theorem 4.2.13], it suffices to construct a probability space $(\Omega, \Sigma, \mathrm{P})$ such that elements $(\epsilon X^\epsilon)_{\epsilon>0}$ on this space with distributions $(\mathrm{P}_x^\epsilon)_{\epsilon>0}$ satisfy the following exponential equivalence property.

$$\limsup_{\epsilon\to 0} \epsilon \log \mathrm{P}\Big(|\epsilon X_{\underline{t}}^\epsilon - \tilde{\epsilon} X_{\underline{t}}^{\tilde{\epsilon}}| > \delta\Big) = -\infty$$

The *scaled-dynamics* realization from Proposition 2.1.9 will do just this.

Cramér's theorem—the tool we leveraged to prove the above principle—is proven by using measure changes induced by densities of the following form.

$$\exp\Big(\langle \underline{u}, X_{\underline{t}}^\epsilon\rangle - \Psi^\epsilon(\underline{t}, \underline{u}, X_0^\epsilon)\Big)$$

Observe that Theorem 1.3.12 gives us a perspective of how this measure depends on the increments; a valid moment $\underline{u} \in \mathbb{V}^{|\underline{t}|}$ for the above expression must satisfy $\underline{u} = U_{\underline{t}}(\underline{\theta})$ for some $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$, and

$$\exp\Big(\langle \underline{u}, X_{\underline{t}}^\epsilon\rangle - \Psi^\epsilon(\underline{t}, \underline{u}, X_0^\epsilon)\Big) = \exp\sum_{k=1}^{|\underline{t}|}\Big(\langle \theta_k, X_{t_k}^\epsilon - X_{t_{k-1}}^\epsilon\rangle - \varphi^\epsilon\big(\Delta t_k, \theta_k, X_{t_{k-1}}^\epsilon\big)\Big) =: \underline{Z}^{\epsilon,\underline{t},\underline{\theta}}$$

We denote these changes of measure like so,

$$\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}(\mathrm{d}\omega) := \underline{Z}^{\epsilon,\underline{t},\underline{\theta}}(\omega) \cdot \mathrm{P}(\mathrm{d}\omega),$$

and observe the nature of how they make our increments distribute.

Proposition 2.3.2. *Fix $\epsilon > 0$, $x \in \mathbb{X}$, $\underline{t} \vdash [0,\infty)$, and $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$. For each $\ell = 1,\dots,|\underline{t}|$ and $\theta \in \mathbb{V}$ with $\theta_\ell + \theta \in \mathcal{D}_\Psi(\Delta t_\ell)$, we have the following moments.*

$$\begin{aligned}&\mathrm{E}_{\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}}\Big(\exp\langle\theta, X^\epsilon_{t_\ell} - X^\epsilon_{t_{\ell-1}}\rangle|\mathscr{F}^\epsilon_{t_{\ell-1}}\Big)\\ &\qquad = \exp\Big(\varphi^\epsilon\big(\Delta t_\ell, \theta_\ell+\theta, X^\epsilon_{t_{\ell-1}}\big) - \varphi^\epsilon\big(\Delta t_\ell, \theta_\ell, X^\epsilon_{t_{\ell-1}}\big)\Big)\end{aligned}$$

This furthermore means we have the following conditional expectations.

$$\mathrm{E}_{\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}}\Big(\epsilon X^\epsilon_{t_\ell} - \epsilon X^\epsilon_{t_{\ell-1}}|\epsilon X^\epsilon_{\underline{t}_{1:\ell-1}} = \underline{x}_{1:\ell-1}\Big) = \nabla_{\theta_\ell}\varphi(\Delta t_\ell, \theta_\ell, x_{\ell-1})$$

Proof. Denote $n := |\underline{t}|$ for brevity. We first show the following conditional expectation for any $m = n-1,\dots,1$.

$$\mathrm{E}_{\mathrm{P}^\epsilon_x}\big(\underline{Z}^{\epsilon,\underline{t},\underline{\theta}}|\mathscr{F}^\epsilon_{t_m}\big) = \prod_{k=1}^{m}\big(\langle\theta_k, X^\epsilon_{t_k} - X^\epsilon_{t_{k-1}}\rangle - \varphi^\epsilon(\Delta t_k, \theta_k, X^\epsilon_{t_{k-1}})\big), \tag{2.3.3}$$

by iteratively projecting onto $\mathscr{F}^\epsilon_{t_n},\dots,\mathscr{F}^\epsilon_{t_{m+1}}$. For any quantity $H \in \mathscr{F}_{t_m}/\mathscr{B}(\mathbb{R}_+)$, we have the following.

$$\begin{aligned}&\mathrm{E}_{\mathrm{P}^\epsilon_x}\big(H\underline{Z}^{\epsilon,\underline{t},\underline{\theta}}\big)\\ &= \mathrm{E}_{\mathrm{P}^\epsilon_x}\Big(H\prod_{k=1}^{n}\exp\big(\langle\theta_k, X^\epsilon_{t_k} - X^\epsilon_{t_{k-1}}\rangle - \varphi^\epsilon(\Delta t_k, \theta_k, X^\epsilon_{t_{k-1}})\big)\Big)\\ &= \mathrm{E}_{\mathrm{P}^\epsilon_x}\Big(H\prod_{k=1}^{n-1}\exp\big(\langle\theta_k, X^\epsilon_{t_k} - X^\epsilon_{t_{k-1}}\rangle - \varphi^\epsilon(\Delta t_k, \theta_k, X^\epsilon_{t_{k-1}})\big)\\ &\qquad\qquad \mathrm{E}_{\mathrm{P}^\epsilon_x}\big(\exp\langle\theta_n, X^\epsilon_{t_n} - X^\epsilon_{t_{n-1}}\rangle|\mathscr{F}^\epsilon_{t_{n-1}}\big)\exp\big(-\varphi^\epsilon(\Delta t_n, \theta_n, X^\epsilon_{t_{n-1}})\big)\Big)\\ &= \mathrm{E}_{\mathrm{P}^\epsilon_x}\Big(H\prod_{k=1}^{n-1}\exp\big(\langle\theta_k, X^\epsilon_{t_k} - X^\epsilon_{t_{k-1}}\rangle - \varphi^\epsilon(\Delta t_k, \theta_k, X^\epsilon_{t_{k-1}})\big)\Big)\\ &\ \ \vdots\\ &= \mathrm{E}_{\mathrm{P}^\epsilon_x}\Big(H\prod_{k=1}^{m}\exp\big(\langle\theta_k, X^\epsilon_{t_k} - X^\epsilon_{t_{k-1}}\rangle - \varphi^\epsilon(\Delta t_k, \theta_k, X^\epsilon_{t_{k-1}})\big)\Big)\end{aligned}$$

which indicates that (2.3.3) is true. Now choosing $H \in \mathscr{F}^\epsilon_{t_{\ell-1}}/\mathscr{B}(\mathbb{R}_+)$, we apply (2.3.3) for $m = \ell$ and $m = \ell - 1$ to see that the following holds.

$$\begin{aligned}&\mathrm{E}_{\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}}\Big(\exp\langle\theta, X^\epsilon_{t_\ell} - X^\epsilon_{t_{\ell-1}}\rangle H\Big)\\ &= \mathrm{E}_{\mathrm{P}^\epsilon_x}\Big(\exp\langle\theta, X^\epsilon_{t_\ell} - X^\epsilon_{t_{\ell-1}}\rangle H\underline{Z}^{\epsilon,\underline{t},\underline{\theta}}\Big)\\ &= \mathrm{E}_{\mathrm{P}^\epsilon_x}\Big(\exp\langle\theta, X^\epsilon_{t_\ell} - X^\epsilon_{t_{\ell-1}}\rangle H\prod_{k=1}^{\ell}\exp\big(\langle\theta_k, X^\epsilon_{t_k} - X^\epsilon_{t_{k-1}}\rangle - \varphi^\epsilon(\Delta t_k, \theta_k, X^\epsilon_{t_{k-1}})\big)\Big)\\ &= \mathrm{E}_{\mathrm{P}^\epsilon_x}\Big(H\prod_{k=1}^{\ell-1}\exp\big(\langle\theta_k, X^\epsilon_{t_k} - X^\epsilon_{t_{k-1}}\rangle - \varphi^\epsilon(\Delta t_k, \theta_k, X^\epsilon_{t_{k-1}})\big)\\ &\qquad\qquad \mathrm{E}_{\mathrm{P}^\epsilon_x}\big(\exp\langle\theta_\ell+\theta, X^\epsilon_{t_\ell} - X^\epsilon_{t_{\ell-1}}\rangle|\mathscr{F}_{t_{\ell-1}}\big)\exp\big(-\varphi^\epsilon(\Delta t_\ell, \theta_\ell, X^\epsilon_{t_{\ell-1}})\big)\Big)\end{aligned}$$

$$= \mathrm{E}_{\mathrm{P}_x^\epsilon}\Big(H\prod_{k=1}^{\ell-1}\exp\big(\langle\theta_k, X^\epsilon_{t_k}-X^\epsilon_{t_{k-1}}\rangle-\varphi^\epsilon(\Delta t_k,\theta_k,X^\epsilon_{t_{k-1}})\big)$$

$$\exp\big(\varphi^\epsilon(\Delta t_\ell,\theta_\ell+\theta,X^\epsilon_{t_{\ell-1}})-\varphi^\epsilon(\Delta t_\ell,\theta_\ell,X^\epsilon_{t_{\ell-1}})\big)\Big)$$

$$= \mathrm{E}_{\mathrm{P}_x^\epsilon}\Big(\exp\big(\varphi^\epsilon(\Delta t_\ell,\theta_\ell+\theta,X^\epsilon_{t_{\ell-1}})-\varphi^\epsilon(\Delta t_\ell,\theta_\ell,X^\epsilon_{t_{\ell-1}})\big)H\underline{Z}^{\epsilon,\underline{t},\underline{\theta}}\Big)$$

$$= \mathrm{E}_{\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}}\Big(\exp\big(\varphi^\epsilon(\Delta t_\ell,\theta_\ell+\theta,X^\epsilon_{t_{\ell-1}})-\varphi^\epsilon(\Delta t_\ell,\theta_\ell,X^\epsilon_{t_{\ell-1}})\big)H\Big)$$

This gives us our first desired identity.

$$\mathrm{E}_{\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}}\Big(\exp\langle\theta,X^\epsilon_{t_\ell}-X^\epsilon_{t_{\ell-1}}\rangle|\mathscr{F}^\epsilon_{t_{\ell-1}}\Big)$$
$$=\exp\big(\varphi^\epsilon(\Delta t_\ell,\theta_\ell+\theta,X^\epsilon_{t_{\ell-1}})-\varphi^\epsilon(\Delta t_\ell,\theta_\ell,X^\epsilon_{t_{\ell-1}})\big)$$

Now that this identity is established, we appeal to Propositions 2.2.7(a), 1.2.9, and 2.1.3 in specifying an open ball $B(\theta_k,\delta)\subseteq\mathcal{D}_\Psi(\Delta t_k)$ on which we may apply derivatives to get the following identity.

$$\begin{aligned}\mathrm{E}_{\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}}\Big(\epsilon X^\epsilon_{t_\ell}-\epsilon X^\epsilon_{t_{\ell-1}}|\mathscr{F}^\epsilon_{t_{\ell-1}}\Big) &= \epsilon\mathrm{E}_{\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}}\Big(X^\epsilon_{t_\ell}-X^\epsilon_{t_{\ell-1}}|\mathscr{F}^\epsilon_{t_{\ell-1}}\Big)\\
&= \epsilon\mathrm{E}_{\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}}\Big(\nabla_\theta\exp\langle\theta,X^\epsilon_{t_\ell}-X^\epsilon_{t_{\ell-1}}\rangle|_{\theta=0}|\mathscr{F}^\epsilon_{t_{\ell-1}}\Big)\\
&= \epsilon\nabla_\theta\mathrm{E}_{\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}}\Big(\exp\langle\theta,X^\epsilon_{t_\ell}-X^\epsilon_{t_{\ell-1}}\rangle|\mathscr{F}^\epsilon_{t_{\ell-1}}\Big)\Big|_{\theta=0}\\
&= \epsilon\nabla_\theta\exp\Big(\varphi^\epsilon\big(\Delta t_\ell,\theta_\ell+\theta,X^\epsilon_{t_{\ell-1}}\big)-\varphi^\epsilon\big(\Delta t_\ell,\theta_\ell,X^\epsilon_{t_{\ell-1}}\big)\Big)\Big|_{\theta=0}\\
&= \nabla_\theta\exp\Big(\varphi\big(\Delta t_\ell,\theta_\ell+\theta,\epsilon X^\epsilon_{t_{\ell-1}}\big)-\varphi\big(\Delta t_\ell,\theta_\ell,\epsilon X^\epsilon_{t_{\ell-1}}\big)\Big)\Big|_{\theta=0}\\
&= \nabla_{\theta_\ell}\varphi\big(\Delta t_\ell,\theta_\ell,\epsilon X^\epsilon_{t_{\ell-1}}\big)\end{aligned}$$

Seeing as the above quantity is measurable with respect to the σ-algebra generated by $X_{\underline{t}_{1:\ell}}$, we have our desired identity.

$$\mathrm{E}_{\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}}\Big(\epsilon X^\epsilon_{t_\ell}-\epsilon X^\epsilon_{t_{\ell-1}}|\epsilon X^\epsilon_{\underline{t}_{1:\ell-1}}=\underline{x}_{1:\ell-1}\Big)=\nabla_{\theta_\ell}\varphi\big(\Delta t_\ell,\theta_\ell,\epsilon X^\epsilon_{t_{\ell-1}}\big)$$

Understanding these measures $\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}$ will be important for our next section, where we relate the seemingly alternative exponential martingale method to studying large deviations. We now proceed from that short tangent back to our large deviation principle.

Proposition 2.3.4. *For each $x\in\mathbb{X}$, the family $(\mathrm{P}^\epsilon_x)_{\epsilon>0}$ is exponentially tight on $\mathbb{D}([0,\infty),\mathbb{X})$.*

Proof. First note that each family $(\pi_{\underline{t}\#}\mathrm{P}^\epsilon_x)_{\epsilon>0}$ of finite-dimensional distributions is exponentially tight by Theorem 2.3.1; for each $\alpha>0$, $K_\alpha=\Psi^*(\underline{t},\cdot,x)^{-1}[0,\alpha]$ is a compact subset of $\mathbb{V}^{|\underline{t}|}$ (by goodness of the rate function) clearly satisfying the following.

$$\limsup_{\epsilon\to0}\epsilon\log\pi_{\underline{t}\#}\mathrm{P}^\epsilon_x(K^c_\alpha)\le-\inf_{\underline{x}\in\overline{K^c_\alpha}}\Psi^*(\underline{t},\underline{x},x)\le-\alpha$$

By [FK06, Theorem 4.1] (rather, adapting to countinuously parameterized family), our desired exponential tightness is obtained if we produce to each $\epsilon,\delta,\lambda,T>0$ a random variable $\gamma_\epsilon(\delta,\lambda,T)$ with the following dominating property over all $t\in[0,T]$, $s\in[0,\delta]$,

$$\mathrm{E}_{\mathrm{P}_x}\Big(\exp\big(\epsilon^{-1}\lambda\big(|\epsilon X^\epsilon_{t+s}-\epsilon X^\epsilon_t|\wedge1\big)|\mathscr{F}^\epsilon_t\Big)\le\mathrm{E}_{\mathrm{P}_x}\Big(\exp\gamma_\epsilon(\delta,\lambda,T)|\mathscr{F}^\epsilon_t\Big),\tag{2.3.5}$$

such that the following equalities are true for all $\lambda > 0$.

$$\lim_{\delta\to 0}\limsup_{\epsilon\to 0}\epsilon\log \mathrm{E}_{\mathrm{P}_x}\exp\left(\epsilon^{-1}\lambda\big(|\epsilon X^\epsilon_\delta - \epsilon X^\epsilon_0|\wedge 1\big)\right) = 0 \tag{2.3.6}$$

$$\lim_{\delta\to 0}\limsup_{\epsilon\to 0}\epsilon\log \mathrm{E}_{\mathrm{P}_x}\exp\gamma_\epsilon(\delta,\lambda,T) = 0 \tag{2.3.7}$$

To show this fact, first note that, by Proposition 2.2.7(d), to each $\lambda > 0$ there exist $\delta_\lambda > 0$ and $C_\lambda > 0$ such that the following inequality holds.

$$\big|\Psi(t,u,x) - \Psi(0,u,x)\big| \le C_\lambda\cdot t\cdot\big(1+|x|\big),\quad t\in[0,\delta_\lambda],\ u\in\overline{B}(0,\lambda\sqrt{d}),\ x\in\mathbb{X}$$

We then define the following function $f_\epsilon(\cdot,\delta,\lambda):\mathbb{X}\to\mathbb{R}$ for each $\epsilon>0$, $\lambda>0$, and $\delta\in[0,\delta_\lambda]$.

$$f_\epsilon(x',\delta,\lambda) \coloneqq \log 2d + \frac{1}{\epsilon}\cdot C_\lambda\cdot\delta\cdot\big(1+|x|\big)$$

We now define $\gamma_\epsilon(\delta,\lambda,T)\coloneqq f_\epsilon\big(\epsilon X^\epsilon_t,\delta,\lambda\big)$.

Note that for all $\lambda,\epsilon,t>0$, $\delta\in[0,\delta_\lambda]$, and $s\in[0,\delta]$, we use Proposition 2.1.3 to get the following.

$$\begin{aligned}
&\mathrm{E}_{\mathrm{P}_x}\Big(\exp\left(\epsilon^{-1}\lambda\big(|\epsilon X^\epsilon_{t+s}-\epsilon X^\epsilon_t|\wedge 1\big)|\mathscr{F}^\epsilon_t\right)\\
&\quad\le \mathrm{E}_{\mathrm{P}_x}\Big(\exp\left(\epsilon^{-1}\lambda\big|\epsilon X^\epsilon_{t+s}-\epsilon X^\epsilon_t\big|\right)|\mathscr{F}^\epsilon_t\Big)\\
&\quad\le \sum_{\ell=0}^{1}\sum_{i=1}^{d}\mathrm{E}_{\mathrm{P}_x}\left(\exp\left(\Big\langle(-1)^\ell\epsilon^{-1}\lambda\sqrt{d}e^i,\epsilon X^\epsilon_{t+s}-\epsilon X^\epsilon_t\Big\rangle\right)|\mathscr{F}^\epsilon_t\right)\\
&\quad= \sum_{\ell=0}^{1}\sum_{i=1}^{d}\exp\left(\epsilon^{-1}\Psi\Big(s,(-1)^\ell\lambda\sqrt{d}e^i,\epsilon X^\epsilon_t\Big)-\epsilon^{-1}\Big\langle(-1)^\ell\lambda\sqrt{d}e^i,\epsilon X^\epsilon_t\Big\rangle\right)\\
&\quad\le 2d\cdot\exp\left(\frac{1}{\epsilon}\cdot C_\lambda\cdot\delta\cdot\big(1+|\epsilon X^\epsilon_t|\big)\right)
\end{aligned}$$

$$\quad= \exp f_\epsilon(\epsilon X^\epsilon_t,\delta,\lambda) \tag{2.3.8}$$

Note that (2.3.8) makes (2.3.5) true.

$$\begin{aligned}
\mathrm{E}_{\mathrm{P}_x}\Big(\exp\left(\epsilon^{-1}\lambda\big(|\epsilon X^\epsilon_{t+s}-\epsilon X^\epsilon_t|\wedge 1\big)|\mathscr{F}^\epsilon_t\right) &\le \exp f_\epsilon(\epsilon X^\epsilon_t,\delta,\lambda)\\
&= \exp\gamma_\epsilon(\delta,\lambda,T)\\
&= \mathrm{E}_{\mathrm{P}_x}\Big(\exp\gamma_\epsilon(\delta,\lambda,T)|\mathscr{F}^\epsilon_t\Big)
\end{aligned}$$

For (2.3.6), we also use (2.3.8).

$$\begin{aligned}
&\lim_{\delta\to 0}\limsup_{\epsilon\to 0}\epsilon\log \mathrm{E}_{\mathrm{P}_x}\exp\left(\epsilon^{-1}\lambda\big(|\epsilon X^\epsilon_\delta-\epsilon X^\epsilon_0|\wedge 1\big)\right)\\
&\quad= \lim_{\delta\to 0}\limsup_{\epsilon\to 0}\epsilon\log \mathrm{E}_{\mathrm{P}_x}\mathrm{E}_{\mathrm{P}_x}\Big(\exp\left(\epsilon^{-1}\lambda\big(|\epsilon X^\epsilon_\delta-\epsilon X^\epsilon_0|\wedge 1\big)\right)|\mathscr{F}^\epsilon_0\Big)\\
&\quad\le \lim_{\delta\to 0}\limsup_{\epsilon\to 0}\epsilon\log \mathrm{E}_{\mathrm{P}_x}\exp f_\epsilon(\epsilon X^\epsilon_0,\delta,\lambda)\\
&\quad= \lim_{\delta\to 0}\limsup_{\epsilon\to 0}\Big(\epsilon\log 2d + C_\lambda\cdot\delta\cdot\big(1+|x|\big)\Big)
\end{aligned}$$

$$= 0$$

Using Proposition 2.2.7(a), to each $\lambda > 0$, there exists $\delta'_\lambda > 0$ such that $B(0, \delta_\lambda) \subseteq \mathcal{D}_\Psi(\delta'_\lambda)$. Now, for any $\delta < \delta_\lambda \wedge \delta'_\lambda/(2C_\lambda\sqrt{d})$, we again use Proposition 2.1.3 to get the following.

$$\begin{aligned}
&\epsilon \log \mathrm{E}_{\mathrm{P}_x}\Big(\exp \gamma_\epsilon(\delta, \lambda, T)\Big) \\
&\quad = \epsilon \log \mathrm{E}_{\mathrm{P}_x}\Big(\exp f_\epsilon(\epsilon X^\epsilon_t, \delta, \lambda)\Big) \\
&\quad = \epsilon \log 2d + C_\lambda \cdot \delta + \epsilon \log \mathrm{E}_{\mathrm{P}_x} \exp\left(\epsilon^{-1} C_\lambda \delta |\epsilon X^\epsilon_t|\right) \\
&\quad \leq \epsilon \log 2d + C_\lambda \cdot \delta + \epsilon \log \sum_{\ell=0}^{1} \sum_{i=1}^{d} \mathrm{E}_{\mathrm{P}_x} \exp \left\langle (-1)^\ell \epsilon^{-1} C_\lambda \delta \sqrt{d} e^i, \epsilon X^\epsilon_t \right\rangle \\
&\quad \leq 2\epsilon \log 2d + C_\lambda \cdot \delta + \max_{\substack{\ell=0,1 \\ i=1,\dots,d \\ k=1,\dots,|\underline{t}|}} \epsilon \log \mathrm{E}_{\mathrm{P}_x} \exp \Big\langle (-1)^\ell \epsilon^{-1} C_\lambda \delta \sqrt{d} e^i, \epsilon X^\epsilon_t \Big\rangle \\
&\quad \leq 2\epsilon \log 2d + C_\lambda \cdot \delta + \max_{\substack{\ell=0,1 \\ i=1,\dots,d \\ k=1,\dots,|\underline{t}|}} \Psi\Big(t, (-1)^\ell C_\lambda \delta \sqrt{d} e^i, x\Big)
\end{aligned}$$

This gives us (2.3.7).

Theorem 2.3.9. *For each $x \in \mathbb{X}^\circ$, the family $(\mathrm{P}^\epsilon_x)_{\epsilon>0}$ satisfies a large deviation principle on $\mathbb{D}([0,\infty), \mathbb{X})$ with good rate function $I_x : \mathbb{D}([0,\infty), \mathbb{X}) \to [0,\infty]$ as follows.*

$$I_x(\xi) = \begin{cases} \sup\limits_{\underline{t} \vdash \Delta^c_\xi} \Psi^*\big(\underline{t}, \xi(\underline{t}), \xi(0)\big) & \xi(0) = x \\ \infty & \textit{otherwise} \end{cases} \tag{2.3.10}$$

Above, $\Delta_\xi \subseteq [0,\infty)$ denotes the points of discontinuity of ξ.

Proof. Provided we have some $\underline{t} \vdash [0,\infty)$ the vector $\hat{\underline{t}}$ associated with prepending 0 to $\underline{t}$,

$$\hat{\underline{t}} = (0, t_1, \dots, t_{|\underline{t}|}),$$

induces the following finite-dimensional distributions.

$$\pi_{\hat{\underline{t}}\#} \mathrm{P}^\epsilon_x = \delta_x \otimes \pi_{\underline{t}\#} \mathrm{P}^\epsilon_x$$

For these partitions, it is easy to see from Theorem 2.3.1 that the large deviations principle for $(\pi_{\hat{\underline{t}}\#} \mathrm{P}^\epsilon_x)_{\epsilon>0}$ has good rate function as below.

$$(x_0, x_1, \dots, x_{|\underline{t}|}) \mapsto \begin{cases} \Psi^*(\underline{t}, \underline{x}, x_0), & x_0 = x \\ \infty & \text{otherwise} \end{cases}$$

By [FK06, Theorem 4.28], we now use Theorem 2.3.1 and Proposition 2.3.4 to get a large deviations principle for the family $(\mathrm{P}^\epsilon_x)_{\epsilon>0}$ with good rate function I_x.

2.4 Exponential martingales

Now that we have our principle, we discuss how various exponential martingales $Z^{\epsilon,h}$ induce changes of measure alternative to those $\underline{Z}^{\epsilon,\underline{t},\underline{\theta}}$ which themselves inspire a different flavor of proof. We start by defining the building blocks of these exponential martingales.

$$\begin{aligned}
&\mathcal{H}_\Lambda \coloneqq \big\{h \in \mathbb{D}([0,\infty),\mathbb{V}) : h \text{ has compact support and finite-variation}\big\}\\
&G : \mathbb{D}([0,\infty),\mathbb{X}) \times \mathcal{H}_\Lambda \to \mathbb{R},\\
&\qquad G(\xi,h) \coloneqq \big\langle \xi(0), h(0)\big\rangle - \int_0^\infty \xi(s-)\mathrm{d}h(s) - \int_0^\infty \Lambda\big(h(s),\xi(s)\big)\mathrm{d}s
\end{aligned}$$

Proposition 2.4.1. *For each $\epsilon > 0$ and $h \in \mathcal{H}_\Lambda$, we have the following identities, where $\mathcal{E}(H)$ denotes the Doléans-Dade exponential of H.*

$$\begin{aligned}
Z^{\epsilon,h} &\coloneqq \exp\left(\frac{1}{\epsilon}G(\epsilon X^\epsilon, h)\right)\\
&= \exp\left(\big\langle X_0^\epsilon, h(0)\big\rangle - \int_0^\infty X_{s-}^\epsilon \mathrm{d}h(s) - \int_0^\infty \Lambda^\epsilon\big(h(s), X_s^\epsilon\big)\mathrm{d}s\right)\\
&= \exp\left(h \bullet X_\infty^\epsilon - \Lambda^\epsilon\big(h, X^\epsilon\big) \bullet \ell_\infty\right)\\
&= \mathcal{E}\left(h \bullet X^{\epsilon,c} - \big(e^{\langle h, \mathrm{id}_\mathbb{V}\rangle} - 1\big) * \tilde{q}^{X^\epsilon}\right)_\infty
\end{aligned}$$

Furthermore, $Z^{\epsilon,h}$ is integrable and $t \mapsto \mathrm{E}_{\mathrm{P}_x}(Z^{\epsilon,h}|\mathscr{F}_t)$ is a martingale.

Proof. We start by establishing the first two identities. Realizing h as a predictable, finite-variation process, we apply integration by parts (see [JS03, Proposition I.4.49(b)]) to get the following identity.

$$\begin{aligned}
\frac{1}{\epsilon}G(\epsilon X^\epsilon, h) &= \frac{1}{\epsilon}\Big(\big\langle \epsilon X_0^\epsilon, h(0)\big\rangle - \int_0^\infty \epsilon X_{s-}^\epsilon \mathrm{d}h(s) - \int_0^\infty \Lambda\big(h(s), \epsilon X_s^\epsilon\big)\mathrm{d}s\Big)\\
&= \big\langle X_0^\epsilon, h(0)\big\rangle - \int_0^\infty X_{s-}^\epsilon \mathrm{d}h(s) - \int_0^\infty \Lambda^\epsilon\big(h(s), X_s^\epsilon\big)\mathrm{d}s\\
&= \big\langle X_0^\epsilon, h(0)\big\rangle - X_-^\epsilon \bullet h_\infty - \Lambda^\epsilon(h, X_-^\epsilon) \bullet \ell_\infty\\
&= h \bullet X_\infty^\epsilon - \Lambda^\epsilon(\theta, X_-^\epsilon) \bullet \ell_\infty
\end{aligned}$$

The remaining identity is a special case of that from [JS03, Theorem III.7.24], but we will perform the Itô calculus here for completion's sake. Note that Theorem 1.4.4 tells us that $\exp(h \bullet X^\epsilon)$ is a $(\mathrm{P}_x, \mathscr{F})$ jump-diffusion, and (A.4.15) from Theorem A.4.13 gives us its special semimartingale decomposition.

$$\begin{aligned}
\exp\left(h \bullet X_t^\epsilon\right) = &\Big(\exp\left(h \bullet X^\epsilon\right) \cdot \Lambda(h, X^\epsilon)\Big) \bullet \ell_t + \Big(\exp\left(h \bullet X_-^\epsilon\right) \cdot h\Big) \bullet X_t^{\epsilon,\mathrm{c}}\\
&+ \exp\left(h \bullet X_-^\epsilon\right)\Big(e^{\langle h, \mathrm{id}_\mathbb{V}\rangle} - 1\Big) * \tilde{q}_t^{X^\epsilon}
\end{aligned}$$

Note that this allows us to define the process H inside the alleged Doléans-Dade exponential.

$$H = h \bullet X^{\epsilon,c} + \big(e^{\langle h, \mathrm{id}_\mathbb{V}\rangle} - 1\big) * \tilde{q}^{X^\epsilon}$$

Observe that, by definition of the Doléans-Dade exponential and the special semimartingale decomposition of X^ϵ,

$$
\begin{aligned}
X^\epsilon_t &= \frac{1}{\epsilon}\beta(\epsilon X^\epsilon)\bullet\ell + X^{\epsilon,c}_t + \mathrm{id}_{\mathbb{V}} * \tilde{q}^{X^\epsilon}_t,\\
\big\langle X^{\epsilon,c,i}, X^{\epsilon,c,j}\big\rangle &= \frac{1}{\epsilon}\alpha(\epsilon X^\epsilon)\bullet\ell,\\
\hat{q}^{X^\epsilon}(\mathrm{d}s,\mathrm{d}v) &= \frac{1}{\epsilon}\mu(\epsilon X^\epsilon,\mathrm{d}v)\mathrm{d}s,
\end{aligned}
$$

we have the following.

$$
\begin{aligned}
\log\mathcal{E}(H)_t &= H_t - H_0 - \frac{1}{2}\langle H^c, H^c\rangle_t + \big(\log(1+\mathrm{id}_{\mathbb{R}}) - \mathrm{id}_{\mathbb{R}}\big) * q^H_t\\
&= h\bullet X^{\epsilon,c}_t + \big(e^{\langle h,\mathrm{id}_{\mathbb{V}}\rangle} - 1\big) * \tilde{q}^{X^\epsilon}_t\\
&\quad - \frac{1}{2}\langle h\bullet X^{\epsilon,c}, h\bullet X^{\epsilon,c}\rangle_t - \big(e^{\langle h,\mathrm{id}_{\mathbb{V}}\rangle} - 1 - \langle h,\mathrm{id}_{\mathbb{V}}\rangle\big) * q^{X^\epsilon}_t\\
&= h\bullet X^{\epsilon,c}_t - \big(e^{\langle h,\mathrm{id}_{\mathbb{V}}\rangle} - 1 - \langle h,\mathrm{id}_{\mathbb{V}}\rangle\big) * \hat{q}^{X^\epsilon}_t\\
&\quad - \frac{1}{2}\langle h, \alpha^\epsilon(X^\epsilon)h\rangle\bullet\ell_t + \langle h,\mathrm{id}_{\mathbb{V}}\rangle * \tilde{q}^{X^\epsilon}_t\\
&= h\bullet X^\epsilon_t - \beta^\epsilon(X^\epsilon)\bullet\ell_t - \langle h,\mathrm{id}_{\mathbb{V}}\rangle * \tilde{q}^{X^\epsilon}_t - \big(e^{\langle h,\mathrm{id}_{\mathbb{V}}\rangle} - 1 - \langle h,\mathrm{id}_{\mathbb{V}}\rangle\big) * \hat{q}^{X^\epsilon}_t\\
&\quad - \frac{1}{2}\langle h, \alpha^\epsilon(X^\epsilon)h\rangle\bullet\ell_t + \langle h,\mathrm{id}_{\mathbb{V}}\rangle * \tilde{q}^{X^\epsilon}_t\\
&= h\bullet X^\epsilon_t - \Lambda^\epsilon(h, X^\epsilon)\bullet\ell_t
\end{aligned}
$$

Note that this is one argument that $Z^{\epsilon,h} = \mathcal{E}(H)_\infty$ corresponds to a $(\mathrm{P},\mathscr{F})$ local martingale, since H is (see [JS03, Theorem I.4.61(b)]). In any case, the martingale nature comes from Theorem 1.4.4.

The above proposition prescribes another change of measure. For each $\epsilon > 0$, $h \in \mathcal{H}_\Lambda$, and $x \in \mathbb{X}$, define

$$\mathrm{Q}^{\epsilon,h}_x(\mathrm{d}\omega) \coloneqq Z^{\epsilon,h}(\omega)\cdot\mathrm{P}^\epsilon_x(\mathrm{d}\omega).$$

Let us explore the distribution of ϵX^ϵ over these spaces.

Proposition 2.4.2. *Fix $\epsilon > 0$, $h \in \mathcal{H}_\Lambda$, and $x \in \mathbb{X}$. The process X^ϵ is a $(\mathrm{Q}^{\epsilon,h}_x, \mathscr{F}^\epsilon)$ special semimartingale with the following decomposition.*

$$X^\epsilon_t = x/\epsilon + \frac{1}{\epsilon}\beta^h(\cdot, \epsilon X^\epsilon_-)\bullet\ell_t + X^{\epsilon,c}_t + \mathrm{id}_{\mathbb{V}} * \tilde{q}^{h,X^\epsilon}_t,$$

where the drift β^h, diffusion α, and jump predictable compensator $\hat{q}^{h,X^\epsilon}$ (above, we have $\tilde{q}^{h,X^\epsilon} = q^{X^\epsilon} - \hat{q}^{h,X^\epsilon}$) are as follows.

$$
\begin{aligned}
\beta^h(s,x) &\coloneqq \beta(x) + \alpha(x)h(s) + \int_{\mathbb{V}} v\big(e^{\langle h(s),v\rangle} - 1\big)\mu(x,\mathrm{d}v)\\
\langle X^{\epsilon,c,i}, X^{\epsilon,c,j}\rangle &= \frac{1}{\epsilon}\alpha_{i,j}(\epsilon X^\epsilon_-)\bullet\ell\\
\hat{q}^{h,X^\epsilon}(\mathrm{d}s,\mathrm{d}v) &\coloneqq e^{\langle h(s),v\rangle}\frac{1}{\epsilon}\mu(\epsilon X^\epsilon_{s-},\mathrm{d}v)\mathrm{d}s
\end{aligned}
$$

Moreover the distributions $\mathrm{Q}_x^{\epsilon,h}$ weakly converge to a degenerate measure δ_{ξ_h} at the solution ξ_h to the following dynamical system.

$$\begin{cases} \dot{\xi}_h(t) = \beta^h\big(t, \xi_h(t)\big) \quad t \geq 0 \\ \xi_h(0) = x \end{cases} \tag{2.4.3}$$

Proof. The $(\mathrm{Q}_x^{\epsilon,h}, \mathscr{F}^\epsilon)$ dynamics of X^ϵ simply come from Theorem 1.4.4. As far as weak convergence is concerned, this is immediate from the exponential tightness of our family and convergence of finite-dimensional distributions.

From here, a large deviation principle may be approached by concentrating probability on a ball and using Chebyshev-like bounds on $Z^{\epsilon,h}$.

$$\begin{aligned} \epsilon \log \mathrm{P}_x^\epsilon\big(\epsilon X^\epsilon \in B(\xi,\delta)\big) &= \epsilon \log \mathrm{E}_{\mathrm{Q}_x^{\epsilon,h}}\big((Z^{\epsilon,h})^{-1} 1_{B(\xi,\delta)}(\epsilon X^\epsilon)\big) \\ &= \epsilon \log \mathrm{E}_{\mathrm{Q}_x^{\epsilon,h}}\Big(\exp\big(-\frac{1}{\epsilon}G(\epsilon X^\epsilon, h)\big) 1_{B(\xi,\delta)}(\epsilon X^\epsilon)\Big) \\ &\leq -\inf_{\xi' \in B(\xi,\delta)} G(\xi', h) \end{aligned}$$

Showing lower semi-continuity of $G(\cdot, h)$ would then result in us being able to say

$$\lim_{\delta \to 0} \epsilon \log \mathrm{P}_x^\epsilon\big(\epsilon X^\epsilon \in B(\xi,\delta)\big) \leq -G(\xi, h),$$

for all $h \in \mathcal{H}_\Lambda$, and so we'd have

$$\lim_{\delta \to 0} \epsilon \log \mathrm{P}_x^\epsilon\big(\epsilon X^\epsilon \in B(\xi,\delta)\big) \leq -\sup_{h \in \mathcal{H}_\Lambda} G(\xi, h)$$

Meanwhile, the lower bound would be approached by showing $\mathcal{H}_\Lambda$ is suitably rich to have a dense family of limit functions ξ_h as in (2.4.3) of Proposition 2.4.2, on which $\sup_{h' \in \mathcal{H}_\Lambda} G(\xi_h, h') = G(\xi_h, h)$, and so a large deviations lower bound is attained from the following bound.

$$\begin{aligned} \epsilon \log \mathrm{P}_x^\epsilon\big(\epsilon X^\epsilon \in B(\xi_h,\delta)\big) &= \epsilon \log \mathrm{E}_{\mathrm{Q}_x^{\epsilon,h}}\big((Z^{\epsilon,h})^{-1} 1_{B(\xi_h,\delta)}(\epsilon X^\epsilon)\big) \\ &= \epsilon \log \mathrm{E}_{\mathrm{Q}_x^{\epsilon,h}}\Big(\exp\big(-\frac{1}{\epsilon}G(\epsilon X^\epsilon, h)\big) 1_{B(\xi_h,\delta)}(\epsilon X^\epsilon)\Big) \\ &\geq -\sup_{\xi' \in B(\xi_h,\delta)} G(\xi', h) + \epsilon \log \mathrm{Q}_x^{\epsilon,h}\big(\epsilon X^\epsilon \in B(\xi_h,\delta)\big) \\ &\geq -G(\xi_h, h) + \epsilon \log \mathrm{Q}_x^{\epsilon,h}\big(\epsilon X^\epsilon \in B(\xi_h,\delta)\big) \\ &= -\sup_{h' \in \mathcal{H}_\Lambda} G(\xi_h, h') + \epsilon \log \mathrm{Q}_x^{\epsilon,h}\big(\epsilon X^\epsilon \in B(\xi_h,\delta)\big) \end{aligned}$$

The benefit of this approach is that our rate function $\xi \mapsto \sup_{h \in \mathcal{H}_\Lambda} G(\xi, h)$ has an integral form. Instead of proving the large deviation principle in this alternative fashion, we reconcile the measure changes that appear in each of the approaches.

It is readily evident that $\mathrm{Q}_x^{\epsilon,h}$ is a generalization of $\underline{\mathrm{Q}}_x^{\epsilon,\underline{t},\underline{\theta}}$, as we are replacing summations in $\underline{Z}^{\epsilon,\underline{t},\underline{\theta}}$ with integrals in $Z^{\epsilon,h}$.

$$\begin{aligned} \underline{Z}^{\epsilon,\underline{t},\underline{\theta}} &= \exp \sum_{k=1}^{|\underline{t}|} \Big(\big\langle \theta_k, X_{t_k}^\epsilon - X_{t_{k-1}}^\epsilon \big\rangle - \varphi^\epsilon\big(\Delta t_k, \theta_k, X_{t_{k-1}}^\epsilon\big)\Big) \\ Z^{\epsilon,h} &= \exp\Big(\int_0^\infty h(s) \mathrm{d}X_s^\epsilon - \int_0^\infty \Lambda^\epsilon\big(h(s), X_s^\epsilon\big) \mathrm{d}s\Big) \end{aligned} \tag{2.4.4}$$

The summand $\langle \theta_k, X^\epsilon_{t_k} - X^\epsilon_{t_{k-1}} \rangle$ relates to the integral term $h(s)\mathrm{d}X^\epsilon_s$, while $\varphi^\epsilon(\Delta t_k, \theta_k, X^\epsilon_{t_{k-1}})$ relates to $\Lambda^\epsilon\big(h(s), X^\epsilon_s\big)\mathrm{d}s$. To explicitly resolve these two expressions, we disambiguate the operations in $\underline{Z}^{\epsilon,\underline{t},\underline{\theta}}$ involving X^ϵ.

$$\underline{G}(\cdot, \underline{t}, \underline{\theta}) : \mathbb{D}([0,\infty), \mathbb{X}) \to \mathbb{R},$$
$$\underline{G}(\xi, \underline{t}, \underline{\theta}) := \sum_{k=1}^{|\underline{t}|} \Big(\big\langle \theta_k, \xi(t_k) - \xi(t_{k-1}) \big\rangle - \Psi\big(\Delta t_k, \theta_k, \xi(t_{k-1})\big) \Big)$$

We now have a common notation for factoring X^ϵ through each density.

$$\underline{Z}^{\epsilon,\underline{t},\underline{\theta}} = \exp\Big(\frac{1}{\epsilon}\underline{G}\big(\epsilon X^\epsilon, \underline{t}, \underline{\theta}\big)\Big), \quad Z^{\epsilon,h} = \exp\Big(\frac{1}{\epsilon}G\big(\epsilon X^\epsilon, h\big)\Big)$$

We now state the main theorem of this section, which resolves the *twisting/tilting* approach of measure changes $\underline{\mathrm{Q}}^{\epsilon,\underline{t},\underline{\theta}}$ in Proposition 2.3.2 with the *exponential martingale* approach of measure changes $\mathrm{Q}^{\epsilon,h}$ in Proposition 2.4.2. It relies on the following parameterization of maps $h(\cdot, \underline{t}, \underline{\theta})$ over $\underline{t} \vdash [0,\infty)$ and $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$.

$$h(t, \underline{t}, \underline{\theta}) = \sum_{k=1}^{|\underline{t}|} 1_{[t_{k-1}, t_k)}(t)\psi(\Delta t_k, \theta_k) \tag{2.4.5}$$

Theorem 2.4.6. *For each $\underline{t} \vdash [0,\infty)$ and $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$, we have $h(\cdot, \underline{t}, \underline{\theta}) \in \mathcal{H}_\Lambda$, and for any semimartingale H,*

$$\underline{G}(H, \underline{t}, \underline{\theta}) = G\big(H, h(\cdot, \underline{t}, \underline{\theta})\big).$$

Thus, for any $\epsilon > 0$ and $x \in \mathbb{X}$, we have the following identities.

$$\underline{Z}^{\epsilon,\underline{t},\underline{\theta}} = Z^{\epsilon, h(\cdot,\underline{t},\underline{\theta})}, \quad \underline{\mathrm{Q}}^{\epsilon,\underline{t},\underline{\theta}}_x = \mathrm{Q}^{\epsilon, h(\cdot,\underline{t},\underline{\theta})}_x$$

Proof. It is clear that $h(\cdot, \underline{t}, \underline{h})$ is compactly supported on the intervals of $\underline{t}$. Meanwhile, Proposition 1.2.5(c) tells us that $h(\cdot, \underline{t}, \underline{\theta})$ is differentiable everywhere but potentially the nodes of $\underline{t}$, and so it is of finite variation. This concludes $h \in \mathcal{H}_\Lambda$.

Seeing as Ψ is $\mathbb{X}$-affine and stochastic integration is linear, we can use a simplified version of Itô's formula below (again using Proposition 1.2.5(c))

$$\begin{aligned}
\underline{G}(H, \underline{t}, \underline{\theta}) &= \sum_{k=1}^{|\underline{t}|} \Big(\big\langle \theta_k, H_{t_k} - H_{t_{k-1}} \big\rangle - \varphi\big(\Delta t_k, \theta_k, H_{t_{k-1}}\big) \Big) \\
&= \sum_{k=1}^{|\underline{t}|} \Big(\Psi\big(t_k - t_k, \theta_k, H_{t_k}\big) - \Psi\big(t_k - t_{k-1}, \theta_k, H_{t_{k-1}}\big) \Big) \\
&= \sum_{k=1}^{|\underline{t}|} \int_{t_{k-1}}^{t_k} \mathrm{d}\Psi(t_k - \cdot, \theta_k, H_\cdot) \\
&= \sum_{k=1}^{|\underline{t}|} \Big(-\int_{t_{k-1}}^{t_k} \dot{\Psi}(t_k - t, \theta_k, H_t)\mathrm{d}t + \int_{t_{k-1}}^{t_k} \psi(t_k - t, \theta_k)\mathrm{d}H_t \Big) \\
&= \sum_{k=1}^{|\underline{t}|} \Big(\int_{t_{k-1}}^{t_k} \psi(t_k - t, \theta_k)\mathrm{d}H_t - \int_{t_{k-1}}^{t_k} \Lambda\big(\psi(t_k - t, \theta_k), H_t\big)\mathrm{d}t \Big)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{|\underline{t}|} \Big(\int_{t_{k-1}}^{t_k} h(t, \underline{t}, \underline{\theta}) \mathrm{d}H_t - \int_{t_{k-1}}^{t_k} \Lambda\big(h(t, \underline{t}, \underline{\theta}), H_t\big) \mathrm{d}t \Big) \\
&= h(\cdot, \underline{t}, \underline{\theta}) \bullet H_\infty - \Lambda\big(h(\cdot, \underline{t}, \underline{\theta}), H\big) \bullet \ell_\infty
\end{aligned}
$$

From here, as we did in Proposition 2.4.1, we apply integration by parts to complete the equality.

$$
\begin{aligned}
\underline{G}(H, \underline{t}, \underline{\theta}) &= h(\cdot, \underline{t}, \underline{\theta}) \bullet H_\infty - \Lambda\big(h(\cdot, \underline{t}, \underline{\theta}), H\big) \bullet \ell_\infty \\
&= \langle h(0, \underline{t}, \underline{\theta}), H_0 \rangle - \int_0^\infty H_{s-} \mathrm{d}h(s, \underline{t}, \underline{\theta}) - \int_0^\infty \Lambda\big(h(\cdot, \underline{t}, \underline{\theta}), H_s\big) \mathrm{d}s \\
&= G\big(H, h(\cdot, \underline{t}, \underline{\theta})\big)
\end{aligned}
$$

Evaluating this identity at ϵX^ϵ now gives us the remaining equalities.

Remark 2.4.7. *Note that a conditional cumulant φ and Lévy-Khintchine map Λ may be defined for any jump-diffusion, despite not generally being $\mathbb{X}$-affine.*

$$
\begin{aligned}
\varphi(t, \theta, x) &= \mathrm{E}_{\mathrm{P}_x} \exp \langle \theta, X - x \rangle, \\
\Lambda(u, x) &= \big\langle u, \beta(x) \big\rangle + \frac{1}{2} \big\langle u, \beta(x) u \big\rangle + \int_{\mathbb{V}} \Big(e^{\langle u, v \rangle} - 1 - \big\langle u, v \big\rangle \Big) \mu(x, \mathrm{d}v)
\end{aligned}
$$

Furthermore, we may always construct (local) measure changes like $\underline{Z}^{\underline{t}, \underline{\theta}}$ and Z^h in (2.4.4) and postulate an equivalence like Theorem 2.4.6.

$$
\exp \sum_{k=1}^{|\underline{t}|} \Big(\big\langle \theta_k, X_{t_k} - X_{t_{k-1}} \big\rangle - \varphi^\epsilon\big(\Delta t_k, \theta_k, X_{t_{k-1}}\big) \Big) = \exp \Big(\int_0^\infty h(s) \mathrm{d}X_s - \int_0^\infty \Lambda^\epsilon\big(h(s), X_s\big) \mathrm{d}s \Big)
$$

This is the very result which would generally *prove that the approaches of twisting/tilting and martingales are in fact identical.*

Corollary 2.4.8. *For each $\underline{t} \vdash [0, \infty)$ and $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$, the distributions $\underline{\mathrm{Q}}_x^{\epsilon, \underline{t}, \underline{\theta}}$ converge weakly to the degenerate measure $\delta_{\xi_{\underline{t}, \underline{\theta}}}$ at the solution $\xi_{\underline{t}, \underline{\theta}}$ to the dynamical system in which $\xi_{\underline{t}, \underline{\theta}}(0) = x$ and for each $k = 1, \ldots, |\underline{t}|$, we have the following equation.*

$$
\begin{aligned}
\dot{\xi}_{\underline{t}, \underline{\theta}}(t) = \beta\big(\xi_{\underline{t}, \underline{\theta}}(t)\big) + \alpha\big(\xi_{\underline{t}, \underline{\theta}}(t)\big) \psi(t_k - t, \theta_k) & \qquad t \in [t_{k-1}, t_k) \\
+ \int_{\mathbb{V}} v \big(e^{\langle \psi(t_k - t, \theta_k), v \rangle} - 1 \big) \mu\big(\xi_{\underline{t}, \underline{\theta}}(t), \mathrm{d}v\big), &
\end{aligned}
$$

Proof. This is simply substituting $h(\cdot, \underline{t}, \underline{\theta})$ for h in Proposition 2.4.2.

2.5 Integral representation of rate function

So far, for each $x \in \mathbb{X}^\circ$, Theorem 2.3.9 provides us a large deviation principle for $(\mathrm{P}_x^\epsilon)_{\epsilon > 0}$ with good rate function I_x as in (2.3.10). This section is concerned with simplifying the nature of I_x to take a more explicit integral form, comparable to existing principles for other families of stochastic processes.

$$
I_x(\xi) = \begin{cases} \displaystyle\int_0^\infty \Lambda^*\big(\dot{\xi}(t), \xi(t)\big) \mathrm{d}t & \xi(0) = x, \ \xi \in \mathbb{A}([0, \infty), \mathbb{X}) \\ \infty & \text{otherwise} \end{cases}
$$

To show this, we start by defining a map I which composes the initial state through our rate functions $(I_x)_{x\in\mathbb{X}}$, to remove the finiteness condition of each I_x in (2.3.10).

$$I(\xi) \coloneqq I_{\xi(0)}(\xi) = \sup_{\underline{t}\vdash\Delta_\xi^c} \Psi^*\big(\underline{t}, \xi(\underline{t}), \xi(0)\big)$$

In the following results, we will–without mention–assume evaluations of $I(\xi)$ for $\xi(0) \in \mathbb{X}^\circ$, so that we can use Theorem 1.3.12. Note that this is at no loss of generality, since we are resolving our rate function I_x for a large deviation principle that already requires $x \in \mathbb{X}^\circ$.

Lemma 2.5.1. *If $\xi \notin \mathbb{A}([0,\infty),\mathbb{X})$, then $I(\xi) = \infty$.*

Proof. Fix some $\xi \in \mathbb{D}([0,\infty),\mathbb{X})$ with $\xi \notin \mathbb{A}([0,\infty),\mathbb{X})$. That is to say, there exists some $\tau > 0$ for which $\xi \notin \mathbb{A}([0,\tau],\mathbb{X})$. For any $\rho > 0$, we now use Proposition 2.2.7(d) to produce some $\delta > 0$ and $C_{\delta,\rho} > 0$ such that the following bound holds.

$$\big|\Psi(t,u,x) - \Psi(0,u,x)\big| \leq C_{\delta,\rho}\cdot t\cdot\big(1+|x|\big), \quad t\in[0,\delta],\ u\in\overline{B}(0,\rho),\ x\in\mathbb{X}$$

Because $\xi \notin \mathbb{A}([0,\tau],\mathbb{X})$, there exists $\epsilon > 0$ and a partition $\underline{t}^\rho \vdash [0,\tau]$ such that

$$\sum_{k=1}^{|\underline{t}^\rho|} \Delta t_k^\rho < \delta \wedge \Big(C_{\delta,\rho}\big(1+\sup_{t\in[0,\tau]}|\xi(t)|\big)\Big)^{-1}$$

$$\sum_{k=1}^{|\underline{t}^\rho|} \big|\xi(t_k^\rho) - \xi(t_{k-1}^\rho)\big| \geq \epsilon$$

The countable nature of Δ_ξ allows us to further impose that $\underline{t}^\rho \vdash \Delta_\xi^c$. This, along with Theorem 1.3.12, results in the following inequality.

$$\begin{aligned}
&\sup_{\underline{t}\vdash[0,\infty)} \Psi^*\big(\underline{t},\xi(\underline{t}),\xi(0)\big)\\
&\quad\geq \sup_{\rho>0}\Psi^*\big(\underline{t}^\rho,\xi(\underline{t}^\rho),\xi(0)\big)\\
&\quad\geq \sup_{\rho>0}\sup_{\underline{\theta}\in\mathcal{D}_\Psi(\underline{t})}\sum_{k=1}^{|\underline{t}^\rho|}\Big(\langle\theta_k,\xi(t_k^\rho)-\xi(t_{k-1}^\rho)\rangle - \varphi\big(\Delta t_k^\rho,\theta_k,\xi(t_{k-1}^\rho)\big)\Big)\\
&\quad= \sup_{\rho>0}\sum_{k=1}^{|\underline{t}^\rho|}\sup_{\theta_k\in\mathcal{D}_\Psi(\Delta t_k^\rho)}\Big(\big\langle\theta_k,\xi(t_k^\rho)-\xi(t_{k-1}^\rho)\big\rangle\\
&\qquad\qquad - \Big(\Psi\big(\Delta t_k^\rho,\theta_k,\xi(t_{k-1}^\rho)\big) - \Psi^*\big(0,\theta_k,\xi(t_{k-1}^\rho)\big)\Big)\Big)\\
&\quad\geq \sup_{\rho>0}\sum_{k=1}^{|\underline{t}^\rho|}\Big(\rho\big|\xi(t_k^\rho)-\xi(t_{k-1}^\rho)\big| - C_{\delta,\rho}\cdot\Delta t_k^\rho\cdot\big(1+|\xi(t_{k-1}^\rho)|\big)\Big)\\
&\quad\geq \epsilon\cdot\sup_{\rho>0}\rho - 1\\
&\quad= \infty
\end{aligned}$$

Lemma 2.5.2. *For each $\xi \in \mathbb{A}([0,\infty),\mathbb{X})$, $\underline{t} \vdash [0,\infty)$ and $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$, we have an identity similar to that of Theorem 2.4.6.*

$$\underline{G}(\xi,\underline{t},\underline{\theta}) = G\big(\xi,h(\cdot,\underline{t},\underline{\theta})\big) = \int_0^\infty\Big(\big\langle h(t,\underline{t},\underline{\theta}),\dot{\xi}(t)\big\rangle - \Lambda\big(h(t,\underline{t},\underline{\theta}),\xi(t)\big)\Big)\mathrm{d}t.$$

Proof. Similar to as in Theorem 2.4.6, we use Propositions 1.2.5(c) and 2.2.7(a) and apply the fundamental theorem of calculus and integration by parts.

$$\begin{aligned}
\underline{G}(\xi,\underline{t},\underline{\theta}) &= \sum_{k=1}^{|\underline{t}|}\Big(\big\langle\theta_k,\xi(t_k)-\xi(t_{k-1})\big\rangle-\varphi\big(\Delta t_k,\theta_k,\xi(t_{k-1})\big)\Big)\\
&= \sum_{k=1}^{|\underline{t}|}\Big(\Psi\big(t_k-t_k,\theta_k,\xi(t_k)\big)-\Psi\big(t_k-t_{k-1},\theta_k,\xi(t_{k-1})\big)\Big)\\
&= \sum_{k=1}^{|\underline{t}|}\int_{t_{k-1}}^{t_k}\frac{\partial}{\partial t}\Psi\big(t_k-t,\theta_k,\xi(t)\big)\mathrm{d}t\\
&= \sum_{k=1}^{|\underline{t}|}\int_{t_{k-1}}^{t_k}\Big(-\dot{\Psi}\big(t_k-t,\theta_k,\xi(t)\big)+\big\langle\psi(t_k-t,\theta_k),\dot{\xi}(t)\big\rangle\Big)\mathrm{d}t\\
&= \sum_{k=1}^{|\underline{t}|}\int_{t_{k-1}}^{t_k}\Big(\big\langle\psi(t_k-t,\theta_k),\dot{\xi}(t)\big\rangle-\Lambda\big(\psi(t_k-t,\theta_k),\xi(t)\big)\Big)\mathrm{d}t\\
&= \int_0^\infty\Big(\big\langle h(t,\underline{t},\underline{\theta}),\dot{\xi}(t)\big\rangle-\Lambda\big(h(t,\underline{t},\underline{\theta}),\xi(t)\big)\Big)\mathrm{d}t\\
&= \big\langle h(0,\underline{t},\underline{\theta}),\xi(0)\big\rangle-\int_0^\infty\xi(t-)\mathrm{d}h(t,\underline{t},\underline{\theta})-\int_0^\infty\Lambda\big(h(t,\underline{t},\underline{\theta}),\xi(t)\big)\mathrm{d}t\\
&= G\big(\xi,h(\cdot,\underline{t},\underline{\theta})\big)
\end{aligned}$$

Proposition 2.5.3. *For each $\xi\in\mathbb{A}([0,\infty),\mathbb{X})$, we have the following upper bound.*

$$I(\xi)\le\int_0^\infty\Lambda^*\big(\dot{\xi}(t),\xi(t)\big)\mathrm{d}t$$

Proof. Fix $\underline{t}\vdash[0,\infty)$ and $\underline{u}\in\mathbb{V}^{|\underline{t}|}$. Observe that if $\Psi\big(\underline{t},\underline{u},\xi(0)\big)=\infty$, we immediately have the following inequality.

$$\big\langle\underline{u},\xi(\underline{t})\big\rangle-\Psi\big(\underline{t},\underline{u},\xi(0)\big)=-\infty\le0=\int_0^\infty\Lambda^*\big(\dot{\xi}(t),\xi(t)\big)\mathrm{d}t$$

Otherwise, Theorem 1.3.12 tells us that $\underline{u}=U_{\underline{t}}(\underline{\theta})$ for some $\underline{\theta}\in\mathcal{D}_\Psi(\underline{t})$. By Lemma 2.5.2, we now see the same inequality.

$$\begin{aligned}
\big\langle\underline{u},\xi(\underline{t})\big\rangle-\Psi\big(\underline{t},\underline{u},\xi(0)\big) &= \sum_{k=1}^{|\underline{t}|}\Big(\big\langle\theta_k,\xi(t_k)-\xi(t_{k-1})\big\rangle-\varphi(\Delta t_k,\theta_k,\xi(t_{k-1}))\Big)\\
&= \underline{G}(\xi,\underline{t},\underline{\theta})\\
&= \int_0^\infty\Big(\big\langle h(t,\underline{t},\underline{\theta}),\dot{\xi}(t)\big\rangle-\Lambda\big(h(t,\underline{t},\underline{\theta}),\xi(t)\big)\Big)\mathrm{d}t\\
&\le \int_0^\infty\Lambda^*\big(\dot{\xi}(t),\xi(t)\big)\mathrm{d}t
\end{aligned}$$

Thus, we have the following upper bound.

$$I(\xi)=\sup_{\underline{t}\vdash\Delta_\xi^c}\Psi^*\big(\underline{t},\xi(\underline{t}),\xi(0)\big)=\sup_{\underline{t}\vdash[0,\infty)}\sup_{\underline{u}\in\mathbb{V}^{|\underline{t}|}}\Big(\big\langle\underline{u},\xi(\underline{t})\big\rangle-\Psi\big(\underline{t},\underline{u},\xi(0)\big)\Big)$$

$$\leq \int_0^\infty \Lambda^*\big(\dot\xi(t), \xi(t)\big)\mathrm{d}t$$

Lemma 2.5.4. *Fix $\tau > 0$ and $h \in \mathbb{C}([0,\tau], \mathbb{X})$. For each $\epsilon > 0$, there exists a partition $\underline{t} \vdash [0,\tau]$ and $\underline{\theta} \in \mathcal{D}_\Psi(\underline{t})$ such that we have the following approximation.*

$$\sup_{t\in[0,\tau)} \big|h(t, \underline{t}, \underline{u}) - h(t)\big| < \epsilon$$

Proof. Denote $M \coloneqq \sup_{t\in[0,\tau]} |h(t)|$ and fix $\epsilon > 0$. We start by using Proposition 2.2.7(d) to guarantee $\delta_0 > 0$ and $C_M > 0$ such that the following bound holds.

$$\big|\Psi(t,u,x) - \Psi(0,u,x)\big| \leq C_M \cdot t \cdot \big(1 + |x|\big), \quad t \in [0,\delta_0],\ u \in \overline{B}(0,M),\ x \in \mathbb{X}$$

From here, we use the affine structure of Ψ to see the following inequality.

$$\big|\psi(t,u) - u\big| = 3C_M\sqrt{d}\cdot t, \quad (t,u) \in [0,\delta_0] \times \overline{B}(0,M)$$

Seeing as $h \in \mathbb{C}([0,\tau], \mathbb{V})$, it is uniformly continuous. Fix $\delta_1 > 0$ such that all $s,t \in [0,\tau]$ with $|t-s| < \delta_1$, we have the following inequality.

$$\big|h(t) - h(s)\big| < \epsilon/2$$

Fix an integer $N \in \mathbb{N}$ large enough to impose the following inequality.

$$\frac{\tau}{N} < \frac{\epsilon}{6C_M\sqrt{d}} \wedge \delta_0 \wedge \delta_1$$

Now define partition $\underline{t}$ by $t_k = k\tau/N$ for $k = 1,\ldots,N$. This way, for each $k = 1,\ldots,N$, and $t \in [t_{k-1}, t_k)$, we have

$$\begin{aligned}
\Big|h\big(t,\underline{t},h(\underline{t})\big) - h(t)\Big| &= \Big|\psi\big(t_k - t, h(t_k)\big) - h(t)\Big| \\
&\leq \Big|\psi\big(t_k - t, h(t_k)\big) - h(t_k)\Big| + \big|h(t_k) - h(t)\big| \\
&< 3C_M\sqrt{d}\cdot(t_k - t) + \epsilon/2 \\
&< \epsilon.
\end{aligned}$$

Proposition 2.5.5. *For each $\xi \in \mathbb{A}([0,\infty), \mathbb{X})$, we have the following inequality.*

$$I(\xi) \geq \int_0^\infty \Lambda^*\big(\dot\xi(t), \xi(t)\big)\mathrm{d}t$$

Proof. We proceed in a way similar to that of [Puh01]. Define a map $f : [0,\infty) \times \mathbb{V} \to \mathbb{R}$ as below.

$$f(t,\theta) \coloneqq \big\langle \theta, \dot\xi(t)\big\rangle - \Lambda\big(\theta, \xi(t)\big)$$

For a fixed $\epsilon > 0$, $t \in [0,\infty)$, we define the following set.

$$\Gamma_t^\epsilon \coloneqq \left\{\theta \in \mathbb{V} : \Big(\sup_{\theta'\in\mathbb{V}} f(t,\theta') - \epsilon\Big)_+ \wedge \frac{1}{\epsilon} \leq f(t,\theta) \leq \frac{1}{\epsilon}\right\}$$

Continuity of $f(t,\cdot)$ and the least upper bound property guarantees Γ_t^ϵ is nonempty and measurable. Thus, we may construct a measurable selection $h:[0,\infty)\to\mathbb{V}$ for Γ^ϵ, which is to say h is Lebesgue measurable and the following holds.

$$f\big(t,h(t)\big)\in\Gamma_t^\epsilon,\quad t\in[0,\infty) \tag{2.5.6}$$

We now use Luzin's theorem to approximate $h|_{[0,1/\epsilon]}$ with $\tilde h\in\mathbb{C}([0,1/\epsilon],\mathbb{V})$ to the following extent.

$$\int_{\tilde h\neq h}\mathrm{d}t<\epsilon^2 \tag{2.5.7}$$

Now, we combine our inequalities from (2.5.6) and (2.5.7) to see that

$$\begin{aligned}
&\int_0^{1/\epsilon}\Big(f\big(t,\tilde h(t)\big)\vee 0\Big)\mathrm{d}t\\
&\quad=\int_0^{1/\epsilon}f\big(t,h(t)\big)\mathrm{d}t+\int_{\tilde h\neq h}\bigg(\Big(f\big(t,\tilde h(t)\big)\vee 0\Big)-f\big(t,h(t)\big)\bigg)\mathrm{d}t\\
&\quad\geq\int_0^{1/\epsilon}\bigg(\Big(\sup_{\theta\in\mathbb{V}}f(t,\theta)-\epsilon\Big)\wedge\frac{1}{\epsilon}\bigg)\mathrm{d}t-\int_{\tilde h\neq h}\frac{1}{\epsilon}\mathrm{d}t\\
&\quad=\int_0^{1/\epsilon}\bigg(\Big(\Lambda^*\big(\dot\xi(t),\xi(t)\big)-\epsilon\Big)\wedge\frac{1}{\epsilon}\bigg)\mathrm{d}t-\epsilon
\end{aligned}\tag{2.5.8}$$

By Lemma 2.5.4, the fact that each $f(t,0)=0$, and continuity of each $f(t,\cdot)$, we may now use Fatou's lemma to guarantee some $\underline{t}\vdash[0,\infty)$ and $\underline{\theta}\in\mathcal{D}_\Psi(\underline{t})$ such that

$$\int_0^{1/\epsilon}f\big(t,h(t,\underline{t},\underline{\theta})\big)\mathrm{d}t\geq\int_0^{1/\epsilon}\Big(f\big(t,\tilde h(t)\big)\vee 0\Big)\mathrm{d}t-\epsilon \tag{2.5.9}$$

Combining (2.5.8) and (2.5.9), we now see that

$$\int_0^{1/\epsilon}f\big(t,h(t,\underline{t},\underline{\theta})\big)\mathrm{d}t\geq\int_0^{1/\epsilon}\bigg(\Big(\Lambda^*\big(\dot\xi(t),\xi(t)\big)-\epsilon\Big)\wedge\frac{1}{\epsilon}\bigg)\mathrm{d}t-2\epsilon.$$

By Theorem 1.3.12 and Lemmas 2.5.1 and 2.5.2, the above inequality gives us the following.

$$\begin{aligned}
I_x(\xi)&\geq\Psi^*\big(\underline{t},\xi(\underline{t}),\xi(0)\big)\\
&\geq\sum_{k=1}^{|\underline{t}|}\Big(\big\langle\theta_k,\xi(t_k)\big\rangle-\Psi\big(\Delta t_k,\theta_k,\xi(t_{k-1})\big)\Big)\\
&=\sum_{k=1}^{|\underline{t}|}\int_{t_{k-1}}^{t_k}\Big(\big\langle\psi(t_k-t,\theta_k),\dot\xi(t)\big\rangle-\Lambda\big(\psi(t_k-t,\theta_k),\xi(t)\big)\Big)\mathrm{d}t\\
&=\int_0^{1/\epsilon}f\big(t,h(t,\underline{t},\underline{\theta})\big)\mathrm{d}t\\
&\geq\int_0^{1/\epsilon}\bigg(\Big(\Lambda^*\big(\dot\xi(t),\xi(t)\big)-\epsilon\Big)\wedge\frac{1}{\epsilon}\bigg)\mathrm{d}t-2\epsilon
\end{aligned}$$

Taking $\epsilon\to 0$ now yields our desired result.

$$I_x(\xi)\geq\int_0^\infty\Lambda^*\big(\dot\xi(t),\xi(t)\big)\mathrm{d}t$$

Theorem 2.5.10. *For each $x \in \mathbb{X}^\circ$, the family $(\mathrm{P}_x^\epsilon)_{\epsilon>0}$ satisfies a large deviation principle on $\mathbb{D}([0,\infty),\mathbb{X})$ with good rate function $I_x : \mathbb{D}([0,\infty),\mathbb{X}) \to [0,\infty]$ as follows.*

$$I_x(\xi) = \begin{cases} \displaystyle\int_0^\infty \Lambda^*\big(\dot{\xi}(t),\xi(t)\big)\mathrm{d}t & \xi \in \mathbb{A}([0,\infty),\mathbb{X}),\ \xi(0) = x \\ \infty & \textit{otherwise} \end{cases} \tag{2.5.11}$$

Proof. Theorem 2.3.9 gives us our large deviation principle with rate function I_x as in (2.3.10). Fix $\xi \in \mathbb{D}([0,\infty),\mathbb{X})$. If $\xi(0) \neq x$, then we already have $I_x(\xi) = \infty$. Otherwise, $I_x(\xi) = I_{\xi(0)}(\xi) = I(\xi)$, and so Lemma 2.5.1 tells us $I_x(\xi) = \infty$ if $\xi \notin \mathbb{A}([0,\infty),\mathbb{X})$. If $\xi \in \mathbb{A}([0,\infty),\mathbb{X})$, then Propositions 2.5.3 and 2.5.5 tell us that

$$I_x(\xi) = \int_0^\infty \Lambda^*\big(\dot{\xi}(t),\xi(t)\big)\mathrm{d}t.$$

This concludes that I can be written as in (2.5.11).

Chapter 3

Large deviation rate functions

We dedicate this final chapter to interpreting the rate function (2.5.11) from Theorem 2.5.10 and deriving a semi-closed form for Λ^* which illustrates more on the nature of our distributions $(\mathrm{P}_x^\epsilon)_{\epsilon>0}$. Before doing this, let us establish some intuition with what our integral form *already* tells us.

Firstly, we recognize that a large deviation principle introduces a first-order exponential asymptotic; the principle shows us that the following limit holds,

$$\lim_{\delta\to 0}\lim_{\epsilon\to 0}\epsilon\log \mathrm{P}_x^\epsilon\Big(\epsilon X^\epsilon \in B(\xi,\delta)\Big) = -I_x(\xi),$$

and so we may introduce a correction term $o(\epsilon,\delta)$,

$$\mathrm{P}_x^\epsilon\Big(\epsilon X^\epsilon \in B(\xi,\delta)\Big) = \exp\Big(-\frac{1}{\epsilon}I_x(\xi) + o(\epsilon,\delta)\Big),$$

which satisfies growth smaller than $1/\epsilon$, in the explicit sense of the following limit.

$$\lim_{\delta\to 0}\lim_{\epsilon\to 0}\epsilon o(\epsilon,\delta) = 0$$

By ignoring this expression $o(\epsilon,\delta)$, we are concerning ourselves with a first-order exponential approximation of the following probabilities.

$$\mathrm{P}_x^\epsilon\Big(\epsilon X^\epsilon \in B(\xi,\delta)\Big) \approx \exp\Big(-\frac{1}{\epsilon}I_x(\xi)\Big) \tag{3.0.1}$$

To this end, I_x interprets properties of ξ that the family $(\epsilon X^\epsilon)_{\epsilon>0}$ is systematically trending towards/against. Should $I_x(\xi)$ be large, the decay of (3.0.1) is faster, demonstrating that ξ exhibits some property that the family $(\epsilon X^\epsilon)_{\epsilon>0}$ does not tend to satisfy. Thus is the notion of I_x governing a *principle* for what behaviors make a path ξ *largely deviate* from the typical behavior of $(\epsilon X^\epsilon)_{\epsilon>0}$.

The integral nature of I_x indicates to us that there are local properties that the family $(\epsilon X^\epsilon)_{\epsilon>0}$ systematically exhibits which lends to faster/slower decay from some ξ. Take, for instance, if we perturbed some ξ only on some interval $[t,\tau]$, coercing the values of $\Lambda^*(\dot\xi(s),\xi(s))$ to be larger on this interval $s\in[t,\tau]$. This will increase the value of $I_x(\xi)$, indicating that it was this behavior on the interval $[t,\tau]$ that produced a deviation from the behavior of our family $(\epsilon X^\epsilon)_{\epsilon>0}$. Knowing this, it is imperative that we derive some form for Λ^*. Doing so will allow us to see local properties that the family $(\epsilon X^\epsilon)_{\epsilon>0}$ exhibits.

Unfortunately, the $\mathbb{X}$-affine nature of Λ provides us no simplification in evaluating Λ^*.

$$\begin{aligned}\Lambda^*(\dot{x}, x) &= \langle u, \dot{x}\rangle - \Lambda(u, x)\\ &= \langle u, \dot{x}\rangle - \big\langle u, \beta^{\chi}(x)\big\rangle - \frac{1}{2}\big\langle u, \alpha(x)\big\rangle - \int_{\mathbb{V}} \big(e^{\langle u,v\rangle} - 1 - \langle u, \chi(v)\rangle\big)\mu(x, \mathrm{d}v),\end{aligned}$$

However, by looking to existing rate functions from the literature [DRL04, DZ10, KK14, GZ18], we will be able to build an intuition and a toolbox for developing a semi-closed form for $\Lambda*$. We find this result remarkable for multiple reasons.

1. The representation of Λ^* will feel familiar to already-known results
2. The representation of Λ^* will generalize the results we list
3. The representation (though not the large deviation principle) of Λ^* and the corresponding rate function I_x exist independently of the affine assumption.

This chapter is very example heavy, and so we find it to be the most intuitive of the chapters. By the end, we hope rate functions, as well as large deviations, will be better understood from a general perspective. All this said, let us now proceed to our final chapter, which is organized as follows.

Section 3.1. Introduces an important large deviation principle for stochastic processes with independent increments.

Section 3.2. Introduces tools which allow us to lift the preceding principles to stochastic processes with state-dependence.

Section 3.3. Introduces a trick which allows one to evaluate rate functions in the case of there being *multiple sources* of randomness.

Section 3.4. Establishes our main result of developing a semi-closed form for Λ^*.

Section 3.5. Discusses the intuition and potential extensions of our result.

3.1 Mogulskii's theorem

A surprisingly powerful theorem in the theory of large deviations of stochastic processes is that of Mogulskii (see [DZ10, Theorems 5.1.2 and 5.1.19 and Exercise 5.122]). Fixing a family $(V_j)_{j\in\mathbb{N}}$ of independent quantities distributing with common distribution κ with light tails,

$$\Lambda_\kappa(u) := \log \int_{\mathbb{V}} e^{\langle u,v\rangle}\kappa(\mathrm{d}v) < \infty, \quad u \in \mathbb{V} \tag{3.1.1}$$

this theorem provides a large deviation principle for the laws associated to quantities Y^ϵ as below.

$$Y_t^\epsilon = \epsilon \sum_{j=1}^{[t/\epsilon]} V_j, \quad t \in [0, \tau]$$

It states that the associated laws $(\mathrm{P}^\epsilon)_{\epsilon>0}$ satisfy a large deviation principle on the space $\mathbb{L}^\infty[0,\tau]$ of bounded functions $[0,\tau]\to\mathbb{V}$, equipped with the supremum norm. The rate function, like ours, is an integral of the Fenchel-Legendre transform of Λ_κ.

$$\xi \mapsto \begin{cases} \displaystyle\int_0^\tau \Lambda_\kappa^*\big(\dot\xi(t)\big)\mathrm{d}t & \xi(0)=0,\ \xi\in\mathbb{A}([0,\tau],\mathbb{V}) \\ \infty & \text{otherwise} \end{cases}$$

Very minor adjustments can actually make this theorem similar to the context of our principle. Firstly, the principle may be lifted to the space $\mathbb{L}^\infty_{\mathrm{loc}}[0,\infty)$ of locally bounded functions $[0,\infty)\to\mathbb{V}$, equipped with the weighted supremum norm,

$$(\xi,\xi')\mapsto \sup_{t\in[0,\infty)} e^{-t}|\xi(t)-\xi'(t)|,$$

for this metric is consistent with $\xi_n\to\xi$ if and only if $\xi_n|_{[0,\tau]}\to\xi|_{[0,\tau]}$ uniformly for all $\tau\geq 0$, which is the same as the projective limit space induced by the restriction maps.

$$(\xi_\tau)_{\tau>0}\in\lim_{\leftarrow\tau}\mathbb{L}^\infty[0,\tau] \quad \xleftrightarrow{\xi_\tau=\xi|_{[0,\tau]}} \quad \xi\in\mathbb{L}^\infty_{\mathrm{loc}}[0,\infty)$$

Applying Dawson-Gärtner [DZ10, Theorem 4.6.1], the rate function over this space is as follows.

$$\xi \mapsto \begin{cases} \displaystyle\sup_{\tau>0}\int_0^\tau \Lambda_\kappa^*\big(\dot\xi(t)\big)\mathrm{d}t & \xi(0)=0,\ \xi\in\mathbb{A}([0,\tau),\mathbb{V}) \text{ for all } \tau>0 \\ \infty & \text{otherwise} \end{cases}$$

From here, we recognize that each process Y^ϵ is càdlàg; if ν is supported on $\mathbb{X}$, the process takes values in $\mathbb{D}([0,\infty),\mathbb{X})$, and so we may restrict our principle (see [DZ10, Lemma 4.1.5(b)]). Our rate function then takes the same form (recall the local definition of absolute continuity $\mathbb{A}([0,\infty),\mathbb{X})$).

$$\xi \mapsto \begin{cases} \displaystyle\int_0^\infty \Lambda_\kappa^*\big(\dot\xi(t)\big)\mathrm{d}t & \xi(0)=0,\ \xi\in\mathbb{A}([0,\infty),\mathbb{X}) \\ \infty & \text{otherwise} \end{cases} \tag{3.1.2}$$

Example 3.1.3 (Brownian motion). *Applying Mogulskii's theorem when our increment distribution κ is* $\mathrm{Normal}(0,\mathrm{id}_\mathbb{V})$, *the integral in our rate function in (3.1.2) becomes the following.*

$$\int_0^\infty \frac{1}{2}\big|\dot\xi(t)\big|^2\mathrm{d}t \tag{3.1.4}$$

Furthermore, for a Brownian motion W, the process $\sqrt{\epsilon}W$ ends up being exponentially equivalent to Y^ϵ,

$$\limsup_{\epsilon\to0}\epsilon\log\mathrm{P}\big(|\sqrt{\epsilon}W-Y^\epsilon|\geq\delta\big)=-\infty,$$

which makes the family $\sqrt{\epsilon}W$ satisfy the large deviation principle with rate function (3.1.4); this result is known as Schilder's theorem (see [DZ10, Theorem 5.2.3]).

Note that $(\sqrt{\epsilon}W)_{\epsilon>0}$ is a family of affine processes covered Theorem 2.5.10. We have $\epsilon X^\epsilon=\sqrt{\epsilon}W$, where the base process X has special differential characteristics $(0,\mathrm{id}_\mathbb{V},0)$.

The easiest way to see this is by considering Proposition 2.1.9 with initial state $x = 0$. Our theorem also immediately resolves (2.5.11) the same rate function.

$$\Lambda^*(\dot{x}, x) = \sup_{u \in \mathbb{V}} \left(\langle u, \dot{x} \rangle - \frac{1}{2} \langle u, \mathrm{id}_{\mathbb{V}} \cdot u \rangle \right) = \frac{1}{2} |\dot{x}|$$

Example 3.1.5 (Poisson). *One may apply a very similar argument for when our increment distribution κ is* Poisson(1)*. In this case, the integral in the rate function in (3.1.2) evaluates to*

$$\int_0^\infty \left(\dot{\xi}(t) \log \left(\dot{\xi}(t) \right) - \dot{\xi}(t) + 1 \right) \mathrm{d}t, \tag{3.1.6}$$

so long as $\xi(t) \geq 0$ for Lebesgue-almost-every $t \geq 0$ (otherwise, it is infinite). In the case that $\xi(t) = 0$, we are taking the continuous extension of the integrand, i.e. $0 \log(0) := 0$. Similar to the work of Schilder's theorem, we may show, for a standard Poisson process N, $\epsilon N_{\cdot/\epsilon}$ is exponentially equivalent to this Y^ϵ, which makes the family satisfy a large deviation principle with rate function (3.1.6). In fact and exercise of our reference text, [DZ10, Exercise 5.2.12], suggests the reader to show just this.

Again, such a family $(\epsilon N_{\cdot/\epsilon})_{\epsilon > 0}$ is covered by Theorem 2.5.10. To see this, consider a base affine process X on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with special differential characteristics as below, where δ_1 denotes the degenerate distribution at $1 \in \mathbb{R}$.

$$\beta(x) = 1, \quad \alpha(x) = 0, \quad \mu(x, \mathrm{d}v) = \delta_1(\mathrm{d}v)$$

Setting the initial state $x = 0$ and looking at Proposition 2.1.9, we may say that ϵX^ϵ can be realized as follows.

$$\begin{aligned}
\epsilon X_t^\epsilon &= t + \epsilon 1_{[0,1]}(\mathrm{id}_{\mathbb{R}}) * \tilde{p}_t^\epsilon \\
&= t + \epsilon 1_{[0,1]}(\mathrm{id}_{\mathbb{R}}) * p_t^\epsilon - \epsilon 1_{[0,1]}(\mathrm{id}_{\mathbb{R}}) * \hat{p}_t^\epsilon \\
&= t + \epsilon p([0, t/\epsilon] \times [0, 1]) - \int_0^{t/\epsilon} \int_{\mathbb{R}} \epsilon 1_{[0,1]}(v) \mathrm{d}v \mathrm{d}s \\
&= \epsilon p([0, t/\epsilon] \times [0, 1])
\end{aligned}$$

As stated in [JS03, Theorem II.4.8], this Poisson random measure p is a Poisson point process with Lebesgue intensity. This means that, for each $t \geq 0$, $N_t := p([0, t] \times [0, 1])$ distributes Poisson(t)*, and $N_t - N_s = p((s, t] \times [0, 1])$ is independent of $N_s = p([0, s], [0, 1])$ for each $0 \leq s < t$. In other words, N is a standard Poisson process and*

$$\epsilon X_t^\epsilon = \epsilon p([0, t/\epsilon] \times [0, 1]) = \epsilon N_{t/\epsilon}.$$

As with the normal increments, our rate function (2.5.11) resolves this immediately.

$$\begin{aligned}
\Lambda^*(\dot{x}, x) = \sup_{u \in \mathbb{V}} \left(u\dot{x} - u - \int_{\mathbb{R}} \left(e^{uv} - 1 - uv \right) \delta_1(\mathrm{d}v) \right) &= \sup_{u \in \mathbb{V}} \left(u\dot{x} - e^u + 1 \right) \\
&= \begin{cases} \dot{x} \log \dot{x} - \dot{x} + 1, & \dot{x} \geq 0 \\ \infty, & \textit{otherwise} \end{cases}
\end{aligned}$$

3.2 Transformations

While Mogulskii's theorem specifies that the processes $(Y^\epsilon)_{\epsilon>0}$—by design—have independent increments, we may use transformation arguments to produce large deviation principles for families of processes with state-dependent increments. The two ways of leveraging this are via the contraction principle or measure-change arguments.

The contraction principle states that mapping the quantities $F(Y^\epsilon)$ via a continuous map F produces a large deviation principle for the family $(F_\# \mathrm{P}^\epsilon)_{\epsilon>0}$ of measures $F_\# \mathrm{P}^\epsilon$ associated with these respective quantities $F(Y^\epsilon)$ (see [DZ10, Theorem 4.2.1]). Seeing as this section serves as a survey for intuition on rate functions, we will digress from discussing the specifics of continuity of F on restricted spaces and/or exponentially equivalent families in our example below.

Example 3.2.1 (Diffusions). *We can leverage Example 3.1.3 to a family of processes* $(\epsilon X^\epsilon)_{\epsilon>0}$*,*

$$\epsilon X^\epsilon = x + \beta(\epsilon X^\epsilon) \bullet \ell + \sqrt{\epsilon} \bullet W,$$

where the drift $\beta : \mathbb{V} \to \mathbb{V}$ *is bounded and Lipschitz. Having a map* F_β *which implicitly solves the equation,*

$$F_\beta(\omega) = \xi, \quad \xi(t) = x + \beta(\xi) \bullet \ell_t + \omega_t, \tag{3.2.2}$$

will make $F_\beta(\sqrt{\epsilon} W) = \epsilon X^\epsilon$ *for each* $\epsilon > 0$*, so the contraction principle states that the distributions of* $(\epsilon X^\epsilon)_{\epsilon>0}$ *satisfy a large deviation principle in which the rate function* I_β *is derived from that* I_W *from Example 3.1.3.*

$$I_\beta(\xi) \coloneqq \inf \Big\{ I_W(\omega) : F_\beta(\omega) = \xi \Big\},$$

$$I_W(\omega) \coloneqq \begin{cases} \displaystyle\int_0^\infty \frac{1}{2} \big|\dot{\omega}(t)\big|^2 \mathrm{d}t, & \omega(0) = 0,\ \omega \in \mathbb{A}([0,\infty), \mathbb{V}), \\ \infty, & \textit{otherwise} \end{cases}$$

When $F_\beta(\omega) = \xi$*, equation (3.2.2) tells us* $\xi(0) = x$ *and* $\dot{\omega} = \dot{\xi} - \beta(\xi)$*, and so we have the following.*

$$I_\beta(\xi) = \begin{cases} \displaystyle\int_0^\infty \frac{1}{2} \Big|\dot{\xi} - \beta\big(\xi(t)\big)\Big| \mathrm{d}t, & \xi(0) = x,\ \xi \in \mathbb{A}([0,\infty), \mathbb{V}), \\ \infty, & \textit{otherwise} \end{cases}$$

Similarly, we may introduce a bounded, Lipschitz diffusion $\alpha = \sigma\sigma^* : \mathbb{V} \to \mathbb{L}(\mathbb{V})$ *in which each* $\alpha(x)$ *is invertible, so that the dynamics become as follows.*

$$\epsilon X^\epsilon = x + \beta(\epsilon X^\epsilon) \bullet \ell + \sqrt{\epsilon} \sigma(\epsilon X^\epsilon) \bullet W, \tag{3.2.3}$$

Having a map $F_{\beta,\alpha}$ *which implicitly solves the equation,*

$$F_{\beta,\alpha}(\omega) = \xi, \quad \xi(t) = \beta(\xi) \bullet \ell_t + \sigma(\xi) \bullet \omega_t,$$

will allow us to repeat the above argument to get a large deviation principle for $(\epsilon X^\epsilon)_{\epsilon>0}$ *with rate function* $I_{\beta,\alpha}$*; when* $\xi(0) = x$*,* $\xi \in \mathbb{A}([0,\infty), \mathbb{V})$*, we get the following.*

$$I_{\beta,\alpha}(\xi) = \int_0^\infty \frac{1}{2} \Big\langle \Big(\dot{\xi}(t) - \beta\big(\xi(t)\big)\Big), \alpha\big(\xi(t)\big)^{-1} \Big(\dot{\xi}(t) - \beta\big(\xi(t)\big)\Big) \Big\rangle \mathrm{d}t$$

The true details of this result, attributed to Freidlin-Wentzel [DZ10, Theorems 5.6.3 and 5.6.7], are rather complicated, and the above argument is just a heuristic. Also, note that this result does not apply to the general class of affine diffusions, as β, α are generally not bounded or Lipschitz, and α need not be invertible. However, [KK14]—a paper which inspires parts of our proof—first proved that affine (jump-)diffusions with special differential characteristics $(\beta, \alpha, 0)$ satisfy a large deviation principle with rate function similar to that above. Our rate function (2.5.11) from Theorem 2.5.10 immediately resolves an identical representation.

$$\begin{aligned}\Lambda^*(\dot{x}, x) &= \sup_{u\in\mathbb{V}}\left(\langle u, \dot{x}\rangle - \big\langle u, \beta(x)\big\rangle - \frac{1}{2}\big\langle u, \alpha(x)u\big\rangle\right)\\ &= \begin{cases} \frac{1}{2}\Big\langle\big(\dot{x}-\beta(x)\big), \alpha(x)^{\dagger}\big(\dot{x}-\beta(x)\big)\Big\rangle, & \dot{x}-\beta(x)\in \operatorname{image}\alpha(x),\\ \infty, & \textit{otherwise}\end{cases}\end{aligned}$$

Above, $a^{\dagger}\in\mathbb{L}(\mathbb{V})$ denotes the pseudoinverse of $a\in\mathbb{L}(\mathbb{V})$.

The above result leveraged a large deviation principle for Brownian motions $(\sqrt{\epsilon}W)_{\epsilon>0}$ to one on state-dependent diffusions $(\epsilon X^{\epsilon})_{\epsilon>0}$ via mappings $\epsilon X^{\epsilon} = F(\sqrt{\epsilon}W)$. The analogue of the Brownian motion W for jump processes—in the sense of homogeneous independent-increments—is the Poisson process N. To introduce state-dependence to our sequence $(\epsilon N_{\cdot/\epsilon})_{\epsilon>0}$ from Example 3.1.5, we may perform a measure-change argument.

Example 3.2.4 (Continuous-branching/Hawkes). *Consider our sequence $(\epsilon N_{\cdot/\epsilon})_{\epsilon>0}$ derived from a Poisson process N, as in Example 3.1.5. Denoting P^{ϵ} and $\mathscr{F}^{\epsilon}$ the distribution and filtration, respectively, of each $J^{\epsilon} := \epsilon N_{\cdot/\epsilon}$, we may construct a measure $\mathrm{Q}^{\epsilon}_{\tau} \ll \mathrm{P}^{\epsilon}|_{\mathscr{F}^{\epsilon}_{\tau}}$ for each $\tau>0$, in which $\frac{1}{\epsilon}J^{\epsilon}$ has Q^{ϵ} intensity $\frac{1}{\epsilon}\lambda(J^{\epsilon})$ for affine function λ.*

$$\lambda : \mathbb{R}_+ \to \mathbb{R}_+, \quad \lambda(x) = l_0 + l_1 x$$

The martingale that induces this change of measure is familiar from Theorem A.4.16, selecting $h(s,x) = \log\lambda(x)$. The Lévy-Khintchine map associated with $\frac{1}{\epsilon}J^{\epsilon}$ is $\Lambda(u,x) = \frac{1}{\epsilon}(e^u - 1)$, and it resolves to the following.

$$Z^{\epsilon} := \exp\left(\frac{1}{\epsilon}\log\lambda(J^{\epsilon}_{-})\cdot J^{\epsilon} + \frac{1}{\epsilon}\big(1-\lambda(J^{\epsilon})\big)\cdot\ell\right)$$

The associated measure $\mathrm{Q}^{\epsilon}_{\tau}(\mathrm{d}\omega) := Z^{\epsilon}(\omega)\cdot \mathrm{P}^{\epsilon}|_{\mathscr{F}^{\epsilon}_{\tau}}(\mathrm{d}\omega)$ makes $\frac{1}{\epsilon}J^{\epsilon}$ have the desired intensity $\frac{1}{\epsilon}\lambda(J^{\epsilon})$. The large deviation principle associated with $(\mathrm{Q}^{\epsilon}_{\tau})_{\epsilon>0}$ then comes from that of $(\mathrm{P}^{\epsilon}|_{\mathscr{F}^{\epsilon}_{\tau}})_{\epsilon>0}$. The change in rate function comes from exponential corrections of our martingale term. Indeed, we observe that if J^{ϵ} is uniformly within $\delta>0$ of some absolutely continuous increasing ξ on some compact interval $[0,\tau]$, then we have the following inequalities.

$$\begin{aligned}\left|\int_0^{\tau}\Big(1-\lambda\big(J^{\epsilon}_t\big)\Big)\mathrm{d}t - \int_0^{\tau}\Big(1-\lambda\big(\xi(t)\big)\Big)\mathrm{d}t\right| &\le \ell_1\delta\tau\\ \left|\int_0^{\tau}\log\lambda\big(J^{\epsilon}_{t-}\big)\mathrm{d}J^{\epsilon}_t - \int_0^{\tau}\log\lambda\big(\xi(t-)\big)\mathrm{d}J^{\epsilon}_t\right| &\le \sup_{t\in[0,\tau]}\Big|\log\lambda\big(J^{\epsilon}_t\big) - \log\lambda\big(\xi(t)\big)\Big|\cdot J^{\epsilon}_T\\ &\le \frac{l_1}{l_0}\delta\Big(\xi(\tau)+\delta\Big),\end{aligned}$$

$$\left| \int_0^\tau \log \lambda\big(\xi(t-)\big) \mathrm{d}J_t^\epsilon - \int_0^\tau \log \lambda\big(\xi(t-)\big) \mathrm{d}\xi(t) \right| \leq \log \lambda\big(\xi(\tau)\big)\delta$$
$$\left| \int_0^\tau \log \lambda\big(J_{t-}^\epsilon\big) \mathrm{d}J_t^\epsilon - \int_0^\tau \log \lambda\big(\xi(t-)\big) \mathrm{d}\xi(t) \right| \leq \frac{l_1}{l_0}\delta\Big(\xi(\tau)+\delta\Big) + \log \lambda\big(\xi(\tau)\big)\delta$$

Then these give us the following identity, revealing the rate function.

$$\begin{aligned}
&\lim_{\delta\to 0}\lim_{\epsilon\to 0} \epsilon \log \mathrm{Q}_\tau^\epsilon\big(J^\epsilon|_{[0,\tau]} \in B(\xi,\delta)\big) \\
&= \lim_{\delta\to 0}\lim_{\epsilon\to 0} \epsilon \log \mathrm{E}_{\mathrm{P}^\epsilon}\left(\exp\Big(\frac{1}{\epsilon}\int_0^\tau \log \lambda(J_{s-}^\epsilon)\mathrm{d}J_s^\epsilon + \frac{1}{\epsilon}\int_0^\tau \big(1-\lambda(J_s^\epsilon)\big)\mathrm{d}s\Big) 1_{J^\epsilon|_{[0,\tau]}\in B(\xi,\delta)} \right) \\
&= \int_0^\tau \log \lambda\big(\xi(s-)\big)\mathrm{d}\xi(s) + \int_0^\tau \Big(1-\lambda\big(\xi(s)\big)\Big)\mathrm{d}s \\
&\quad + \lim_{\delta\to 0}\Big(l_1\delta\tau + \frac{l_1}{l_0}\delta\big(\xi(\tau)+\delta\big) + \log\lambda\big(\xi(\tau)\big)\delta\Big) + \lim_{\delta\to 0}\lim_{\epsilon\to 0}\epsilon \log \mathrm{P}\Big(\epsilon N_{\cdot/\epsilon}|_{[0,\tau]} \in B(\xi,\delta)\Big) \\
&= \int_0^\tau \Big(\dot\xi(s)\log\lambda\big(\xi(s)\big) - \lambda\big(\xi(s)\big) + 1\Big)\mathrm{d}s - \int_0^\tau \Big(\dot\xi(t)\log\dot\xi(t) - \dot\xi(t) + 1\Big)\mathrm{d}s \\
&= -\int_0^\tau \left(\dot\xi(s)\log\Big(\frac{\dot\xi(s)}{\lambda\big(\xi(s)\big)}\Big) - \dot\xi(s) + \lambda\big(\xi(s)\big)\right)\mathrm{d}s
\end{aligned}$$

From here, we lift our large deviation principle back to the interval $[0,\infty)$ *by another application of Dawson-Gärtner.*

Of course, many technical details are missing in the above argument, but we again reiterate that this is merely to gather intuition on rate functions. For a rigorous proof involving such analysis, we refer the reader to [GZ18], where they prove a large deviation principle for an asymptotic family of nonlinear Hawkes processes. In any case, we again note that Theorem 2.5.10 covers the large deviation principle and rate function mentioned above. A continuous-branching process X *with intensity* $\lambda(X)$ *has special semimartingale decomposition as below,*

$$X = 1 * q^X = 1 * \hat q^X + 1 * \tilde q^X = \lambda(X) \bullet \ell + 1 * \tilde q^X$$

which lends itself to the following special differential characteristics (β,α,μ),

$$\beta(x) = \lambda(x), \quad \alpha(x) = 0, \quad \mu(x,\mathrm{d}v) = \lambda(x)\delta_1(\mathrm{d}v)$$

and so the corresponding integrand in (2.5.11) evaluates the following function.

$$\begin{aligned}
\Lambda^*\big(\dot x, x\big) &= \sup_{u\in\mathbb{R}}\left(u\dot x - u\lambda(x) - \int_{\mathbb{R}}\big(e^{uv} - 1 - uv\big)\lambda(x)\delta_1(\mathrm{d}v)\right) \\
&= \sup_{u\in\mathbb{R}}\Big(u\dot x - \lambda(x)e^u + \lambda(x)\Big) \\
&= \begin{cases} \dot x \log\Big(\dfrac{\dot x}{\lambda(x)}\Big) - \dot x + \lambda(x), & \dot x \geq 0,\ \lambda(x)\geq 0 \\ \infty, & \textit{otherwise} \end{cases}
\end{aligned}$$

We may even extend this to Hawkes processes $(\epsilon X^\epsilon)_{\epsilon>0}$ *induced by base process,*

$$X = r\big(\mu - X_t\big)\bullet\ell + N, \quad N \textit{ intensity } \lambda(X),$$

for this process X has affine special differential characteristics.

$$\beta(x) = r(\mu - x) + \lambda(x), \quad \alpha(x) = 0, \quad \mu(x, \mathrm{d}v) = \lambda(x)\delta_1(\mathrm{d}v)$$

The rate function then involves the following expression,

$$\begin{aligned}
\Lambda^*(\dot{x}, x) &= \sup_{u \in \mathbb{R}} \left(u\dot{x} - u\Big(r(\mu - x) - \lambda(x)\Big) - \int_{\mathbb{R}} \big(e^{uv} - 1 - uv\big)\lambda(x)\delta_1(\mathrm{d}v) \right) \\
&= \sup_{u \in \mathbb{R}} \left(u\Big(\dot{x} - r(\mu - x)\Big) - \lambda(x)e^u + \lambda(x) \right) \\
&= \begin{cases} \dot{x} \log\left(\dfrac{\dot{x} - r(\mu - x)}{\lambda(x)}\right) - \dot{x} + \lambda(x), & \lambda(x) > 0,\ \dot{x} - r(\mu - x) \geq 0, \\ \infty, & \textit{otherwise} \end{cases}
\end{aligned}$$

which is similar to the linear case of [GZ18].

3.3 Coupling

The previous result did not include any examples in which the jump distribution was non-degenerate. Intuitively speaking, there is no way to naturally transform a Poisson process to a *compound*-Poisson process, for the distribution of the jumps introduces a new source of randomness. This intuition coincides with our difficulties in evaluating our rate function (2.5.11) for such processes.

Consider the simple example of a compound-Poisson process X driven by standard Poisson process N and independent jumps $(V_k)_{k\in\mathbb{N}}$ distributing with common distribution κ satisfying the light-tails condition (3.1.1).

$$X_t = \sum_{k=1}^{N_t} V_k$$

This process has special differential characteristics (β, α, μ) as below,

$$\beta(x) = \overline{\kappa}, \quad \alpha(x) = 0, \quad \mu(x, \mathrm{d}v) = \kappa(\mathrm{d}v),$$

where $\overline{\kappa} := \int_{\mathbb{V}} v\kappa(\mathrm{d}v)$ denotes the mean of κ. Associating an asymptotic family $(\epsilon X^\epsilon)_{\epsilon>0}$ with this base process X will result in the large deviation principle of Theorem 2.5.10, where the rate function (2.5.11) will involve the following expression,

$$\begin{aligned}
\Lambda^*(\dot{x}, x) &= \sup_{u \in \mathbb{V}} \left(\langle u, \dot{x} \rangle - \langle u, \overline{\kappa} \rangle - \int_{\mathbb{V}} \big(e^{\langle u,v \rangle} - 1 - \langle u, v \rangle\big)\kappa(\mathrm{d}v) \right) \\
&= \sup_{u \in \mathbb{V}} \left(\langle u, \dot{x} \rangle - e^{\Lambda_\kappa(u)} + 1 \right),
\end{aligned}$$

where we recall Λ_κ is the cumulant generating function associated with κ. The arbitrary nature of this function means that resolving even a semi-closed form for the above expression is a difficult task.

Our expression Λ^* is determined by the special differential characteristics (β, α, μ) associated with X, which—if we are familiar with the theory of semimartingales—serve as the predictable projections of a semimartingale. This somehow suggests to us that Λ^* is insufficient in understanding the deviations of X, since the jump times of N are totally inaccessible. This raises the question of if somehow coupling (X, N) will provide us *more* information. From the technical perspective of σ-algebras, the answer is *no*, since X determines N. However, as moot of a discussion as this is from a technical perspective, it turns out to head us in the right direction.

We will see in the below examples that coupling (X, N) will give us semi-closed forms for our rate function. For illustrative purposes, these examples will again include heuristics on how to prove a principle and derive the rate function without Theorem 2.5.10. However, these arguments are no longer backed by results in literature, for we are now entering uncharted territory. We reiterate that these results need not extend past heuristics, for Theorem 2.5.10 already provides us the principle.

Example 3.3.1 (Compound-Poisson). *Returning to the example above, let us consider the following family $(\epsilon X^\epsilon)_{\epsilon>0}$ determined by base path X.*

$$\epsilon X_t^\epsilon = \epsilon \sum_{k=1}^{N_{t/\epsilon}} V_k$$

To understand the deviations of this object, we observe how it appears to be a composition of two processes.

$$\epsilon X_t^\epsilon = \epsilon A^\epsilon_{\epsilon N_t^\epsilon} = \Big((\epsilon A^\epsilon) \circ (\epsilon N^\epsilon) \Big)_t, \quad \epsilon A_t^\epsilon := \epsilon \sum_{k=1}^{[t/\epsilon]} V_k, \quad \epsilon N_t^\epsilon := \epsilon N_{t/\epsilon}$$

Note that the deviations of ϵA^ϵ are understood by Mogulskii's theorem, while ϵN^ϵ are understood from Example 3.1.5. Moreover, these processes are independent, so it is easy to see that $(\epsilon A^\epsilon, \epsilon N^\epsilon)_{\epsilon>0}$ satisfies a large deviation principle with rate function in which the finite points evaluate as the following integral.

$$I_{AN}(\gamma, \eta) = \int_0^\infty \Big(\dot\eta(t) \log \dot\eta(t) - \dot\eta(t) + 1 + \Lambda_\kappa^*\big(\dot\gamma(t)\big) \Big) \mathrm{d}t$$

From here, $(\epsilon X^\epsilon, \epsilon N^\epsilon)_{\epsilon>0}$ should satisfy a large deviation principle via the contraction principle of $\epsilon X^\epsilon = \epsilon A^\epsilon \circ \epsilon N^\epsilon$. The rate function is then as follows.

$$I_{XN}(\xi, \eta) = \inf \big\{ I_{AB}(\gamma, \eta) : \xi = \gamma \circ \eta \big\}$$

From here, we recognize that the condition $\xi = \gamma \circ \eta$ implies $\dot\gamma = \frac{\dot\xi \circ \eta^{-1}}{\dot\eta \circ \eta^{-1}}$ and so substituting this in I_{AB} and making a time-change $t \leftarrow \eta^{-1}(t)$ (on only the final term) will reduce to the following expression for finite points.

$$I_{XN}(\xi, \eta) = \int_0^\infty \left(\dot\eta(t) \log \dot\eta(t) - \dot\eta(t) + 1 + \dot\eta(t) \Lambda_\kappa^* \Big(\frac{\dot\xi(t)}{\dot\eta(t)} \Big) \right) \mathrm{d}t \tag{3.3.2}$$

This integrand should make intuitive sense when we recognize that (ξ, η) serves as a proxy for (X, N). The first three terms accumulate rate for deviations in the arrivals of N, while the last term is accumulating rate for deviations in the jump sizes. The term $\dot\xi(t)/\dot\eta(t)$ is a deviation of X normalized against one in N, which is effectively a deviation from a jump

V_k. Deviations of a jump V_k are measured by Λ^*_κ, and they are arriving at a rate of N, hence the final expression in the rate function. For more detail on how a large deviation principle could be proven with this type of argument, we refer the reader to [DRL04]. In fact, Theorem 4 in this paper inspired us to use coupling as an argument.

In any case, we may still refer to Theorem 2.5.10 for a rigorous argument. Denote $(\Omega, \Sigma, \mathrm{P})$ the space that includes all these quantities, $\mathscr{F}^N$ the filtration associated with N, and $\hat{X}^\epsilon = (X^\epsilon, N_{\cdot/\epsilon})$. It is clear that $\hat{X}^\epsilon$ is a pure jump process,

$$\hat{X^\epsilon}_t = \mathrm{id}_{\mathbb{V}\times\mathbb{R}_+} * q_t^{\hat{X}^\epsilon},$$

and that any predictable $H : \Omega \times \mathbb{R}_+ \times (\mathbb{V} \times \mathbb{R}_+) \to \mathbb{R}_+$ is such that the following holds.

$$\begin{aligned}
\mathrm{E}_\mathrm{P}\Big(H * q_\infty^{\hat{X}^\epsilon}\Big) &= \mathrm{E}_\mathrm{P} \int_0^\infty H(\cdot, s, \Delta X_s, 1)\mathrm{d}N_{s/\epsilon} \\
&= \mathrm{E}_\mathrm{P} \int_0^\infty H(\cdot, \epsilon s, \Delta X_{\epsilon s}, 1)\mathrm{d}N_s \\
&= \mathrm{E}_\mathrm{P}\bigg(\mathrm{E}_\mathrm{P}\bigg(\int_0^\infty H(\cdot, \epsilon s, \Delta X_{\epsilon s}, 1)\mathrm{d}N_s|\mathscr{F}_\infty^N\bigg)\bigg) \\
&= \mathrm{E}_\mathrm{P}\bigg(\int_\mathbb{V}\int_0^\infty H(\cdot, \epsilon s, v_1, 1)\mathrm{d}N_s\kappa(\mathrm{d}v_1)\bigg) \\
&= \int_0^\infty \int_\mathbb{V} H(\cdot, \epsilon s, v_1, 1)\kappa(\mathrm{d}v_1)\mathrm{d}s \\
&= \int_0^\infty \int_\mathbb{V} H(\cdot, s, v_1, v_2)\frac{1}{\epsilon}\kappa(\mathrm{d}v_1)\delta_1(\mathrm{d}v_2)\mathrm{d}s, \qquad\qquad s \leftarrow \epsilon s
\end{aligned}$$

This means that we have the following predictable compensator,

$$\hat{q}^{\hat{X}^\epsilon}(\mathrm{d}s, \mathrm{d}\hat{v}) = \frac{1}{\epsilon}\kappa(\mathrm{d}v_1)\delta_1(\mathrm{d}v_2)\mathrm{d}s$$

which by our light-tails (3.1.1) and Proposition 1.4.2 means that $\hat{X}^\epsilon$ is special and has the following decomposition.

$$\hat{X}^\epsilon = \mathrm{id}_{\mathbb{V}\times\mathbb{R}_+} * \hat{q}_t^{\hat{X}^\epsilon} + \mathrm{id}_{\mathbb{V}\times\mathbb{R}_+} * \tilde{q}_t^{\hat{X}^\epsilon} = \frac{1}{\epsilon}\begin{pmatrix}\overline{\kappa}\\ 1\end{pmatrix} \bullet \ell_t + \mathrm{id}_{\mathbb{V}\times\mathbb{R}_+} * \tilde{q}_t^{\hat{X}^\epsilon}$$

This means that the special differential characteristics $(\beta^\epsilon, \alpha^\epsilon, \mu^\epsilon)$ of $\hat{X}^\epsilon$ are as follows.

$$\beta^\epsilon(\hat{x}) = \frac{1}{\epsilon}\begin{pmatrix}\overline{\kappa}\\ 1\end{pmatrix}, \quad \alpha^\epsilon(\hat{x}) = 0, \quad \mu^\epsilon(\hat{x}, \mathrm{d}\hat{v}) = \frac{1}{\epsilon}\kappa(\mathrm{d}v_1)\delta_1(\mathrm{d}v_2)$$

This parameterization, along with our light-tails (3.1.1) means that we have the hypotheses of Theorem 2.5.10. Note that the associated Lévy-Khintchine map Λ is as follows.

$$\begin{aligned}
\Lambda(\hat{u}, \hat{x}) &= \langle u_1, \overline{\kappa}\rangle + u_2 + \int_{\mathbb{V}\times\mathbb{R}_+}\Big(e^{\langle u_1, v_1\rangle + u_2} - 1 - \langle u_1, v_1\rangle - u_2 v_2\Big)\kappa(\mathrm{d}v_1)\delta_1(\mathrm{d}v_2) \\
&= e^{\Lambda_\kappa(u_1) + u_2} - 1
\end{aligned}$$

The integrand of our rate function 2.5.11 thus involves the following expression.

$$\Lambda^*\big(\hat{\dot{x}}, \hat{x}\big) = \sup_{\hat{u}\in\mathbb{V}\times\mathbb{R}_+}\Big(\langle u_1, \dot{x}_1\rangle - u_2\dot{x}_2 - e^{\Lambda_\kappa(u_1)+u_2} + 1\Big)$$

Setting the gradients to zero results in the system of equations.

$$\begin{aligned}0 &= \dot{x}_1 - e^{\Lambda_\kappa(u_1)+u_2}\nabla\Lambda_\kappa(u_1)\\ 0 &= \dot{x}_2 - e^{\Lambda_\kappa(u_1)+u_2}\end{aligned}$$

Some immediate results come from this; firstly, the second equation gives us that $\dot{x}_2 > 0$ and the following.

$$\dot{x}_2 = e^{\Lambda_\kappa(u_1)+u_2}, \quad \dot{x}_2\Lambda_\kappa(u_1) = \dot{x}_2\log\dot{x}_2 - u_2\dot{x}_2 \tag{3.3.3}$$

Dividing the common expressions between the two equations results to the following.

$$\nabla\Lambda_\kappa(u_1) = \frac{\dot{x}_1}{\dot{x}_2}$$

Seeing as Λ is convex, this is the unique solution to its Fenchel-Legendre transform.

$$\Lambda^*_\kappa\Big(\frac{\dot{x}_1}{\dot{x}_2}\Big) = \langle u_1, \dot{x}_1/\dot{x}_2\rangle - \Lambda(u_1)$$

Scooting the expressions around, we get

$$\langle u_1, \dot{x}_1\rangle = \dot{x}_2\Lambda^*_\kappa\Big(\frac{\dot{x}_1}{\dot{x}_2}\Big) + \dot{x}_2\Lambda(u_1),$$

which when combined with (3.3.3), we get the following.

$$\langle u_1, \dot{x}_1\rangle - u_2\dot{x}_2 - e^{\Lambda_\kappa(u_1)+u_2} + 1 = \dot{x}_2\log\dot{x}_2 - \dot{x}_2 + 1 + \Lambda^*_\kappa\Big(\frac{\dot{x}_1}{\dot{x}_2}\Big)$$

Observe that this results in the same integral as (3.3.2).

Example 3.3.4 (Compound-Hawkes). *We could similarly come up with a family $(\epsilon X^\epsilon)_{\epsilon>0}$ of compound-Hawkes processes with affine intensity $\lambda(x) = l_0 + l_1 x$,*

$$\epsilon X^\epsilon_t = x + r\big(\mu - \epsilon X^\epsilon\big)\bullet\ell_t + \epsilon\sum_{k=1}^{N^\epsilon_t} V_k,$$

$$N^\epsilon \text{ intensity } \frac{1}{\epsilon}\lambda(\epsilon X^\epsilon),$$

$$(V_k)_{k\in\mathbb{N}} \text{ independent and commonly distributed by } \kappa \text{ on } \mathscr{B}(\mathbb{R}_+)$$

A proof for a large deviation principle of $(\epsilon X^\epsilon, \epsilon N^\epsilon)_{\epsilon>0}$ could use the steps of Examples 3.2.4 and 3.3.1, summarized as below.

1. *Use Example 3.3.1 to establish a large deviation principle for the family $(\epsilon K^\epsilon, \epsilon N^\epsilon)_{\epsilon>0}$ where N^ϵ is a Poisson process with rate $1/\epsilon$ and $K^\epsilon = \sum_{k=1}^{N^\epsilon} V_k$ is the accumulated jump process. The integral in the rate function will be as (3.3.2).*

$$I^{pois}_{KN}(\gamma, \eta) = \int_0^\infty \Big(\dot{\eta}(t)\log\dot{\eta}(t) - \dot{\eta}(t) + 1 + \dot{\eta}(t)\Lambda^*_\kappa\Big(\frac{\dot{\gamma}(t)}{\dot{\eta}(t)}\Big)\Big)\mathrm{d}t$$

2. *We observe that defining the map F,*

$$F\gamma(t) = \mu + (x-\mu)e^{-rt} + \int_0^t e^{-r(t-s)}\mathrm{d}\gamma(s),$$

will be such that $\epsilon X^\epsilon = F(\epsilon K^\epsilon)$ gives us our path-properties as above. To see this, fix an arbitrary finite-variation γ with $\gamma(0) = 0$ and define $\xi = F(\gamma)$. Observe that applying the fundamental theorem of calculus and Fubini's theorem provides the following equality.

$$\begin{aligned}
\xi(t) &= \mu + (x-\mu)e^{-rt} + \int_0^t e^{-r(t-s)}\mathrm{d}\gamma(s) \\
&= x + \big((x-\mu)e^{-rt} - (x-\mu)\big) + \int_0^t \Big(1 + \big(e^{-r(t-s)} - e^{-r(s-s)}\big)\Big)\mathrm{d}\gamma(s) \\
&= x - r(x-\mu)e^{-r\ell} \bullet \ell_t + \gamma(t) - \int_0^t \int_s^t re^{-r(\tau-s)}\mathrm{d}\tau\mathrm{d}\gamma(s) \\
&= x + r\Big(\mu - \big(\mu + (x-\mu)e^{-r\ell}\big)\Big) \bullet \ell_t + \gamma(t) - \int_0^t \int_0^\tau re^{-r(\tau-s)}\mathrm{d}\gamma(s)\mathrm{d}\tau \\
&= x + \int_0^t r\Big(\mu - \big(\mu + (x-\mu)e^{-r\tau} - \int_0^\tau e^{-r(\tau-s)}\mathrm{d}\gamma(s)\big)\Big)\mathrm{d}\tau + \gamma(t) \\
&= x + \int_0^t r\big(\mu - \xi(\tau)\big)\mathrm{d}\tau + \gamma(t) \\
&= x + r(\mu - \xi) \bullet \ell_t + \gamma(t)
\end{aligned} \tag{3.3.5}$$

Thus, if we use the contraction map $(\gamma, \eta) \mapsto (F\gamma, \eta)$, we will get a large deviation principle for $(\epsilon X^\epsilon, \epsilon N^\epsilon)$. Note that if an absolutely continuous γ is such that $F\gamma = \xi$, then (3.3.5) tells us

$$\dot{\xi}(t) = r\big(\mu - \xi(t)\big) + \dot{\gamma}(t),$$

and so the integral in our rate function now evaluates as follows.

$$I_{XN}^{pois}(\xi, \eta) = \int_0^\infty \Big(\dot{\eta}(t)\log\dot{\eta}(t) - \dot{\eta}(t) + 1 + \dot{\eta}(t)\Lambda_\kappa^*\Big(\frac{\dot{\xi}(t) - r(\mu - \xi(t))}{\dot{\eta}(t)}\Big)\Big)\mathrm{d}t$$

3. *At last, we apply our measure-change argument as in Example 3.2.4 to change the intensity of N^ϵ to $\frac{1}{\epsilon}\lambda(\epsilon X^\epsilon)$. The work is the exact same, and so our rate function associated with $(\epsilon X^\epsilon, \epsilon N^\epsilon)_{\epsilon>0}$ will become the following.*

$$I_{XN}(\xi, \eta) = \int_0^\infty \Big(\dot{\eta}(t)\log\Big(\frac{\dot{\eta}(t)}{\lambda\big(\xi(t)\big)}\Big) - \dot{\eta}(t) + \lambda\big(\xi(t)\big) + \dot{\eta}(t)\Lambda_\kappa^*\Big(\frac{\dot{\xi}(t) - r(\mu - \xi(t))}{\dot{\eta}(t)}\Big)\Big)\mathrm{d}t \tag{3.3.6}$$

Instead of working through the numerous details of the lengthy argument above, we can again use Theorem 2.5.10 and its rate function (2.5.11). By work nearly identical to the end of Example 3.3.1, we can see that the special differential characteristics of $\hat{X}^\epsilon = (X^\epsilon, N^\epsilon)$ are as below.

$$\beta(\hat{x}) = \begin{pmatrix} r(\kappa - x_1) + \lambda(x_1)\overline{\kappa} \\ \lambda(x_1) \end{pmatrix}, \quad \beta(\hat{x}) = 0, \quad \mu(\hat{x}, \hat{v}) = \lambda(x_1)\kappa(\mathrm{d}v_1)\delta_1(\mathrm{d}v_2)$$

Our associated Lévy-Khintchine map is then the following Λ.

$$\begin{aligned}\Lambda(\hat{u},\hat{x}) &= u_1\big(r(\mu-x_1)+\overline{\kappa}\big)+u_2\lambda(x_1)+\int_{\mathbb{R}_+^2}\Big(e^{u_1v_1+u_2}-1-u_1v_1-u_2v_2\Big)\kappa(\mathrm{d}v_1)\delta_1(\mathrm{d}v_2)\\ &= u_1r(\mu-x_1)+e^{\Lambda_\kappa(u_1)+u_2}-1\end{aligned}$$

Almost identical calculus to Example 3.3.1 also provides us with the following expression, which agrees with 3.3.6.

$$\Lambda^*(\hat{\dot{x}},\hat{x})=\dot{x}_2\log\Big(\frac{\dot{x}_2}{\lambda(x_1)}\Big)-\dot{x}_2+\lambda(x_1)+\dot{x}_2\Lambda_\kappa^*\Big(\frac{\dot{x}_1-r(\mu-x_1)}{\dot{x}_2}\Big)$$

3.4 Semi-closed form of rate function

We have now gathered enough familiarity with rate functions that appear in the large deviations literature and the various tools we may use to lift existing principles to new ones. This will allow us to more easily understand the nature and derivation of our semi-closed form for (2.5.11) in Theorem 2.5.10. Let us now elaborate on the family on which we derive our semi-closed form; it is assumed that X is selected with our assumptions of Section 2.2 that we needed for Theorem 2.5.10.

Firstly, in order for us to perform the tricks of coupling we explored in the previous section, we need to couple our arrivals and/or accumulated jumps. We will need these objects to be locally integrable, so we impose that our base process X is *locally countable* in the sense of our definition in Appendix A.3. In particular, suppose that our base process X has special differential characteristics (β,α,μ), where μ has the finiteness condition below.

(3.4.1) $$\mu(x,\mathbb{V})<\infty,\quad x\in\mathbb{X}$$

Note that Proposition 1.4.3 thus provides us with a factoring

(3.4.2) $$\mu(x,\mathrm{d}v)=\lambda(x)\kappa(x,\mathrm{d}v)$$

into an intensity function $\lambda\in\mathscr{B}(\mathbb{X})/\mathscr{B}(\mathbb{R}_+)$ and conditional jump distribution κ from $(\mathbb{X},\mathscr{B}(\mathbb{X}))$ to $(\mathbb{V},\mathscr{B}(\mathbb{V}))$. Denote to each $\mu(x,\cdot)$ the notation from Sections 3.1 and 3.3.

$$\overline{\kappa(x,\cdot)}:=\int_{\mathbb{V}}v\kappa(x,\mathrm{d}v),\quad \Lambda_{\kappa(x,\cdot)}(u):=\int_{\mathbb{V}}e^{\langle u,v\rangle}\kappa(x,\mathrm{d}v)$$

Note that our finiteness assumption (3.4.1) and our light-tails assumption $\mathcal{D}_\Lambda=\mathbb{V}$ (along with Lemma A.4.2) give us the following properties.

$$\overline{\kappa(x,\cdot)}<\infty,\quad \Lambda_{\kappa(x,\cdot)}(u)<\infty,\quad x\in\mathbb{X}$$

We also have local integrability of the following process by our finiteness assumption (3.4.1).

(3.4.3) $$N^X:=1*q^X$$

Note that this is a stronger restriction than local integrability of the compensated jumps

(3.4.4) $$V^X:=\mathrm{id}_{\mathbb{V}}*\tilde{q}^X,$$

which we already have from the special property of X.

The remainder of this section will assume $\hat{X} := (X, X^{\mathrm{c}}, V^X, N^X)$ is specified from an affine process X with the properties discussed above. We will also factor our kernel $(\mathrm{P}_x)_{x\in\mathbb{X}}$ even further to one $(\hat{\mathrm{P}}_{\hat{x}})_{\hat{x}\in\hat{\mathbb{X}}}$ derived from the regular conditional P_x-distribution of $(\mathrm{P}_x, \mathscr{F})$ time-homogeneous Markov process $(X^{\mathrm{c}}, V^X, N^X)$. We will see in the coming results that $\hat{X}$ and its associated kernel $(\hat{\mathrm{P}}_{\hat{x}})_{\hat{x}\in\hat{\mathbb{X}}}$ and filtration $\mathscr{F}$ still satisfy the assumptions of Section 2.2 needed to apply Theorem 2.5.10; furthermore, its coupling structure will allow us to derive a semi-closed form for (2.5.11). This is a process on

$$\hat{\mathbb{X}} := \mathbb{X} \times \mathbb{V} \times \mathbb{V} \times \mathbb{R}_+,$$

which is a subset of vector space

$$\hat{\mathbb{V}} := \mathbb{V} \times \mathbb{V} \times \mathbb{V} \times \mathbb{R},$$

when equipped with componentwise operations. We henceforth denote elements $\hat{v} \in \hat{\mathbb{V}}$ with components,

$$\hat{v} := (v_1, v_2, v_3, v_4),$$

and equip $\hat{\mathbb{V}}$ with the inner-product below.

$$\big\langle \hat{v}, \hat{w} \big\rangle := \langle v_1, w_1 \rangle + \langle v_2, w_2 \rangle + \langle v_3, w_3 \rangle + v_4 w_4$$

We will also indifferently treat such tuples as column vectors, so that operators on $\mathbb{L}(\hat{\mathbb{V}})$ can be seen as 4×4 block-matrices of operators on $\mathbb{L}(\mathbb{V})$ and $\mathbb{L}(\mathbb{R})$ (with the natural blocking). If we want to go a level deeper within coordinates, we will use the following notation.

$$\hat{v} := \big(v^{11}, \ldots, v^{1d}, v^{21}, \ldots, v^{2d}, v^{31}, \ldots, v^{3d}, v^{41}\big)$$

Lemma 3.4.5. *The space $\hat{\mathbb{X}}$ is a cone in $\hat{\mathbb{V}}$ with operations taken componentwise.*

Proof. This is obvious, as each of the factors $\mathbb{X}$, $\mathbb{V}$, $\mathbb{R}_+$ that appear in our product is also a cone.

Proposition 3.4.6. *For each $\hat{x} \in \hat{\mathbb{X}}$, the process $\hat{X}$ is a $(\hat{\mathrm{P}}_{\hat{x}}, \mathscr{F})$ special jump-diffusion with affine special differential characteristics $(\hat{\beta}, \hat{\alpha}, \hat{\mu})$ as below.*

$$\hat{\beta}(\hat{x}) = \begin{pmatrix} \beta(x_1) \\ 0 \\ 0 \\ \lambda(x_1) \end{pmatrix}, \quad \hat{\alpha}(\hat{x}) = \begin{pmatrix} \alpha(x_1) & \alpha(x_1) & 0 & 0 \\ \alpha(x_1) & \alpha(x_1) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{3.4.7}$$
$$\hat{\mu}(\hat{x}, \mathrm{d}\hat{v}) = \lambda(x_1)\kappa(x_1, \mathrm{d}v_1)\delta_{0,v_1,1}(\mathrm{d}v_2, \mathrm{d}v_3, \mathrm{d}v_4),$$

Moreover, denoting $\hat{\Lambda}$ the associated Lévy-Khintchine map, we have $\mathcal{D}_{\hat{\Lambda}} = \hat{\mathbb{V}}$ and the following.

$$\begin{aligned}\hat{\Lambda}(\hat{u}, \hat{x}) = \langle u_1, \beta(x_1) \rangle + \frac{1}{2}\big\langle u_1 + u_2, \alpha(x)(u_1 + u_2)\big\rangle \\ + \lambda(x_1) \exp\Big(\Lambda_{\kappa(x_1,\cdot)}(u_1 + u_3) + u_4\Big) - \lambda(x_1) - \big\langle u_1 + u_3, \lambda(x_1)\overline{\kappa(x_1,\cdot)}\big\rangle\end{aligned}$$

Proof. The fact that $\hat{X}$ is special simply comes from the following decomposition.

$$\begin{aligned}
\hat{X} = \begin{pmatrix} X \\ X^{\mathrm{c}} \\ V^X \\ N^X \end{pmatrix} &= \begin{pmatrix} x_1 + \beta(X) \bullet \ell_t + X^{\mathrm{c}} + \mathrm{id}_{\mathbb{V}} * \tilde{q}_t^X \\ x_2 + X^{\mathrm{c}} \\ x_3 + \mathrm{id}_{\mathbb{V}} * \tilde{q}_t^X \\ x_4 + 1 * q_t^X \end{pmatrix} \\
&= \begin{pmatrix} x_1 + \beta(X) \bullet \ell_t + X^{\mathrm{c}} + \mathrm{id}_{\mathbb{V}} * \tilde{q}_t^X \\ x_2 + X^{\mathrm{c}} \\ x_3 + \mathrm{id}_{\mathbb{V}} * \tilde{q}_t^X \\ x_4 + 1 * \hat{q}_t^X + 1 * \tilde{q}_t^X \end{pmatrix} \\
&= \hat{x} + \begin{pmatrix} \beta(X) \\ 0 \\ 0 \\ \lambda(X) \end{pmatrix} \bullet \ell_t + \begin{pmatrix} X^{\mathrm{c}} \\ X^{\mathrm{c}} \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \mathrm{id}_{\mathbb{V}} \\ 0 \\ \mathrm{id}_{\mathbb{V}} \\ 1 \end{pmatrix} * \tilde{q}_t^X
\end{aligned}$$

Observe that the first term establishes the drift $\hat{\beta}$ as in (3.4.7). The diffusion $\hat{\alpha}$ in (3.4.7) is verified by the following identities.

$$\begin{aligned}
&\left\langle \hat{X}^{\mathrm{c},1i}, \hat{X}^{\mathrm{c},1j} \right\rangle = \left\langle X^{\mathrm{c},i}, X^{\mathrm{c},j} \right\rangle = \alpha_{ij}(X) \bullet \ell, && \left\langle \hat{X}^{\mathrm{c},1i}, \hat{X}^{\mathrm{c},2j} \right\rangle = \left\langle X^{\mathrm{c},i}, X^{\mathrm{c},j} \right\rangle = \alpha_{ij}(X) \bullet \ell, \\
&\left\langle \hat{X}^{\mathrm{c},1i}, \hat{X}^{\mathrm{c},3j} \right\rangle = \left\langle X^{\mathrm{c},i}, 0 \right\rangle = 0, && \left\langle \hat{X}^{\mathrm{c},1i}, \hat{X}^{\mathrm{c},4j} \right\rangle = \left\langle X^{\mathrm{c},i}, 0 \right\rangle = 0, \\
&\left\langle \hat{X}^{\mathrm{c},2i}, \hat{X}^{\mathrm{c},2j} \right\rangle = \left\langle X^{\mathrm{c},i}, X^{\mathrm{c},j} \right\rangle = \alpha_{ij}(X) \bullet \ell, && \left\langle \hat{X}^{\mathrm{c},2i}, \hat{X}^{\mathrm{c},3j} \right\rangle = \left\langle X^{\mathrm{c},i}, 0 \right\rangle = 0, \\
&\left\langle \hat{X}^{\mathrm{c},2i}, \hat{X}^{\mathrm{c},4j} \right\rangle = \left\langle X^{\mathrm{c},i}, 0 \right\rangle = 0, && \left\langle \hat{X}^{\mathrm{c},3i}, \hat{X}^{\mathrm{c},4j} \right\rangle = \left\langle 0, 0 \right\rangle = 0,
\end{aligned}$$

Lastly, for any predictable $H : \Omega \times \mathbb{R}_+ \times \hat{\mathbb{V}} \to \mathbb{R}_+$, we have the following.

$$\begin{aligned}
\mathrm{E}_{\mathrm{P}_x}\left(H * q_\infty^{\hat{X}}\right) &= \mathrm{E}_{\mathrm{P}_x} \int_{\mathbb{R}_+ \times \mathbb{V}} H(\cdot, s, v_1, 0, v_1, 1) q^X(\mathrm{d}s, \mathrm{d}\hat{v}) \\
&= \mathrm{E}_{\mathrm{P}_x} \int_0^\infty \int_{\mathbb{V}} H(\cdot, s, v_1, 0, v_1, 1) \mu(X_s, \mathrm{d}v_1) \mathrm{d}s \\
&= \mathrm{E}_{\mathrm{P}_x} \int_0^\infty \int_{\mathbb{V}} H(\cdot, s, v_1, 0, v_1, 1) \lambda(X_s) \kappa(X_s, \mathrm{d}v_1) \mathrm{d}s \\
&= \mathrm{E}_{\mathrm{P}_x} \int_0^\infty \int_{\hat{\mathbb{V}}} H(\cdot, s, \hat{v}) \lambda(X_s) \kappa(X_s, \mathrm{d}v_1) \delta_{0, v_1, 1}(\mathrm{d}v_2, \mathrm{d}v_3, \mathrm{d}v_3) \mathrm{d}s
\end{aligned}$$

This concludes that, $\hat{q}^{\hat{X}}(\mathrm{d}s, \mathrm{d}\hat{v}) = \hat{\mu}(\hat{X}_s, \mathrm{d}\hat{v})\mathrm{d}s$, where $\hat{\mu}$ is as in (3.4.7). This work shows that $\hat{X}$ is a $(\hat{\mathrm{P}}_{\hat{x} \in \hat{\mathbb{X}}}, \mathscr{F})$ special jump-diffusion; its special differential characteristics are also affine.

$$\begin{aligned}
\hat{\beta}(\hat{x}) &= \begin{pmatrix} b_0 \\ 0 \\ 0 \\ l_0 \end{pmatrix} + \sum_{i=1}^d x^{1i} \begin{pmatrix} b_i \\ 0 \\ 0 \\ l_i \end{pmatrix} \\
\hat{\alpha}(\hat{x}) &= \begin{pmatrix} a_0 & a_0 & 0 & 0 \\ a_0 & a_0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \sum_{i=1}^d x^{1i} \begin{pmatrix} a_0 & a_0 & 0 & 0 \\ a_0 & a_0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}
\end{aligned}$$

$$\hat{\mu}(\hat{x}, \mathrm{d}\hat{v}) = m_0(\mathrm{d}v_1)\delta_{0,v_1,1}(\mathrm{d}v_2, \mathrm{d}v_3, \mathrm{d}v_4) + \sum_{i=1}^{d} x^{1i} m_i(\mathrm{d}v_1)\delta_{0,v_1,1}(\mathrm{d}v_2, \mathrm{d}v_3, \mathrm{d}v_4)$$

The associated Lévy-Khintchine map $\hat{\Lambda}$ associated with it satisfies $\mathcal{D}_{\hat{\Lambda}} = \hat{\mathbb{V}}$. To see this, observe the following inequality for any $\hat{x} \in \hat{\mathbb{X}}$ and $\hat{u} \in \hat{\mathbb{V}}$.

$$\begin{aligned}\int_{|\hat{v}|>1} e^{\langle \hat{u}, \hat{v}\rangle} \hat{\mu}(\hat{x}, \mathrm{d}\hat{v}) &= \int_{\mathbb{V}} e^{\langle u_1 + u_3, v_1\rangle + u_4} 1_{|\hat{v}|>1}(v_1, 0, v_1, 1)\mu(x_1, \mathrm{d}v_1) \\ &= e^{u_4} \int_{\mathbb{V}} e^{\langle u_1 + u_3, v\rangle} \mu(x_1, \mathrm{d}v)\end{aligned}$$

Because we have $\mu(x_1, \mathbb{V}) < \infty$ and $\mathcal{D}_\Lambda = \mathbb{V}$, Lemma A.4.2 tells us (using it twice) that the above expression is finite and that $\mathcal{D}_{\hat{\Lambda}} = \hat{\mathbb{V}}$. We may also simplify $\hat{\Lambda}$ to the desired expression.

$$\begin{aligned}\hat{\Lambda}(\hat{u}, \hat{x}) &= \big\langle \hat{u}, \hat{\beta}(\hat{x})\big\rangle + \frac{1}{2}\big\langle \hat{u}, \hat{\alpha}(\hat{x})\hat{u}\big\rangle + \int_{\hat{\mathbb{V}}} \Big(e^{\langle \hat{u}, \hat{v}\rangle} - 1 - \big\langle \hat{u}, \hat{v}\big\rangle\Big)\hat{\mu}(\hat{x}, \mathrm{d}\hat{v}) \\ &= \langle u_1, \beta(x_1)\rangle + u_4\lambda(x_1) \\ &\quad + \frac{1}{2}\big\langle u_1, \alpha(x_1)u_1\big\rangle + \frac{1}{2}\big\langle u_1, \alpha(x_1)u_2\big\rangle + \frac{1}{2}\big\langle u_2, \alpha(x_1)u_1\big\rangle + \frac{1}{2}\big\langle u_2, \alpha(x_1)u_2\big\rangle \\ &\quad + \int_{\mathbb{V}} \Big(e^{\langle u_1, v\rangle + \langle u_3, v\rangle + u_4} - 1 - \langle u_1, v_1\rangle - \langle u_3, v\rangle - u_4\Big)\lambda(x_1)\kappa(x_1, \mathrm{d}v) \\ &= \langle u_1, \beta(x_1)\rangle + \frac{1}{2}\big\langle u_1 + u_2, \alpha(x_1)(u_1 + u_2)\big\rangle \\ &\quad + \lambda(x_1)\Big(\int_{\mathbb{V}} e^{\langle u_1 + u_3, v\rangle}\kappa(x_1, \mathrm{d}v)\Big)e^{u_4} - \lambda(x_1) - \big\langle u_1 + u_3, \lambda(x_1)\overline{\kappa(x_1, \cdot)}\big\rangle \\ &= \langle u_1, \beta(x_1)\rangle + \frac{1}{2}\big\langle u_1 + u_2, \alpha(x)(u_1 + u_2)\big\rangle \\ &\quad + \lambda(x_1)\exp\Big(\Lambda_{\kappa(x_1,\cdot)}(u_1 + u_3) + u_4\Big) - \lambda(x_1) - \big\langle u_1 + u_3, \lambda(x_1)\overline{\kappa(x_1, \cdot)}\big\rangle\end{aligned}$$

Unfortunately, despite $\hat{X}$ having affine characteristics, this is not enough to ensure it is an affine process. We technically need existence of a solution $\hat{\Psi}$ to its associated system $\partial_t\hat{\Psi}(t, \hat{u}, \hat{x}) = \Lambda\big(\hat{\psi}(t, \hat{u}), \hat{x}\big)$ as in Proposition 1.4.5. It is partially solved by the state-independence of Λ in the coordinates x_2, x_3, x_4, but we cannot find a way to leverage the equation associated with Λ. Thus is the following assumption.

Assumption 3.4.8. *Assume that X is such that the associated special differential characteristics (β, α, μ) and the associated parts b_i, a_i, m_i, l_i for $i = 0, \ldots, d$ induce a solution to the following equation, for each $\hat{u} \in \mathcal{U}_{\hat{\mathbb{X}}}$.*

$$\begin{aligned}&i = 0, \ldots, d \\ &\partial_t\hat{\psi}_i(t, \hat{u}) = \Big\langle \hat{\psi}(t, \hat{u}), b_i\Big\rangle + \frac{1}{2}\Big\langle \hat{\psi}(t, \hat{u}) + u_2, a_i\big(\hat{\psi}(t, \hat{u}) + u_2\big)\Big\rangle \\ &\qquad\qquad + \int_{\mathbb{V}} \Big(e^{\langle \hat{\psi}(t,\hat{u}) + u_3, v\rangle + u_4} - 1 - \Big\langle \hat{\psi}(t, \hat{u}) + u_3, v\Big\rangle\Big)m_i(\mathrm{d}v) \\ &\psi_0(t, u) = 0,\ \ \psi_i(t, u) = u_i\end{aligned}$$

Remark 3.4.9. *1. When we are dealing with the canonical space $\mathbb{X} = \mathbb{R}_+^m \times \mathbb{R}^n$, then $\hat{\mathbb{X}} = \mathbb{R}_+^{\hat{m}} \times \mathbb{R}^{\hat{n}}$, up to reordering the factors. In this scenario, we have admissibility conditions on the special differential characteristics which impose existence of the solution, for each $\hat{u} \in \mathcal{U}_{\hat{\mathbb{X}}}$.*

2. Note that we already have a solution to the following equation for each $u \in \mathcal{U}_{\mathbb{X}}$, by existence of X as an affine process and Theorem 1.1.4.

$$\begin{aligned}
&i = 0, \dots, d \\
&\dot{\psi}_i(t,u) = \langle \psi(t,u), b_i \rangle + \frac{1}{2}\langle \psi(t,u), a_i \psi(t,u)\rangle + \int_{\mathbb{V}} \big(e^{\langle \psi(t,u), v\rangle} - 1 - \langle \psi(t,u), v\rangle\big) m_i(\mathrm{d}v) \\
&\psi_0(t,u) = 0, \;\; \psi_i(t,u) = u_i
\end{aligned}$$

Theorem 3.4.10. *Let X be an affine process as introduced above with Assumption 3.4.8 satisfied. Then $\hat{X}$ is a base affine process with which we may parameterize $\epsilon \hat{X}^\epsilon$ with distribution as in Section 2.1. Fixing $\hat{x} \in \hat{\mathbb{X}}^\circ$ and denoting $\hat{\mathrm{P}}^\epsilon_{\hat{x}}$ the distribution associated with $\epsilon \hat{X}^\epsilon$ starting at $\hat{x}$, we have a large deviation principle for $(\hat{\mathrm{P}}^\epsilon_{\hat{x}})_{\epsilon>0}$. The rate function $I_{\hat{x}}$ simplifies to the following semi-closed form, where we denote the components of an arbitrary function $\hat{\xi}$ by $\hat{\xi} := (\xi, \omega, \gamma, \eta)$.*

$$\begin{aligned}
I_{\hat{x}}(\hat{\xi}) = &\int_0^\infty \frac{1}{2}\Big\langle \dot{\omega}(t), \alpha\big(\xi(t)\big)^\dagger \dot{\omega}(t)\Big\rangle \mathrm{d}t + \int_0^\infty \left(\dot{\eta}(t)\log\Big(\frac{\dot{\eta}(t)}{\lambda\big(\xi(t)\big)}\Big) - \dot{\eta}(t) + \lambda\big(\xi(t)\big)\right)\mathrm{d}t \\
&+ \int_0^\infty \dot{\eta}(t)\Lambda^*_{\kappa(\xi(t),\cdot)}\left(\frac{\dot{\gamma}(t) + \lambda\big(\xi(t)\big)\overline{\kappa(\xi(t),\cdot)}}{\dot{\eta}(t)}\right)\mathrm{d}t
\end{aligned}$$

In the evaluation above, we are insisting $0\log 0 = 0$, $0 \cdot \Lambda_{\kappa(\cdot,\cdot)}(\cdot) = 0$, and that $\hat{\xi}$ satisfies the following properties below, where statements involving t are taken Lebesgue-almost-everywhere; otherwise $I_{\hat{x}}(\hat{\xi}) = \infty$.

- *$\hat{\xi}(0) = \hat{x}$,*
- *$\xi \in \mathbb{A}([0,\infty), \mathbb{X})$, $\omega \in \mathbb{A}([0,\infty), \mathbb{V})$, $\gamma \in \mathbb{A}([0,\infty), \mathbb{V})$, and $\eta \in \mathbb{A}([0,\infty), \mathbb{R}_+)$,*
- *$\dot{\xi}(t) = \beta\big(\xi(t)\big) + \dot{\omega}(t) + \dot{\gamma}(t)$,*
- *$\dot{\omega}(t) \in \mathrm{image}\,\big(\alpha(\xi(t))\big)$,*
- *$\lambda\big(\xi(t)\big) > 0$.*
- *$\dot{\eta}(t) \geq 0$,*

Proof. Assumption 3.4.8 is another way of rewriting the system in Proposition 1.4.5, as in Remark 1.1.6. This makes $\hat{X}$ an affine process, so Lemma 3.4.5 and Proposition 3.4.6 tell us that $\hat{X}$ satisfies the assumptions of Section 2.2. Thus, the parameterization $\epsilon \hat{X}^\epsilon$ is a family described by Theorem 2.5.10, and so we have a large deviation principle for $(\hat{\mathrm{P}}^\epsilon_{\hat{x}})_{\epsilon>0}$. Again using Proposition 3.4.6, we see that the integrand evaluates the following function.

$$\hat{\Lambda}^*(\dot{\hat{x}}, \hat{x}) = \sup_{\hat{u} \in \hat{\mathbb{V}}} \left(\big\langle \hat{u}, \dot{\hat{x}}\big\rangle - \hat{\Lambda}\big(\hat{u}, \hat{x}\big)\right)$$

$$
\begin{aligned}
&= \sup_{\hat{u}\in\hat{\mathbb{V}}} \Bigg(\langle u_1, \dot{x}_1\rangle + \langle u_2, \dot{x}_2\rangle + \langle u_3, \dot{x}_3\rangle + u_4\dot{x}_4 - \langle u_1, \beta(x_1)\rangle \\
&\qquad - \frac{1}{2}\big\langle u_1 + u_2, \alpha(x_1)(u_1+u_2)\big\rangle - \lambda(x_1)\exp\Big(\Lambda_{\kappa(x_1,\cdot)}(u_1+u_3) + u_4\Big) \\
&\qquad + \lambda(x_1) + \big\langle u_1 + u_3, \lambda(x_1)\overline{\kappa(x_1,\cdot)}\big\rangle \Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&= \sup_{\hat{u}\in\hat{\mathbb{V}}} \Bigg(\big\langle u_1, \dot{x}_1 - \beta(x_1) - \dot{x}_2 - \dot{x}_3\big\rangle + \Big\langle u_1 + u_2, \dot{x}_2 - \frac{1}{2}\alpha(x_1)(u_1+u_2)\Big\rangle \\
&\qquad + \Big\langle u_1 + u_3, \dot{x}_3 + \lambda(x_1)\overline{\kappa(x_1,\cdot)}\Big\rangle + u_4\dot{x}_4 \\
&\qquad - \lambda(x_1)\exp\Big(\Lambda_{\kappa(x_1,\cdot)}(u_1+u_3) + u_4\Big) + \lambda(x_1)\Bigg)
\end{aligned} \tag{3.4.11}
$$

We now evaluate the expression depending on different scenarios.

1. Suppose we had $\dot{x}_1 \neq \beta(x_1) + \dot{x}_2 + \dot{x}_3$. Observe that the expression in (3.4.11) can have all expressions but the first summand canceled, if $u_1 = -u_2 = -u_3$ and $u_4 = 0$. This shows us the following.

$$
\begin{aligned}
\hat{\Lambda}^*(\hat{\dot{x}}, \hat{x}) &= \sup_{\hat{u}\in\hat{\mathbb{V}}} \Big(\big\langle \hat{u}, \hat{\dot{x}}\big\rangle - \hat{\Lambda}\big(\hat{u}, \hat{x}\big)\Big) \\
&\geq \sup_{\substack{\hat{u}=(u,-u,-u,0)\\ u=\rho(\dot{x}_1-\beta(x_1)-\dot{x}_2-\dot{x}_3)\\ \rho\in\mathbb{R}}} \Big(\big\langle \hat{u}, \hat{\dot{x}}\big\rangle - \hat{\Lambda}\big(\hat{u}, \hat{x}\big)\Big) \\
&= \sup_{\rho\in\mathbb{R}} \rho\Big|\dot{x}_1 - \beta(x_1) - \dot{x}_2 - \dot{x}_3\Big|^2 \\
&= \infty
\end{aligned}
$$

2. Now assume $\dot{x}_2 \notin \operatorname{image}\alpha(x_1)$. Observe that the expression in (3.4.11) can have all expressions other than the quadratic term canceled, if $u_1 = u_3 = u_4 = 0$. Denote $w_2, w_2^{\perp} \in \mathbb{V}$ the projections of $\dot{x}_2$ onto $\operatorname{image}(\alpha(x_1))$ and its orthogonal complement, respectively. Select $\tilde{w}_2 \in \mathbb{V}$ so that $w_2 = \frac{1}{2}\alpha(x_1)\tilde{w}_2$,

$$
\dot{x}_2 = w_2 + w_2^{\perp} = \frac{1}{2}\alpha(x_1)\tilde{w}_2 + w_2^{\perp},
$$

and observe the following.

$$
\begin{aligned}
\hat{\Lambda}^*(\hat{\dot{x}}, \hat{x}) &= \sup_{\hat{u}\in\hat{\mathbb{V}}} \Big(\big\langle \hat{u}, \hat{\dot{x}}\big\rangle - \hat{\Lambda}\big(\hat{u}, \hat{x}\big)\Big) \\
&\geq \sup_{\substack{\hat{u}=(0,u,0,0)\\ u=\rho w_2^{\perp}\\ \rho\in\mathbb{R}}} \Big(\big\langle \hat{u}, \hat{\dot{x}}\big\rangle - \hat{\Lambda}\big(\hat{u}, \hat{x}\big)\Big) \\
&= \sup_{\rho\in\mathbb{R}} \Big\langle \rho w_2^{\perp}, \dot{x}_2 - \frac{1}{2}\alpha(x_1)\cdot\rho w_2^{\perp}\Big\rangle \\
&= \sup_{\rho\in\mathbb{R}} \Bigg(\frac{1}{2}\Big\langle \rho w_2^{\perp}, \alpha(x_1)(\tilde{w}_2 - \rho w_2^{\perp})\Big\rangle + \Big\langle \rho w_2^{\perp}, w_2^{\perp}\Big\rangle \Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&= \sup_{\rho\in\mathbb{R}} \rho\left|w_2^\perp\right|^2 \\
&= \infty
\end{aligned}
$$

3. Now suppose $\lambda(x_1) = 0$. Then, (3.4.11) shows us that selecting $u_1 = u_2 = u_4 = 0$ will cancel all terms not involving $\dot{x}_3$.

$$
\begin{aligned}
\hat{\Lambda}^*(\hat{\dot{x}}, \hat{x}) = \sup_{\hat{u}\in\hat{\mathbb{V}}} \left(\left\langle \hat{u}, \hat{\dot{x}} \right\rangle - \hat{\Lambda}\left(\hat{u}, \hat{x}\right)\right) \geq \sup_{\substack{\hat{u}=(0,0,u,0)\\ u=\rho\dot{x}_3\\ \rho\in\mathbb{R}}} \left(\left\langle \hat{u}, \hat{\dot{x}} \right\rangle - \hat{\Lambda}\left(\hat{u}, \hat{x}\right)\right) &= \sup_{\rho\in\mathbb{R}} \rho\left|\dot{x}_3\right|^2 \\
&= \infty
\end{aligned}
$$

4. Now suppose $\lambda(x_1) < 0$. It is clear that selecting $u_1 = u_2 = u_3 = 0$ in (3.4.11) cancels many of the terms unrelated to $\dot{x}_4$, which lends us to the following argument.

$$
\begin{aligned}
\hat{\Lambda}^*(\hat{\dot{x}}, \hat{x}) = \sup_{\hat{u}\in\hat{\mathbb{V}}} \left(\left\langle \hat{u}, \hat{\dot{x}} \right\rangle - \hat{\Lambda}\left(\hat{u}, \hat{x}\right)\right) &\geq \sup_{\substack{\hat{u}=(0,0,0,\rho)\\ \rho\in\mathbb{R}}} \left(\left\langle \hat{u}, \hat{\dot{x}} \right\rangle - \hat{\Lambda}\left(\hat{u}, \hat{x}\right)\right) \\
&= \sup_{\rho\in\mathbb{R}} \left(\rho\dot{x}_4 + \lambda(x_1)(1-e^\rho)\right) \\
&= \lim_{\rho\to-\infty} \left(\rho\dot{x}_4 + \lambda(x_1)(1-e^\rho)\right) \\
&= \infty
\end{aligned}
$$

5. Now suppose $\lambda(x_1) > 0$ and $\dot{x}_4 < 0$. Again using the same selection as the preceding part, we have another infinite value.

$$
\begin{aligned}
\hat{\Lambda}^*(\hat{\dot{x}}, \hat{x}) = \sup_{\hat{u}\in\hat{\mathbb{V}}} \left(\left\langle \hat{u}, \hat{\dot{x}} \right\rangle - \hat{\Lambda}\left(\hat{u}, \hat{x}\right)\right) &\geq \sup_{\substack{\hat{u}=(0,0,0,\rho)\\ \rho\in\mathbb{R}}} \left(\left\langle \hat{u}, \hat{\dot{x}} \right\rangle - \hat{\Lambda}\left(\hat{u}, \hat{x}\right)\right) \\
&= \sup_{\rho\in\mathbb{R}} \left(\rho\dot{x}_4 + \lambda(x_1)(1-e^\rho)\right) \\
&= \lim_{\rho\to\infty} \left(\rho\dot{x}_4 + \lambda(x_1)(1-e^\rho)\right) \\
&= \infty
\end{aligned}
$$

6. Now suppose $\dot{x}_1 = \beta(x_1) + \dot{x}_2 + \dot{x}_3$, $\dot{x}_2 \in \text{image}(\alpha(x_1))$, $\lambda(x_1) > 0$, and $\dot{x}_4 > 0$. Using Lemma A.4.4, we are able to take the gradients of the expression in (3.4.11) to get the following system of equations.

$$
\begin{aligned}
0 &= \dot{x}_1 - \beta(x_1) - \alpha(x_1)(u_1+u_2) \\
&\qquad - \lambda(x_1)\exp\left(\Lambda_{\kappa(x_1,\cdot)}(u_1+u_3)+u_4\right)\nabla\Lambda_{\kappa(x_1,\cdot)}(u_1+u_3) + \lambda(x_1)\overline{\kappa(x_1,\cdot)} \\
0 &= \dot{x}_2 - \alpha(x_1)(u_1+u_2) \\
0 &= \dot{x}_3 - \lambda(x_1)\exp\left(\Lambda_{\kappa(x_1,\cdot)}(u_1+u_3)+u_4\right)\nabla\Lambda_{\kappa(x_1,\cdot)}(u_1+u_3) + \lambda(x_1)\overline{\kappa(x_1,\cdot)}
\end{aligned}
$$

$$0 = \dot{x}_4 - \lambda(x_1)\exp\Big(\Lambda_{\kappa(x_1,\cdot)}(u_1+u_3)+u_4\Big)$$

These equations yield the following via simple algebra.

$$\begin{aligned} u_1+u_2 &= \alpha(x_1)^\dagger \dot{x}_2 \\ \dot{x}_4 &= \lambda(x_1)\exp\Big(\Lambda_{\kappa(x_1,\cdot)}(u_1+u_3)+u_4\Big) \\ \dot{x}_4\log\Big(\frac{\dot{x}_4}{\lambda(x_1)}\Big) &= \dot{x}_4\Lambda_{\kappa(x_1,\cdot)}(u_1+u_3)+u_4\dot{x}_4 \\ \frac{\dot{x}_3+\lambda(x_1)\overline{\kappa(x_1,\cdot)}}{\dot{x}_4} &= \nabla\Lambda_{\kappa(x_1,\cdot)}(u_1+u_3) \end{aligned} \tag{3.4.12}$$

Seeing as $\Lambda_{\kappa(x_1,\cdot)}$ is a convex function, the last equality of (3.4.12) corresponds to the unique extreme point of the Fenchel-Legendre transform.

$$\Big\langle u_1+u_3, \frac{\dot{x}_3+\lambda(x_1)\overline{\kappa(x_1,\cdot)}}{\dot{x}_4}\Big\rangle - \Lambda_{\kappa(x_1,\cdot)}(u_1+u_3) = \Lambda^*_{\kappa(x_1,\cdot)}\Big(\frac{\dot{x}_3+\lambda(x_1)\overline{\kappa(x_1,\cdot)}}{\dot{x}_4}\Big)$$

Combining this equality with those of (3.4.12) now gives us the following identity.

$$\begin{aligned} \big\langle \hat{u},\hat{\dot{x}}\big\rangle - \hat{\Lambda}\big(\hat{u},\hat{x}\big) &= \langle u_1,\dot{x}_1\rangle + \langle u_2,\dot{x}_2\rangle + \langle u_3,\dot{x}_3\rangle + u_4\dot{x}_4 - \langle u_1,\beta(x_1)\rangle \\ &\quad - \frac{1}{2}\big\langle u_1+u_2, \alpha(x)(u_1+u_2)\big\rangle \\ &\quad - \lambda(x_1)\exp\Big(\Lambda_{\kappa(x_1,\cdot)}(u_1+u_3)+u_4\Big) \\ &\quad + \lambda(x_1) + \big\langle u_1+u_3, \lambda(x_1)\overline{\kappa(x_1,\cdot)}\big\rangle \\ &= \frac{1}{2}\Big\langle \dot{x}_2, \alpha(x_1)^\dagger\dot{x}_2\Big\rangle + \dot{x}_4\log\Big(\frac{\dot{x}_4}{\lambda(x_1)}\Big) - \dot{x}_4 + \lambda(x_1) \\ &\quad + \Lambda^*_{\kappa(x_1,\cdot)}\Big(\frac{\dot{x}_3+\lambda(x_1)\overline{\kappa(x_1,\cdot)}}{\dot{x}_4}\Big) \end{aligned}$$

Observe that since $\hat{\Lambda}(\cdot,\hat{x})$ convex, the critical point we have solved is a global extremum, evaluating $\hat{\Lambda}^*(\hat{\dot{x}},\hat{x})$.

7. The last case we must consider is when $\dot{x}_1 = \beta(x_1)+\dot{x}_2+\dot{x}_3$, $\dot{x}_2 \in \mathrm{image}(\alpha(x_1))$, $\lambda(x_1) > 0$, and $\dot{x}_4 = 0$. For this, we start by greedily optimizing in the u_4 coordinate in (3.4.11).

$$\begin{aligned} \hat{\Lambda}^*(\hat{\dot{x}},\hat{x}) &= \sup_{\hat{u}\in\hat{\mathbb{V}}}\Big(\big\langle\hat{u},\hat{\dot{x}}\big\rangle - \hat{\Lambda}\big(\hat{u},\hat{x}\big)\Big) \\ &= \sup_{u_1,u_2,u_3\in\mathbb{V}}\Big[\big\langle u_1, \dot{x}_1-\beta(x_1)-\dot{x}_2-\dot{x}_3\big\rangle + \Big\langle u_1+u_2, \dot{x}_2 - \frac{1}{2}\alpha(x_1)(u_1+u_2)\Big\rangle \\ &\qquad + \Big\langle u_1+u_3, \dot{x}_3+\lambda(x_1)\overline{\kappa(x_1,\cdot)}\Big\rangle + \lambda(x_1) \\ &\qquad + \sup_{u_4\in\mathbb{R}}\Big(-\lambda(x_1)\exp\Big(\Lambda_{\kappa(x_1,\cdot)}(u_1+u_3)+u_4\Big)\Big)\Big] \\ &= \sup_{u_1,u_2,u_3\in\mathbb{V}}\Big[\big\langle u_1, \dot{x}_1-\beta(x_1)-\dot{x}_2-\dot{x}_3\big\rangle + \Big\langle u_1+u_2, \dot{x}_2 - \frac{1}{2}\alpha(x_1)(u_1+u_2)\Big\rangle \\ &\qquad + \Big\langle u_1+u_3, \dot{x}_3+\lambda(x_1)\overline{\kappa(x_1,\cdot)}\Big\rangle + \lambda(x_1)\Big] \end{aligned}$$

From here, we are left with yet another convex function we intend to optimize; observe that the critical point now easily evaluates our expression.

$$\begin{aligned}
0 &= \dot{x}_1 - \beta(x_1) - \alpha(x_1)(u_1 + u_2) + \lambda(x_1)\overline{\kappa(x_1, \cdot)} \\
0 &= \dot{x}_2 - \alpha(x_1)(u_1 + u_2) \\
0 &= \dot{x}_3 + \lambda(x_1)\overline{\kappa(x_1, \cdot)} \\
&\implies \quad \langle \hat{u}, \hat{\dot{x}} \rangle - \hat{\Lambda}(\hat{u}, \hat{x}) = \frac{1}{2}\big\langle \dot{x}_2, \alpha(x_1)^{\dagger}\dot{x}_2 \big\rangle + \lambda(x_1)
\end{aligned}$$

Thus, this expression solves $\Lambda^*(\hat{\dot{x}}, \hat{x})$. Note that this expression is identical to the previous case, when taking convention that $0 \log 0 = 0$ and $0 \cdot \Lambda_{\kappa(\cdot,\cdot)}(\cdot) = 0$

These cases simplify the nature of Λ^* in (2.5.11), finishing the proof.

3.5 Final discussion

1. Discuss how contraction maps or an alternative proof can deal with the **overdetermined** nature, at the expense of a nastier rate function.
   (a) **determine-continuous-noise flavor.** Normal term gets messy
   (b) **determine-arrivals flavor.** Poisson and jump-term-denominator gets messy
   (c) **determine-jumps flavor.** This is the one we have already presented; the jump-term-numerator gets messy.
2. Discuss how the deviations of X from the dynamical system $X = \beta(X) \bullet \ell$ are imposed from *continuous deviations* X^{c} and *discontinuous deviations* $\mathrm{id}_{\mathbb{V}} * \tilde{q}^X$.
3. Discuss how four quantities $X, X^{\mathrm{c}}, N^X, V^X$ heuristically relate in simple infinitesimal equality $(X, X^{\mathrm{c}}, N^X, V^X) \approx (\xi, \omega, \eta, \gamma)$.
$$\dot{\xi}(t) = \beta\big(\xi(t)\big) + \dot{\omega}(t) + \dot{\eta}(t) \cdot \dot{\gamma}(t)$$
4. Each of the primitive deviations have a simple analogy, when we think of infinitesimals.
$$\begin{aligned}
&\dot{\omega}(t) && \textit{normal deviations of covariance } \alpha\big(\xi(t)\big) \\
&\dot{\eta}(t) && \textit{Poisson deviations of rate } \lambda\big(\xi(t)\big) \\
&\dot{\gamma}(t) && \textit{jump deviations from } \kappa\big(\xi(t), \mathrm{d}v\big) \\
\xi(t) = \beta\big(\xi(t)\big) + &\dot{\omega}(t) + \dot{\eta}(t) \cdot \dot{\gamma}(t) && \textit{all combined deviations}
\end{aligned}$$
5. Think of results from first section in this regard.
   (a) birth is $\dot{\xi}(t) = \dot{\eta}(t)$, so we only need $\xi \approx X$.
   (b) diffusion is $\dot{\xi}(t) = \beta\big(\xi(t)\big) + \dot{\omega}(t)$, and so we only need $\xi \approx X$ and rate function includes $\xi(t) - \beta\big(\xi(t)\big)$ where $\dot{\omega}$ is.
   (c) compound Poisson is $\dot{\xi}(t) = \dot{\eta}(t) \cdot \dot{\gamma}(t)$, so we choose one of the following pairs $(\xi, \eta) \approx (X, N^X)$, $(\xi, \gamma) \approx (X, V^X)$, or $(\eta, \gamma) \approx (N^X, V^X)$.
   (d) compound linear Hawkes is $\dot{\xi}(t) = \beta\big(\xi(t)\big) + \dot{\eta}(t) \cdot \dot{\gamma}(t)$, so we can choose $(\xi, \eta) \approx (X, N^X)$ or $(\xi, \dot{\gamma}) \approx (X, V^X)$.

Appendix A

Jump-diffusions

In order to discuss jump-diffusions on a finite-dimensional real vector space, one must have a decent understanding of semimartingales. A great text for a comprehensive study of this is [JS03], which we will refer to in our proofs. In terms of notational differences, we choose our probability space $(\Omega, \Sigma, \mathrm{P})$ and filtration $\mathscr{F} = (\mathscr{F}_t)_{t\geq 0}$, where $\mathscr{F}_\infty \subseteq \Sigma$ denotes the joined space. Furthermore, we do not explicitly write processes to take values in $\mathbb{R}^d$, but rather some vector space $\mathbb{V}$ with dimension $d := \dim \mathbb{V}$ and inner-product $\langle \cdot, \cdot \rangle$. Surely—due to our isometric isomorphism $\mathbb{V} \equiv \mathbb{R}^d$—any componentwise or linear notion, such as integration or differentiation may be taken as equivalent. Furthermore, we sometimes specify that a stochastic process X has a Borel state space $\mathbb{X} \subseteq \mathbb{V}$, as this is the case when studying affine processes. We find it important to highlight the following important notation of objects introduced in [JS03, Chapters I-II].

- Given $(\mathrm{P}, \mathscr{F})$ locally square-integrable martingales $M, N : \Omega \times \mathbb{R}_+ \to \mathbb{R}$, denote $\langle M, N \rangle$ the predictable quadratic covariation.

- Given $H, X : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ with H being $\mathscr{F}$ predictable and $(\mathrm{P}, \mathscr{F})$ locally bounded and X a $(\mathrm{P}, \mathscr{F})$ semimartingale, denote the stochastic integral process as follows.

$$H \bullet X_t = \int_0^t H_s \mathrm{d}X_s$$

  We may lift this concept componentwise and linearly. This allows us to choose the codomains of H, X to various combinations of $\mathbb{V}$ and $\mathbb{L}(\mathbb{V}, \mathbb{W})$ when evaluating $H \bullet X$, so long as such a combination allows for $H_t \cdot X_t$ to make sense.

- Denote $\ell : \mathbb{R}_+ \to \mathbb{R}_+$ the identity map to allow a concise notation for Lebesgue integration.

$$H \bullet \ell_t = \int_0^t H_s \mathrm{d}s$$

  Throughout, we will often use the following fact without mention; any càdlàg process H is such that $H_t = H_{t-}$ at all but a countable amount of times $t \in \mathbb{R}$, and so

$$H_- \bullet \ell_t = \int_0^t H_{s-} \mathrm{d}s = \int_0^t H_s \mathrm{d}s = H \bullet \ell_t$$

- Given a random measure $q : \Omega \times \mathscr{B}(\mathbb{R}_+ \times \mathbb{V}) \to [0, \infty]$, denote the stochastic integral process against some suitably integrable process $H : \Omega \times \mathbb{R}_+ \times \mathbb{V} \to \mathbb{R}$ as follows.

$$H * q_t = \int_{[0,t]\times\mathbb{V}} H_s(v) q(\mathrm{d}s, \mathrm{d}v)$$

  Denote its $(\mathrm{P}, \mathscr{F})$ predictable projection by $\hat{q}$ and the compensated measure $\tilde{q} = q - \hat{q}$. We will frequently use, without mention, that $\hat{q}$ is characterized by the property that any $\mathscr{F}$ predictable $H : \Omega \times \mathbb{R}_+ \times \mathbb{V} \to \mathbb{R}_+$ is such that

$$\mathrm{E}_{\mathrm{P}}\big(H * q_\infty\big) = \mathrm{E}_{\mathrm{P}}\big(H * \hat{q}_\infty\big)$$

  Also denote $H * \tilde{q}$ the compensated local martingale process for suitable $H \in G_{\mathrm{loc}}(q)$, as constructed in [JS03, Definition II.1.27]. Lift these integration notions to vector-valued H componentwise. Instead of choosing a canonical variable for integrating expressions in this form, we use the identity maps $\mathrm{id}_{\mathbb{V}}$ or ℓ.

$$f(\ell, \mathrm{id}_{\mathbb{V}}) * q_t = \int_{[0,t]\times\mathbb{V}} f(s, v) q(\mathrm{d}s, \mathrm{d}v)$$

- Given $(\mathrm{P}, \mathscr{F})$ semimartingales $X, Y : \Omega \times \mathbb{R}_+ \to \mathbb{R}$, denote $[X, Y]$ the quadratic covariation.

- Given a semimartingale X, denote X^{c} its continuous local martingale component and q^X its jump measure.

A.1 Formulation

As in [JS03, Definition III.2.18], a $(\mathrm{P}, \mathscr{F})$ jump-diffusion X on state space $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ is a $(\mathrm{P}, \mathscr{F})$ semimartingale in which the χ-characteristics $(B^\chi, A, \hat{q}^X)$ have the following decompositions.

$$B_t^\chi = \int_0^t \beta^\chi(X_s)\mathrm{d}s, \quad A_t = \int_0^t \alpha(X_s)\mathrm{d}s, \quad \hat{q}^X(\mathrm{d}s, \mathrm{d}v) = \mu(X_s, \mathrm{d}v)\mathrm{d}s, \tag{A.1.1}$$

where the functions have the following properties.

- $\beta^\chi : \mathbb{X} \to \mathbb{V}$ is Borel measurable, $\beta^\chi \in \mathscr{B}(\mathbb{X})/\mathscr{B}(\mathbb{V})$.

- $\alpha : \mathbb{X} \to \mathbb{L}(\mathbb{V})$ is Borel measurable, $\alpha \in \mathscr{B}(\mathbb{X})/\mathscr{B}(\mathbb{L}(\mathbb{V}))$, and $\alpha(x)$ is self-adjoint and non-negative for each $x \in \mathbb{X}$.

- $\mu : \mathbb{X} \times \mathscr{B}(\mathbb{V}) \to [0, \infty]$ is a transition kernel from $\mathbb{X}$ to $\mathbb{V}$, and it satisfies the following properties for each $x \in \mathbb{X}$.

$$\mu\big(x, \{0\}\big) = 0, \quad \int_{\mathbb{V}} \big(1 \wedge |v|^2\big)\mu(x, \mathrm{d}v) < \infty \tag{A.1.2}$$

In other words, our jump-diffusion X has the following canonical semimartingale representation (see [JS03, Theorem II.2.34] for definition).

$$\begin{gathered} X = X_0 + \beta^\chi(X) \boldsymbol{\cdot} \ell + X^{\mathrm{c}} + \chi * \tilde{q}^X + (\mathrm{id}_{\mathbb{V}} - \chi) * q^X \\ \langle X^{\mathrm{c},i}, X^{\mathrm{c},j} \rangle = \alpha_{ij}(X) \boldsymbol{\cdot} \ell \\ \hat{q}^X(\mathrm{d}s, \mathrm{d}v) = \mu(X_s, \mathrm{d}v)\mathrm{d}s \end{gathered} \tag{A.1.3}$$

Remark A.1.4. *(a) Note that we differ slightly from the definition we reference by imposing a time-homogeneity formulation. There is no loss of generality in doing so, because we may always extend the state to* $\mathbb{R}_+ \times \mathbb{X}$ *via* $\hat{X}_t = (t, X_t)$.

(b) Note that (A.1.1) can be written concisely by using the identity ℓ *on* $\mathbb{R}_+$.

$$B^\chi_t = \beta^\chi(X) \bullet \ell_t, \quad A_t = \alpha(X) \bullet \ell_t, \quad \hat{q}^X([0,t], \mathrm{d}v) = \mu(X, \mathrm{d}v) \bullet \ell_t$$

(c) If we have a jump-diffusion with χ*-characteristics in (A.1.1), we call* $(\beta^\chi, \alpha, \mu)$ *the* differential χ-characteristics. *We see from (A.1.3) that* β^χ *and* $\beta^{\hat{\chi}}$ *relate between different truncation functions* $\chi, \hat{\chi}$ *with the simple identity.*

$$\beta^{\hat{\chi}}(x) = \beta^\chi(x) + \int_{\mathbb{V}} \big(\hat{\chi}(v) - \chi(v)\big)\mu(x, \mathrm{d}v) \tag{A.1.5}$$

(d) The conditions on $\alpha(x)$ *and* $\mu(x, \mathrm{d}v)$ *are immediate consequences of (A.1.1). For the most general setting, see the corresponding result for* any *semimartingale, in [JS03, Proposition II.2.9].*

Example A.1.6. *Fix a probability space* $(\Omega, \Sigma, \mathrm{P})$ *and filtration* $\mathscr{F} = (\mathscr{F})_{t\geq 0}$.

Just as with $(\mathbb{R}^d, \mathscr{B}(\mathbb{V}))$*, we say that* W *is a standard* $(\mathrm{P}, \mathscr{F})$ *Brownian motion on* $(\mathbb{V}, \mathscr{B}(\mathbb{V}))$ *if it is a continuous* $(\mathrm{P}, \mathscr{F})$ *martingale with predictable quadratic covariation as follows.*

$$\langle W^i, W^j \rangle_t = \begin{cases} t & i = j \\ 0 & \textit{otherwise} \end{cases}$$

It is clear that W *is a* $(\mathrm{P}, \mathscr{F})$ *jump-diffusion with differential* χ*-characteristics* $(0, \alpha, 0)$*, where* $\alpha(x) = \mathrm{id}_{\mathbb{V}}$ *for all* $x \in \mathbb{X}$.

Similarly, we say that p *is a standard* $(\mathrm{P}, \mathscr{F})$ *Poisson random measure on* $\mathscr{B}(\mathbb{R}_+ \times \mathbb{V})$ *if its* $(\mathrm{P}, \mathscr{F})$ *predictable projection is the Lebesgue measure* $\hat{p}(\mathrm{d}s, \mathrm{d}v) = \mathrm{d}s \otimes \mathrm{d}v$ *(identifying measures on* $\mathscr{B}(\mathbb{R}^d)$ *as those on* $\mathscr{B}(\mathbb{V})$*). By [JS03, Theorem II.4.8], this* p *is the same as a Poisson point process with Lebesgue intensity, which has infinitely many jumps on any nonempty interval of time. The accumulated jumps* $\mathrm{id}_{\mathbb{V}} * p$ *form a* $(\mathrm{P}, \mathscr{F})$ *jump-diffusion with parameters as follows.*

$$\beta^\chi(x) = \int_{\mathbb{V}} \chi(v)\mathrm{d}v, \quad \alpha(x) = 0, \quad \mu(x, \mathrm{d}v) = \mathrm{d}v,$$

because we have the following decomposition.

$$\begin{aligned} \mathrm{id}_{\mathbb{V}} * p &= \chi * p + (\mathrm{id}_{\mathbb{V}} - \chi) * p \\ &= \chi * \hat{p} + \chi * \tilde{p} + (\mathrm{id}_{\mathbb{V}} - \chi) * p \\ &= \beta^\chi \bullet \ell + \chi * \tilde{p} + (\mathrm{id}_{\mathbb{V}} - \chi) * p \end{aligned}$$

Note that the infinite activity of p *means that the last term cannot be compensated.*

We will see at the end of this section that these two objects W *and* p *are the fundamental building blocks of all jump-diffusions.*

The following Lemma will be repeatedly used as a shortcut of Itô's formula and various identities that always apply with jump-diffusions.

Lemma A.1.7. *Let X be a jump-diffusion with differential χ-characteristics $(\beta^{\chi}, \alpha, \mu)$ and $f \in \mathbb{C}^2(\mathbb{V}, \mathbb{R})$. The composition $f(X)$ has the following semimartingale representation.*

$$\begin{aligned} f(X_t) = f(X_0) &+ \Big(\mathrm{D}f(X) \cdot \beta^{\chi}(X)\Big) \bullet \ell_t + \frac{1}{2} \operatorname{tr} \Big(\mathrm{D}^2 f(X) \circ \alpha(X)\Big) \bullet \ell_t + \mathrm{D}f(X_-) \bullet X^{\mathrm{c}} \\ &+ \Big(\mathrm{D}f(X_-) \cdot \chi\Big) * \tilde{q}_t^X + \Big(f(X_- + \mathrm{id}_{\mathbb{V}}) - f(X_-) - \mathrm{D}f(X_-) \cdot \chi\Big) * q_t^X \end{aligned}$$

Proof. Apply Itô's formula [JS03, Theorem I.4.57] and use the predictable covariation identity in (A.1.3) to get the following.

$$\begin{aligned} f(X_t) &= f(X_0) + \sum_{i=1}^{d} \mathrm{D}_i f(X_-) \bullet X_t^i + \frac{1}{2} \sum_{i,j=1}^{d} \mathrm{D}_{ij} f(X_-) \bullet \langle X^{\mathrm{c},i}, X^{\mathrm{c},j} \rangle_t \\ &\quad + \sum_{0<s\le t} \Big(f(X_s) - f(X_{s-}) - \sum_{i=1}^{d} \mathrm{D}f_i(X_{s-}) \Delta X_s\Big) \\ &= f(X_0) + \mathrm{D}f(X_-) \bullet X_t + \frac{1}{2} \sum_{i,j=1}^{d} \mathrm{D}_{ij} f(X_-) \bullet \big(\alpha_{ij}(X) \bullet \ell\big)_t \\ &\quad + \Big(f(X_- + \mathrm{id}_{\mathbb{V}}) - f(X_-) - \mathrm{D}f(X_-) \cdot \mathrm{id}_{\mathbb{V}}\Big) * q_t^X \end{aligned}$$

Using the iterated stochastic integral formula [JS03, Remark I.4.37], we may simplify the above equation to the following.

$$\begin{aligned} f(X_t) = f(X_0) &+ \mathrm{D}f(X_-) \bullet X_t + \frac{1}{2} \operatorname{tr} \Big(\mathrm{D}_{ij} f(X_-) \circ \alpha(X)\Big) \bullet \ell_t \\ &+ \Big(f(X_- + \mathrm{id}_{\mathbb{V}}) - f(X_-) - \mathrm{D}f(X_-) \cdot \mathrm{id}_{\mathbb{V}}\Big) * q_t^X \end{aligned}$$

Now substitute our representation of (A.1.3) and repeat the iterated stochastic integral to get the following.

$$\begin{aligned} f(X_t) &= f(X_0) + \mathrm{D}f(X_-) \bullet \big(X_0 + \beta^{\chi}(X) \bullet \ell + X^{\mathrm{c}} + \chi * \tilde{q}^X + (\mathrm{id}_{\mathbb{V}} - \chi) * q^X\big)_t \\ &\quad + \frac{1}{2} \operatorname{tr} \Big(\mathrm{D}^2 f(X_-) \circ \alpha(X)\Big) \bullet \ell_t + \Big(f(X_- + \mathrm{id}_{\mathbb{V}}) - f(X_-) - \mathrm{D}f(X_-) \cdot \mathrm{id}_{\mathbb{V}}\Big) * q_t^X \\ &= f(X_0) + \Big(\mathrm{D}f(X_-) \cdot \beta^{\chi}(X)\Big) \bullet \ell_t + \frac{1}{2} \operatorname{tr} \Big(\mathrm{D}^2 f(X_-) \circ \alpha(X)\Big) \bullet \ell_t + \mathrm{D}f(X_-) \bullet X^{\mathrm{c}} \\ &\quad + \mathrm{D}f(X_-) \bullet \big(\chi * \tilde{q}^X\big)_t + \Big(f(X_- + \mathrm{id}_{\mathbb{V}}) - f(X_-) - \mathrm{D}f(X_-) \cdot \chi\Big) * q_t^X \end{aligned}$$

Furthermore, since $X_- = X$ on all but a countable amount of jumps, we may rewrite the Lebesgue integrals.

(A.1.8)
$$\begin{aligned} f(X_t) = f(X_0) &+ \Big(\mathrm{D}f(X) \cdot \beta^{\chi}(X)\Big) \bullet \ell_t + \frac{1}{2} \operatorname{tr} \Big(\mathrm{D}^2 f(X) \circ \alpha(X)\Big) \bullet \ell_t + \mathrm{D}f(X_-) \bullet X^{\mathrm{c}} \\ &+ \mathrm{D}f(X_-) \bullet \big(\chi * \tilde{q}^X\big)_t + \Big(f(X_- + \mathrm{id}_{\mathbb{V}}) - f(X_-) - \mathrm{D}f(X_-) \cdot \chi\Big) * q_t^X \end{aligned}$$

For the remaining equality, we construct localizing sequence $(T_n)_{n\in\mathbb{N}}$ of $\mathscr{F}$ stopping times,

(A.1.9)
$$T_n(\omega) \coloneqq \inf \big\{t > 0 : X_t(\omega) > n\big\} \wedge n, \quad \omega \in \Omega,\ n \in \mathbb{N},$$

to see that $\mathrm{D}f(X_-)$ is $(\mathrm{P},\mathscr{F})$ locally bounded.

$$\left|\mathrm{D}f(X^{T_n}_{s-})\right| \leq \sup_{|x|\leq n}\left|\mathrm{D}f(x)\right|$$

Thus, by [JS03, Proposition II.1.30], we may rewrite the following.

$$\mathrm{D}f(X_-) \bullet \big(\chi * \tilde{q}^X\big)_t = \big(\mathrm{D}f(X_-)\cdot\chi\big) * \tilde{q}^X_t,$$

which when substituted into (A.1.8) gives us our desired identity.

In the above lemma, the final term in the semimartingale decomposition of $f(X)$ is typically not able to be compensated into a local martingale. If we did have local integrability of the following quantity,

$$\Big|f(X_- + \mathrm{id}_{\mathbb{V}}) - f(X_-) + \mathrm{D}f(X_-)\cdot\chi\Big| * \hat{q}^X,$$

then by [JS03, Proposition II.1.28] we could rewrite $f(X)$ into a canonical special semimartingale decomposition.

$$\begin{aligned} f(X_t) &= f(X_0) + \mathcal{L}f(X)\bullet\ell_t + \mathrm{D}f(X_-)\bullet X^{\mathrm{c}} + \big(f(X_- + \mathrm{id}_{\mathbb{V}}) - f(X_-)\big) * \tilde{q}^X_t \\ \mathcal{L}f(x) &\coloneqq \mathrm{D}f(x)\cdot\beta^\chi(x) + \frac{1}{2}\operatorname{tr}\Big(\mathrm{D}^2 f(x)\circ\alpha(x)\Big) \\ &\qquad + \int_{\mathbb{V}}\Big(f(x+v) - f(x) - \mathrm{D}f(x)\cdot\chi(v)\Big)\mu(x,\mathrm{d}v) \end{aligned} \tag{A.1.10}$$

So long as f is bounded, we can guarantee this special semimartingale property.

Proposition A.1.11. *Let X and f as in Lemma A.1.7, and further impose f is bounded. Then the composition $f(X)$ is a special semimartingale with the decomposition as in (A.1.10).*

Proof. Seeing as f is bounded, [JS03, Lemma I.4.24] tells us that $f(X)$ is a special semimartingale. By [JS03, Proposition I.4.23], it is then the case that the following term is locally integrable.

$$\Big(f(X_- + \mathrm{id}_{\mathbb{V}}) - f(X_-) - \mathrm{D}f(X_-)\cdot\chi\Big) * q^X_t$$

By our discussion above, this suffices to conclude (A.1.10).

This operator $\mathcal{L}$ in (A.1.10) gives a nice closed from for suitable $f(X)$, and so we reserve it the term of *generator* associated with X. Note that we do not mark dependence on χ, as any other truncation function $\hat{\chi}$ will produce the same operator; see Remark A.1.4(c) and note that the displacement from β^χ and $\beta^{\hat{\chi}}$ would be the same as that in the integral term. One particular setting in which this result is useful is establishing a Lévy-Khintchine formula for jump-diffusions.

Proposition A.1.12. *Fix a jump-diffusion X with differential χ-characteristics (β^χ,α,μ). Then, for each $u\in\mathrm{i}\mathbb{V}$, the process* $\exp\big(\langle u,X\rangle - \Lambda(u,X)\bullet\ell\big)$ *is a complex-valued* $(\mathrm{P},\mathscr{F})$ *local martingale, where $\Lambda : \mathrm{i}\mathbb{V}\times\mathbb{X}\to\mathbb{R}$ is the associated* Lévy-Khintchine map.

$$\Lambda(u,x) = \big\langle u,\beta^\chi(x)\big\rangle + \frac{1}{2}\big\langle u,\alpha(x)\big\rangle + \int_{\mathbb{V}}\big(e^{\langle u,v\rangle} - 1 - \langle u,\chi(v)\rangle\big)\mu(x,\mathrm{d}v),$$

Proof. For a fixed $u \in \mathrm{i}\mathbb{V}$, note that the map f_u, defined by $f_u(v) = \exp\langle u, v\rangle$ is bounded. Thus, by Proposition A.1.11, we have

$$f_u(X_t) = f_u(X_0) + \mathcal{L}f_u(X) \bullet \ell_t + M_t,$$

where M is a $(\mathrm{P}, \mathscr{F})$ local martingale. Observe that the partial derivatives of f are as follows,

$$\mathrm{D}_i f_u(x) = f_u(x) u_i, \qquad \mathrm{D}_{ij} f_u(x) = f_u(x) u_i u_j, \tag{A.1.13}$$

so we have the following equation.

$$\begin{aligned}
\mathcal{L}f_u(x) &= \mathrm{D}f_u(x) \cdot \beta^\chi(x) + \frac{1}{2}\operatorname{tr}\Big(\mathrm{D}^2 f_u(x) \circ \alpha(x)\Big) \\
&\qquad + \int_{\mathbb{V}} \Big(f_u(x+v) - f_u(x) - \mathrm{D}f(x)\cdot\chi(v)\Big)\mu(x, \mathrm{d}v) \\
&= f_u(x)\big\langle u, \beta^\chi(x)\big\rangle + \frac{1}{2} f_u(x)\big\langle u, \alpha(x) u\big\rangle + f_u(x)\int_{\mathbb{V}} \Big(f_u(v) - 1 - \big\langle u, \chi(v)\big\rangle\Big)\mu(x, \mathrm{d}v) \\
&= f_u(x) \cdot \Lambda(u, x)
\end{aligned}$$

Denoting $A = f_u(X) = \exp\langle u, X\rangle$ and $B = \exp\big(-\Lambda(u, X)\bullet\ell\big)$, we now use the fact that B is $\mathscr{F}$ predictable and of finite-variation, so [JS03, Proposition I.4.49(b)] gives us the following.

$$\begin{aligned}
&\exp\Big(\langle u, X\rangle - \Lambda(u, X)\bullet\ell\Big) \\
&= A_t B_t \\
&= A_0 B_0 + A_- \bullet B_t + B \bullet A_t \\
&= \exp\langle u, X_0\rangle + A_- \bullet \Big(\big(-B\cdot\Lambda(u, X)\big)\bullet\ell\Big)_t + B \bullet \Big(f_u(X_0) + \mathcal{L}f_u(X)\bullet\ell + M\Big)_t \\
&= \exp\langle u, X_0\rangle - \Big(A \cdot B \cdot \Lambda(u, X)\Big)\bullet\ell_t + \Big(B\cdot f_u(X)\cdot\Lambda(u, X)\Big)\bullet\ell_t + B\bullet M_t \\
&= \exp\langle u, X_0\rangle + B \bullet M_t
\end{aligned}$$

This identity and [JS03, Remark I.4.34(b)] concludes the proof.

It turns out that each of the preceding results is sufficient in characterizing a semimartingale X as a jump-diffusion.

Theorem A.1.14. *The following statements are equivalent for a stochastic process X on state space $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$.*

(a) X is a $(\mathrm{P}, \mathscr{F})$ jump-diffusion with differential χ-characteristics $(\beta^\chi, \alpha, \mu)$.

(b) For each bounded $f \in \mathbb{C}^2(\mathbb{V}, \mathbb{R})$, the process $f(X_t) - \mathcal{L}f(X_t)\bullet\ell_t$ is a $(\mathrm{P}, \mathscr{F})$ local martingale, where

$$\mathcal{L}f(x) \coloneqq \mathrm{D}f(x)\cdot\beta^\chi(x) + \frac{1}{2}\operatorname{tr}\Big(\mathrm{D}^2 f(x)\circ\alpha(x)\Big) + \int_{\mathbb{V}}\Big(f(x+v) - f(x) - \mathrm{D}f(x)\cdot\chi(v)\Big)\mu(x, \mathrm{d}v)$$

(c) For each $u \in \mathrm{i}\mathbb{V}$, the process $\exp\Big(\langle u, X\rangle - \Lambda(u, X)\bullet\ell\Big)$ is a $(\mathrm{P}, \mathscr{F})$ local martingale, where Λ is our Lévy-Khintchine map.

$$\Lambda(u, x) = \big\langle u, \beta^\chi(x)\big\rangle + \frac{1}{2}\big\langle u, \alpha(x)\big\rangle + \int_{\mathbb{V}}\big(e^{\langle u, v\rangle} - 1 - \langle u, \chi(v)\rangle\big)\mu(x, \mathrm{d}v),$$

(d) *Denoting* $(\mathrm{P}_x)_{x\in\mathbb{X}}$ *the* P*-conditional distributions of* X *factored through the initial state* X_0 *and selecting Borel functions* σ, c *to satisfy,*

$$\begin{aligned} &\sigma : \mathbb{X} \to \mathbb{L}(\mathbb{V}) \quad \sigma\sigma^*(x) = \alpha(x) \\ &c : \mathbb{X}\times\mathbb{V} \to \mathbb{V} \quad \mu(x,\Gamma) = \int_{\mathbb{V}} 1_\Gamma\big(c(x,v)\big)\mathrm{d}v \end{aligned} \tag{A.1.15}$$

each P_x *is a solution to the equation associated with a standard Brownian motion* W *and Poisson random measure* p*, where* $\chi' = \mathrm{id}_{\mathbb{V}} - \chi$*.*

$$X_t = x + \beta^\chi(X)\bullet \ell_t + \sigma(X_-)\bullet W_t + \big(\chi\circ c(X_-,\mathrm{id}_{\mathbb{V}})\big) * \tilde{p}_t + \big(\chi'\circ c(X_-,\mathrm{id}_{\mathbb{V}})\big) * p_t$$

Proof. This is simply restating [JS03, Theorems II.2.42, II.2.49, and III.2.26] in terms of our identities from the previous propositions and lemmas. The choice of standard intensity $\mathrm{d}t\otimes\mathrm{d}v$ for the Poisson random measure is such that the jump factor $\mathrm{d}v$ satisfies the atom-free and infinite properties in [JS03, Remark III.2.28(3)].

Remark A.1.16. *In the final part above, the push-forward map* c *may put mass on* 0,

$$\int_{\mathbb{V}} 1_{\{0\}}\big(c(x,v)\big)\mathrm{d}v > 0,$$

to thin *or* delete *jumps coming from* p *(of which there are infinitely many). However, this contradicts the condition (A.1.2) that* $\mu(x,\{0\}) = 0$ *for all* $x\in\mathbb{X}$*. Explicitly, the push-forward in (A.1.15) happens on the space* $\mathbb{V}_0 \coloneqq \mathbb{V} - \{0\}$,

$$\mu(x,\Gamma) = \int_{\mathbb{V}} 1_\Gamma\big(c(x,v)\big)\mathrm{d}v, \quad \Gamma\in\mathscr{B}(\mathbb{V}_0)$$

to allow for such thinning.

A.2 Special jump-diffusions

We now turn our focus to $(\mathrm{P},\mathscr{F})$ jump-diffusions which are additionally *special* in the sense of them having a semimartingale decomposition in which the finite-variation term is predictable. When looking at the canonical representation of a jump-diffusion X with χ-characteristics (β^χ,α,μ), it is clear how to make this predictable.

$$\begin{aligned} X_t &= X_0 + \beta^\chi(X)\bullet\ell_t + X_t^{\mathrm{c}} + \chi * \tilde{q}^X + (\mathrm{id}_{\mathbb{V}} - \chi) * q^X \\ &= X_0 + \beta^\chi(X)\bullet\ell_t + (\mathrm{id}_{\mathbb{V}} - \chi) * \hat{q}^X + X_t^{\mathrm{c}} + \mathrm{id}_{\mathbb{V}} * \tilde{q}^X \\ &= X_0 + \left(\beta^\chi(X) + \int_{\mathbb{V}}\big(v-\chi(v)\big)\mu(X,\mathrm{d}v)\right)\bullet\ell_t + X_t^{\mathrm{c}} + \mathrm{id}_{\mathbb{V}} * \tilde{q}^X \end{aligned} \tag{A.2.1}$$

In such a case, it is nice to define the function $\beta : \mathbb{X}\to\mathbb{V}$,

$$\beta(x) \coloneqq \beta^\chi(x) + \int_{\mathbb{V}}\big(v-\chi(v)\big)\mu(x,\mathrm{d}v), \tag{A.2.2}$$

so that (A.2.1) may be simplified to a concise special semimartingale decomposition.

$$X_t = X_0 + \beta(X)\bullet\ell + X^{\mathrm{c}} + \mathrm{id}_{\mathbb{V}} * \tilde{q}_t^X$$

We call the triplet (β, α, μ) that results from (A.2.2) the *special differential characteristics* and its components β, α, μ the *drift*, *diffusion*, and *jump kernel*, respectively.

The calculus of (A.2.1) begs the question that $(\mathrm{id}_{\mathbb{V}} - \chi) * q^X$ can be compensated which is not generally the case—otherwise, the term *special* would be a misnomer! The next result specifies conditions on which we may perform the above calculus.

Lemma A.2.3. *Let X be a $(\mathrm{P}, \mathscr{F})$ jump-diffusion with differential χ-characteristics $(\beta^\chi, \alpha, \mu)$, such that μ satisfies the following condition.*

$$x \mapsto \int_{\mathbb{V}} \big|v - \chi(v)\big| \mu(x, \mathrm{d}v) \textit{ is bounded on compact subsets}$$

Then, X is special with drift β as in (A.2.2).

Proof. By choosing a $\mathscr{F}$ localizing sequence $(T_n)_{n\in\mathbb{N}}$ as in (A.1.9), our hypothesis gives us the following integrability.

$$\mathrm{E}_{\mathrm{P}} \big|\mathrm{id}_{\mathbb{V}} - \chi\big| * \hat{q}^X_{T_n} = \mathrm{E}_{\mathrm{P}} \int_0^{T_n} \int_{\mathbb{V}} \big|v - \chi(v)\big| \mu(X_t, \mathrm{d}v) \mathrm{d}t \le n \cdot \sup_{|x| \le n} \int_{\mathbb{V}} \big|v - \chi(v)\big| \mu(x, \mathrm{d}v) < \infty$$

Now, [JS03, Proposition II.1.28] allows us to compensate as we did in (A.2.1)

Seeing as $(\mathrm{id}_{\mathbb{V}} - \chi) * q^X$ may be compensated for special jump-diffusions X, all the characterizing objects of Theorem A.1.14 may be rewritten in terms of our drift β—effectively, χ becomes the identity.

$$\mathcal{L}f(x) := \mathrm{D}f(x) \cdot \beta(x) + \frac{1}{2} \operatorname{tr} \Big(\mathrm{D}^2 f(x) \circ \alpha(x)\Big) + \int_{\mathbb{V}} \Big(f(x+v) - f(x) - \mathrm{D}f(x) \cdot v\Big) \mu(x, \mathrm{d}v)$$
$$\Lambda(u, x) = \big\langle u, \beta(x) \big\rangle + \frac{1}{2} \big\langle u, \alpha(x) \big\rangle + \int_{\mathbb{V}} \big(e^{\langle u, v\rangle} - 1 - \langle u, v\rangle\big) \mu(x, \mathrm{d}v),$$
$$X_t = x + \beta(X) \bullet \ell_t + \sigma(X_-) \bullet W_t + c(X_-, \mathrm{id}_{\mathbb{V}}) * \tilde{p}_t$$

A.3 Locally countable jump-diffusions

We see that a $(\mathrm{P}, \mathscr{F})$ jump-diffusion X is special if the accumulated large jumps $(\mathrm{id}_{\mathbb{V}} - \chi) * q^X$ may be compensated. To this end, being special is a condition on the jumps *away* from the origin. We now turn our focus to jump-diffusions X in which the jumps *near* the origin behave nicely. For any jump-diffusion X, we may count the jumps with the jump process N^X.

$$N^X_t := \sum_{0 < s \le t} 1_{\Delta X_s \neq 0} = 1 * q^X_t \tag{A.3.1}$$

For many jump diffusions, it may be the case that we P-almost-surely have $N^X_t = \infty$ for all $t > 0$. We way that X has $(\mathrm{P}, \mathscr{F})$ *locally countable*, so long as N^X is $(\mathrm{P}, \mathscr{F})$ locally integrable. Below, we state how to verify this using the differential characteristics.

Lemma A.3.2. *Fix a $(\mathrm{P}, \mathscr{F})$ jump-diffusion X with differential χ-characteristics $(\beta^\chi, \alpha, \mu)$ satisfying*

$$x \mapsto \mu\big(x, \mathbb{V}\big) \textit{ is bounded on compact sets,}$$

then X is locally countable. Moreover, we may define $\lambda : \mathbb{X} \to \mathbb{R}_+$ and probability kernel $\kappa : \mathbb{X} \times \mathscr{B}(\mathbb{V}) \to [0,1]$ by the following factoring.

$$\lambda(x) \coloneqq \mu(x, \mathbb{V}), \quad \mu(x, \mathrm{d}v) =: \lambda(x)\kappa(x, \mathrm{d}v)$$

Also, N has $(\mathrm{P}, \mathscr{F})$ intensity $\lambda(X)$.

Proof. Select the sequence $(T_n)_{n\in\mathbb{N}}$ as in (A.1.9). Note now that, since the constant function 1 is predictable,

$$\mathrm{E}_\mathrm{P} N^X_{T_n} = \mathrm{E}_\mathrm{P} 1 * q^X_{T_n} = \mathrm{E}_\mathrm{P} 1 * \hat{q}^X_{T_n} = \mathrm{E}_\mathrm{P} \int_0^{T_n} \mu(X_t, \mathbb{V})\mathrm{d}t \le n \cdot \sup_{|x|\le n} \mu(x, \mathbb{V}) < \infty$$

This means that N^X is locally integrable, making X locally countable. Moreover, by [JS03, Theorem II.1.8],

$$N^X - \int_0^t \lambda(X_s)\mathrm{d}s = 1 * q^X - \int_0^t \int_\mathbb{V} \mu(X_s, \mathrm{d}v)\mathrm{d}s = 1 * q^X - 1 * \hat{q}^X$$

is a $(\mathrm{P}, \mathscr{F})$ local martingale, which finishes the proof.

Remark A.3.3. *(a) Such objects λ, κ always exist with our assumption of the Lemma. Seeing as μ is a transition kernel from $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ to $(\mathbb{V}, \mathscr{B}(\mathbb{V}))$, we have our desired measurability.*

$$\lambda \coloneqq \mu(\cdot, \mathbb{V}) \in \mathscr{B}(\mathbb{X})/\mathscr{B}(\mathbb{R}_+)$$

Constructing κ should be obvious algebra, so long as we have no zero measures; otherwise, we may define

$$\kappa(x, \Gamma) \coloneqq \delta_{e_1}(\Gamma) \cdot 1_{\lambda^{-1}\{0\}}(x) + \frac{\mu(x, \Gamma)}{\lambda(x)} 1_{\mathbb{X}-\lambda^{-1}\{0\}}(x),$$

where δ_{e_1} is the degenerate measure at $e_1 \in \mathbb{V}$. This ensures that any $\kappa(\cdot, \Gamma) \in \mathscr{B}(\mathbb{X})/\mathscr{B}([0,1])$ and any $\kappa(x, \cdot)$ a probability measure on $\mathscr{B}(\mathbb{V})$. Also, when $\mu(x, \cdot)$ is the zero measure,

$$\mu(x, \mathrm{d}v) = 0 = \lambda(x) \cdot \delta_{e_1}(\mathrm{d}v) = \lambda(x)\kappa(x, \mathrm{d}v),$$

and otherwise,

$$\mu(x, \mathrm{d}v) = \mu(x, \mathbb{V})\frac{\mu(x, \mathrm{d}v)}{\mu(x, \mathbb{V})} = \lambda(x)\kappa(x, \mathrm{d}v).$$

(b) We call λ the intensity map *and κ the* (conditional) jump distribution

(c) As far as we know, there is no widely accepted source which explores jump-diffusions to the extent of declaring a notion like locally countable*, as we have. This means that there is likely some clash of terminology, should such a concept already exist.*

A.4 Real moments of jump-diffusions

We now turn our focus to the real moments of $(\mathrm{P}, \mathscr{F})$ jump-diffusions and the extension of our Lévy-Khintchine map Λ to real moments.

$$\Lambda(u,x) = \big\langle u, \beta^{\chi}(x)\big\rangle + \frac{1}{2}\big\langle u, \alpha(x)\big\rangle + \int_{\mathbb{V}} \big(e^{\langle u,v\rangle} - 1 - \langle u, \chi(v)\rangle\big)\mu(x, \mathrm{d}v), \quad u \in \mathbb{V},\ x \in \mathbb{X}$$

The above expression may be infinite, as the final term includes an unbounded integral over a possibly infinite measure. That said, we find it imperative to denote the following sets of finiteness.

$$\mathcal{D}_\Lambda(x) \coloneqq \Big\{u \in \mathbb{V} : \Lambda(u,x) < \infty\Big\}, \qquad \mathcal{D}_\Lambda \coloneqq \bigcap_{x\in\mathbb{X}} \mathcal{D}_\Lambda(x) \tag{A.4.1}$$

The following results will explore the nature of the maps $\Lambda(\cdot, x) : \mathcal{D}_\Lambda(x) \to \mathbb{R}$ for fixed differentiable χ-characteristics $(\beta^\chi, \alpha, \mu)$, where our truncation function χ is defined by $\chi(v) = v 1_{|v|\le 1}$. Note that there is no loss of generality in selecting this truncation function, since they all evaluate Λ identically.

Lemma A.4.2. *For any $x \in \mathbb{X}$, we have $u \in \mathcal{D}_\Lambda(x)$ if and only if $\int_{|v|>1} e^{\langle u,v\rangle}\mu(x, \mathrm{d}v) < \infty$.*

Proof. To each $u, v \in \mathbb{V}$, Taylor's theorem gives us $\gamma_{u,v} \in [0,1]$ such that

$$e^{\langle u,v\rangle} = 1 + \langle u,v\rangle + \frac{1}{2}e^{\gamma_{u,v}\langle u,v\rangle}\langle u,v\rangle^2.$$

This allows us to see that, for each $x \in \mathbb{X}$, $\Lambda(u,x)$ and $\int_{|v|>1} e^{\langle u,v\rangle}\mu(x,\mathrm{d}v)$ differ by finite expressions.

$$\begin{aligned}
&\left|\Lambda(u,x) - \int_{|v|>1} e^{\langle u,v\rangle}\mu(x,\mathrm{d}v)\right| \\
&= \left|\langle u, \beta^\chi(x)\rangle + \frac{1}{2}\langle u, \alpha(x)u\rangle + \int_{|v|\le 1}\big(e^{\langle u,v\rangle} - 1 - \langle u,v\rangle\big)\mu(x,\mathrm{d}v) - \int_{|v|>1}\mu(x,\mathrm{d}v)\right| \\
&\le \left|\langle u, \beta^\chi(x)\rangle + \frac{1}{2}\langle u, \alpha(x)u\rangle\right| + \left|\int_{|v|\le 1}\frac{1}{2}e^{\gamma_{u,v}\langle u,v\rangle}\langle u,v\rangle^2\mu(x,\mathrm{d}v)\right| + \int_{|v|>1}\mu(x,\mathrm{d}v) \\
&\le \left|\langle u, \beta^\chi(x)\rangle + \frac{1}{2}\langle u, \alpha(x)u\rangle\right| + \left(\frac{1}{2}e^{|u|} + 1\right)\int_{\mathbb{V}}\big(1 \wedge |v|^2\big)\mu(x,\mathrm{d}v)
\end{aligned}$$

Thus, one can be defined as a finite displacement of the other.

Lemma A.4.3. *For each $x \in \mathbb{X}$, $\mathcal{D}_\Lambda(x)$ is convex.*

Proof. We use our characterization of $\mathcal{D}_\Lambda(x)$ from Lemma A.4.2. Let $u, u' \in \mathcal{D}_\Lambda(x)$, $\gamma \in (0,1)$, and use Hölder's inequality to see the following.

$$\begin{aligned}
&\int_{|v|>1} e^{\langle u'+\gamma(u-u'),v\rangle}\mu(x,\mathrm{d}v) \\
&\quad= \int_{|v|>1}\big|(e^{\langle u,v\rangle})^\gamma \cdot (e^{\langle u',v\rangle})^{1-\gamma}\big|\mu(x,\mathrm{d}v)
\end{aligned}$$

$$
\begin{aligned}
&\leq \left(\int_{|v|>1} \left|(e^{\langle u,v\rangle})^{\gamma}\right|^{\frac{1}{\gamma}} \mu(x,\mathrm{d}v)\right)^{\gamma} \left(\int_{|v|>1} \left|(e^{\langle u',v\rangle})^{1-\gamma}\right|^{\frac{1}{1-\gamma}} \mu(x,\mathrm{d}v)\right)^{1-\gamma} \\
&= \left(\int_{|v|>1} e^{\langle u,v\rangle} \mu(x,\mathrm{d}v)\right)^{\gamma} \left(\int_{|v|>1} e^{\langle u',v\rangle} \mu(x,\mathrm{d}v)\right)^{1-\gamma} \\
&< \infty
\end{aligned}
$$

An arbitrary convex combination now satisfies $\gamma u + (1-\gamma)u' \in \mathcal{D}_\Lambda(x)$.

Lemma A.4.4. *For each $x \in \mathbb{X}$, the map $\Lambda(\cdot, x)$ is continuously differentiable on $\mathcal{D}_\Lambda(x)^\circ$, with derivative $\mathrm{D}\Lambda(\cdot, x) : \mathcal{D}_\Lambda(x)^\circ \to \mathbb{L}(\mathbb{V}, \mathbb{R})$ as follows.*

$$
\mathrm{D}\Lambda(u,x)w = \Big\langle \beta^{\chi}(x) + \alpha(x)u + \int_{\mathbb{V}} \big(e^{\langle u,v\rangle} v - \chi(v)\big)\mu(x,\mathrm{d}v), w \Big\rangle, \quad u \in \mathcal{D}_\Lambda(x)^\circ \tag{A.4.5}
$$

Proof. Fix $x \in \mathbb{X}$, $u \in \mathcal{D}_\Lambda(x)^\circ$. Let $\epsilon > 0$ such that $B(u,\epsilon) \subseteq \mathcal{D}_\Lambda(x)$. For all $0 < \delta < \epsilon$ and $i = 1, \ldots, d$, we now have the following identity

$$
\begin{aligned}
\frac{\Lambda(u+\delta e_i, x) - \Lambda(u,x)}{\delta} &= \langle e_i, \beta^{\chi}(x)\rangle + \langle e_i, \alpha(x)u\rangle + \frac{1}{2}\langle \delta e_i, \alpha(x)u\rangle \\
&\quad + \int_{|v|\leq 1} \frac{1}{\delta}\Big(e^{\langle u+\delta e_i, v\rangle} - e^{\langle u,v\rangle} - \langle \delta e_i, v\rangle\Big)\mu(x,\mathrm{d}v) \\
&\quad + \int_{|v|>1} \frac{1}{\delta}\Big(e^{\langle u+\delta e_i, v\rangle} - e^{\langle u,v\rangle}\Big)\mu(x,\mathrm{d}v)
\end{aligned} \tag{A.4.6}
$$

Evaluating the limit of (A.4.6) as $\delta \to 0$ is now a matter of exchanging the limit with integration; we will do this by using the dominated convergence theorem.

For the first integral, Taylor's theorem provides us $\gamma_0, \gamma_1 \in [0,1]$ such that the following hold.

$$
\begin{aligned}
e^{\langle u+\delta e_i, v\rangle} &= 1 + \langle u+\delta e_i, v\rangle + \frac{1}{2}\langle u+\delta e_i, v\rangle^2 e^{\gamma_0\langle u+\delta e_i, v\rangle} \\
e^{\langle u,v\rangle} &= 1 + \langle u,v\rangle + \frac{1}{2}\langle u,v\rangle^2 e^{\gamma_1\langle u,v\rangle}
\end{aligned}
$$

This shows us that, for all $0 < \delta < \epsilon$ and $|v| \leq 1$,

$$
\begin{aligned}
\left|\frac{1}{\delta}\Big(e^{\langle u+\delta e_i, v\rangle} - e^{\langle u,v\rangle} - \langle \delta e_i, v\rangle\Big)\right| &= \left|\frac{1}{2}\langle u+\delta e_i, v\rangle^2 e^{\gamma_0\langle u+\delta e_i, v\rangle} + \frac{1}{2}\langle u,v\rangle^2 e^{\gamma_1\langle u,v\rangle}\right| \\
&\leq \Big(\big(|u|+\epsilon\big)^2 e^{|u|+\epsilon}\Big)|v|^2.
\end{aligned}
$$

This dominating function is integrable,

$$
\int_{|v|\leq 1} \Big(\big(|u|+\epsilon\big)^2 e^{|u|+\epsilon}\Big)|v|^2 \mu(x,\mathrm{d}v) \leq \Big(\big(|u|+\epsilon\big)^2 e^{|u|+\epsilon}\Big)\int_{\mathbb{V}} (1 \wedge |v|^2)\mu(x,\mathrm{d}v) < \infty,
$$

so we may apply the dominated convergence theorem.

$$
\lim_{\delta\to 0} \int_{|v|\leq 1} \frac{1}{\delta}\Big(e^{\langle u+\delta e_i, v\rangle} - e^{\langle u,v\rangle} - \langle \delta e_i, v\rangle\Big)\mu(x,\mathrm{d}v)
$$

$$= \int_{|v|\le 1} \lim_{\delta\to 0} \frac{1}{\delta}\Big(e^{\langle u+\delta e_i, v\rangle} - e^{\langle u,v\rangle} - \langle \delta e_i, v\rangle\Big)\mu(x,\mathrm{d}v)$$

$$= \int_{|v|\le 1} \Big(e^{\langle u,v\rangle} v_i - v_i\Big)\mu(x,\mathrm{d}v) \tag{A.4.7}$$

For the second integral, we again use Taylor's theorem to establish for each $0 < \delta < \epsilon/2$, some $\gamma_\delta \in [0,\delta]$ such that

$$e^{\langle u+\delta e_i, v\rangle} = e^{\langle u,v\rangle} + \langle \delta e_i, v\rangle e^{\langle u+\gamma_\delta e_i, v\rangle}$$

This way, we have the following dominating function.

$$\left|\frac{1}{\delta}\Big(e^{\langle u+\delta e_i, v\rangle} - e^{\langle u,v\rangle}\Big)\right| \le \left|\langle e_i, v\rangle e^{\langle u+\gamma_\delta e_i, v\rangle}\right| \le |v_i| e^{\langle u,v\rangle + \epsilon |v_i|/2}$$

The claim is that this dominating function is integrable. To see this, first note that because we have the following limit,

$$\lim_{|v|\to\infty} \frac{|v_i| e^{\langle u,v\rangle+\epsilon|v_i|/2}}{e^{\langle u,v\rangle + 2\epsilon|v_i|/3}} = \lim_{|v|\to\infty} \frac{|v_i|}{e^{\epsilon|v_i|/6}} = 0$$

There exists $M > 0$ such that for all $|v| > M$,

$$|v_i| e^{\langle u,v\rangle+\epsilon|v_i|/2} < e^{\langle u,v\rangle+2\epsilon|v_i|/3}.$$

We now see that

$$\begin{aligned}
&\int_{|v|>1} |v_i| e^{\langle u,v\rangle+\epsilon|v_i|/2}\mu(x,\mathrm{d}v)\\
&= \int_{1<|v|\le M} |v_i| e^{\langle u,v\rangle+\epsilon|v_i|/2}\mu(x,\mathrm{d}v) + \int_{|v|>M} |v_i| e^{\langle u,v\rangle+\epsilon|v_i|/2}\mu(x,\mathrm{d}v)\\
&\le \int_{1<|v|\le M} M e^{(|u|+\epsilon/2)M}\mu(x,\mathrm{d}v) + \int_{|v|>M} e^{\langle u,v\rangle+2\epsilon|v_i|/3}\mu(x,\mathrm{d}v)\\
&\le M e^{(|u|+\epsilon/2)M}\int_{\mathbb{V}} (1\wedge |v|^2)\mu(x,\mathrm{d}v) + \sum_{\ell=0}^{1}\int_{|v|>1} e^{\langle u+2\epsilon e_i/3, v\rangle}\mu(x,\mathrm{d}v)\\
&< \infty.
\end{aligned}$$

We again use the dominated convergence theorem to deduce the following.

$$\lim_{\delta\to 0}\int_{|v|>1} \frac{1}{\delta}\Big(e^{\langle u+\delta e_i, v\rangle} - e^{\langle u,v\rangle}\Big)\mu(x,\mathrm{d}v)$$

$$= \int_{|v|>1} \lim_{\delta\to 0}\frac{1}{\delta}\Big(e^{\langle u+\delta e_i, v\rangle} - e^{\langle u,v\rangle}\Big)\mu(x,\mathrm{d}v)$$

$$= \int_{|v|>1} e^{\langle u,v\rangle} v_i \mu(x,\mathrm{d}v) \tag{A.4.8}$$

Combining equations (A.4.6), (A.4.7), and (A.4.8) now yields our desired identity.

$$\mathrm{D}_i\Lambda(u,x) = \Big\langle e_i, \beta^\chi(x) + \alpha(x)u + \int_{\mathbb{V}} \Big(e^{\langle u,v\rangle} v - \chi(v)\Big)\mu(x,\mathrm{d}v)\Big\rangle$$

Continuity of $\mathrm{D}_i\Lambda(u,x)$ for $u \in \mathcal{D}_\Lambda(x)^\circ$ involves very similar dominated convergence theorem arguments as above. From here, it is clear that Λ is continuously differentiable with the form in (A.4.5).

As we have seen in Lemmas A.2.3 and A.3.2, if we have local boundedness of certain integrals of a jump kernel μ, we can leverage these to $(\mathrm{P},\mathscr{F})$ local conditions of the associated jump-diffusion X. Throughout the remainder of this section, we impose the following uniform-boundedness principle for the kernel μ.

$$\begin{aligned} &f \in \mathscr{B}(\mathbb{V})/\mathscr{B}(\mathbb{R}), \quad \int_{\mathbb{V}} |f(v)|\mu(x,\mathrm{d}v) < \infty \text{ for all } x \in \mathbb{X} \\ &\implies \quad x \mapsto \int_{\mathbb{V}} |f(v)|\mu(x,\mathrm{d}v) \text{ bounded on compact sets} \end{aligned} \tag{A.4.9}$$

With this assumption, we get some nice results on finite exponential moments of X.

Proposition A.4.10. *Fix a $(\mathrm{P},\mathscr{F})$ jump-diffusion X with differential χ-characteristics (β^χ,α,μ). Suppose we have the regularity condition (A.4.9) above. If $0 \in \mathcal{D}_\Lambda^\circ$, then X is special.*

Proof. If $0 \in \mathcal{D}_\Lambda^\circ$, then there exists some $\delta > 0$ such that $\overline{B}(0,\delta) \subseteq \mathcal{D}_\Lambda$. Observe the following implication of this fact, for each $x \in \mathbb{X}$.

$$\begin{aligned} \int_{\mathbb{V}} \big|v - \chi(v)\big|\mu(x,\mathrm{d}v) &= \int_{|v|>1} |v|\mu(x,\mathrm{d}v) \\ &\le \int_{|v|>1} \frac{\sqrt{d}}{\delta} \exp\Big(\frac{\delta|v|}{\sqrt{d}}\Big)\mu(x,\mathrm{d}v) \\ &\le \frac{\sqrt{d}}{\delta} \int_{|v|>1} \exp\Big(\max_{i=1}^{d}\max_{\ell=0}^{1}\big\langle(-1)^\ell\delta e^i, v\big\rangle\Big)\mu(x,\mathrm{d}v) \\ &\le \frac{\sqrt{d}}{\delta} \sum_{i=1}^{d}\sum_{\ell=0}^{1} \int_{|v|>1} \exp\big\langle(-1)^\ell\delta e^i, v\big\rangle\mu(x,\mathrm{d}v) \\ &< \infty \end{aligned}$$

Our regularity condition (A.4.9) now allows us to apply Lemma A.2.3 to conclude X is special.

Proposition A.4.11. *Fix a $(\mathrm{P},\mathscr{F})$ special jump-diffusion X with special differential characteristics (β,α,μ). Suppose we have the regularity condition (A.4.9) above. If $u \in \mathcal{D}_\Lambda$, then $\exp\langle u,X\rangle$ is special, and $\exp\big(\langle u,X\rangle - \Lambda(u,X)\bullet\ell\big)$ is a $(\mathrm{P},\mathscr{F})$ local martingale.*

Proof. Using Lemma A.1.7 for the function $f_u(v) = \exp\langle u,v\rangle$ and its derivative identities as in (A.1.13), we get the following.

$$\begin{aligned} \exp\langle u,X_t\rangle = {}& \exp\langle u,X_0\rangle + \exp\langle u,X_t\rangle\Big(\langle u,\beta(X)\rangle + \frac{1}{2}\langle u,\alpha(X)u\rangle\Big)\bullet\ell_t \\ &+ \mathrm{D}f_u(X_-)\bullet X^{\mathrm{c}} + \Big(\exp\langle u,X_-\rangle\langle u,\mathrm{id}_{\mathbb{V}}\rangle\Big) * \tilde{q}^X_t \\ &+ \exp\langle u,X_-\rangle\cdot\Big(\exp\langle u,\mathrm{id}_{\mathbb{V}}\rangle - 1 - \langle u,\mathrm{id}_{\mathbb{V}}\rangle\Big) * q^X \end{aligned} \tag{A.4.12}$$

Note that localizing our final term on the sequence $(T_n)_{n\in\mathbb{N}}$ of stopping times in (A.1.9), we get the following.

$$\begin{aligned}
&\mathrm{E}_{\mathrm{P}}\Big|\exp\langle u, X_-\rangle\Big(\exp\langle u,\mathrm{id}_{\mathbb{V}}\rangle - 1 - \langle u,\mathrm{id}_{\mathbb{V}}\rangle\Big)\Big| * \hat{q}^X_{T_n}\\
&= \mathrm{E}_{\mathrm{P}}\int_0^{T_n}\int_{\mathbb{V}}\Big|\exp\langle u, X_s\rangle\Big(\exp\langle u,v\rangle - 1 - \langle u,v\rangle\Big)\Big|\mu(X_s,\mathrm{d}v)\mathrm{d}s\\
&\le n\cdot\sup_{|x|\le n}\bigg(e^{\langle u,x\rangle}\int_{\mathbb{V}}\big|e^{\langle u,v\rangle} - 1 - \langle u,v\rangle\big|\mu(x,\mathrm{d}v)\bigg)
\end{aligned}$$

Seeing as $u\in\mathcal{D}_\Lambda$, the integral in the above quantity is finite, and so (A.4.9) gives us finiteness of the supremum. Using [JS03, Proposition II.1.28] now allows us to compensate the jump term in (A.4.12).

$$\exp\langle u, X_t\rangle = \exp\langle u, X_0\rangle + \Big(\exp\langle u, X\rangle\cdot\Lambda(u,X)\Big)\bullet\ell_t + \mathrm{D}f_u(X_-)\bullet X^{\mathrm{c}} + \Big(\exp\langle u, X_-\rangle\langle u,\mathrm{id}_{\mathbb{V}}\rangle\Big)*\tilde{q}^X_t$$

This is a representation of $\exp\langle u, X\rangle$ as an initial term, predictable term of finite variation, and a local martingale. Thus, it is a special semimartingale. From here, we may perform the product rule on $\exp\big(\langle u, X\rangle - \Lambda(u,X)\bullet\ell\big)$ as we did in Proposition A.1.12 to show that the process is a local martingale.

Theorem A.4.13. *Fix a* $(\mathrm{P},\mathscr{F})$ *special jump-diffusion* X *with special differential characteristics* (β,α,μ)*. Suppose we have the regularity condition (A.4.9) above. For each* $(\mathrm{P},\mathscr{F})$ *predictable* H *of finite-variation with image contained in* $\mathcal{D}^\circ_\Lambda$*, the process* $\exp(H\bullet X)$ *is special and*

$$\exp\Big(H\bullet X - \Lambda(H,X)\bullet\ell\Big)$$

is a $(\mathrm{P},\mathscr{F})$ *local martingale.*

Proof. Perform Itô's formula [JS03, Theoerem I.4.57] in addition to its jump-diffusion variant in Lemma A.1.7 and various stochastic integral identities [JS03, Remarks I.4.36, I.4.37, Theorem I.4.40(d), Proposition II.1.30(b)].

$$\begin{aligned}
&\exp\big(H\bullet X_t\big)\\
&= \exp\big(H\bullet X_-\big)\bullet\big(H\bullet X\big)_t + \frac{1}{2}\exp\big(H\bullet X_-\big)\bullet\big\langle(H\bullet X)^{\mathrm{c}},(H\bullet X)^{\mathrm{c}}\big\rangle_t\\
&\quad + \sum_{0<s\le t}\Big(\exp\big(H\bullet X_{s-} + \Delta\big(H\bullet X\big)_s\big) - \exp\big(H\bullet X_{s-}\big) - \exp\big(H\bullet X_{s-}\big)\Delta\big(H\bullet X\big)_s\Big)\\
&= \Big(\exp\big(H\bullet X_-\big)\cdot H\Big)\bullet X_t + \frac{1}{2}\exp\big(H\bullet X\big)\big\langle H,\alpha(X)H\big\rangle\bullet\ell_t\\
&\quad + \exp\big(H\bullet X_-\big)\Big(e^{\langle H,\mathrm{id}_{\mathbb{V}}\rangle} - 1 - \langle H,\mathrm{id}_{\mathbb{V}}\rangle\Big)*q^X_t
\end{aligned}$$

(A.4.14)

$$\begin{aligned}
&= \Big(\exp\big(H\bullet X\big)\cdot\langle H,\beta\rangle + \frac{1}{2}\exp\big(H\bullet X\big)\big\langle H,\alpha(X)H\big\rangle\Big)\bullet\ell_t + \Big(\exp\big(H\bullet X_-\big)\cdot H\Big)\bullet X^{\mathrm{c}}_t\\
&\quad + \exp\big(H\bullet X_-\big)\langle H,\mathrm{id}_{\mathbb{V}}\rangle*\tilde{q}^X_t\\
&\quad + \exp\big(H\bullet X_-\big)\Big(e^{\langle H,\mathrm{id}_{\mathbb{V}}\rangle} - 1 - \langle H,\mathrm{id}_{\mathbb{V}}\rangle\Big)*q^X_t
\end{aligned}$$

Now, choosing our $(\mathrm{P},\mathscr{F})$ localizing sequence $(T_n)_{n\in\mathbb{N}}$ as in A.1.9, we have the following bound.

$$\begin{aligned}
&\mathrm{E}_{\mathrm{P}}\Big|\exp\big(H\bullet X_-\big)\Big(e^{\langle H,\mathrm{id}_{\mathbb{V}}\rangle}-1-\langle H,\mathrm{id}_{\mathbb{V}}\rangle\Big)*\hat{q}^X_{T_n}\Big|\\
&=\mathrm{E}_{\mathrm{P}}\int_0^{T_n}\int_{\mathbb{V}}\Big|\exp\big(H\bullet X_s\big)\Big(e^{\langle H(s),v\rangle}-1-\langle H(s),v\rangle\Big)\Big|\mu(X_s,\mathrm{d}v)\mathrm{d}s\\
&\le n\cdot\sup_{|x|\le n}\sup_{s\in[0,n]}e^{|x|\cdot|H(s)|}\int_{\mathbb{V}}\big|e^{\langle H(s),v\rangle}-1-\langle H(s),v\rangle\big|\mu(x,\mathrm{d}v)
\end{aligned}$$

Seeing as $\Lambda(\cdot,x)$ is continuously differentiable, it is uniformly bounded on $\mathcal{D}^\circ_\Lambda$. This, along with the fact that H is bounded (it has finite variation) and assumption (A.4.9) allow us to conclude that the preceding expression is finite. Thus, we may compensate the final jump integral in (A.4.14).

$$\begin{aligned}
\text{(A.4.15)}\qquad \exp\big(H\bullet X_t\big)&=\Big(\exp\big(H\bullet X\big)\cdot\Lambda(H,X)\Big)\bullet\ell_t+\Big(\exp\big(H\bullet X_-\big)\cdot H\Big)\bullet X^{\mathrm{c}}_t\\
&\quad+\exp\big(H\bullet X_-\big)\Big(e^{\langle H,\mathrm{id}_{\mathbb{V}}\rangle}-1\Big)*\tilde{q}^X_t
\end{aligned}$$

The decomposition of $\exp(H\bullet X)$ into a predictable finite-variation process and a local martingale implies that it is special. Now, we write M as the local martingale term above, $A=\exp(H\bullet X)$, and $B=\exp(-\Lambda(H,X)\bullet\ell)$. We now recognize that B is predictable and finite-variation and use [JS03, Proposition I.4.49(b)] to conclude our proof.

$$\begin{aligned}
\exp\big(H\bullet X_t-\Lambda(H,X)\bullet\ell_t\big)&=A_tB_t\\
&=A_-\bullet B_t+B\bullet A_t\\
&=\big(A\cdot B\cdot-\Lambda(H,X)\big)\bullet\ell_t\\
&\qquad+B\bullet\Big(\big(\exp(H\bullet X)\cdot\Lambda(H,X)\big)\bullet\ell+M\Big)_t\\
&=\big(A\cdot B\cdot-\Lambda(H,X)\big)\bullet\ell_t+\big(B\cdot A\cdot\Lambda(H,X)\big)\bullet\ell_t+B\bullet M_t\\
&=B\bullet M_t
\end{aligned}$$

In the case that the local martingale in the preceding theorem satisfies sufficient integrability, we are able to introduce a measure change with which the dynamics of X are still understood. Note that the following theorem states this, with careful language delineating the fact that special differential characteristics depend on the underlying measure.

Theorem A.4.16. *Fix a $(\mathrm{P},\mathscr{F})$ special jump-diffusion X with $(\mathrm{P},\mathscr{F})$ special differential characteristics (β,α,μ). Fix $h:[0,\infty)\times\mathbb{X}\to\mathbb{V}$ such that $h(\ell,X_-)$ is of finite variation. Define*

$$Z^h:=\exp\Big(h(\ell,X_-)\bullet X-\Lambda\big(h(\ell,X),X\big)\bullet\ell\Big),$$

and assume $(Z^h_t)_{t\in[0,\tau]}$ is uniformly integrable for each $\tau>0$. Then, to each $\tau>0$, we may define Q^h on $(\Omega,\mathscr{F}_\tau)$, via

$$\mathrm{Q}^h_\tau(\mathrm{d}\omega):=Z^h_\tau(\omega)\cdot\mathrm{P}(\mathrm{d}\omega),$$

such that X is a $(\mathrm{Q}^h_\tau,(\mathscr{F}_t)_{t\in[0,\tau]})$ special jump-diffusion with $(\mathrm{Q}^h_\tau,(\mathscr{F}_t)_{t\in[0,\tau]})$ special differential characteristics (β^h,α,μ^h).

$$\beta^h(s,x)=\beta(x)+\alpha(x)h(s,x)+\int_{\mathbb{V}}v\big(e^{\langle h(s,x),v\rangle}-1\big)\mu(x,\mathrm{d}v),$$

$$\mu^h(s,x,\mathrm{d}v) = e^{\langle h(s,x),v\rangle}\mu(x,\mathrm{d}v)\mathrm{d}s$$

Proof. Note that Theorem A.4.13 gives us the $(\mathrm{P},\mathscr{F})$ local martingale property, so our P uniform-integrability assumption serves to give us a $(\mathrm{P},\mathscr{F})$ uniformly-integrable martingale Z^h on $[0,\tau]$. Now, evaluating the $(\mathrm{Q}^h_\tau,(\mathscr{F}_t)_{t\in[0,\tau]})$ dynamics of X amount to applying Girsanov's theorem for semimartingales. Specifically we use [JS03, Theorem III.3.24] and verify that our proposed differential characteristics align with the identities of [JS03, III.3.28]. For any $\mathscr{F}$ predictable $H:\Omega\times\mathbb{R}_+\times\mathbb{V}\to\mathbb{R}_+$, we apply [JS03, Remark I.4.36] to see that the following identities hold.

$$\begin{aligned}
\mathrm{E}_{\mathrm{P}}\big((Z^hH)*q^X\big) &= \mathrm{E}_{\mathrm{P}}\int_{\mathbb{R}_+\times\mathbb{V}} Z^h_s H(\cdot,s,v)q^X(\mathrm{d}s,\mathrm{d}v)\\
&= \mathrm{E}_{\mathrm{P}}\int_{\mathbb{R}_+\times\mathbb{V}} \exp\Big(h(\ell,X_-)\bullet X_s - \Lambda\big(h(\ell,X),X\big)\bullet \ell_s\Big)H(\cdot,s,v)q^X(\mathrm{d}s,\mathrm{d}v)\\
&= \mathrm{E}_{\mathrm{P}}\int_{\mathbb{R}_+\times\mathbb{V}} \exp\Big(h(\ell,X_-)\bullet X_{s-} + \Delta\big(h(\ell,X_-)\bullet X\big)_s - \Lambda\big(h(\ell,X),X\big)\bullet \ell_s\Big)\\
&\qquad\qquad \cdot H(\cdot,s,v)q^X(\mathrm{d}s,\mathrm{d}v)\\
&= \mathrm{E}_{\mathrm{P}}\int_{\mathbb{R}_+\times\mathbb{V}} Z^h_{s-}e^{\langle h(s,X_{s-}),v\rangle}H(\cdot,s,v)q^X(\mathrm{d}s,\mathrm{d}v)\\
&= \mathrm{E}_{\mathrm{P}}\int_{\mathbb{R}_+\times\mathbb{V}} Z^h_{s-}e^{\langle h(s,X_{s-}),v\rangle}H(\cdot,s,v)\hat{q}^X(\mathrm{d}s,\mathrm{d}v)\\
&= \mathrm{E}_{\mathrm{P}}\int_{\mathbb{R}_+\times\mathbb{V}} Z^h_{s-}e^{\langle h(s,X_{s-}),v\rangle}H(\cdot,s,v)\mu(X_s,\mathrm{d}v)\mathrm{d}s\\
&= \mathrm{E}_{\mathrm{P}}\int_{\mathbb{R}_+\times\mathbb{V}} Z^h_{s}e^{\langle h(s,X_{s}),v\rangle}H(\cdot,s,v)\mu(X_s,\mathrm{d}v)\mathrm{d}s
\end{aligned}$$

This first part indicates that our μ^h is as described. Now, let us establish the drift by looking at the continuous predictable projections. First, we recall (A.4.15) and the calculations below it to note that Z^h has the following continuous local martingale component.

$$\begin{aligned}
Z^{h,\mathrm{c}}_t &= \exp\Big(-\Lambda\big(h(\ell,X),X\big)\bullet\ell\Big)\bullet\bigg(\exp\big(h(\ell,X_-)\bullet X_-\big)h(\ell,X_-)\bullet X^{\mathrm{c}}_t\bigg)\\
&= Z^h_- h(\ell,X_-)\bullet X^{\mathrm{c}}_t
\end{aligned}$$

Note that we applied our usual calculus operation [JS03, Remark I.4.37] above. This identity gives us our desired predictable quadratic covariation when using the previously referenced calculus in addition to that from [JS03, Theorem I.4.40].

$$\begin{aligned}
\big\langle Z^{h,\mathrm{c}},X^{\mathrm{c},j}\big\rangle &= \Big\langle Z^h_- h(\ell,X_-)\bullet X^{\mathrm{c}},X^{\mathrm{c},j}\Big\rangle\\
&= \Big\langle \sum_{i=1}^d Z^h_- h_i(\ell,X_-)\bullet X^{\mathrm{c},i},X^{\mathrm{c},j}\Big\rangle\\
&= \sum_{i=1}^d Z^h_- h_i(\ell,X_-)\bullet\Big\langle X^{\mathrm{c},i},X^{\mathrm{c},j}\Big\rangle\\
&= \sum_{i=1}^d Z^h_- h_i(\ell,X_-)\bullet\big(\alpha_{ij}(X)\bullet\ell\big)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{d} Z_{-}^{h} h_i(\ell, X)\alpha_{ij}(X) \bullet \ell \\
&= Z^{h}\alpha(X)h(\ell, X) \bullet \ell
\end{aligned}
$$

Bibliography

[Cuc11] Christina Cuchiero. Affine and polynomial processes, 2011.

[DFS03] D. Duffie, D. Filipovic, and W. Schachermayer. Affine processes and applications in finance. *The Annals of Applied Probability*, 13(3):984–1053, aug 2003.

[DRL04] Ken Duffy and Mark Rodgers-Lee. Some Useful Functions for Functional Large Deviations. *Stochastics and Stochastic Reports*, 76(3):267–279, jun 2004.

[DZ10] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*, volume 38. Springer Berlin Heidelberg, 2 edition, 2010.

[FK06] Jin Feng and Thomas G Kurtz. *Large deviations for stochastic processes.* American Mathematical Soc., 2006.

[GZ18] Fuqing Gao and Lingjiong Zhu. Some asymptotic results for nonlinear Hawkes processes. *Stochastic Processes and their Applications*, 128(12):4051–4077, 2018.

[JS03] Jean Jacod and Albert N. Shiryaev. *Limit Theorems for Stochastic Processes*, volume 288. Springer Berlin Heidelberg, 2003.

[Kal02] Olav Kallenberg. *Foundations of Modern Probability.* Springer New York, 2002.

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

[KRM15] Martin Keller-Ressel and Eberhard Mayerhofer. Exponential moments of affine processes. *The Annals of Applied Probability*, 25(2):714–752, apr 2015.

[Puh01] A Puhalskii. *Large Deviations and Idempotent Probability.* Monographs and Surveys in Pure and Applied Mathematics. CRC Press, 2001.

[Sko56] A. V. Skorokhod. Limit theorems for stochastic processes. *Theory of Probability & Its Applications*, 1:261–290, 1 1956.

[SV10] Peter Spreij and Enno Veerman. The affine transform formula for affine jump-diffusions with a general closed convex state space. *arXiv preprint arXiv:1005.1099*, 2010.

[Wal98] Wolfgang Walter. *Ordinary Differential Equations*, volume 182. Springer New York, 1998.