Large deviations for multi-scale jump-diffusion processes

Rohini Kumar*
Department of Mathematics
Wayne State University

and

Lea Popovic
Department of Mathematics and Statistics
Concordia University

March 23, 2015

Abstract

We obtain large deviation results for a two time-scale model of jump-diffusion processes. The processes on the two time scales are fully inter-dependent, the slow process has small perturbative noise and the fast process is ergodic. Our results extend previous large deviation results for diffusions. We provide concrete examples in their applications to finance and biology, with an explicit calculation of the large deviation rate function.

Contents

1 Introduction 2

2 Two time-scale jump-diffusion 3
  2.1 Examples ................................................................. 5

3 Large deviation principle 6
  3.1 Convergence of viscosity solutions of PIDEs .................. 10
  3.2 Comparison Principle ..................................................... 13
  3.3 Exponential tightness ..................................................... 16

4 Examples 16
  4.1 Model for stock price with stochastic volatility ............... 16
  4.2 Model for self-regulating protein production ................... 18

*Work partially supported by National Science Foundation grant DMS 1209363
# Introduction

For a number of processes in finance and biology the appropriate stochastic modelling is done in terms of multi-scale Markov processes with fully dependent slow and fast fluctuating variables. The most common examples of such multi-scale processes (random evolutions, diffusions, state dependent Markov chains) are all particular cases of jump-diffusions. The law of large numbers limit, central limit theorem, and the corresponding large deviations behaviour of these models are all of interest in applications.

One case of their use in finance is in multi-factor stochastic volatility models, which are used to capture the smiles and skews of implied volatility. The separation of time scales is helpful for calibration, since it allows one to reduce the number of group parameters. The rate function from the large deviation principle for the stock price process can be used to obtain the price of short maturity options, as well as the limit of the at-the-money implied volatility. These have been explicitly calculated for models in which the logarithm of the stock price and the stochastic volatility are driven by diffusions ([FFK12], [FFF10]). However, much of the empirical evidence ([B-NS01], [Kou02]) suggests that mean-reverting jump-diffusions would be a more appropriate model for the problem.

In biology one case of their use is in models of intracellular biochemical reactions. Due to low copy numbers of various key molecular types and varying strengths in chemical bonds, normalized copy numbers of different types of molecules are processes on multiple time-scales (see [BKRP05], [KK13] for references to the biology literature). Changes in molecular compositions are modelled by state-dependent Markov chains, and on the slower time scale are well approximated by diffusions with small noise or piecewise deterministic Markov chains ([KKP12]). The rate function from the large deviation principle for slowly fluctuating molecular species is used to calculate the propensity for switching in a network that has multiple stable equilibria. Since intracellular processes are also subject to other sources of ‘extrinsic’ noise, multiple time-scale diffusions may include jumps from additional sources. For example, there can be errors during cell division ([Pau10], [Pau11]); a stochastic model combining both reactions and cell division was analyzed in [McSP14].

Large deviation results for multi-scale diffusions have been studied by Freidlin (see [F98] Chapter 7), Veretennikov ([Ver00]), Dupuis et al. ([Dup12]), and Puhalskii ([Puh15]). For the multi-scale Markov chains where the slow process is a piecewise deterministic Markov processes and the fast process is a Markov chain on a finite state space explicit results were obtained by Faggionato et al. ([Fagg09], [Fagg10]). For jump-diffusions there are very few large deviation results. On a single time scale, there are results by Imkeller et al. ([Imk09]) for first exit times for SDEs driven by symmetric stable and exponentially light-tail symmetric Lévy processes. An approach based on control theory and the variational representation was developed by Budhiraja et al. in [Bud11] and extended to infinite dimensional versions [Bud13] (that is, SPDEs rather than SDEs driven by a Poisson random measure). It is not easy to see how to...
use these results in a multi-scale model of jump-diffusions.

A general method for Markov processes based on non-linear semigroups and viscosity methods was developed by Feng and Kurtz in [FK06]. However, verifying the abstract conditions needed to apply this method to multi-scale jump-diffusions is a non-trivial task. In this paper we give a proof of large deviations for two time-scale jump-diffusions, using a technique developed by Feng et al. in [FFK12]. The advantage of this method is that it is constructive and, with some effort, can be tailored to different multi-scale processes. Our proof follows the steps of [FFK12], extending it to processes with jumps and full dependence of the slow and fast components. It is based on viscosity solutions to the Cauchy problem for a sequence of partial integro-differential equations and uses a construction of the sub- and super-solutions to related Cauchy problems as in [FFK12]. Our results hold for slow and fast jump-diffusions which are fully inter-dependent, and where the fast processes is ergodic but not necessarily symmetric. In case the evolution of both processes is spatially homogeneous in the slow variables, we can also provide a more explicit (than a solution to a variational problem) formula for the rate function.

2 Two time-scale jump-diffusion

Consider the following system of stochastic differential equations:

\[
\begin{align*}
\frac{dX_{\epsilon,t}}{dt} &= b(X_{\epsilon,t-}, Y_{\epsilon,t-})dt + \epsilon b_0(X_{\epsilon,t-}, Y_{\epsilon,t-})dt + \sqrt{\epsilon} \sigma(X_{\epsilon,t-}, Y_{\epsilon,t-})dW_{t}^{(1)} \\
&\quad + \epsilon \int k(X_{\epsilon,t-}, Y_{\epsilon,t-}, z)\hat{N}_{\epsilon,t}^{(1)}(dz, dt), \\
\frac{dY_{\epsilon,t}}{dt} &= \frac{1}{\epsilon} b_1(X_{\epsilon,t-}, Y_{\epsilon,t-})dt + \frac{1}{\epsilon} \sigma_1(X_{\epsilon,t-}, Y_{\epsilon,t-}) \left( \rho dW_{t}^{(1)} + \sqrt{1 - \rho^2} dW_{t}^{(2)} \right) \\
&\quad + \int k_1(X_{\epsilon,t-}, Y_{\epsilon,t-}, z)\hat{N}_{\epsilon,t}^{(2)}(dz, dt),
\end{align*}
\]

(1a)

where \(\hat{N}_{\epsilon,t}^{(1)}(\cdot, \cdot), \hat{N}_{\epsilon,t}^{(2)}(\cdot, \cdot)\) are independent Poisson random measures with intensity measures \(\nu(dz) \times \frac{1}{\epsilon} dt, \nu_1(dz) \times \frac{1}{\epsilon} dt\); the Lévy measures \(\nu_1\) and \(\nu_2\) satisfy \(\int_{\mathbb{R}} (1 \wedge z^2) \nu_1(dz) < \infty\) and \(\int_{\mathbb{R}} (1 \wedge z^2) \nu_2(dz) < \infty\); the centered versions are defined as

\[
\hat{N}_{\epsilon,t}^{(1)}(\cdot, \cdot) = N_{\epsilon,t}^{(1)}(\cdot, \cdot) - \nu(dz) \times \frac{1}{\epsilon} dt, \quad \hat{N}_{\epsilon,t}^{(2)}(\cdot, \cdot) = N_{\epsilon,t}^{(2)}(\cdot, \cdot) - \nu_1(dz) \times \frac{1}{\epsilon} dt
\]

and \(W^{(1)}, W^{(2)}\) are independent Brownian motions independent of \(N_{\epsilon,t}^{(1)}(\cdot, \cdot), N_{\epsilon,t}^{(2)}(\cdot, \cdot)\).

To ensure existence and uniqueness of solutions to the system (1) we assume

**Assumption 2.1 (Lipschitz condition).** There exists \(K_1 > 0\) such that \(\forall (x_1, y_1), (x_2, y_2) \in \mathbb{R}^2\)

\[
\begin{align*}
&|b(x_2, y_2) - b(x_1, y_1)|^2 + |b_0(x_2, y_2) - b_0(x_1, y_1)|^2 + |b_1(x_2, y_2) - b_1(x_1, y_1)|^2 \\
&\quad + |\sigma(x_2, y_2) - \sigma(x_1, y_1)|^2 + |\sigma_1(x_2, y_2) - \sigma_1(x_1, y_1)|^2 \\
&\quad + \int |k(x_2, y_2, z) - k(x_1, y_1, z)|^2 \nu(z) dz + \int |k_1(x_2, y_2, z) - k_1(x_1, y_1, z)|^2 \nu_1(z) dz \\
&\leq K_1(|x_2 - x_1|^2 + |y_2 - y_1|^2).
\end{align*}
\]

(2)
Define
\[ V(y; x, p) := b(x, y)p + \frac{1}{2}\sigma^2(x, y)p^2 + \int (e^{pk(x, y, z)} - 1 - pk(x, y, z)) \nu(z)dz. \] (3)

**Assumption 2.2 (Growth condition).** There exists \( K_2 > 0 \) such that \( \forall (x, y) \in \mathbb{R}^2 \)
\[ |b(x, y)|^2 + |b_0(x, y)|^2 + |b_1(x, y)|^2 + |\sigma(x, y)|^2 + |\sigma_1(x, y)|^2 \]
\[ + \int |k_1(x, y, z)|^2 \nu(z)dz + \int |k(x, y, z)|^2 \nu(z)dz \leq K_2(1 + x^2 + y^2). \] (4)
And, for each \( x \) and \( p \) in \( \mathbb{R} \) there exists \( K_{x,p} > -\infty \) such that
\[ V(y; x, p) \geq K_{x,p} \quad \forall y \in \mathbb{R}. \] (5)

If existence and uniqueness of solutions to (1a)+(1b) can be established by other means, we will only assume the coefficients are continuous, and the lower bound (5) on \( V \).

The infinitesimal generator of \((X_s, Y_t)\) is for \( f \in C_b^2(\mathbb{R} \times \mathbb{R}) \) defined by
\[ \mathcal{L}_\epsilon f(x, y) = b(x, y)\partial_x f(x, y) + \rho \sigma(x, y)\sigma_1(x, y)\partial_{xx} f(x, y) \]
\[ + \epsilon b_0(x, y)\partial_x f(x, y) + \frac{\epsilon}{2}\sigma^2(x, y)\partial_{xx} f(x, y) \]
\[ + \frac{1}{\epsilon} \int (f(x + \epsilon k(x, y, z), y) - f(x, y) - k(x, y, z)\partial_x f(x, y)) \nu(z)dz \]
\[ + \frac{1}{\epsilon} \left[ b_1(x, y)\partial_y f(x, y) + \frac{\epsilon}{2}\sigma_1^2(x, y)\partial_{yy} f(x, y) \right. \]
\[ + \left. \int \left( f(x, y + k_1(x, y, z)) - f(x, y) - k_1(x, y, z)\partial_y f(x, y) \right) \nu_1(z)dz \right]. \] (6)

Fix \( x \in \mathbb{R} \) and let \( Y^x \) denote the process satisfying the SDE
\[ dY_t = b_1(x, Y_{t-})dt + \sigma_1(x, Y_{t-}) \left( \rho dW_t^{(1)} + \sqrt{1 - \rho^2}dW_t^{(2)} \right) \]
\[ + \int k_1(x, Y_{t-}, z)\tilde{N}^{(2)}(dz, dt), \quad Y_0^x = y_0. \] (7)
This is the SDE (1b) where \( \epsilon \) is set equal to 1 and \( X_{\epsilon,t} \) is set equal to \( x \). Let \( \mathcal{L}_1^x \) denote the generator of \( Y^x \), then, for \( f \in C_b^2(\mathbb{R}) \),
\[ \mathcal{L}_1^x f(y) := b_1(x, y)\partial_y f(y) + \frac{\epsilon}{2}\sigma_1^2(x, y)\partial_{yy} f(y) \]
\[ + \int \left( f(y + k_1(x, y, z)) - f(y) - k_1(x, y, z)\partial_y f(x, y) \right) \nu_1(z)dz. \] (8)

For fixed \( p \in \mathbb{R} \) define the perturbed \( \mathcal{L}_1^{x,p} \) generator for \( f \in C_b^2(\mathbb{R}^2) \) by
\[ \mathcal{L}_1^{x,p} f(y) := [\rho \sigma(x, y)\sigma_1(x, y)p + b_1(x, y)]\partial_y f(y) + \frac{\epsilon}{2}\sigma_1^2(x, y)\partial_{yy} f(y) \]
\[ + \int \left( f(y + k_1(x, y, z)) - f(y) - k_1(x, y, z)\partial_y f(x, y) \right) \nu_1(z)dz, \] (9)
and let \( Y^{x,p} \) be the process corresponding to the generator \( \mathcal{L}_1^{x,p} \). For each \( x, p \in \mathbb{R} \) we assume the following about \( Y^{x,p} \)
Assumption 2.3 (Ergodicity condition). The process $Y^{x,p}$ is Feller continuous with transition probability $p_{t}^{x,p}(y_0,dy)$, which at $t = 1$ has a positive density $p_{1}^{x,p}(y_0,y)$ with respect to some reference measure $\alpha(dy)$.

Assumption 2.4 (Lyapunov condition). There exists a positive function $\zeta(\cdot) \in C^{2}(\mathbb{R})$, such that $\zeta$ has compact finite level sets, and for each compact set $\Gamma \subset \mathbb{R}^2$ and $l \in \mathbb{R}$ there exists a compact set $A_l \subset \mathbb{R}$ such that

$$\{ y \in \mathbb{R} : \frac{e^{-\zeta} \mathcal{L}_{1}^{x,p} e^{\zeta}(y)}{1 + |V(y,x,p)| + |b_0(x,y)p| + \sigma^2(x,y)} \geq -l \} \subset A_l, \quad \forall x, p \in \Gamma.$$ (10)

Remark 2.1. Some arguments are simpler in the special case $Y^{x,p}$ in addition has a unique invariant probability measure $\pi^p(x, \cdot)$ with respect to which $p_{t}^{x,p}(y_0, y)$ is symmetric and $\pi^p(x, \cdot)$ is reversible, that is

$$\int_{y \in \mathbb{R}} \mathcal{L}_{1}^{x,p} f(y) \pi^p(x, y) dy = 0, \quad \forall f \in C_c^\infty(\mathbb{R}).$$

and

$$\int f(y) \mathcal{L}_{1}^{x,p} g(y) \pi^p(x, y) dy = \int g(y) \mathcal{L}_{1}^{x,p} f(y) \pi^p(x, y) dy, \quad \forall f, g \in C^2(\mathbb{R})$$

2.1 Examples

We give some examples of $Y$ satisfying Assumption 2.3 and Assumption 2.4, the last two are used in our examples in Section 4

Example 2.1. Let $\rho = 0$, $b_1(x, y) = -b_1(x)y$, $\sigma_1(x, y) = \sigma_1(x)$ and $k_1(x, y, z) = \frac{\sigma_1(x)}{\sqrt{b_1(x)}} z - y$, where $b_1(x), \sigma_1(x) > 0$ are continuous. Let $\nu_1(z) = \exp\{-z^2\}$, then $\hat{N}^{(2)} = N^{(2)}$, and for each $x \in \mathbb{R}$, the solution to

$$dY^x_t = -b_1(x)Y^x_t dt + \sigma_1(x) dW^2_t + \int_{R - \{0\}} \left( \frac{\sigma_1(x)}{\sqrt{b_1(x)}} z - Y^x_t \right) N^{(2)}(dz, dt)$$

has unique invariant probability distribution $\pi(x, dy) = \sqrt{\frac{b_1(x)}{\pi(x)}} \exp\{-\frac{b_1(x)y^2}{\sigma_1(x)}\}$ and $Y^x$ is symmetric with respect to it. Assumption 2.4 is satisfied by $\zeta(y) := \frac{b_1(x)}{2\sigma_1(x)} y^2$.

Example 2.2. Take $\rho = 0$ and let $\alpha \in (1, 2)$. Let $Z_t$ be a 1-dimensional symmetric Levy process whose Levy measure is $\nu_1(z) dz = |z|^{-(1+\alpha)} 1_{|z| > 1} dz$. Its infinitesimal generator is the truncated fractional Laplacian $-(\Delta)^{\alpha/2}_{1+}$ defined as

$$-(-\Delta)^{\alpha/2}_{1+} f(y) = \int_{|z| > 1} (f(y + z) - f(y)) \frac{1}{|z|^{1+\alpha}} dz, \quad \text{for } f \in C^2_c(\mathbb{R}).$$

Let $\sigma_1(x, y) := a(x) \sigma_1(y)$ where $a(\cdot), \sigma_1(\cdot) > 0$ are such that $a(\cdot)$ is continuous and $\sigma_1(\cdot)$ is locally $1/\alpha$-Hölder continuous and $\liminf_{|y| \to \infty} \frac{\sigma_1(y)}{|y|^{\alpha}} > 0$. Let

$$dY^{x}_t = \sigma_1(x, Y^{x}_t) dZ_t.$$

Then from Theorem 1.7(i) in [CWI14], $\pi(x, dy) := \frac{\sigma_1(y)^{-\alpha} dy}{\int \sigma_1(y)^{-\alpha} dy}$ is the unique invariant probability measure for the $Y^x$ process and $Y^x$ is $\pi(x, \cdot)$-reversible. From Lemma 3.2 in [CWI14], we get $\zeta(y) := \ln(1 + |y|^\theta)$ for $\theta \in (0, 1)$ satisfies Assumption 2.4.
Example 2.3. Let $c(z, z')$ be a non-symmetric function such that $0 < c_0 \leq c(z, z') \leq c_1$, $c(z, z') = c(z, -z')$ and $|c(z, z'') - c(z', z'')| \leq c_2 |z - z'|^\beta$ for some $\beta \in (0, 1)$. Let $\alpha \in (0, 2)$, and $Z_t$ be a 1-dimensional non-symmetric process whose infinitesimal generator is defined by

$$L_c^\alpha f(y) = \lim_{\delta \to 0} \int_{|z| > \delta} (f(y + z) - f(y)) \frac{c(y, y + z)}{|z|^{1+\alpha}} dz,$$

for $f \in C^2_c(\mathbb{R})$. Let

$$dY_t = -Y_t dt + dZ_t.$$

Heat kernel estimates from [CZ13] imply this non-symmetric jump diffusion is Feller continuous with a positive transition density $p_t(y_0, y), \forall t > 0$

Example 2.4. Let $Y^x$ be a birth-death Markov chain with birth rate $r_+(y) = \lambda(x)$ and death rate $r_-(y) = \mu(x)y$, satisfying $\lambda(x), \mu(x) > 0$. Since its state space is countable its transition density is positive, with a unique reversible invariant distribution $\pi(x, y) = e^{-\lambda(x)/\mu(x)(\lambda(x)/\mu(x))y}$, $y \in \{0, 1, \ldots \}$.

3 Large deviation principle

We prove a large deviation principle for $\{X_{\epsilon, t}\}_{\epsilon > 0}$ as $\epsilon \to 0$ using the viscosity solution approach to verify convergence of a sequence of exponential generators. Define

$$u^h_\epsilon(t, x, y) := \epsilon \ln E \left[ e^{\frac{h(X_{\epsilon, t})}{\epsilon}} |X_{\epsilon, 0} = x, Y_{\epsilon, 0} = y \right],$$

where $h \in C_b(\mathbb{R})$, the space of bounded uniformly continuous functions on $\mathbb{R}$. It can be shown (see [FK06]) that for each $h \in C_b(\mathbb{R})$, $u^h_\epsilon$ is a viscosity solution of the Cauchy problem:

$$\partial_t u = H_\epsilon u \quad \text{in} \ (0, T] \times \mathbb{R} \times \mathbb{R},$$

$$u(0, x, y) = h(x), \quad \text{for} \ (x, y) \in \mathbb{R} \times \mathbb{R},$$

where the non-linear operator is the exponential generator:

$$H_\epsilon u(x, y) := e^{-u/\epsilon} L_\epsilon e^{u/\epsilon}$$

$$= b(x, y) \partial_x u(x, y) + \rho \sigma(x, y) \sigma_1(x, y) \partial^2_{xy} u(x, y) + \frac{1}{2} \sigma^2(x, y) (\partial_x u(x, y))^2$$

$$+ \epsilon \left[ b_0(x, y) \partial_x u(x, y) + \frac{1}{2} \sigma^2(x, y) \partial^2_{xx} u(x, y) \right]$$

$$+ \int \left( e^{\frac{u(x + k(x, y), z) - u(x, y)}{\epsilon}} - 1 - k(x, y, z) \partial_x u(x, y) \right) \nu(z) dz$$

$$+ \frac{1}{\epsilon} \left[ \rho \sigma(x, y) \sigma_1(x, y) \partial_x u(x, y) \partial_y u(x, y) + b_1(x, y) \partial_y u(x, y) + \frac{1}{2} \sigma_1^2(x, y) \partial^2_{yy} u(x, y) \right]$$

$$+ \int \left( e^{\frac{u(x + k_1(x, y, z), z) - u(x, y)}{\epsilon}} - 1 - \frac{k_1(x, y, z)}{\epsilon} \partial_y u(x, y) \right) \nu_1(z) dz + \frac{1}{2 \epsilon^2} \sigma_1^2(x, y) (\partial_y u(x, y))^2.$$  

(13)
In systems with averaging under the law of large number scaling we can identify the limiting non-linear operator $\overline{H}_0$ as the solution to an eigenvalue problem for the driving process $Y^\epsilon$ obtained from $Y_\epsilon$ with $X_\epsilon = x$ and $\epsilon = 1$.

We first identify $u_0$, the limit of $u_\epsilon$ as $\epsilon \to 0$, using heuristic arguments. Assume

$$u_\epsilon(t, x, y) = u_0(t, x) + \epsilon u_1(t, x, y) + \epsilon^2 u_2(t, x, y) + \ldots. \quad (14)$$

Using the $\epsilon$ expansion of $u_\epsilon$, ($14$), in equation ($12$), and collecting terms of $O(1)$, we get

$$\partial_t u_0(t, x) = b(x, y)\partial_x u_0(t, x) + \frac{1}{2}\sigma^2(x, y)(\partial_x u_0(t, x))^2$$

$$+ \int \left( e^{\partial_x u_0(t, x)k(x, y, z)} - 1 - k(x, y, z)\partial_x u_0(t, x) \right) \nu(z)dz$$

$$+ \rho\sigma(x, y)\sigma_1(x, y)\partial_x u_0(t, x)\partial_y u_1(t, x, y) + b_1(x, y)\partial_y u_1(t, x, y) + \frac{1}{2}\sigma_1^2(x, y)\partial_{yy}^2 u_1(t, x, y)$$

$$+ \int \left( e^{u_1(t, x, y + k_1(x, y, z) - u_1(t, x, y))} - 1 - k_1(x, y, z)\partial_y u_1(t, x, y) \right) \nu_1(z)dz$$

$$+ \frac{1}{2}\sigma_1^2(x, y)(\partial_y u_1(t, x, y))^2. \quad (15)$$

Denote $\partial_x u_0(t, x)$ by $p$ and $\partial_t u_0(t, x)$ by $\lambda$. Fix $t, x$ and hence $p$ and $\lambda$. Using the perturbed $\mathcal{L}_1$ generator ($9$)

$$\mathcal{L}_1^{x,p}f(y) = [\rho\sigma(x, y)\sigma_1(x, y)p + b_1(x, y)]\partial_y f(y) + \frac{1}{2}\sigma_1^2(x, y)\partial_{yy}^2 f(y)$$

$$+ \int \left( f(y + k_1(x, y, z)) - f(y) - k_1(x, y, z)\partial_y f(y) \right) \nu_1(z)dz.$$ 

and

$$V(y; x, p) := b(x, y)p + \frac{1}{2}\sigma^2(x, y)p^2 + \int \left( e^{p(k(x, y, z))} - 1 - k(x, y, z)p \right) \nu(z)dz.$$ 

the equation ($15$) can be written as an eigenvalue problem:

$$(\mathcal{L}_1^{x,p} + V(y; x, p))e^{u_1} = \lambda e^{u_1}, \quad (16)$$

Note that the eigenvalue $\lambda$ depends on $x$ and $p$, and that if we write $\overline{H}_0(x, p) := \lambda$ then $u_0$ satisfies

$$\partial_t u_0(t, x) = \overline{H}_0(x, \partial_x u_0(t, x)),$$

where $\overline{H}_0(x, p)$ is the eigenvalue $\lambda$ in ($16$). By the expansion ($14$), it is clear that $u_0(0, x) = h(x)$.

The approach of [FK06] for obtaining the large deviation principle is to prove convergence of nonlinear semigroups associated with the nonlinear operators $H_\epsilon$. In [FK06] the first step is identifying the limit operator $\overline{H}_0$. Existence and uniqueness of the limiting semigroup is obtained by verifying the ‘range condition’ for the limit operator. This amounts to showing existence of solutions to the equation $(I - \alpha \overline{H}_0) f = h$ for small enough $\alpha > 0$ and sufficiently large class of functions $h$. Since the range condition is difficult to verify, a viscosity method
approach is adopted and the range condition is replaced with a comparison principle condition for \((I - \alpha H_0)f = h\). In the viscosity method, existence of the limiting semigroup is by construction, while uniqueness is obtained via the comparison principle.

The approach in this paper uses convergence of viscosity solutions to the Cauchy problem for PIDEs \((12)\), and to show existence and uniqueness of the limit one then needs to verify the comparison principle for the Cauchy problem \(\partial_t u_0(t, x) = \overline{H}_0(x, \partial_x u_0(t, x))\), with \(u_0(0, x) = h(x)\).

In the proof of the comparison principle we will also use a Donsker-Varadhan variational representation \((18)\) for \(\overline{H}_0\) as follows. Let \(\mathcal{P}(\mathbb{R})\) denote the space of probability measures on \(\mathbb{R}\). Define the rate function \(J(\mu; x, p) : \mathcal{P}(\mathbb{R}) \to \mathbb{R} \cup \{+\infty\}\) by

\[
J(\mu; x, p) := - \inf_{g \in D^{++}(\mathcal{L}_1^{x,p})} \int_{\mathbb{R}} \frac{\mathcal{L}_1^{x,p} g}{g} d\mu,
\]

where \(D^{++}(\mathcal{L}_1^{x,p}) \subset C_0(\mathbb{R})\) denotes the domain of \(\mathcal{L}_1^{x,p}\) with functions that are strictly bounded below by a positive constant. Then \([DV75]\) implies that the principal eigenvalue \(\overline{H}_0(x, p) = \lambda\) in \([16]\) is also given by

\[
\overline{H}_0(x, p) = \sup_{\mu \in \mathcal{P}(\mathbb{R})} \left( \int V(y; x, p) d\mu(y) - J(\mu; x, p) \right),
\]

where \(V(y; x, p) = b(x, y)p + \frac{1}{2} \sigma^2(y)p^2 + \int (e^{p_k(x,y,z)} - 1) \nu(z) dz\).

**Remark 3.1.** In the special case \(Y^{x,p}\) also has a reversible invariant measure \(\pi^p(x, \cdot)\), we can use the Dirichlet form representation for \(J\). Define the Dirichlet form associated with \(Y^{x,p}\) by

\[
\mathcal{E}^{x,p}(f, g) := \int f(y) \mathcal{L}_1^{x,p} g(y) d\pi^p(x, dy).
\]

Then, Theorem 7.44 in Stroock \([Stro84]\) implies that

\[
J(\mu; x, p) = \begin{cases} 
\mathcal{E}^{x,p} \left( \sqrt{\frac{d\mu}{d\pi^p(x, \cdot)}}, \sqrt{\frac{d\mu}{d\pi^p(x, \cdot)}} \right) & \text{if } \mu(\cdot) \ll \pi^p(x, \cdot) \\
+\infty & \text{if } \mu(\cdot) \not\ll \pi^p(x, \cdot).
\end{cases}
\]

The variational formula \((15)\) then reduces to the classical Rayleigh-Ritz formula

\[
\overline{H}_0(x, p) = \sup_{f \in L^2(\pi^p), \|f\| = 1} \left( \int V(y; x, p) f(y) d\pi^p(x, y) dy + \langle \mathcal{L}_1^{x,p} f, f \rangle \right).
\]

To sum up, we will prove that:

**Lemma 1.** Let \(\overline{H}_0\) be as defined in \((18)\), and suppose the comparison principle holds for the nonlinear Cauchy problem:

\[
\begin{align*}
\partial_t u_0(t, x) &= \overline{H}_0(x, \partial_x u_0(t, x)), \quad \text{for} \quad (t, x) \in (0, T] \times \mathbb{R}; \\
u_0(0, x) &= h(x)\end{align*}
\]

Under the Assumptions \((2.3)\) the sequence of functions \(\{u^\epsilon_0\}_{\epsilon > 0}\) defined in \((11)\) converges uniformly over compact subsets of \([0, T] \times \mathbb{R} \times \mathbb{R}\) as \(\epsilon \to 0\) to the unique continuous viscosity solution \(u^\epsilon_0\) of \((21)\).
Lemma 2. The sequence of processes \( \{X_{\epsilon,t}\}_{\epsilon>0} \) is exponentially tight.

Theorem 3. Let \( X_{\epsilon,0} = x_0 \), and suppose all the Assumptions from Lemma 1 hold. Then, \( \{X_{\epsilon,t}\}_{\epsilon>0} \) satisfies a large deviation principle with speed \( 1/\epsilon \) and good rate function

\[
I(x,x_0,t) = \sup_{h \in C_b(\mathbb{R})} \{ h(x) - u_0^h(t,x_0) \}.
\] (22)

Proof. By Bryc’s theorem (Theorem 4.4.2 in [DeZ98]), Lemmas 1 and 2 give us a large deviation principle for \( \{X_{\epsilon,t}\}_{\epsilon>0} \) as \( \epsilon \to 0 \) with speed \( 1/\epsilon \) and good rate function given by (22). \( \square \)

Remark 3.2. Standard theory of comparison principles for viscosity solutions (Theorem 3.7 and Remark 3.8 in Chapter II of [BD97]) implies that the comparison principle holds for (21) as soon as \( \overline{H}_0 \) is uniformly continuous in \( x,p \) on compact sets (see Lemma 9 of the Appendix). In some cases \( \overline{H}_0 \) can be explicitly calculated (see Example 4.2) and continuity directly verified. In other cases one may need to resort to proving that the expression as on the right-hand side of (A.37) is non-positive, using the specifics for the case at hand.

In very special cases, we can further simplify the expression for the rate function:

Corollary 4. If the coefficients in the SDE (11) are independent of \( x \), then \( \overline{H}_0(x,p) \) becomes \( \overline{H}_0(p) \) and by Lemma D.1 in [FFK12], we get

\[
I(x;x_0,t) = t \overline{\mathcal{L}}_0 \left( \frac{x_0 - x}{t} \right), \tag{23}
\]

where \( \overline{\mathcal{L}}_0(\cdot) \) is the Legendre transform of \( \overline{\mathcal{H}}_0(\cdot) \).

The proof of Lemma 1 takes up the bulk of the paper, and consists of the following steps.

(Sec 3.1) • By taking appropriate limits of solutions \( u_h^\epsilon \) to the Cauchy problem (12) we construct upper-semicontinuous and lower-semicontinuous functions \( \overline{u}^h \) and \( \underline{u}^h \), respectively;
• Using an indexing set \( \alpha \in \Lambda \), we construct a family of operators \( H_0(\cdot;\alpha) \) and \( H_1(\cdot;\alpha) \), in such a way that the upper-semicontinuous function \( \overline{u}^h \) is a subsolution to the Cauchy problem for the operator \( \inf_{\alpha \in \Lambda} \{H_0(\cdot;\alpha)\} \), and the lower-semicontinuous function \( \underline{u}^h \) is a supersolution to the Cauchy problem for the operator \( \sup_{\alpha \in \Lambda} \{H_1(\cdot;\alpha)\} \).

(Sec 3.2) • We prove a comparison principle between subsolutions of \( \inf_{\alpha \in \Lambda} \{H_0(\cdot;\alpha)\} \) and supersolutions \( \sup_{\alpha \in \Lambda} \{H_1(\cdot;\alpha)\} \); above;
• We show that this comparison principle implies convergence of solutions \( u_h^\epsilon \) to the Cauchy problem (12) for \( H_\epsilon \) to solutions \( u_0^h \) to the Cauchy problem (21) for \( \overline{H}_0 \).

The proof of Lemma 2 uses the estimates obtained in the proof of Lemma 1 (Section 3.3).
3.1 Convergence of viscosity solutions of PIDEs

In Lemma 1 we use notions of viscosity solutions, subsolutions and supersolutions. For the standard meaning of these terms, as well as for the definition of the comparison principle, we refer the reader to Definition 4.1 in [FFK12]. Their extension to partial integro-differential equations (PIDEs) was obtained already in [Alv96] and can be found in [Bar08].

The proof of convergence of $u_h^\epsilon$ to $u_h^0$ follows the same steps as Lemma 4.1 in [FFK12] which carries over directly to viscosity solutions of PIDEs. Because we will need to verify that the conditions there are met, we restate Lemma 4.1 from [FFK12] for viscosity solutions of PIDEs.

Consider a class of compact sets in $\mathbb{R} \times \mathbb{R}$ defined by

$$Q := \{ K \times \tilde{K} : \text{compact } K, \tilde{K} \subset \subset \mathbb{R} \}.$$ 

Let $\Lambda$ be some indexing set, and suppose we have a family of continuous non-linear operators

$$H_i(x,p,P;\alpha) : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}, \quad \alpha \in \Lambda, i = 0, 1;$$

$$H_\epsilon(z,p,P) : \mathbb{R}^2 \times \mathbb{R}^2 \times M_{2\times 2} \mapsto \mathbb{R}.$$ 

For each $f \in C^2(\mathbb{R}^2)$, let $\nabla f(x,y) \in \mathbb{R}^2$ denote its gradient, $D^2 f(x,y) \in M_{2\times 2}$ the Hessian matrix evaluated at $(x,y)$. Define a sequence of integro-differential operators

$$H_\epsilon f(x,y) := H_\epsilon(x,y,\nabla f(x,y),D^2 f(x,y)),$$

on the domains

$$D_{\epsilon,+} := \{ f : f \in C^2(\mathbb{R}^2), \lim_{r \to \infty} \inf_{|z| > r} f(z) = +\infty \}$$

$$D_{\epsilon,-} := \{ -f : f \in C^2(\mathbb{R}^2), \lim_{r \to \infty} \inf_{|z| > r} f(z) = +\infty \},$$

and define domains $D_+, D_-$ analogously replacing $\mathbb{R}^2$ by $\mathbb{R}$.

Let $u_h^\epsilon$ be the viscosity solution of the Cauchy problem $\partial_t u = H_\epsilon u$ for the above operator $H_\epsilon$, with initial value $h$, and define

**Definition 3.1.**

$$u_h^\epsilon(t,x) := \sup\{ \limsup_{\epsilon \to 0^+} u_h^\epsilon(t_\epsilon,x_\epsilon,y_\epsilon) : \exists (t_\epsilon,x_\epsilon,y_\epsilon) \in [0,T] \times K \times \tilde{K}, \quad (t_\epsilon,x_\epsilon) \to (t,x), K \times \tilde{K} \subset \subset Q \},$$

$$u_h^0(t,x) := \inf\{ \liminf_{\epsilon \to 0^+} u_h^\epsilon(t_\epsilon,x_\epsilon,y_\epsilon) : \exists (t_\epsilon,x_\epsilon,y_\epsilon) \in [0,T] \times K \times \tilde{K}, \quad (t_\epsilon,x_\epsilon) \to (t,x), K \times \tilde{K} \subset \subset Q \}.$$ 

Define $\overline{u}^h$ to be the upper semicontinuous regularization of $u_h^\epsilon$, and $\underline{u}^h$ the lower semicontinuous regularization of $u_h^\epsilon$.

Finally, define the limiting operators for the family $\{H_i(\cdot;\alpha)\}_{\alpha \in \Lambda, i=0,1}$

$$H_0(x,p,P) := \inf_{\alpha \in \Lambda} H_0(x,p,P;\alpha),$$

$$H_1(x,p,P) := \sup_{\alpha \in \Lambda} H_1(x,p,P;\alpha).$$ 

Suppose the following conditions hold:
Condition 3.1 (limsup convergence of operators). For each $f_0 \in D_+$ and $\alpha \in \Lambda$, there exists $f_{0, \epsilon} \in D_{\epsilon, +}$ (which may depend on $\alpha$) such that

1. for each $c > 0$, there exists $K \times \tilde{K} \in Q$ satisfying
   \[
   \{(x, y) : H_\epsilon f_{0, \epsilon}(x, y) \geq -c\} \cap \{(x, y) : f_{0, \epsilon}(x, y) \leq c\} \subset K \times \tilde{K};
   \]

2. for each $K \times \tilde{K} \in Q$,
   \[
   \lim_{\epsilon \to 0} \sup_{(x, y) \in K \times \tilde{K}} |f_{0, \epsilon}(x, y) - f_0(x)| = 0; \tag{24}
   \]

3. whenever $(x_\epsilon, y_\epsilon) \in K \times \tilde{K} \in Q$ satisfies $x_\epsilon \to x$,
   \[
   \lim_{\epsilon \to 0} \sup_{(x, y) \in K \times \tilde{K}} |f_1(x) - f_{1, \epsilon}(x, y_\epsilon)| = 0; \tag{25}
   \]

Condition 3.2 (liminf convergence of operators). For each $f_1 \in D_-$ and $\alpha \in \Lambda$, there exists $f_{1, \epsilon} \in D_{\epsilon, -}$ (which may depend on $\alpha$) such that

1. for each $c > 0$, there exists $K \times \tilde{K} \in Q$ satisfying
   \[
   \{(x, y) : H_\epsilon f_{1, \epsilon}(x, y) \leq c\} \cap \{(x, y) : f_{1, \epsilon}(x, y) \geq -c\} \subset K \times \tilde{K};
   \]

2. for each $K \times \tilde{K} \in Q$,
   \[
   \lim_{\epsilon \to 0} \sup_{(x, y) \in K \times \tilde{K}} |f_1(x) - f_{1, \epsilon}(x, y)| = 0; \tag{26}
   \]

3. whenever $(x_\epsilon, y_\epsilon) \in K \times \tilde{K} \in Q$, and $x_\epsilon \to x$,
   \[
   \liminf_{\epsilon \to 0} H_\epsilon f_{1, \epsilon}(x_\epsilon, y_\epsilon) \geq H_1(x, \nabla f_1(x), D^2 f_1(x); \alpha). \tag{27}
   \]

In this case the following convergence results for $u_\epsilon$ as $\epsilon \to 0$ hold.

Lemma 5. Suppose the viscosity solutions $u_\epsilon^h$ to the partial integro-differential equation

\[
\partial_t u = H_\epsilon u, \quad u(0, x) = h(x)
\]

are uniformly bounded, $\sup_{\epsilon > 0} \|u_\epsilon^h\| < \infty$. Then, under Condition 3.1, $u_\epsilon^h$ is a subsolution of

\[
\partial_t u(t, x) \leq H_0(x, \nabla u(t, x), D^2 u(t, x)) \tag{26}
\]

and, under Condition 3.2, $u_\epsilon^h$ is a supersolution of

\[
\partial_t u(t, x) \geq H_1(x, \nabla u(t, x), D^2 u(t, x)). \tag{27}
\]

with the same initial conditions.
As the proof is the same as the proof of Lemma 4.1 in [FKK12] we omit it here. We do need to check Conditions 3.1 and 3.2 hold for our problem. This involves identifying the right indexing set $\Lambda$, the family of operators $H_0(\cdot;\alpha)$ and $H_1(\cdot;\alpha)$, and the appropriate test functions $f_{0,\varepsilon}$ and $f_{1,\varepsilon}$, for each given $f_0$ and $f_1$, respectively.

**Verifying Condition 3.1** As in [FKK12], we let

$$\Lambda := \{ (\xi, \theta) : \xi \in C^2_c(\mathbb{R}), 0 < \theta < 1 \}$$

and we define the sequence of operators $H_\varepsilon$ as in (13) on the domain

$$D_+ := \{ f \in C^2(\mathbb{R}) : f(x) = \phi(x) + \gamma \log(1 + x^2); \phi \in C^2_c(\mathbb{R}), \gamma > 0 \}.$$  

Define the family of operators $H_0(x, p; \xi, \theta)$ for $(\xi, \theta) \in \Lambda$ by

$$H_0(x, p; \xi, \theta) := \sup_{y \in \mathbb{R}} \{ b(y, p) + 1 \sigma^2(x, y)p^2 + \int (e^{\nu(x, y)} - 1 - pk(x, y, z)) \nu(z)dz \}$$

$$+ (1 - \theta)e^{-\xi}L^{x,p}_1e^{\xi}(y) + \theta e^{-\xi}L^{x,p}_1e^{\xi}(y).$$

(28)

For any $f \in D_+$ and $(\xi, \theta) \in \Lambda$ define a sequence of functions

$$f_{0,\varepsilon}(x, y) := f(x) + \varepsilon g(y), \quad \text{where} \quad g(y) := (1 - \theta)\xi(y) + \theta \zeta(y),$$

and $\zeta$ is the Lyapunov function on $\mathbb{R}$ satisfying Assumption 2.4. Then,

$$H_\varepsilon f_{0,\varepsilon}(x, y) = b(x, y)\partial_x f(x) + \frac{1}{2} \sigma^2(x, y)(\partial_x f(x))^2 + \varepsilon \left( b_0(x, y)\partial_x f(x) + \frac{1}{2} \sigma^2(x, y)\partial_{xx}^2 f(x) \right)$$

$$+ \int \left( e^{\frac{f(x + k(x, y, z)) - f(x, y)}{\varepsilon}} - 1 - pk(x, y, z)\partial_x f(x) \right) \nu(z)dz + e^{-\xi}L^{x,p}_1e^{\xi}(y)$$

$$\leq b(x, y)\partial_x f(x) + \frac{1}{2} \sigma^2(x, y)(\partial_x f(x))^2 + \varepsilon \left( b_0(x, y)\partial_x f(x) + \frac{1}{2} \sigma^2(x, y)\partial_{xx}^2 f(x) \right)$$

$$+ \int \left( e^{\frac{f(x + k(x, y, z)) - f(x, y)}{\varepsilon}} - 1 - pk(x, y, z) \right) \nu(z)dz$$

$$+ (1 - \theta)e^{-\xi}L^{x,p}_1e^{\xi}(y) + \theta e^{-\xi}L^{x,p}_1e^{\xi}(y).$$

(29)

so, for any sequence $(x_\varepsilon, y_\varepsilon)$ such that $x_\varepsilon \to x$

$$\limsup_{\varepsilon \to 0} H_\varepsilon f_{0,\varepsilon}(x_\varepsilon, y_\varepsilon) \leq H_0(x, \partial_x f(x); \xi, \theta),$$

thus verifying Condition 3.1 holds.

By choice of $D_+$, $f \in D_+$ has compact level sets in $\mathbb{R}$. Also note that $||\partial_x f|| + ||\partial_{xx} f|| < \infty$. Assumption 2.4 ensures that $-H_\varepsilon f(x, \cdot)$ has compact level sets for all $x$ in compact sets. This proves Condition 3.1 holds. Condition 3.1 is obvious by choice of functions $f_{0,\varepsilon}$.

**Verifying Condition 3.2** is exactly the same as verifying Condition 3.1 except that the sequence of operators $H_\varepsilon$ are now defined on the domain

$$D_- := \{ f \in C^2(\mathbb{R}) : f(x) = \phi(x) - \gamma \log(1 + x^2); \phi \in C^2_c(\mathbb{R}), \gamma > 0 \};$$
the family of operators $H_1(x,p;\xi,\theta)$ for $(\xi, \theta) \in \Lambda$ is defined by

$$H_1(x,p;\xi,\theta) := \inf_{y \in \mathbb{R}} \{ b(x,y)p + \frac{1}{2}\sigma^2(x,y)p^2 + \int \left( e^{pk(x,y,z)} - 1 - pk(x,y,z) \right) \nu(z)dz + (1+\theta)\epsilon^{-\xi}\mathcal{L}_{1}^{p,\epsilon}(y) - \theta\epsilon^{-\xi}\mathcal{L}_{1}^{p,\epsilon}(y) \};$$

and for any $f \in D_-$ and $\xi, \theta \in \Lambda$ the sequence $f_{1,\epsilon}$ is defined as

$$f_{1,\epsilon}(x,y) := f(x) + \epsilon g(y), \text{ for } g(y) := (1+\theta)\xi(y) - \theta\xi(y),$$

so that for any sequence $(x_\epsilon, y_\epsilon)$ such that $x_\epsilon \to x$ we now have

$$\liminf_{\epsilon \to 0} H_{1,\epsilon}(x_\epsilon, y_\epsilon) \geq H_1(x, \partial_x f(x); \xi, \theta)$$

verifies Condition 3.2.3 holds. Conditions 3.2.1 and 3.2.2 hold by the same arguments as above.

### 3.2 Comparison Principle

The rest of the claim of Lemma 1 requires proving uniqueness of solutions to $\partial_t u = \overline{H}_0 u$, with initial value $h$. This can be verified using the comparison principle on the subsolutions and supersolutions of the constructed limiting operators $H_0$ and $H_1$, and the variational representation of $\overline{H}_0$ from (18). We use the following Lemma 4.2 from [FFK12].

**Lemma 6.** Let $\underline{u}^h$ and $\overline{u}^h$ be defined as in Definition 3.4. If a comparison principle between subsolutions of (26) and supersolutions of (27) holds, that is, if every subsolution $v_1$ of (26) and every supersolution $v_2$ of (27) satisfy $v_1 \leq v_2$, then $\underline{u}^h = \overline{u}^h$ and $u^h_{\epsilon}(t,x,y) \to u^h_{0}(t,x)$, where $u^h_{0} := \underline{u}^h = \overline{u}^h$, as $\epsilon \to 0$, uniformly over compact subsets of $[0,T] \times \mathbb{R} \times \mathbb{R}$.

**Proof.** The comparison principle gives $\underline{u}^h \leq \overline{u}^h$, while by construction we have $\underline{u}^h \leq \overline{u}^h$. This gives uniform convergence of $u^h_{\epsilon} \to u_0 := \overline{u}^h = \underline{u}^h$ over compact subsets of $[0,T] \times \mathbb{R} \times \mathbb{R}$. \qed

We next prove the comparison principle for subsolutions of (26) and supersolutions of (27), that is every sub solution of

$$\partial_t u(t,x) \leq H_0(x,p) := \inf_{0<\theta<1,\xi \in C^2_\mathbb{R}} H_0(x,p;\theta,\xi),$$

where $H_0$ is as defined in (28), is less than or equal to every super solution of

$$\partial_t u(t,x) \geq H_1(x,p) := \sup_{0<\theta<1,\xi \in C^2_\mathbb{R}} H_1(x,p;\theta,\xi)$$

where $H_1$ is as defined in (30). We follow the steps in Section 5.2 in [FFK12] with some modifications. The key step is proving

**Operator Inequality:**

$$\inf_{0<\theta<1,\xi \in C^2_\mathbb{R}} H_0(x,p;\theta,\xi) \leq \overline{H}_0(x,p) \leq \sup_{0<\theta<1,\xi \in C^2_\mathbb{R}} H_1(x,p;\theta,\xi), \quad (31)$$

where $\overline{H}_0(x,p)$ is as defined in (18).
Recall the definition of the rate function $J$ from (17) and variational representation of $\overline{H}_0$ as

$$\overline{H}_0(x, p) = \sup_{\mu \in \mathcal{P}(\mathbb{R})} \left( \int V(y; x, p) d\mu(y) - J(\mu; x, p) \right).$$

Following steps of Lemma 11.35 of [FK06] (which relies on Assumption 2.3) we get that

$$\inf_{0 < \theta < 1, \xi \in C^2_{c}(\mathbb{R})} H_0(x, p : \theta, \xi) \leq \overline{H}_0(x, p).$$

From the proof of Lemma B.10 in [FK06], we have

$$\sup_{0 < \theta < 1, \xi \in C^2_{c}(\mathbb{R})} H_1(x, p : \theta, \xi) \geq \inf_{\mu \in \mathcal{P}(\mathbb{R})} \liminf_{t \to \infty} t^{-1} \ln E^\mu \left[ e^{\int_0^t V(Y_{s}^{x,p}; x,p)ds} \right].$$

Thus, we need to show that, irrespective of the initial distribution,

$$\liminf_{t \to \infty} t^{-1} \ln E \left[ e^{\int_0^t V(Y_{s}^{x,p}; x,p)ds} \right] \geq \overline{H}_0(x, p).$$

The proof of this claim depends on the Assumption 2.3. We define the occupation measures of the $Y^{x,p}$ process:

$$\mu_{t}^{x,p}(\cdot) := \frac{1}{t} \int_0^t 1_{Y_{s}^{x,p}}(\cdot)ds.$$  

Recall that $\mathcal{P}(\mathbb{R})$ is a separable metric space under the Prokhorov metric and that weak convergence of measures is equivalent to convergence in the Prokhorov metric. Let $Q_{t,y_0}$ denote the probability measure on $\mathcal{P}(\mathbb{R})$ induced by the occupation measure $\mu_t$ of $Y$ when $Y_0 = y_0$. In other words, for $A \in \mathcal{B}(\mathcal{P}(\mathbb{R}))$ (the borel sigma-algebra on $\mathcal{P}(\mathbb{R}))$,

$$Q_{t,y_0}(A) = P(\mu_t(\cdot) \in A| Y_0 = y_0).$$

**Lemma 7.** $\inf_{\mu \in \mathcal{P}(\mathbb{R})} \liminf_{t \to \infty} t^{-1} \ln E^\mu \left[ e^{\int_0^t V(Y_{s}^{x,p}; x,p)ds} \right] \geq \overline{H}_0(x, p).$

**Proof.** Define $\phi : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ by $\phi(\mu) = \int V(y; x, p)\mu(dy)$. Take $\tilde{\nu} \in \mathcal{P}(\mathbb{R})$, and let $B(\tilde{\nu}, r)$ denote the open ball in $\mathcal{P}(\mathbb{R})$ of radius $r$, centered at $\tilde{\nu}$. Fix $\nu \in \mathcal{P}(\mathbb{R})$, then there exists a compact set $K$ in $\mathbb{R}$ such that $\nu(K) > 0$. The key ingredient in the proof is the uniform LDP lower bound for the occupation measures:

$$\liminf_{t \to \infty} \frac{1}{t} \log \left[ \inf_{y_0 \in K} Q_{t,y_0}(B(\tilde{\nu}, r)) \right] \geq -J(\tilde{\nu}; x, p).$$  

This is obtained from Theorem 5.5 in [DV83] under Assumption 2.3. While the statement of Theorem 5.5 in [DV83] is in terms of a process level LDP, by the contraction principle it ensures the uniform LDP lower bound (32) for the occupation measures $\mu_{t}^{x,p}$.  

---

14
We now compute
\[
\liminf_{t \to \infty} \frac{1}{t} \log E^\nu \left[ e^{\int_0^t V(Y^x; x, p) \, ds} \right] = \liminf_{t \to \infty} \frac{1}{t} \log E^\nu \left[ e^{\phi(x, p)} \right]
\]
\[
\geq \liminf_{t \to \infty} \frac{1}{t} \log E^\nu \left[ e^{\phi(x, p)} 1_{\{Y_0 \in K\}} \right]
\]
\[
\geq \liminf_{t \to \infty} \frac{1}{t} \log \inf_{y_0 \in K} E^{y_0} \left( e^{\phi(x, p)} \right) + \liminf_{t \to \infty} \frac{1}{t} \log \nu(K)
\]
\[
= \liminf_{t \to \infty} \frac{1}{t} \log \inf_{y_0 \in K} \mu \left( \int_{\mu \in \mathcal{P}([0, \infty])} e^{\phi(\mu)} dQ_{t,y_0}(\mu) \right)
\]
\[
\geq \inf_{\mu \in B(\nu, r)} \phi(\mu) + \liminf_{t \to \infty} \frac{1}{t} \log \inf_{y_0 \in K} Q_{t,y_0} (B(\nu, r))
\]
\[
\geq \inf_{\mu \in B(\nu, r)} \phi(\mu) - J(\nu; x, p)
\]
by (32). By Lemma 8 (see Appendix), \( \phi \) is a lower semi-continuous function, and so \( \phi(\tilde{\nu}) \leq \lim_{r \to 0} \inf_{\mu \in B(\tilde{\nu}, r)} \phi(\mu) \). Thus taking limit as \( r \to 0 \) we get
\[
\liminf_{t \to \infty} \frac{1}{t} \log E^\nu \left[ e^{\int_0^t V(Y^x; x, p) \, ds} \right] \geq \phi(\tilde{\nu}) - J(\tilde{\nu})
\]
(note that since \( V \) is bounded below, \( \phi(\mu) > -\infty \), and so \( \phi(\tilde{\nu}) - J(\tilde{\nu}; x, p) \) is well defined and not \(-\infty + \infty\)). Since \( \tilde{\nu} \) is arbitrary, we get
\[
\liminf_{t \to \infty} \frac{1}{t} \log E^\nu \left[ e^{\int_0^t V(Y^x; x, p) \, ds} \right] \geq \sup_{\nu \in \mathcal{P}([0, \infty])} \{ \phi(\tilde{\nu}) - J(\tilde{\nu}; x, p) \}.
\]
This holds for every \( \nu \in \mathcal{P}([0, \infty]) \) and so
\[
\inf_{\nu \in \mathcal{P}([0, \infty])} \liminf_{t \to \infty} \frac{1}{t} \log E^\nu \left[ e^{\int_0^t V(Y^x; x, p) \, ds} \right] \geq \sup_{\nu \in \mathcal{P}([0, \infty])} \{ \phi(\tilde{\nu}) - J(\tilde{\nu}; x, p) \}.
\]
This concludes the proof of the Operator Inequality (31). \( \square \)

**Remark 3.3.** In the special case \( Y^{x, p} \) also has a reversible invariant measure \( \pi^p(x, \cdot) \) we could also follow the arguments for Lemma 5.4 in [FFK12] using the Dirichlet form representation of \( J \) (19).

**Proof of Lemma 1.** By Lemma 5 and Operator Inequality (31), it follows that \( \overline{u}^h \) is a subsolution and \( u^h \) a supersolution of the Cauchy problem (21): \( \partial_t u(t, x) = \overline{\mathcal{D}}(x, \partial_x u(t, x)) \) with \( u(0, x) = h(x) \). If the comparison principle holds for the Cauchy problem (21), then Lemma 5 gives us \( u^h = \overline{u}^h \) and that \( u^h_t \to u^h_0 = u^h = \overline{u}^h \) uniformly over compact subsets of \([0, T] \times \mathbb{R} \times \mathbb{R}\). \( \square \)
3.3 Exponential tightness

Proof of Lemma 2. We prove exponential tightness using the convergence of $H_\epsilon$ and appealing to supermartingale arguments (see Section 4.5 of [FK06]).

Let $f(x) := \ln(1 + x^2)$, so $f(x) \to \infty$ as $|x| \to \infty$, and also $||f'|| + ||f''|| < \infty$. Define $f_\epsilon(x, y) := f(x) + \epsilon \zeta(y)$ where $\zeta$ is the positive Lyapunov function satisfying Assumption 2.4. Then, for any $c > 0$, there exists a compact $K_c \subset \mathbb{R}$ such that $f_\epsilon(x, y) > c$, $\forall y \in \mathbb{R}$, $\forall x \notin K_c$.

Observe that by (29) (with $\theta = 1$)

\[
H_\epsilon f_\epsilon(x, y) = e^{-(f_\epsilon / \epsilon)} L_\epsilon e^{f_\epsilon / \epsilon} \\
\leq b(x, y) \partial_x f(x) + \frac{1}{2} \sigma^2(x, y)(\partial_x f(x))^2 + \epsilon \left( b_0(x, y) \partial_x f(x) + \frac{1}{2} \sigma^2(x, y) \partial^2_{xx} f(x) \right) \\
+ \int \left( e^{f(x + \epsilon(x, y), z) - f(x, y)} - 1 - \epsilon k(x, y, z) \right) \nu(z) dz + e^{-\zeta \partial_x f(x)} e^{\zeta(y)}.
\]

By boundedness of $||f'||$ and $||f''||$ and Assumption 2.4, we get there exists $C > 0$ such that

\[
\sup_{x \in \mathbb{R}, y \in \mathbb{R}} H_\epsilon f_\epsilon(x, y) \leq C < \infty, \forall \epsilon > 0.
\]

Since $e^{(f_\epsilon(X_{\epsilon,t}, Y_{\epsilon,t}) - f_\epsilon(X_{\epsilon,0}, Y_{\epsilon,0}) / \epsilon - \int_0^t H_\epsilon f_\epsilon(X_{\epsilon,s}, Y_{\epsilon,s})ds}$ is a non-negative local martingale, by optional stopping

\[
P(X_{\epsilon,t} \notin K_c) e^{(c - f_\epsilon(x_0, y_0) - tC) / \epsilon} \\
\leq E \left\{ \exp \left\{ \frac{f_\epsilon(X_{\epsilon,t}, Y_{\epsilon,t})}{\epsilon} - \frac{f_\epsilon(x_0, y_0)}{\epsilon} - \int_0^t H_\epsilon f_\epsilon(X_{\epsilon,s}, Y_{\epsilon,s}) ds \right\} \right\} \leq 1.
\]

Therefore for each $c > 0$

\[
e \ln P(X_{\epsilon,t} \notin K_c) \leq tC - f_\epsilon(x_0, y_0) - c
\]

As $C$ is fixed and independent of $c$ (which we can choose), $\{X_{\epsilon,t}\}_{\epsilon > 0}$ is exponentially tight.

Remark 3.4. A similar argument can be used to verify the exponential compact containment condition in Corollary 4.17 in [FK06], which would give us $\{X_{\epsilon,}\}_{\epsilon > 0}$ is exponentially tight.

4 Examples

4.1 Model for stock price with stochastic volatility

We consider the stochastic volatility model for stock price suggested by Barndorff-Nielson and Shephard [B-NS01]. Let $X_t$ denote the logarithm of stock price and $Y_t$ the stochastic volatility.

\[
dX_t = (r - \frac{1}{2} Y_t)dt + \sqrt{Y_t}dW_t \\
dY_t = -\frac{Y_t}{\delta}dt + dZ^{1/\delta}_t,
\]
where $W_t$ is a standard Brownian motion and $Z_t^{1/\delta}$ is an independent non-Gaussian Lévy process with intensity $\frac{1}{\delta} \nu(dz)$; the parameter $0 < \delta \ll 1$ denotes the mean-reversion time scale in stochastic volatility. The process $Z$ is often referred to as the background driving Lévy process (BDLP). If we are interested in pricing options on the stock which are close to maturity, we will only be interested in small-time asymptotics of the model. We thus scale time by a parameter $0 < \epsilon \ll 1$, to get

$$
\begin{align*}
&dX_{\epsilon,t} = \epsilon (r - \frac{1}{2} Y_{\epsilon,t}) dt + \sqrt{\epsilon} \sqrt{Y_{\epsilon,t}} dW_t \\
&dY_{\epsilon,t} = -\frac{\epsilon}{\delta} Y_{\epsilon,t} dt + dZ_{\epsilon,t}^{1/\delta},
\end{align*}
$$

(33)

The multi scale structure comes from the fast mean reversion in stochastic volatility and the small time to maturity. We are interested in the situation where time to maturity ($\epsilon$) is small, but large compared to mean-reversion time ($\delta$) of stochastic volatility. The interesting regime as seen in [FFK12] is when $\delta = \epsilon^2$. The generator of $(X_{\epsilon}, Y_{\epsilon})$ is given by:

$$
\mathcal{L}_\epsilon f(x, y) = \epsilon \left( (r - \frac{1}{2} y) \partial_x f(x, y) + \frac{1}{2} y \partial_x^2 f(x, y) \right) + \frac{1}{\epsilon} \left( -y \partial_y f(x, y) + \int (f(x, y + z) - f(x, y)) \nu(dz) \right),
$$

for $f \in C^2_b(\mathbb{R}^2)$.

For this example, since the coefficients are $x$-independent, the perturbed operator $\mathcal{L}_1^{x,p}$ is the same as $\mathcal{L}_1$, the generator of $Y$:

$$
\mathcal{L}_1 f(y) = -y f'(y) + \int (f(y + z) - f(y)) \nu(dz), \quad \text{for } f \in C^2_b(\mathbb{R}).
$$

We can obtain the limiting Hamiltonian $\bar{H}_0$ by solving the eigenvalue problem \cite{16}. Here $V(y; x, p) \equiv V(y; p) = \frac{1}{2} y p^2$. $\bar{H}_0(p)$ is the eigenvalue $\lambda$ of the eigenvalue problem

$$
-y f'(y) + \int (f(y + z) - f(y)) \nu(dz) + \frac{1}{2} y p^2 f(y) = \lambda f(y).
$$

Note that $f(y) = \frac{\sqrt{2}}{\lambda} y$ and $\lambda = \int \left( e^{\frac{p^2}{2\lambda^2}} - 1 \right) \nu(dz)$ satisfy the eigenvalue problem. So $\bar{H}_0(p) = \int \left( e^{\frac{p^2}{2\lambda^2}} - 1 \right) \nu(dz)$. In Barndorff-Nielsen and Shephard [B-NS01], the BDLP, $Z$, is assumed to have only positive increments. A simple example of such a Lévy process is a jump process taking finitely many jumps that is the Lévy measure is $\nu(z_i) > 0$ where $z_i > 0$, $i = 1, 2, \ldots, k$. (In this case $\zeta(y) = y^2$ satisfies the Lyapunov function assumption [24].) We can then explicitly compute $\bar{H}_0(p)$ and its Legendre transform $\bar{L}(p)$. As seen in [FFK12] (Lemma D.1 in [FFK12]), since $\bar{H}_0(p)$ is not state dependent, we get the rate function to be $I(x, x_0, t) = t \bar{L} \left( \frac{x_0 - x}{\sqrt{t}} \right)$. In finance, a common example is where $Z$ is a gamma process, in which case $\nu(dz) = \frac{a}{2} e^{-bz} dz$, $a, b > 0$. Then

$$
\bar{H}_0 = \begin{cases}
\alpha \ln \left( 1 + \frac{p^2}{2b-p^2} \right) & \text{if } -\sqrt{2b} < p < \sqrt{2b} \\
\infty & \text{if } p^2 > 2b,
\end{cases}
$$
Markov chain models that apply in stochastic reaction kinetics. This diffusion solves two possible values: it will be 0 when $y = 0$, and the rate function is given by $I(x) = t \tilde{L} \left( \frac{x_0 - x}{t} \right)$, where
\[
\tilde{L}(q) = \begin{cases} 
-a + \sqrt{a^2 + 2bq^2} - a \ln 2b + a \ln \left( \frac{-2a^2}{q^2} + \frac{2a}{q} \sqrt{a^2 + 2bq^2} \right) & \text{if } q > 0 \\
0 & \text{if } q = 0 \\
-a - \sqrt{a^2 + 2bq^2} - a \ln 2b - a \ln \left( \frac{-2a^2}{q^2} - \frac{2a}{q} \sqrt{a^2 + 2bq^2} \right) & \text{if } q < 0.
\end{cases}
\]
This rate function then gives the asymptotic behavior of a European Call option on the stock. Let $K$ denote the strike price and $S_{\epsilon,t} = e^{X_{\epsilon,t}}$, then for $S_0 = e^{x_0} < K$ (out-of-the-money call),
\[
\lim_{\epsilon \to 0} \epsilon \log E \left[ S_{\epsilon,t} - K \right]^- = -I(\log K; x_0, t),
\]
where maturity time $T = \epsilon t$. This follows from Corollary 1.3 in [FFF10].

4.2 Model for self-regulating protein production

The simplest model for translation of protein from DNA with is the system below, with a gene that is either in its “on” state $G_1$, or in its “off” state $G_0$, and in which the protein activates the changes from “off” to “on” state:
\[
\begin{align*}
(1) & \quad G_0 + P \xrightarrow{\kappa_1} G_1 + P \\
(2) & \quad G_1 + P \xrightarrow{\kappa_1} G_0 + P \\
(3) & \quad G_1 \xrightarrow{\kappa_3} G_1 + P \\
(4) & \quad P \xrightarrow{\kappa_3} 0
\end{align*}
\]
Suppose the amount of protein $P$ is of order $1/\epsilon$, whose rate of production $\kappa_2 = 1/\epsilon \kappa_2$, while its rate of degradation $\kappa_3 = \kappa_3$: where $\kappa_2, \kappa_3$ are of $O(1)$. The amount of genes in the “on”- and “off”-state is $\in \{0, 1\}$, their total amount always equaling 1, and suppose the rates of changes of the gene from the “on”-state to the “off”-state and back are very rapid due to its regulation by the large amounts of protein $\kappa_1 = \kappa_1, \kappa_{1-1} = \kappa_{-1}$, where $\kappa_1, \kappa_{-1}$ are of $O(1)$. This system is characteristic of eukaryotes, where the gene switching noise dominates over the transcriptional and translational noise. We can represent the changes in the system using the process $X_\epsilon$ for the count of protein molecules normalized by $\epsilon$, and $Y_\epsilon$ for the (unnormalized) count of “on”-gene molecules. A diffusion process is a good approximation for the evolution of $X_\epsilon$ as long as the count of proteins is not too small, that is, the unnormalized count is $\gg \epsilon$ and $X_\epsilon \sim O(1)$ ([KKP12] gives a rigorous justification of diffusion approximations for Markov chain models that apply in stochastic reaction kinetics). This diffusion solves $dX_{\epsilon,t} = b(X_{\epsilon,t}, Y_{\epsilon,t})dt + \sqrt{\sigma(X_{\epsilon,t}, Y_{\epsilon,t})}dW_t$ with drift $b(x, y) = \kappa_2 y - \kappa_3 x$ (protein production has only two possible values: it will be 0 when $y = 0$, or $\kappa_2$ when $y = 1$), with diffusion coefficient $\sigma^2(x, y) = \kappa_2 y + \kappa_3 x$, and initial value $X_{\epsilon,0} = x_0 > 0$. Changes in the amount of proteins due to other independent sources of noise, such as errors after cell splitting, can be modelled by an additional jump term for $X_\epsilon$ where the jump measure $\nu(dx)$ can be as simple as $\nu(z) = \frac{1}{2} \delta_{-1}(z) + \frac{1}{2} \delta_{+1}(z)$, producing
\[
dX_{\epsilon,t} = (\kappa_2 Y_{\epsilon,t} - \kappa_3 X_{\epsilon,t})dt + \sqrt{\epsilon(\kappa_2 Y_{\epsilon,t} + \kappa_3 X_{\epsilon,t})}dW_t + \epsilon \int 1_{X_{\epsilon,t} > z} \tilde{N}_1^z(dz, dt)
\]

The amount of genes $G_1$ in the “on”-state is a rapidly fluctuating two-state Markov chain $Y$ on $\{0, 1\}$ with rates $r_{0\to1}(x) = \frac{1}{\epsilon} \kappa_1 x$ and $r_{1\to0}(x) = \frac{1}{\epsilon} \kappa_{-1} x$ that depend on the normalized
amount of protein (note that the amount of genes $G_0$ in the “off”-state is $1 - Y$). This chain is reversible, and for each $x > 0$ it has a unique stationary distribution $\pi^x(1) = 1 - \pi^x(0) = \kappa_1/(\kappa_1 + \kappa_{-1})$.

Signalling proteins such as morphogens have to be in the right range of concentrations to avoid triggering the expression of genes at the wrong times. The probabilities of their amounts being out of range are given by the Large Deviation Principle for $X_\epsilon$ as $\epsilon \to 0$, for which we need to obtain the solution to the eigenvalue problem for the operator $V(y; x, p) + L^x$ where $L^x f(y) = r_{0 \to 1}(x)(f(y + 1) - f(y))1_{y=0} + r_{1 \to 0}(x)(f(y - 1) - f(y))1_{y=1}$.

In order to solve $(V(y; x, p) + L^x)e^{u_1} = \lambda e^{u_1}$ for $\lambda$, let $e^{u_1(x,1)} = a_1(x), e^{u_1(x,0)} = a_0(x)$, for some $a_1, a_0$ strictly positive functions. Then

$$(\kappa_2 - \kappa_3 x)pa_1(x) + (\kappa_2 + \kappa_3 x)p^2a_1(x) + \frac{1}{2}(e^p + e^{-p} - 2)a_1(x) + \kappa_{-1}x(a_0(x) - a_1(x)) = \lambda a_1(x)$$

$$-\kappa_3 xpa_0(x) + \kappa_3 xpa_1(x) + \frac{1}{2}(e^p + e^{-p} - 2)a_0(x) + \kappa_1x(a_1(x) - a_0(x)) = \lambda a_0(x)$$

equivalently, with $a(x) = a_1(x)/a_0(x)$,

$$(\kappa_2 - \kappa_3 x)p + (\kappa_2 + \kappa_3 x)p^2 + \kappa_{-1}x\left(\frac{1}{a(x)} - 1\right) = -\kappa_3xp + \kappa_3xp^2 + \kappa_1x(a(x) - 1)$$

which, since $a(x)$ has to be positive, gives

$$a(x) = \frac{-B + \sqrt{B^2 - 4AC}}{2A}, \quad A = \kappa_1x, \quad B = -\kappa_2p - \kappa_2p^2 + (\kappa_{-1} - \kappa_1)x, \quad C = -\kappa_{-1}x$$

and consequently, using notation above,

$$\tilde{H}_0(x, p) = -\kappa_3xp + \kappa_3xp^2 + \kappa_1x(a(x) - 1) + \frac{1}{2}(e^p + e^{-p} - 2)$$

Note that when $\kappa_{-1} = \kappa_1$ then

$$a(x) = \frac{\kappa_2p(1 + p) + \sqrt{(\kappa_2p(1 + p))^2 + (2\kappa_1x)^2}}{2\kappa_1x}$$

and

$$\tilde{H}_0(x, p) = -\kappa_3p(1 - p)x + \frac{1}{2}\kappa_2p(1 + p) + \frac{1}{2}\sqrt{(\kappa_2p(1 + p))^2 + (2\kappa_1x)^2} - \kappa_1x + \frac{1}{2}(e^p + e^{-p} - 2).$$

Note that $\tilde{H}_0(x, p)$ is both convex in $p$ and continuous in $x$.

If one were to use an approximation of the evolution of the normalized protein amount $X_\epsilon$ by a piecewise deterministic process then (without additional noise)

$$dX^{\text{PDMP}} = (\kappa_2Y_{\epsilon,t} - \kappa_3X_{\epsilon,t})dt$$

while $Y_\epsilon$ is the same fast Markov chain on $\{0, 1\}$. In this case $V(y; x, p) = (\kappa_2 - \kappa_3 x)p$ and the Hamiltonian (when $\kappa_1 = \kappa_{-1}$) becomes

$$\tilde{H}_0^{\text{PDMP}}(x, p) = -\kappa_3px + \frac{1}{2}\kappa_2p + \frac{1}{2}\sqrt{(\kappa_2p)^2 + (2\kappa_1x)^2} - \kappa_1x.$$

which is easy to compare to the Hamiltonian $\tilde{H}_0$ of the diffusion process $X_\epsilon$ taking into account the small perturbative noise arising from randomness in the timing of chemical reactions and from randomness in the outcomes of cell splitting.
A Appendix

Lemma 8. Fix \( x, p \in \mathbb{R} \) and let \( \phi: \mathcal{P}(\mathbb{R}) \to \mathbb{R} \) be defined by \( \phi(\mu) = \int V(y; x, p)\mu(dy) \). Then, \( \phi \) is a lower semi-continuous (l.s.c.) function on \( \mathcal{P}(\mathbb{R}) \).

Proof. For the rest of the proof, we will write \( V(y) \) for \( V(y; x, p) \). Let \( V_M := V \cdot 1_{V \leq M} + M \cdot 1_{V \geq M} \), for \( M \geq \inf_y V(y) \). To show that \( \phi(\mu) \) is l.s.c, it is sufficient to show that if \( \mu_n \rightharpoonup \mu \) weakly, then \( \phi(\mu) \leq \liminf_{n \to \infty} \phi(\mu_n) \). Assume \( \mu_n \rightharpoonup \mu \) weakly. Then
\[
\int V_M d\mu = \lim_{n \to \infty} \int V_M d\mu_n,
\]
by definition of weak convergence of measures, since \( V_M \) is a bounded function. By the monotone convergence theorem we get
\[
\phi(\mu) = \int V d\mu = \lim_{M \to \infty} \int V_M d\mu = \sup_{M} \lim_{n \to \infty} \int V_M d\mu_n = \liminf_{n \to \infty} \sup_{M} \int V_M d\mu_n = \liminf_{n \to \infty} \int V d\mu_n \]
by Monotone convergence theorem
\[
= \liminf_{n \to \infty} \phi(\mu_n)
\]
\[\square\]

Lemma 9. Let \( u_1 \) be a bounded, upper semicontinuous (u.s.c.), viscosity subsolution and \( u_2 \) a bounded, lower semicontinuous (l.s.c.), viscosity supersolution of \( \partial_t u(t, x) = H_0(x, \partial_x u(t, x)) \) respectively. If \( u_1(0, \cdot) \leq u_2(0, \cdot) \), and \( H_0 \) is uniformly continuous on compact sets, then \( u_1 \leq u_2 \) on \( [0, T] \times \mathbb{R} \) for any \( T > 0 \).

Proof. Suppose
\[
\sup_{t \leq T, x} \{u_1(t, x) - u_2(t, x)\} > A \geq \delta > 0. \tag{A.34}
\]
Let \( g(t, x) = x^2 + t^2 \). Define
\[
\psi(t, x, y) = u_1(t, x) - u_2(s, y) - \frac{1}{2} \ln \left(1 + \frac{|x - y|^2 + |t - s|^2}{\epsilon}\right) - \beta (g(t, x) + g(s, y)) - At.
\]
Fix \( \beta > 0 \) and let \((\bar{t}_\epsilon, \bar{x}_\epsilon, \bar{s}_\epsilon, \bar{y}_\epsilon)\) denote the point of maximum of \( \psi \) in \([0, T] \times \mathbb{R} \times [0, T] \times \mathbb{R}\) for \( \epsilon > 0 \). Since \( u_1, u_2 \) are bounded, for fixed \( \beta > 0 \), there exists an \( R_\beta > 0 \) such that \( |\bar{x}_\epsilon|, |\bar{y}_\epsilon| \leq R_\beta \) for all \( \epsilon > 0 \).
Using
\[ \psi(\bar{t}_\epsilon, \bar{x}_\epsilon, \bar{t}_\epsilon, \bar{x}_\epsilon) + \psi(\bar{s}_\epsilon, \bar{y}_\epsilon, \bar{s}_\epsilon, \bar{y}_\epsilon) \leq 2\psi(\bar{t}_\epsilon, \bar{x}_\epsilon, \bar{s}_\epsilon, \bar{y}_\epsilon), \]
we get
\[ \frac{1}{2} \ln \left( 1 + \frac{|\bar{x}_\epsilon - \bar{y}_\epsilon|^2 + |ar{t}_\epsilon - \bar{s}_\epsilon|^2}{\epsilon} \right) \leq A(\bar{s}_\epsilon - \bar{t}_\epsilon) + u_1(\bar{t}_\epsilon, \bar{x}_\epsilon) - u_1(\bar{s}_\epsilon, \bar{y}_\epsilon) + u_2(\bar{t}_\epsilon, \bar{x}_\epsilon) - u_2(\bar{s}_\epsilon, \bar{y}_\epsilon) \leq 2AT + 2||u_1|| + 2||u_2|| =: C < \infty, \]
which gives us
\[ |\bar{x}_\epsilon - \bar{y}_\epsilon|^2 + |\bar{t}_\epsilon - \bar{s}_\epsilon|^2 \leq \epsilon e^2C. \]

Therefore $|\bar{x}_\epsilon - \bar{y}_\epsilon|, |\bar{s}_\epsilon - \bar{t}_\epsilon| \to 0$ as $\epsilon \to 0$.

Let
\[ \phi_1(t, x) := u_2(\bar{s}_\epsilon, \bar{y}_\epsilon) + \frac{1}{2} \ln \left( 1 + \frac{|x - \bar{y}_\epsilon|^2 + |t - \bar{s}_\epsilon|^2}{\epsilon} \right) + \beta (g(t, x) + g(\bar{s}_\epsilon, \bar{y}_\epsilon)) + At \]
and
\[ \phi_2(s, y) := u_1(\bar{t}_\epsilon, \bar{x}_\epsilon) - \frac{1}{2} \ln \left( 1 + \frac{|x - y|^2 + |t - s|^2}{\epsilon} \right) - \beta (g(\bar{t}_\epsilon, \bar{x}_\epsilon) + g(s, y)) - A\bar{t}_\epsilon. \]

Then $(\bar{t}_\epsilon, \bar{x}_\epsilon)$ is a point of maximum of $u_1(t, x) - \phi_1(t, x)$ and $(\bar{s}_\epsilon, \bar{y}_\epsilon)$ is a point of minimum of $u_2(s, y) - \phi_2(s, y)$. Since $u_1$ and $u_2$ are sub and super solutions respectively, by the definition of sub and super solutions we get
\[ \frac{\bar{t}_\epsilon - \bar{s}_\epsilon}{1 + |x - \bar{y}_\epsilon|^2 + |t - \bar{s}_\epsilon|^2} + A + 2\beta \bar{t}_\epsilon \leq \bar{H}_0 \left( \bar{x}_\epsilon, \frac{x - \bar{y}_\epsilon}{1 + |x - \bar{y}_\epsilon|^2 + |t - \bar{s}_\epsilon|^2} + 2\beta \bar{x}_\epsilon \right), \tag{A.35} \]
and
\[ \frac{\bar{t}_\epsilon - \bar{s}_\epsilon}{1 + |x - \bar{y}_\epsilon|^2 + |t - \bar{s}_\epsilon|^2} - 2\beta \bar{s}_\epsilon \geq \bar{H}_0 \left( \bar{y}_\epsilon, \frac{\bar{x}_\epsilon - \bar{y}_\epsilon}{1 + |\bar{x}_\epsilon - \bar{y}_\epsilon|^2 + |\bar{t}_\epsilon - \bar{s}_\epsilon|^2} - 2\beta \bar{y}_\epsilon \right). \tag{A.36} \]

Subtracting (A.36) from (A.35), we get
\[ A + 2\beta (\bar{t}_\epsilon + \bar{s}_\epsilon) \leq \bar{H}_0 \left( \bar{x}_\epsilon, \frac{\bar{x}_\epsilon - \bar{y}_\epsilon}{1 + |\bar{x}_\epsilon - \bar{y}_\epsilon|^2 + |\bar{t}_\epsilon - \bar{s}_\epsilon|^2} + 2\beta \bar{x}_\epsilon \right) - \bar{H}_0 \left( \bar{y}_\epsilon, \frac{\bar{x}_\epsilon - \bar{y}_\epsilon}{1 + |\bar{x}_\epsilon - \bar{y}_\epsilon|^2 + |\bar{t}_\epsilon - \bar{s}_\epsilon|^2} - 2\beta \bar{y}_\epsilon \right). \tag{A.37} \]

Since $\bar{H}_0(\cdot, \cdot)$ is uniformly continuous over compact sets, and since $|\bar{x}_\epsilon - \bar{y}_\epsilon| \to 0$ as $\epsilon \to 0$ (for fixed $\beta$), the right-hand side of the above inequality goes to 0 (note that the term $\frac{\bar{x}_\epsilon - \bar{y}_\epsilon}{1 + |\bar{x}_\epsilon - \bar{y}_\epsilon|^2 + |\bar{t}_\epsilon - \bar{s}_\epsilon|^2}$ is bounded and that $|\bar{x}_\epsilon|, |\bar{y}_\epsilon| \leq R_\beta$). This gives
\[ A + 2\beta \inf_{\epsilon} \{\bar{t}_\epsilon + \bar{s}_\epsilon\} \leq 0. \]

Now taking $\beta \to 0$, we get
\[ A \leq 0, \]
which contradicts (A.34). Therefore we must have

$$\sup_{t,x} \{u_1(t,x) - u_2(t,x)\} \leq 0$$

which gives us $u_1 \leq u_2$.

References

[Alv96] Alvarez, O. and Tourin, A. “Viscosity Solutions of nonlinear integro-differential equations.” *Annales de l’Institut Henri Poincaré (C) Analyse Non Linéaire*, Vol. 13(3) (1996): 293-317.

[BKR05] Ball, K., Kurtz, T.G., Rempala, G. and Popovic, L. “Asymptotic Analysis of Multiscale Approximations to Reaction Networks.” *Annals of Applied Probability*, Vol. 16(4) (2005): 1925-1961.

[BD97] M. Bardi and I. Capuzzo Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Systems & Control: Foundations & Applications. Birkhäuser, Boston, 1997.

[Bar08] Barles, G. and Imbert, C. “Second-order elliptic integro-differential equations: viscosity solutions’ theory revisited.” *Annales de l’Institut Henri Poincaré (C) Analyse Non Linéaire* Vol. 25(3) (2008): 567-585.

[B-NS01] Barndorff-Nielsen, O.E. and Shephard, N. “Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics.” *Journal of Royal Statistical Society (B)*, Vol. 63 (2001): 167-241.

[Bud11] Budhiraja, A., Dupuis, P. and Maroulas, V. “Variational representations for continuous time processes.” *Annales de l’Institut Henri Poincaré (B) Probabilités et Statistiques*, Vol. 47(3) (2011): 725-747.

[Bud13] Budhiraja, A., Jiang C., and Dupuis, P. “Large deviations for stochastic partial differential equations driven by a Poisson random measure.” *Stochastic Processes and their Applications* Vol. 123(2) (2013): 523-560.

[CW14] Chen, Z.-Q. and Wang, J. “Ergodicity for time-changed symmetric stable processes.” *Stochastic Processes and their Applications*, Vol. 124(9) (2014): 2799-2823.

[CZ13] Chen, Z.-Q. and Zhang, X. “Heat kernel estimates and analyticity of non-symmetric jump diffusion semigroups.” *arXiv:1306.5015*, preprint (2013).

[DeZ98] Dembo, A. and Zeitouni, O. *Large Deviation Techniques and Applications*. Second Edition, Springer, 1998.

[DV75] Donsker, M.D. and Varadhan, S.R.S. “On a variational formula for the principal eigenvalue for operators with a maximal principle.” *Proceedings of the National Academy of Sciences*, Vol. 72(3) (1975): 780-783.
[DV83] Donsker, M.D. and S.R.S. Varadhan, “Asymptotic evaluation of Markov process expectations for large time, IV, Comm. Pure Appl. Math. 36:2 (1983), pp. 183-212.

[Dup12] Dupuis, P. and Spiliopolous, D. “Large deviations for multiscale diffusion via weak convergence methods.” Stochastic Processes and Their Applications, Vol. 122(4) (2012): 1947-1987.

[EK86] Ethier, S.N. and Kurtz, T.G. Markov Processes: characterization and convergence. John Wiley and Sons, New York, 1986.

[Fagg09] A. Faggionato, D. Gabrielli, and M. Ribezzi Crivellari. “Non-equilibrium thermodynamics of piecewise deterministic Markov Processes”. J Stat Phys, Vol. 137 (2009): 259-304.

[Fagg10] A. Faggionato, D. Gabrielli and M. Crivellari. “Averaging and large deviation principles for fully-coupled piecewise deterministic markov processes and applications to molecular motors”. Markov Processes and Related Fields, Vol. 16 (2010): 497-548.

[FFF10] J. Feng, M. Forde and J.-P. Fouque, Short maturity asymptotics for a fast mean reverting Heston stochastic volatility model, SIAM Journal on Financial Mathematics, Vol. 1, 2010 (p. 126-141)

[FFK12] Feng, J., Fouque, J.P. and Kumar, R. “Small-time asymptotics for fast mean-reverting stochastic volatility models.” Annals of Applied Probability, Vol. 22(4) (2012): 1541-1575.

[FK06] Feng, J. and Kurtz, T.G. Large Deviation for Stochastic Processes. Mathematical Surveys and Monographs, Vol. 131, American Mathematical Society, 2006.

[F98] Freidlin, Mark I. and Wentzell, A.D. Random perturbations of dynamical systems. Vol. 260 of Grundlehren der Mathematischen Wissenschaften, 1998.

[Imk09] Imkeller, P., Pavlyukevich, I. and Wetzel, T. “First exit times for Lévy driven diffusions with exponentially light jumps.” Annals of Applied Probability”, Vol 37(2), (2009): 530-564.

[KK13] Kang, H.-W., and Kurtz, T.G. ”Separation of time-scales and model reduction for stochastic reaction networks.” Annals of Applied Probability, Vol. 23(2), (2013): 529-583.

[KKP12] Kang, H.-W., Kurtz, T.G. and Popovic, L. “Central Limit Theorems and Diffusion Approximations for Multiscale Markov chain Models”, Annals of Applied Probability, Vol. 24(2), (2014): 721-759.

[KM05] Kontoyiannis, I. and Meyn, S.P. “Large Deviations Asymptotics and the Spectral Theory of Multiplicatively Regular Markov Processes”, Electronic Journal of Probability, Vol. 10, (2005): 61-123.

[Kou02] Kou, S.G. “A jump-diffusion model for option pricing.” Management Science, Vol. 48(8) (2002): 1086-1101.
[McSP14] McSweeney, J.K. and Popovic, L. “Stochastically-induced bistability in chemical reaction Ssstems.”, *Annals of Applied Probability*, Vol 24(3) (2014): 1225-1267.

[PR02] Palmowski, Z. and Rolski, T. “A technique for exponential change of measure for Markov processes.” *Bernoulli*, Vol. 8(6) (2002): 767-785.

[Pau10] Huh, D. and Paulsson, J. “Random partitioning of molecules at cell division.” *Proceedings of the National Academy of Sciences*, Vol. 108(36) (2011): 15004-15009.

[Pau11] Huh, D. and Paulsson, J. “Non-genetic heterogeneity from stochastic partitioning at cell division.” *Nature Genetics*, Vol. 43(2) (2011): 95-100.

[Puh15] Puhalskii, A. A. “On large deviations of coupled diffusions with time scale separation.” *Annals of Probability*, to appear.

[Sand13] Sandri, N. “Long-time behavior of stable-like processes.” *Stochastic Processes and their Applications*, Vol. 123(4) (2013): 1276-1300.

[Stroo84] Stroock, D. *An introduction to the Theory of Large Deviations*. Universitext, Springer-Verlag 1984, New York.

[Ver00] Veretennikov, A. Yu. “On large deviations for SDEs with small diffusion and averaging.” *Stochastic processes and their applications*, Vol. 89(1) (2000): 69-79.