

Moment bounds and geometric ergodicity of diffusions with random switching and unbounded transition rates

Xin T Tong and Andrew J Majda

October 19, 2016

Abstract

A diffusion with random switching is a Markov process consists of a stochastic differential equation part X_t and a continuous Markov jump process part Y_t . Such systems have a wide range of applications, where the transition rates of Y_t may not be bounded or Lipschitz. A new analytical framework is developed to understand the stability and ergodicity of these processes, and allows for genuinely unbounded transition rates. Assuming the averaged dynamics is dissipative, the first part of this paper explicitly demonstrates how to construct a polynomial Lyapunov function and furthermore moment bounds. When the transition rates have multiple scales, this construction comes interestingly as a dual process of the averaging of fast transitions. The coefficients of the Lyapunov function can be seen as the potential dissipation of each regime in different scales, and a comparison principle comes naturally under this interpretation. On the basis of these results, the second part of this paper establishes geometric ergodicity for the joint processes. This can be achieved in two scenarios. If there is a commonly accessible regime that satisfies the minorization condition, the geometric convergence to the ergodic measure takes place in the total variation distance. If there is contraction on average, the geometric convergence takes place in a proper Wasserstein distance, and is proved through an application of the asymptotic coupling framework.

1 Introduction

Diffusions with random switching are stochastic processes consisting of two components: a diffusion process X_t in \mathcal{R}^d and a continuous jump process Y_t on a finite set F . The dynamics of X_t follows a stochastic differential equation (SDE)

$$dX_t = b(X_t, Y_t)dt + \sigma(X_t, Y_t)dW_t. \quad (1.1)$$

Throughout, we assume $b(x, y)$ is $\mathcal{C}^{1+\delta}$ and $\sigma(x, y)$ is $\mathcal{C}^{2+\delta}$ in x for some $\delta > 0$. The behavior of Y_t can be described by a transition rate function $\lambda(x, y, \tilde{y})$, in other words

$$\mathbb{P}(Y_{t+h} = \tilde{y} | X_t, Y_t) = \begin{cases} \lambda(X_t, Y_t, \tilde{y})h + o(h), & \tilde{y} \neq Y_t, \\ 1 - \bar{\lambda}(X_t, Y_t)h + o(h), & \tilde{y} = Y_t. \end{cases}$$

We denote the total transition rate as $\bar{\lambda}(x, y) = \sum_{\tilde{y} \neq y} \lambda(x, y, \tilde{y})$, and the joint process as $Z_t = (X_t, Y_t)$, which takes place in the space $E = \mathcal{R}^d \times F$.

Diffusion with random switching is widely used for modelling purpose in many areas [1], and is becoming particularly prominent in the following directions recently:

- 1) For stochastic lattice models in climate science [2, 3, 4, 5, 6, 7, 8, 9, 10], X_t represents the dry atmosphere, so (1.1) is the spatial discretization of a fluid equation. Meanwhile, Y_t represents the unresolved behavior of moisture and clouds.
- 2) In material science [11, 12, 13, 14] and molecular biology [15, 16], X_t represents some macroscopic quantities such as the transmembrane electronic potential, and Y_t stands for the behavior of some particular clusters, proteins, channels, and cells.
- 3) As a simulation strategy for complex processes [2, 17], a Markov jump process can be used as a stochastic parameterization of some sub-grid scale processes. It reduces the model dimension and preserves most of the statistical quantities. It can be seen as the Y_t part in our joint process.
- 4) In filtering and predictive modelling [18, 19, 20], diffusions with random switching are used as test-beds to quantify the uncertainty from model errors.

The popularity of diffusions with random switching comes from its complexity. Even if (1.1) is an Ornstein Uhlenbeck (OU) process with each fixed valued of Y_t , that is

$$dX_t = -\gamma(Y_t)X_t dt + \sigma(Y_t)dW_t, \quad (1.2)$$

the switching of Y_t can generate very rich nonlinear properties, such as polynomial heavy tails [21, 22] and non-regular invariant measures [23, 24]. In the applications mentioned above, this flexibility is exploited to capture natural phenomena, while the equation in each regime is simple enough for intuitive understanding.

With so many applications of diffusions with random switching, the following questions naturally arise:

- When does the joint process $Z_t = (X_t, Y_t)$ possess an invariant measure? What kind of statistics, for example moments of $|X_t|$, are integrable under the invariant measures?
- Is the invariant measure unique? How does it attract other statistical states?

These questions in practice are often imposed as important sanity checks for stochastic models. This is because parameters inference and model validation often require the matching of statistics between nature and simulations, while the well-posedness of these operations require existence of invariant measures, ergodicity and finite moments for the models [25, 26].

In the last decade, a series of works have been devoted for the questions above [27, 21, 22, 28, 29, 30, 31, 32]. In the simplest setting, the transition rates are constants, $\lambda(x, y, y') = \lambda(y, y')$, and X_t is driven by the linear equation (1.2). Both questions above are relatively well understood in this setting, thanks to an application of Perron Frobenius theorem [21, 22]. These intuitive results can be extended to non-constant transition rates through a probabilistic coupling argument [29, 32]. However, for this argument to work, the transition

rates have to be globally bounded and Lipschitz. This restriction excludes many important applications [5, 33], or imposes additional nonphysical compact requirements on the model space [15]. This paper intends to bridge this gap by developing a new analytical framework.

In order for the joint process to have an invariant measure and finite moments for $|X_t|$, the rough requirement is that the averaged dynamics is dissipative. In the simplest setting of [21, 22], this condition can be defined as $\mathbb{E}\gamma(Y_t) > 0$ with γ as in (1.2). It is difficult to generalize this condition to settings with non-constant, unbounded transition rates. We will explore two directions for generalization:

- 1) Inspired by the formulation in [9, 10], we assume there is a $|X_t|$ controlled multiscale structure in the transition rates λ , while the fast averaging procedures induce a dissipation. It is important to note that the multiple scales here are not introduced by an auxiliary variable ϵ as in other standard settings [34, 35].
- 2) There is a comparison principle in favor of dissipation.

Interestingly, in both directions, the averaged dissipation can be demonstrated by constructing a polynomial Lyapunov functions, $V(x, y) = \sum a_i(y)|x|^{m_i}$. In the multiscale setting, the construction comes as a dual process of the fast averaging procedures, and $a_i(y)$ represents the potential dissipation of regime $y \in F$ in a particular transition scale. To the best of the authors' knowledge, this is the first explicit connection between averaging and Lyapunov functions. Since the $a_i(y)$ is interpreted as the potential dissipation, which can roughly be defined as a renormalization of $\mathbb{E}^y \exp(-\int_0^t \gamma(Y_s) ds)$ for large t , the comparison principle has an intuitive formulation, and the verification is much more straightforward and general than the coupling approach used in [29, 32]. This idea of Lyapunov functions construction can be traced back to [27], while our results generalize and offer new probabilistic interpretations to the working conditions.

The second part of this paper discusses geometric ergodicity assuming the existence of a Lyapunov function. Following the frameworks of [36, 37, 38], it suffices to show a version of the small set argument under a proper distance. This can be achieved in two scenarios:

- 1) If there is a commonly reachable regime that satisfies the minorization condition, Theorem 3.4 proves geometrically ergodicity in the total variation distance.
- 2) If there is contraction on average, and the transition rates and their first derivatives are bounded by the Lyapunov function, Theorem 3.6 shows geometrical ergodicity in a proper Wasserstein distance.

The unbounded transition rates appear to be an major obstacle in the second scenario as the coupling method of [29, 32] fail to work. It is resolved by viewing a diffusion with random switching as an annealed piecewise deterministic Markov processes, and then applying the asymptotic coupling framework of [37, 38] to the underlying densities. This strategy was exploited by the authors in [33] to study the piecewise contractive stochastic lattice models in [9, 10], now we generalize it to the contractive on average setting.

The remainder of this paper is arranged as follow. Section 2 discusses criterions that lead to dissipation on average, and how to construct a Lyapunov function in different scenarios. Section 3 gives the precise statements of geometric ergodicity when there is a hypoelliptic

regime or there is contraction on average. Conditions leading to the second scenario are briefly discussed and compared with results in Section 2. The proofs of geometric ergodicity are contained in Section 4, where we also discuss how to verify the accessibility of one regime. Section 5 summarizes the results and discusses some related questions.

2 Dissipation on average and Lyapunov functions

A simple way to generalize (1.2) to a nonlinear setting is to assume a rate function $\gamma : F \mapsto \mathcal{R}$ measures the dissipation and inflation of each regime in F .

Assumption 2.1. *With some strictly positive constants $K, \epsilon > 0$,*

$$\langle b(x, y), x \rangle \leq -\gamma(y)|x|^2 + K, \quad \|\sigma(x, y)\|^2 \leq K|x|^{2-\epsilon}. \quad (2.1)$$

Notice that $\gamma(y)$ could be negative, which introduces an inverse dissipation or inflation, and makes the global dissipation problem nontrivial. Given a *transition dissipation pair*, (λ, γ) , the main objective of this section is to find intuitive criteria that lead to dissipation on average in different scenarios, and show that there exists a polynomial like Lyapunov function.

Lyapunov functions are good tools to illustrate dissipation. In this paper, we say a function $V : E \mapsto [0, \infty)$ is a *Lyapunov function* if it has compact sub-level sets, and for some strictly positive constants $\bar{\gamma}$ and K

$$\mathcal{L}V(z) \leq -\bar{\gamma}V(z) + K, \quad \forall z = (x, y) \in E. \quad (2.2)$$

By Dynkin's formula, Grönwall's inequality and possibly a localization argument, (2.2) leads to

$$\mathbb{E}^z V(Z_t) \leq e^{-\bar{\gamma}t} V(z) + K/\bar{\gamma}. \quad (2.3)$$

If in addition $V(z) \geq |x|^m$ for all sufficiently large x , the m -th moment of $|X_t|$ is bounded under each invariant measure. Note also that by replacing V with $V + 1$, (2.2) holds with a different K . So without loss of generality we can assume $V \geq 1$.

To continue our discussion, note that the infinitesimal generator of a diffusion with random switching is given by [29]

$$\begin{aligned} \mathcal{L}f(x, y) &= b(x, y) \cdot \nabla_x f(x, y) + \frac{1}{2} \text{tr}[\sigma^t(x, y) \nabla_x^2 f(x, y) \sigma(x, y)] \\ &\quad + \sum_{y' \in F} \lambda(x, y, y') (f(x, y') - f(x, y)). \end{aligned} \quad (2.4)$$

One naive choice of Lyapunov function that leads to a moment bound is simply $V(x, y) = |x|^m$. Since Assumption 2.1 indicates $\mathcal{L}V(z) \leq -m(\gamma(y) - \epsilon')V(z) + K'$ for some positive constants ϵ' and K' , so with this choice of V , (2.2) holds if $\gamma(y) > 0$ for all y . But this is too restrictive. The limit of this naive choice comes from its ignorance of the Y part, so the averaging effect from transitions are missing. A natural way to incorporate the information of Y is considering the following monomial or polynomial form

$$V(z) = a(y)|x|^m, \quad \text{or} \quad V(z) = \sum_{i \leq I} a_i(y)|x|^{m_i}. \quad (2.5)$$

The coefficients $a_i(y)$ are strictly positive numbers that represent the potential dissipation, of which the meaning will be discussed in Remark 2.3. By incorporating this information of Y_t , V captures the global dissipation although \mathcal{L} is a local operator.

We will adopt three simplified notations in the following exposition. First, for Lyapunov functions we only need to be concerned with large x and the constant terms are usually ignorable. The precise statement is given by Lemma A.2. So we write $f(z) \lesssim g(z)$ if there is a constant K such that $f(z) \leq g(z) + K$. Second, we often identify a function $a : F \mapsto \mathcal{R}$ as a vector in $\mathcal{R}^{|F|}$, with the y -th coordinate being $[a]_y = a(y)$. We also use $\Lambda(x)$ for the Markov transition matrix on F , with entries $[\Lambda(x)]_{y,y'} = \mathbf{1}_{y' \neq y} \lambda(x, y, y') - \mathbf{1}_{y=y'} \bar{\lambda}(x, y)$. With this notation, we can separate the first and second lines of (2.4) into the form

$$\mathcal{L}f(x, y) = \mathcal{L}_X(y)f(\cdot, y) + [\Lambda(x)f(x, \cdot)]_y, \quad (2.6)$$

where \mathcal{L}_X and $\Lambda(x)$ represent the dynamics of the diffusion part and transition part respectively. Third, $|x|^m$ is not \mathcal{C}^2 when $m < 2$, and not well defined when $m < 0$, so rigorously speaking \mathcal{L} can not apply to it. However, as proposed in [22], $f_m(x) = |x|^{m+n}/(1 + |x|^n)$ with sufficiently large n is \mathcal{C}^2 , and carries essentially the same dissipation property of $|x|^m$, that is for any $\delta > 0$

$$\mathcal{L}_X(y)f_m(x) \lesssim (-m\gamma(y) + \delta)f_m(x),$$

see Lemma A.3. So without lost of generality, we assume all $|x|^m$ are well defined and \mathcal{C}^2 , else we just use f_m in its place.

2.1 Constant transition rates

In order to build up the intuition, let us first review the classical case studied in [21, 22] where $\Lambda(x) = \Lambda$ is a constant irreducible matrix, and X_t is driven by the linear equation (1.2). Y_t then is an ergodic Markov process on F that does not depend on X_t . Let π be the unique ergodic measure of Y_t , the dissipation on average can be formulated as

$$\sum_{y \in F} \pi(y)\gamma(y) > 0. \quad (2.7)$$

Suppose that $\sigma \equiv 0$, $|X_t|^m = \exp(-m \int_0^t \gamma(Y_s) ds) |X_0|^m$. Then by Jensen's inequality and Birkhoff ergodic theorem, we see that (2.7) is necessary for the whole dynamics to be dissipative. In fact, it is also sufficient, due to the following theorem, which is a translation of Theorem 1.5 [22] in our context.

Theorem 2.2. *Suppose that X_t follows (1.2) and Y_t is an ergodic Markov jump process with constant transition rate Λ . Suppose also the average dissipation is positive $\sum \pi(y)\gamma(y) > 0$, with π being the ergodic measure of Y_t . Let Γ be the diagonal matrix with entries $\gamma(y)$ on the y, y -th component, then there is an $m > 0$ such that the spectrum of $-m\Gamma + \Lambda$ lies in the negative half plane, and $V(x) = a(y)|x|^m$ is a Lyapunov function. Here a as a vector is the Perron eigenvector of $-m\Gamma + \Lambda$.*

Proof. According to Assumption 2.1 and Lemma A.3, for $V(x) = a(y)|x|^m$ and any fixed $\delta > 0$,

$$\begin{aligned}\mathcal{L}V(z) &= a(y)\mathcal{L}_X(y)|x|^m + |x|^m[\Lambda a]_y \\ &\lesssim [-(m\gamma(y)a_y - \delta) + [\Lambda a]_y]|x|^m \\ &= [(-m\Gamma + \Lambda)a]_y - \delta |x|^m.\end{aligned}\tag{2.8}$$

Based on (2.8), if a is a right eigenvector of the matrix $-m\Gamma + \Lambda$ associated with a negative eigenvalue, while a is strictly positive component wise, then $V(x)$ will be a Lyapunov function in the sense of (2.2). Such a can be found by the following two observations from [22].

First, from the Feynman Kac formula, we find that the y, y' -th component of matrix $\exp(-m\Gamma t + \Lambda t)$ is

$$\mathbb{E}^y \mathbb{1}_{Y_t=y'} \exp\left(-\int_0^t m\gamma(Y_s)ds\right) > 0,\tag{2.9}$$

so Perron-Frobenius theorem applies to $\exp(-m\Gamma t + \Lambda t)$. As a consequence, if a is the Perron eigenvector, which is the eigenvector associated to the eigenvalue with maximum real part, a is strictly positive component wise. Since the spectrum of $\exp(-m\Gamma t + \Lambda t)$ and the spectrum of $-m\Gamma + \Lambda$ clearly has an one to one relation, so a is also the eigenvector of $-m\Gamma + \Lambda$ associated with the eigenvalue of the maximum real part.

Second, at $m = 0$, the Perron eigenvalue is 0. Through a perturbation analysis of m to the positive direction, one can show the spectrum of $-m\Gamma + \Lambda$ lies in the negative half plane for small enough m . The details of this results can be found in proposition 4.2 of [22].

Combining these two arguments, we find a strictly positive m , such that the spectrum of $-m\Gamma + \Lambda$ is in the negative half plane, and the Perron eigenvector a of $\exp(-m\Gamma t + \Lambda t)$ is an eigenvector of $-m\Gamma + \Lambda$ associated with a negative eigenvalue, while all the components of a are strictly positive. \square

Remark 2.3. Let $\mathbb{1}$ be the vector with one on each component, since a is the Perron eigenvector, we can approximate a by normalizing $\exp(-m\Gamma t + \Lambda t)\mathbb{1}$ with $t \rightarrow \infty$. Because of the Feynman Kac's formulation (2.9),

$$\frac{a(y)}{a(y')} = \lim_{t \rightarrow \infty} \frac{\mathbb{E}^y \exp\left(-\int_0^t m\gamma(Y_s)ds\right)}{\mathbb{E}^{y'} \exp\left(-\int_0^t m\gamma(Y_s)ds\right)}.\tag{2.10}$$

In other words, $a(y)$ measures the potential dissipation along the whole future. This explains why $V(z)$ captures the global dissipation: if γ is negative for certain y , then y produces a weaker potential dissipation comparing with average states, so the transition part in (2.8), $[\Lambda(x)a]_y = \sum_{y'} \lambda(x, y, y')(a(y') - a(y))$ could be negative and compensate the inflation in $m\gamma(y)$.

2.2 Multiscale transitions: one fast scale

When the transition rates are coupled with the diffusion part, so are the trajectories of Y_t and X_t . This makes the notion of dissipation on average no longer as simple as (2.7).

One way to manifest it is finding multiscale structures controlled by X_t in the transition rates. Such structures rise naturally in many physical models, as part of X_t represents the temperature of environment [9, 10] or the electronic potential [15], and controls the speed of moisture connectivity or chemical reaction. For the simplicity of discussion, we assume in the following sections that the transition rate matrix has polynomial dependence over $|x|$:

$$\Lambda(x) = \sum_{0 \leq j \leq J} \Lambda_j |x|^{n_j}, \quad (2.11)$$

where Λ_j are constant $|F| \times |F|$ transition matrices, and $0 \leq n_0 < \dots < n_J$. We remark that it is possible to generalize our methods below to cases where $\Lambda(x)$ is the sum of other functions of separate orders. While the discussion below may appear to be abstract in the first read, the ideas are rather elementary. A concrete example will be illustrated by Figure 2.1 in Section 2.4, the readers can read that section first for the intuition.

Since Lyapunov functions concern only large $|x|$, see Lemma A.2, it is intuitive that the highest order transition Λ_J plays a dominating role; and if over the invariant measure of Λ_J the average of γ is positive, then there should be a Lyapunov function that quantifies dissipation on average. The complications to this argument may come from two aspects: 1) the support of each Λ_j may not be the whole state space F , so different subsets of F may have different transition scales; 2) on the support of each Λ_j , Λ_j may not induce an irreducible Markov chain. Here $F' \subset F$ is the *support* of a transition rate matrix Λ if F' is the minimal subset such that $\lambda(y, y') = 0$ if y and y' are not both in F' . We will leave the first issue to the next subsection, and focus first on the averaging phenomenon from multiscale transitions and possible reducible structures.

Irreducibility was a necessary condition in Theorem 2.2, when the transition matrix was constant, but not anymore if the transition is genuinely fast. Consider the following simple example on two states

$$F = \{-2, 1\}, \quad \gamma(y) = y, \quad \lambda(-2, 1) = 1, \quad \lambda(1, -2) = 0, \quad dX_t = -\gamma(Y_t)X_t dt.$$

Clearly δ_1 is the invariant measure for Y_t and there is dissipation on average over this measure. However, $\mathbb{E}^{x,y} X_t = x \mathbb{E}^y \exp(-\int_0^t Y_s ds)$, and if we start from $y = -2$,

$$\mathbb{E}^y \exp\left(-\int_0^t Y_s ds\right) \geq \exp(2t) \mathbb{P}^y(Y_s = -2, s \leq t) = \exp(t),$$

so $\mathbb{E}^{x,y} |X_t|$ diverges to infinity. On the other hand, if we replace the transition rates by $\lambda(-2, 1) = \lambda(1, -2) = |x| + 1$, the time that Y_t spend in -2 is much shorter as $|X_t|$ gets large, so the dynamics is dissipative on average.

To continue our discussion, we need the notion of connected components. Given a transition rate matrix $\Lambda(x)$, we say $F' \subset F$ is an *order n maximal connected component* of F if the following hold

- 1) There is a constant $|F'| \times |F'|$ matrix $\Lambda_{F'}$ such that $\Lambda(x)|_{F'} - |x|^n \Lambda_{F'}$ is of order $|x|^{n-\delta}$ for some $\delta > 0$. Here $\Lambda(x)|_{F'}$ is the sub-diagonal matrix of $\Lambda(x)$ with indices in F' .
- 2) For any $y, y' \in F'$, there is a path $y = y_0, y_1, \dots, y_m = y'$ such that for each i , either $\lambda_{F'}(y_i, y_{i+1}) > 0$ or $\lambda_{F'}(y_{i+1}, y_i) > 0$.

3) F' is maximal as there is no strict superset of F' that also satisfies the conditions above.

When such an F' and $\Lambda_{F'}$ exist, because $\Lambda_{F'}|x|^n$ consists of the leading order terms in the stochastic matrix $\Lambda(x)$, $\Lambda_{F'}$ itself must also be a stochastic matrix. As a consequence, there is a Markov jump process with constant rate $\Lambda_{F'}$ on F' . This is a nice mechanism we will exploit in our discussion later.

Next we define the *irreducible components*, which is also called the *closed communicating classes* in the literature like [39]. A subset $G \subset F'$ is an *irreducible component*, if 1) for all $y, y' \in G$, there is a path $y = y_0, \dots, y_n = y'$ in G such that $\lambda_{F'}(y_i, y_{i+1}) > 0$; 2) for all $y \in G, y' \notin G$ such a path does not exist. We will use G^c to denote the transient set, which consists of states in F' not being in any irreducible components.

Now we consider the simple case where the highest order component is F itself.

Theorem 2.4. *Suppose the whole state space F is a maximal connected component of order $n > 0$. Let $\{G_k\}$ be the irreducible components, and π_k be the ergodic measure generated by $\Lambda_F|_{G_k}$. Let γ be the linear dissipation rate function satisfying Assumption 2.1, while $\sum_{y \in G_k} \pi_k(y)\gamma(y) > 0$ for each G_k , then for any $m > 0$ there is a Lyapunov function of form*

$$V(x, y) = |x|^m + a(y)|x|^{m-n}.$$

Proof. Let $V_0(z) = |x|^m$. Directly apply the generator, by Lemma A.3 for any $\delta > 0$

$$\mathcal{L}V_0(z) \lesssim (-m\gamma(y) + \delta)|x|^m.$$

The right hand side is a polynomial of order m , and for any irreducible component G_k , $\sum_{y \in G_k} \pi_k(y)\gamma(y) > 0$

$$\sum_{y \in G_k} \pi_k(y)(-m\gamma(y) + \delta)|x|^m \lesssim -2\delta|x|^m,$$

if δ is sufficiently small. Applying a Fredholm alternative type of argument, which is Lemma 2.5 directly below, there is a monomial $Q(z) = a(y)|x|^{m-n}$ with $a(y) \geq 0$, such that

$$[\Lambda(x)Q(z)]_y + (-m\gamma(y) + \delta)|x|^m \lesssim -\delta|x|^m.$$

Then because $V(z) = V_0(z) + Q(z)$ is of order m , and $\mathcal{L}_X(y)Q(z)$ is of order $m - n$ by Lemma A.3, so $\mathcal{L}V(z) \lesssim -\delta V(z)$, and $V(z)$ is a Lyapunov function. \square

Lemma 2.5. *Let F' be a maximal connected component of order $n > 0$, and $P(x, y)$ be a polynomial of $|x|$ of order $m > 0$ such that $P(x, y) = 0$ if $y \notin F'$. Suppose the following holds*

$$\sum_{y \in G_k} \pi_k(y)P(x, y) \lesssim 0, \quad \forall x \in \mathcal{R}^d.$$

Then we can find a positive monomial $Q(z) = q(y)|x|^{m-n}$ with $q(y) \geq 0$ and $q(y) = 0$ if $y \in F/F'$, such that for any $\epsilon > 0$

$$\Lambda(x)Q(z) + P(z) \lesssim \epsilon|x|^m, \quad y \in F'.$$

Proof. First, we will specify the value of $q(y)$ for each non transient y . We assume $y \in G_k$ in the discussion below. Let $p(y)|x|^m$ be the maximum order term in $P(z)$, then clearly $\sum_{y' \in G_k} \pi_k(y')p(y') \leq 0$. By the Fredholm alternative, there is a vector q_k with nonzero components only for indices in G_k , so that

$$p|_{G_k} + \Lambda_{F'} q_k = \left(\sum_{y' \in G_k} \pi_k(y')p(y') \right) \mathbf{1}_{G_k}. \quad (2.12)$$

$\mathbf{1}_{G_k}$ stands for the indicator vector of set G_k . Note that $\Lambda_{G_k} \mathbf{1}_{G_k} = 0$, we can always replace q_k with $q_k + \kappa \mathbf{1}_{G_k}$ with a proper κ , so it is still a solution to (2.12), but with strictly positive components. So we can assume $q_k > 0$ component wise. We let $q(y) = q_k(y)$ for non transient y .

Next, for the transient states $y \in G^c$, consider a Markov jump process Y'_t on F' , driven by the transition matrix $\Lambda_{F'}$. Denote the expected time of hitting any one of G_k from any y as $T(y)$. Clearly $T(y) = 0$ for $y \in G_k$. By running the one step analysis for the jumps, we find for $y \in G^c$,

$$T(y) = \bar{\lambda}_{F'}^{-1}(y) + \sum_{y'} \frac{\lambda_{F'}(y, y')}{\bar{\lambda}_{F'}(y)} T(y'),$$

recall that $\bar{\lambda}_{F'}(y) = \sum_{y' \neq y} \lambda_{F'}(y, y')$. As a consequence, for $y \in G^c$,

$$\Lambda_{F'} T(y) = \sum_{y' \in F'} \lambda_{F'}(y, y') (T(y') - T(y)) = -1.$$

We will let $q(y) = \beta T(y)$ with

$$\beta > \max_{y \in F'} |p(y)| + (\max_{y \in F'} \bar{\lambda}_{F'}(y)) (\max_{y \in G_k, \forall k} q(y)).$$

Now we verify our claim, it suffices to show $\Lambda(x)Q(z) + P(z)$ has its order m term being less than 0. Since $(\Lambda(x) - \Lambda_{F'}|x|^n)Q(z)$ is of order strictly less than m , it suffices to show

$$[\Lambda_{F'} q]_y + p(y) \leq 0, \quad \forall y \in F'.$$

This is clearly the case when $y \in G_k$ because it is implied by (2.12). And for $y \in G^c$, this holds because

$$\begin{aligned} [\Lambda_{F'} q]_y + p(y) &= \sum_{y' \in F'} \lambda_{F'}(y, y') (q(y') - q(y)) + p(y) \\ &= \sum_{y' \in G^c} \lambda_{F'}(y, y') \beta (T(y') - T(y)) + \sum_k \sum_{y' \in G_k} \lambda_{F'}(y, y') (q(y') - \beta T(y)) + p(y) \\ &\leq \beta \Lambda_{F'} T(y) + \bar{\lambda}_{F'}(y) \max_{y' \in G_k, \forall k} q(y') + p(y) < 0. \end{aligned}$$

□

2.3 Multiscale transitions: multiple scaling structures

When F is not the maximal connected component, the transitions inside a maximal connected component F' of highest order will be significantly faster than transitions outside. These fast transitions will average the dissipation of each irreducible component G_k inside F' , and also the manner how Y_t leaves F' . In the perspective of the states outside F' , each G_k is essentially a single point, and its dissipation rate is the averaged dissipation over π_k ; and each transient state $y \in G^c$ is an intermediate state that can jump to any of the G_k , while the time Y_t spent on it is ignorable, as long as the rates towards non-irreducible parts, $\lambda(x, y, y')$ with $y' \in G^c \cup F/F'$, are not too strong.

To be more specific, given a transition dissipation pair $(\Lambda(x), \gamma)$ with $\Lambda(x)$ being of order $|x|^n$, let F' be a maximal connected component, G_1, \dots, G_k be the irreducible sets, and G^c be the transient set. We define a new structure $(\tilde{\Lambda}(x), \tilde{\gamma})$ on the averaged space $\tilde{F} = (F/F') \cup \{g_1, \dots, g_k\}$ as the average of the original structure on F' . Intuitively, the new rates are the same for states outside F' ,

$$\tilde{\gamma}(y) = \gamma(y), \quad \tilde{\lambda}(x, y, y') = \lambda(x, y, y'), \quad y, y' \in F/F'.$$

The rates related to g_k are given by the following averages:

$$\begin{aligned} \tilde{\gamma}(g_k) &= \sum_{y \in G_k} \pi_k(y) \gamma(y), & \tilde{\lambda}(x, g_k, y) &= \sum_{y' \in G_k} \pi_k(y') \lambda(x, y', y), \quad y \in F/F', \\ \tilde{\lambda}(x, y, g_k) &= \sum_{y' \in G_k} \lambda(x, y, y') + \sum_{y' \in G^c} \lambda(x, y, y') p_{F'}(y, g_k), \quad y \in F/F', \\ \tilde{\lambda}(x, g_j, g_k) &= \sum_{y \in G_j} \pi_j(y) \sum_{y' \in G_k} \lambda(x, y, y') + \sum_{y \in G_j} \pi_j(y) \sum_{y' \in G^c} \lambda(x, y, y') p_{F'}(y', g_k). \end{aligned} \tag{2.13}$$

In above, π_k is the ergodic measure on G_k induced by the matrix $\Lambda_{F'}$, and $p_{F'}(y, g_k)$ is the probability that Y'_t ends up in G_k if Y'_t is a Markov chain driven by $\Lambda_{F'}$ and starts from y .

Note that in these averaging procedures, the transition rates from $y \in G^c$ to $y' \in G^c \cup F/F'$ are completely wiped out. So we need these rates to be not too strong, else the averaged structure cannot represent this information. In particular we have the following non-dominating condition,

Assumption 2.6. *For any transient state y and any $y' \in G^c \cup F/F'$, suppose that $p_{F'}(y, g_k) > 0$, then there is a $y'' \in G_k$ such that $\lambda(x, y, y')$ has at most the same polynomial order in $|x|$ as $\lambda(x, y'', y')$.*

Since there are only finitely many states, there are only finitely many, say m_J , connected components with the highest order n_J in (2.11). After applying an averaging step on one of these components, F' , the transition rates related to F' are of order strictly less than n_J , and the states after averaging has a smaller cardinality $|\tilde{F}| \leq |F|$. So after m_J steps, the transition rates are of order n_{J-1} . We can repeat this argument J times, and finally end up with an averaged transition matrix $\tilde{\Lambda}$ being a constant matrix. Intuitively, this constant matrix dictates whether the original system is dissipative on average.

Theorem 2.7. *Let state space \tilde{F} , constant transition rates $\tilde{\Lambda}$ and dissipation rates $\tilde{\gamma}$ be the final result of a sequence of averaging procedures. Suppose at each averaging step, the transient transition rates follow the non-dominating condition, Assumption 2.6. Then the original system has a polynomial like Lyapunov function of some order $m > 0$, if \tilde{F} consists of only irreducible components of $\tilde{\Lambda}$, while on each of them the average dissipation of $\tilde{\gamma}$ is positive. If in addition $\tilde{\gamma}(y) > 0$ for all $y \in \tilde{F}$, m can be any positive number.*

Theorem 2.4 was a special one averaging step case of the theorem above, and the conditions there were not optimal. But we keep Theorem 2.4 for its simpler intuition.

Proof of Theorem 2.7. Based on Theorem 2.2 it is clear how to find a Lyapunov function $\tilde{V}(z)$ for the final averaged dynamics $(\tilde{\Lambda}, \tilde{\gamma})$. In particular, if $\tilde{\gamma}(y) > 0$ for all y , then for any $\delta > 0, m > 0$,

$$\mathcal{L}|x|^m \lesssim (-m \min_y \{\tilde{\gamma}(y)\} + \delta)|x|^m,$$

so $|x|^m$ is a Lyapunov function. Then by the induction principle, it suffices to show that, given an averaging step

$$(F, \Lambda, \gamma) \Rightarrow (F/F' \cup \{g_1, \dots, g_K\}, \tilde{\Lambda}, \tilde{\gamma})$$

and a polynomial like Lyapunov function $\tilde{V}(z) = \sum \tilde{a}_i(y)|x|^{n_i}$ for the averaged structure, how to construct a new polynomial Lyapunov function $V(z)$.

One thing that requires special attention is the order of Lyapunov function \tilde{V} and the detailed transitions, defined as

$$\tilde{\lambda}(x, y, y')(\tilde{V}(x, y') - \tilde{V}(x, y)).$$

These polynomials are clearly of order at most m in the final state, since $\tilde{\lambda}$ are constants and \tilde{V} is of order m . We will show this polynomial order is inherited by the constructed Lyapunov function $V(z)$ of the pre-averaged dynamics.

Denote the maximal order term in $\tilde{V}(z)$ as $\tilde{a}_m(y)|x|^m$. Because \tilde{V} is a Lyapunov function, there is a $\gamma_0 > 0$ such that the order m terms in $\mathcal{L}\tilde{V}(z)$ are

$$[\tilde{\Lambda}(x)\tilde{V}(x)]_y - m\tilde{\gamma}(y)\tilde{a}_m(y)|x|^m \lesssim -\gamma_0\tilde{V}(z), \quad y \in F/F' \cup \{g_1, \dots, g_K\}, \quad (2.14)$$

because $\mathcal{L}_X|x|^{m-\delta}$ is of order less than m based on Lemma A.3. To continue, we notice the dual of the averaging step produces the following function on $\mathcal{R}^d \times F$ based on \tilde{V} :

$$V_0(x, y) = \begin{cases} \tilde{V}(x, y), & y \in F/F'; \\ \tilde{V}(x, g_k), & y \in G_k; \\ \sum_k p_{F'}(y, g_k)\tilde{V}(x, g_k), & y \in G^c. \end{cases}$$

Decompose the transition rates into two parts $\Lambda(x) = \Lambda_{F'}|x|^n + \Lambda^c(x)$, where Λ^c is of order $n - \delta$ for some $\delta > 0$. With some technical verification in Lemma A.4, the following duality equations hold

$$\begin{cases} [\tilde{\Lambda}(x)\tilde{V}(x, \cdot)]_y = [\Lambda(x)V_0(x, \cdot)]_y, & y \in F/F'; \\ [\tilde{\Lambda}(x)\tilde{V}(x, \cdot)]_{g_k} = \sum_{y \in G_k} \pi_k(y)[\Lambda(x)V_0(x, \cdot)]_y. \end{cases} \quad (2.15)$$

We claim that the detail transitions of V_0 induced by $\Lambda(x)$ is of order at most m . The detail transition of V_0 from a $y \in G_k$ to a $y' \in F/F'$ is $\lambda(x, y, y')(\tilde{V}(x, g_k) - \tilde{V}(x, y'))$. A combination of $y' \in G_k$ is

$$\tilde{\lambda}(x, g_k, y')(\tilde{V}(x, g_k) - \tilde{V}(x, y')) = \left(\sum_{y \in G_k} \pi_k(y) \lambda(x, y, y') \right) (\tilde{V}(x, g_k) - \tilde{V}(x, y')).$$

The left hand side is of order at most m from inductions; on the right hand, the coefficients follow $\lambda(x, y, y') \geq 0, \pi_k(y) > 0$. So $\lambda(x, y, y')(\tilde{V}(x, g_k) - \tilde{V}(x, y'))$ is of order at most m . Likewise, because

$$\begin{aligned} & \tilde{\lambda}(x, g_k, g_j)(\tilde{V}(x, g_j) - \tilde{V}(x, g_k)) \\ &= \sum_{y \in G_k} \pi_k(y) \left[\sum_{y' \in G^c} \lambda(x, y, y') \sum_j p_{F'}(y', g_j) + \sum_{y' \in G_j} \lambda(x, y, y') \right] (\tilde{V}(x, g_j) - \tilde{V}(x, g_k)), \end{aligned}$$

we can conclude the detail transition of V_0 from $y \in G_k$ to $y' \in G_j$ or G^c is of order at most m . With the same argument, the detail transition of V_0 from $y \in F/F'$ to other y' are all of order at most m . For $y \in G^c$, the detail transition to any y' will be

$$\lambda(x, y, y')(V_0(x, y') - V_0(x, y)) = \lambda(x, y, y') \sum_k p_{F'}(y, g_k)(V_0(x, y') - V_0(x, g_k)),$$

where y_k is any element in G_k . Then because of the non-dominating condition Assumption 2.6, and that $\lambda(x, y_k, y')(V_0(x, y') - V_0(x, y_k))$ is of order at most m , so is $\lambda(x, y, y')(V_0(x, y') - V_0(x, y))$.

If we let $V(z) = V_0(x, y)$, the Lyapunov dissipation will be inherited for $y \in F/F'$, but there will be an order m error term for $y \in F'$, and we will apply Lemma 2.5 to fix this with a monomial. In particular, the image of V_0 through \mathcal{L} is

$$\mathcal{L}V_0(z) \lesssim P(x, y) := -m\gamma(y)\tilde{a}_m(y)|x|^m + [\Lambda(x)V_0(x)]_y.$$

The average of $P(x, y)$ over any G_k , by the second duality equation, is

$$\sum_{y \in G_k} \pi_k(y) ([\Lambda(x)V_0(x, \cdot)]_y - \gamma(y)\tilde{a}_m(g_k)|x|^m) = [\tilde{\Lambda}(x)\tilde{V}(x)]_{g_k} - \tilde{\gamma}(g_k)\tilde{a}_m(g_k)|x|^m$$

which is bounded by (2.14). Since the $P(x, y)$ is of order at most m , by the Fredholm alternative Lemma 2.5, there is a positive monomial $Q(z)$ of order less than $m - n$, such that if we let $V(z) = V_0(z) + Q(z)$, then for all $y \in F'$

$$-\gamma(y)\tilde{a}_m(g_k)|x|^m + [\Lambda(x)(Q(z) + V_0(z))]_y \lesssim -\gamma_0 V_0(z).$$

Since the order of $Q(z)$ is less than $m - n$, and $\mathcal{L}_X(y)Q(z)$ produces a term of order at most $m - n$, we find that $\mathcal{L}V(z) \lesssim -\gamma_0 V(z)$ if $y \in F'$.

As for $y \in F/F'$,

$$\mathcal{L}V(z) \lesssim -\tilde{\gamma}(y)\tilde{a}_m(g_k)|x|^m + [\tilde{\Lambda}(x)\tilde{V}(x)]_y + [\Lambda(x)Q(z)]_y.$$

Notice the first two parts are bounded by (2.14). $[\Lambda(x)Q(z)]_y = [\Lambda^c(x)Q(z)]_y$ because F' is the connected component of Λ_F . Then notice $Q(z)$ is of order at most $m - n$, while $\Lambda^c(x)$ is of order strictly less than n , therefore $[\Lambda^c(x)Q(z)]_y$ is of order strictly less than m , so $\mathcal{L}V(z) \lesssim -\frac{1}{2}\gamma_0 V(z)$.

Lastly, we notice the detail transition of $V(z)$ is the sum of detail transition of $V_0(z)$ and $Q(z)$, the first part is of order m from previous discussion, and $\lambda(x, y, y')(Q(x, y') - Q(x, y))$ is of order at most m as well. So the detail transitions of $V(z)$ are of order at most m . \square

Remark 2.8. *Once we finish the construction of $V = \sum a_i(y)|x|^{m_i}$ and look back, we can see $a_i(y)$ captures the potential dissipation with the transition rates of order $|x|^{m_I - m_i}$, and within a maximal connected component of that order. And for y in this maximal connected components, $a_j(y)$ are of identical value for $j \geq i$. In other words, from the value of the sequence $\{a_i(y)\}_{i \leq I}$, we can actually tell which connected component of what order does y belong to. The following subsection gives a simple and concrete example.*

2.4 A multiscale transition example

In this section, we consider one concrete example with multiscale transitions, where the averaging steps mentioned in the previous section can be discussed explicitly. In subplot 1) of Figure 2.1, a Markov process is defined on four states $F = \{a, b, c, d\}$ with the transition rates given along the arrows. The dissipation rates are given by

$$\gamma(a) = -1, \quad \gamma(b) = 2, \quad \gamma(c) = -1, \quad \gamma(d) = -1.$$

F is the maximal connected component of order 2. It has two irreducible components $\{a, b\}$ and $\{d\}$. The induced invariant measure on $\{a, b\}$ is $\pi(a) = \frac{1}{3}, \pi(b) = \frac{2}{3}$. c is the only transient state. Starting from c and driven by the maximal order transition Λ_F , it is equal likely to end up in $\{a, b\}$ and $\{d\}$.

After one averaging step, we have a two-state Markov chain in subplot 2). The states represent the irreducible components in the original system. The dissipation rates are given by

$$\tilde{\gamma}(ab) = \pi(a)\gamma(a) + \pi(b)\gamma(b) = 1, \quad \tilde{\gamma}(d) = \gamma(d) = -1.$$

The transition rates are given by

$$\tilde{\lambda}(x, ab, d) = \pi(b)\lambda(x, b, d) = |x|, \quad \tilde{\lambda}(x, d, ab) = p_F(c, ab)\lambda(x, d, c) = 2|x|.$$

So the chain is of order 1, while invariant measure driven by the dynamics of this order is $\tilde{\pi}(ab) = \frac{2}{3}, \tilde{\pi}(d) = \frac{1}{3}$.

With the final step of averaging, we end up with one state in 3), so the transition matrix is the constant matrix of zero. The dissipation rate is given by $\tilde{\pi}(ab)\tilde{\gamma}(ab) + \tilde{\pi}(d)\tilde{\gamma}(d) = \frac{1}{3}$. So the whole system is dissipative on average, Theorem 2.7 applies, and the Lyapunov function can be of any order.

In particular, for the trivial Markov process described by subplot 3), $|x|^m$ with any $m > 0$ is a Lyapunov function. The procedure in the proof of Theorem 2.7 indicates this Lyapunov

function can be pulled back into Lyapunov functions for the Markov processes described by 2) and 1):

$$|x|^m + a_2(y)|x|^{m-1} + a_1(y)|x|^{m-2} \iff |x|^m + \tilde{a}_2(y)|x|^{m-1} \iff |x|^m.$$

Where $\tilde{a}_2(ab) = 0, \tilde{a}_2(d) = \frac{2}{3}m$; $a_2(a) = a_2(b) = 0, a_2(c) = \frac{1}{3}m, a_2(d) = \frac{2}{3}m, a_1(a) = \frac{2}{3}m, a_1(b) = 0, a_1(c) = \frac{3}{2}m, a_1(d) = 0$, assuming $m \geq 4$ so $|x|^{m-2}$ is \mathcal{C}^2 .

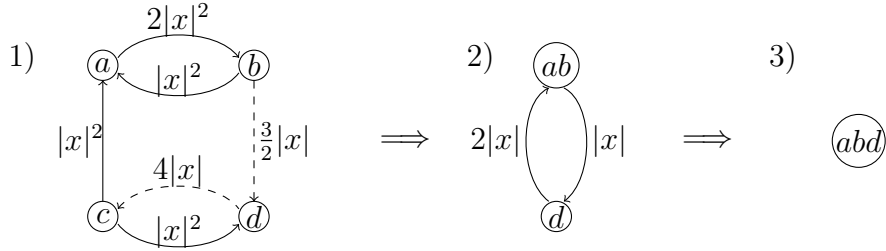


Figure 2.1: Averaging multiscale transition rates. Subplot 1) is the original system, 2) and 3) are the systems after one and two averaging steps. The leading order transitions are marked by solid arrows, and the lower order transitions are marked by dashed arrows.

In order to illustrate Assumption 2.6, we consider one modification of 1) in Figure 2.1 that violates Assumption 2.6. Suppose there is another state e , and

$$\lambda(x, a, e) = \lambda(x, b, e) = \lambda(x, d, e) = 1, \quad \lambda(x, c, e) = |x|.$$

If one averages again like Figure 2.1, the strong transition from c to e will be ignored, because c is a transient state in the averaging procedure.

2.5 Comparison principle

The other way to deal with non-constant transition rate is through a comparison principle. To be specific, suppose $(\Lambda(x), \gamma)$ is dissipative on average, which may be established by Theorem 2.2 or Theorems 2.4, 2.7. Suppose also in another transition dissipation pair $(\tilde{\Lambda}(x), \tilde{\gamma})$, the dissipation is stronger in all regimes, while the regime transitions are more favorable for dissipation, then intuitively $(\tilde{\Lambda}(x), \tilde{\gamma})$ would also admits a dissipation on average.

The only vagueness of the previous argument is how to determine that the regime transitions are more favorable for dissipation. In our contexts, since the coefficients $a_i(y)$ in (2.5) characterize the potential dissipation, see Remark 2.3, so intuitively, we would say $(\tilde{\Lambda}(x), \tilde{\gamma})$ has more favorable dissipation than $(\Lambda(x), \gamma)$ if

$$\tilde{\gamma}(y) \geq \gamma(y), \quad (\tilde{\lambda}(x, y, y') - \lambda(x, y, y'))(a_i(y') - a_i(y)) \leq 0, \quad \forall i.$$

In this interpretation, the comparison principle is straightforward and can be generalized to dynamics on different spaces, where the state space F can be countable.

Theorem 2.9. Let P be a mapping from F to F' . Suppose the transition dissipation pair $(\Lambda'(x), \gamma')$ on F' admits a Lyapunov function of polynomial form $V(x, y) = \sum_{i \leq I} a_i(y) |x|^{m_i}$ with $a_i \geq 0$. Suppose $(\Lambda(x), \gamma)$ on F is more favorable for dissipation in the sense that for any $i \leq I$ and $q \in F, y \in F'$

$$\gamma(q) \geq \gamma(P(q)), \quad \left(\sum_{P(q')=y} \lambda(x, q, q') - \lambda'(x, P(q), y) \right) (a_i(y) - a_i(P(q))) \leq 0. \quad (2.16)$$

Then $V(x, P(q))$ will be a Lyapunov function for $(\Lambda(x), \gamma)$.

Proof. Let m_I be the maximum polynomial order, the fact that V is a Lyapunov function for $(\Lambda'(x), \gamma')$ indicates that $\forall y \in F'$

$$-m_I a_I(y) \gamma'(y) |x|^{m_I} + \sum_{i \leq I} [\Lambda'(x) a_i]_y |x|^{m_i} \lesssim -\bar{\gamma} a_I(y) |x|^{m_I} \quad (2.17)$$

for a $\bar{\gamma} > 0$. According to Assumption 2.1 and Lemma A.3, for any $\delta > 0$ the following holds

$$\mathcal{L}V(x, P(q)) \lesssim (-m_I \gamma(q) + \delta) a_I(P(q)) |x|^{m_I} + \sum_{q' \in F'} \lambda(x, q, q') \sum_{i \leq I} (a_i(P(q')) - a_i(P(q))) |x|^{m_i}.$$

Note that

$$\begin{aligned} \sum_{q' \in F'} \lambda(x, q, q') (a_i(P(q')) - a_i(P(q))) &= \sum_{y \in F} \left(\sum_{P(q')=y} \lambda(x, q, q') \right) (a_i(y) - a_i(P(q))) \\ &\leq \sum_{y \in F} \lambda'(x, P(q), y) (a_i(y) - a_i(P(q))). \end{aligned}$$

Combine both inequalities above and compare it with (2.17), $\mathcal{L}V \lesssim (-\bar{\gamma} a_I(y) + \delta) |x|^{m_I}$. Since δ can be arbitrarily small, it can further be bounded by $-\bar{\gamma}' a_I(y) |x|^{m_I}$ hence also $-\bar{\gamma}' V$ for some $\bar{\gamma}' > 0$. \square

The simplicity of this proof comes from our interpretation of dissipation on average through polynomial Lyapunov functions. On the other hand, comparison principles can also be demonstrated by coupling methods when the transition rates are bounded. [29, 32] have shown dissipation on average in the following birth-death scenario through proofs of considerate length, while it is only a special case of Theorem 2.9.

Corollary 2.10 (Birth-death type criterion). *Suppose there is a partition of regimes $F = F_1 \cup \dots \cup F_n$, and there is an increasing sequence of dissipation rates β_1, \dots, β_n . Suppose*

$$\lambda(x, y, y') = 0 \quad \text{if } y, y' \text{ are not in neighboring } F_i,$$

and $\gamma(y) \geq \beta_k$ if $y \in F_k$, where γ is the linear dissipation rate as in Assumption 2.1. For $k = 1, \dots, n$, denote

$$b_k = \inf_{x \in E} \inf_{y \in F_k} \sum_{y' \in F_{k+1}} \lambda(x, y, y'), \quad d_k = \sup_{x \in E} \sup_{y \in F_k} \sum_{y' \in F_{k-1}} \lambda(x, y, y'), \quad \nu_k = \prod_{i=1}^k \frac{b_i}{d_i}.$$

Then the transition dissipation pair $(\Lambda(x), \gamma)$ has a monomial Lyapunov function if $\sum_{k=1}^n \beta_k \nu_k > 0$.

Proof. Let $F' = \{1, \dots, n\}$, and Y'_t be a birth-death process on F' with birth rate b_k and death rate d_k . Using the detailed balance relation, it is easy to see ν_k is a multiple of the invariant measure of Y'_t . So if X'_t follows (1.2) with $\gamma(Y'_t) = \beta_{Y'_t}$, Theorem 2.2 applies, and there is a monomial Lyapunov function $a(y)|x|^m$. Meanwhile, let Y''_t be another birth-death process on F' with the same rates, we couple Y'_t and Y''_t so that they jump independently before they are of the same value, then we make them do the same jumps simultaneously. Then if $Y'_0 \geq Y''_0$, we have for all $t > 0$, $\beta_{Y'_t} \geq \beta_{Y''_t}$. Since the ratio $a(k)/a(k')$ can be given by (2.10), it is clear that $a(k) \geq a(k')$ if $k \leq k'$. Then we apply Theorem 2.9 with $P : y \rightarrow k$ if $y \in F_k$ to see the original transition dissipation pair has a Lyapunov function. Condition (2.16) holds because if we let

$$b(q, k) = \left(\sum_{P(q')=k} \lambda(x, q, q') - \lambda'(x, P(q), k) \right) (a_i(k) - a_i(P(q))),$$

then $b(q, k) = 0$ if $q \in F_j, k \notin \{j-1, j+1\}$. And if $q \in F_j$

$$b(q, j-1) = \left(\sum_{P(q')=j-1} \lambda(x, q, q') - d_j \right) (a_{j-1} - a_j) \leq 0$$

and likewise $b(q, j+1) \leq 0$. □

3 Geometric ergodicity

In Section 2, Lyapunov functions are constructed for diffusions with random switching when the dynamics is dissipative on average. Because these Lyapunov function are in the form of polynomials, $\mathbb{E}|X_t|^m$ is bounded uniformly for a proper m . Then by the Krylov-Bogoliubov Theorem[40], there is at least one invariant measure for the joint process $Z_t = (X_t, Y_t)$. It is natural to ask whether this invariant measure is unique, and how does the law of Z_t converge to the unique invariant measure π .

For many stochastic processes, this question is answered by geometric ergodicity. Namely, let d be a distance for probability measures, and $P_t^* \mu$ be the law of Z_t given that $Z_0 \sim \mu$, then there is a $\gamma > 0, C_{\mu, \nu}$ such that

$$d(P_t^* \mu, P_t^* \nu) \leq e^{-\gamma t} C_{\mu, \nu}.$$

By letting μ be an invariant measure, the bound above indicates there is only one invariant measure and all other statistical states are attracted to it geometrically fast.

In a series of important works on this subject [36, 41, 37, 38], a general framework has been developed to verify geometric ergodicity, assuming that a Lyapunov function exists. In the following we will apply this framework and show Z_t is geometric ergodic

- 1) in total variation distance if there is a commonly reachable minorization regime;
- 2) in a proper Wasserstein distance if there is contraction on average.

The Lyapunov function V will play a regularizing role in our discussion. In order to show geometric ergodicity for unbounded rates, we will replace the globally bounded or Lipschitz conditions in [29, 32] with weaker requirements that the transition rates or their derivatives are bounded by the Lyapunov function V . One important consequence is that an explosion, that is $|X_t|$ reaches infinity or Y_t has infinitely many jumps in finite time, is no longer possible. To see this, let T_t be the number of jumps in $Y_{s \leq t}$, then a space-time version of (2.4) shows that $\mathcal{L}T_t = \bar{\lambda}(z)$. If $\bar{\lambda}(z) \leq MV(z)$ for a constant M , then applying Dynkin's formula to the stopping time τ_N , which is the first time either $|X_t| = N$ or $T_t = N$,

$$\begin{aligned} \mathbb{E} [T_{t \wedge \tau_N} + V(Z_{t \wedge \tau_N})] &\leq \mathbb{E} \int_0^{t \wedge \tau_N} (\bar{\lambda}(Z_s) - \gamma V(Z_s) + K) ds \\ &\leq \int_0^t \mathbb{E}(MV(Z_s) + K) ds \leq tM(\mathbb{E}V(Z_0) + K/\gamma) + Kt. \end{aligned}$$

By letting $N \rightarrow \infty$, we find $T_t < \infty, V(Z_t) < \infty$ a.s. Since V has compact sub-level sets, $|X_t| < \infty$ a.s. We have not mentioned this technical issue till now for two reasons: 1) if it is mentioned in the beginning of Section 2, it is unclear how to find the Lyapunov function V ; 2) both \mathcal{L} and (2.2) are well defined without this condition, so it was safe for Section 2 to work without this condition.

3.1 Convergence in total variation with a minorization regime

The classical notion of ergodicity is often illustrated in the total variation norm, which is defined as

$$\|\mu - \nu\|_{tv} = \sup_{|f| \leq 1} \int f d\mu - \int f d\nu.$$

This norm is also called the L^1 distance, because if μ and ν have densities p and q , $\|\mu - \nu\|_{tv} = \int |p - q| dx$. Geometric ergodicity in total variation is well studied and understood in the finite dimensional Markov chain and SDE [36, 41]. A classical formulation can be found in Theorem 1.5 of [38] or Theorem 2.10 of [29], here we present it using our notation:

Theorem 3.1. *Let P_t be a Markov semigroup admitting a Lyapunov function V . Suppose the minorization condition holds, in the sense that for a sufficiently large C , there is a probability measure ν and $\epsilon, t_0 > 0$, such that*

$$\|P_{t_0}^* \delta_z - P_{t_0}^* \delta_{z'}\|_{tv} > 2 - \epsilon, \quad \forall z, z' : V(z), V(z') \leq C.$$

Then $(P_t)_{t \geq 0}$ has a unique invariant measure π , and for some positive constants D and β ,

$$\|P_t^* \delta_z - \pi\|_{tv} \leq D e^{-\beta t} (1 + V(z)).$$

When F consists of only one regime, X_t is simply an SDE on \mathcal{R}^d . In this context, following the arguments in [41], the minorization conditions of Theorem 3.1 can be verified by the hypoellipticity and reachability conditions below:

Assumption 3.2. *Let $dX_t = f(X_t)dt + \sigma(X_t) \circ dW_t$ be a diffusion process in \mathcal{R}^d , where \circ denotes the Stratonovich integral,*

1. *Hypoellipticity condition: let \mathcal{L} be the Lie algebra generated by the vector fields*

$$\{f, \sigma_1, \dots, \sigma_m\}$$

with σ_i being the columns of σ , and \mathcal{L}_0 is the ideal in \mathcal{L} generated by $\{\sigma_1, \dots, \sigma_m\}$, assuming \mathcal{L}_0 spans \mathcal{R}^d at all points.

2. *Reachability condition: there is a point $x_h \in \mathcal{R}^d$ such that for any compact set C and $\epsilon > 0$, there is a t_0 such that from any $x \in C$ there is a piecewise constant process w_t such that the solution to the following ODE*

$$dx_t = [b(x_t) + \sigma(x_t)w_t]dt, \quad x_0 = x,$$

satisfies $|x_{t_0} - x_h| \leq \epsilon$.

Theorem 3.4 below indicates that for diffusions with random switching, it suffices to check minorization condition for one particular regime, using say Assumption 3.2, and show this regime is commonly accessible and satisfies a mild growth condition for $V(Z_t)$. In particular, we define

Definition 3.3. *A regime $y^* \in F$ is commonly accessible, if for all $z \in \mathcal{R}^d \times F$ there is a $t > 0$ such that $\mathbb{P}^z(Y_t = y^*) > 0$. We say it has polynomial growth for a function V , if there is a constant K_t with polynomial growth in t , such that the following holds for the SDE $dX'_t = b(X'_t, y^*)dt + \sigma(X'_t, y^*)dW_t$,*

$$\mathbb{E}^x V(X'_t, y^*) \leq K_t(V(x, y^*) + 1). \quad (3.1)$$

Theorem 3.4. *Let $Z_t = (X_t, Y_t)$ be a diffusion with random switching that admits a Lyapunov function V . Suppose the transition rates satisfy $\bar{\lambda}(z) \leq MV(z)$, moreover there is a regime $y_h \in F$ such that it is commonly accessible and has polynomial growth for V . Then if the SDE given by $dX'_t = b(X'_t, y_h)dt + \sigma(X'_t, y_h)dW_t$ satisfies the minorization condition in Theorem 3.1, the diffusion Z_t has an invariant measure π and is geometrically ergodic under the total variation distance.*

The proof is located in Section 4, where we will also show a simple way to verify the common accessibility of one regime.

3.2 Wasserstein metric convergence with contraction on average

Another mechanism that may generate geometric ergodicity is contraction. Contraction can be formulated through the Lyapunov exponents of stochastic flows. Recall that in the diffusion part, b is required to be $\mathcal{C}^{2+\delta}$ and σ is required to be $\mathcal{C}^{1+\delta}$ in x , and there is no explosion. As a consequence, the solution to the SDE with $Y_t = y$ can be written as $X_t = \Psi_t^{y,\omega} X_0$, where $\Psi_t^{y,\omega}$ is a diffeomorphism [42]. We say $\rho : F \mapsto \mathcal{R}$ is a *contraction rate function*, if

$$\|D_x \Psi_t^{y,\omega} x\| \leq \exp(-\rho(y)t), \quad \forall (x, y) \in \mathcal{R}^d \times F, \quad a.s. \quad (3.2)$$

Here ω can be seen as a realization of the Wiener process in (1.1), and we denote its law as P_W . One way to verify (3.2) is imposing the following requirement on (1.1):

$$(x - x') \cdot (b(x, y) - b(x', y)) \leq -\rho(y)|x - x'|^2, \quad \sigma(x, y) = \sigma(y).$$

[29, 32] and the references within also suggest other method to establish (3.2) under possibly a different norm, our method below is possible for generalization in those scenarios as well. We will say the joint process admits a *contraction on average*, if there are constants $m, \bar{\rho}, C_\rho > 0$ such that

$$\mathbb{E}^z \exp \left(-m \int_0^t \rho(Y_s) ds \right) \leq C_\rho \exp(-\bar{\rho}t). \quad (3.3)$$

The next subsection will discuss how to verify (3.3) given the contraction rates ρ .

The total variation norm is often too stringent to capture contraction. For example, consider a trivial deterministic process in \mathcal{R} , $dX_t = -\rho X_t dt$. The invariant measure is obviously δ_0 , a point mass at the origin, and it attracts other points. Yet, starting from any nonzero point, the distribution of X_t is a point mass at $e^{-\rho t} X_0$, which has total variation distance 2 from δ_0 .

A more suitable distance for our purpose is the Wasserstein distance, which is also used in previous works for models with bounded or Lipschitz transition [22, 28, 29, 33]. For any distance d on a space E , the associated Wasserstein distance between two measures μ, ν on E is defined as:

$$d(\mu, \nu) := \inf_{\Gamma \in \mathcal{C}(\mu, \nu)} \int d(x, x') \Gamma(dx, dx') \quad (3.4)$$

Here $\mathcal{C}(\mu, \nu)$ is the set of all coupling measures between μ and ν .

The distance function here can be very flexible. One remarkable discovery in [37, 38] is that by properly incorporating the Lyapunov function into d , the corresponding Wasserstein distance can characterize relatively weak convergence. This is known as the asymptotic coupling framework. As for diffusions with random switching, this framework allows us to generalize geometric ergodicity results to cases where the transition rates and their derivatives are bounded only by the Lyapunov function.

Similar to the situation with dissipation, the notion of contraction on average is essential for our discussion. The precise statement is the following:

Theorem 3.5. *Let $Z_t = (X_t, Y_t)$ be a diffusion with random switching that admits a Lyapunov function V . Suppose the following four conditions hold*

1) $V(x, y)$ has polynomial growth in x , so there are n, M such that

$$V(x + u, y) \leq M(V(x) + |u|^n) \quad \text{and} \quad V(x, y) \geq \frac{1}{M}|x|^{\frac{1}{n}}.$$

2) The transition rates and their derivatives are bounded by MV with a constant $M > 0$

$$\bar{\lambda}(z) \leq MV(z), \quad \sum_{y'} |\nabla_x \lambda(x, y, y')| \leq MV(z).$$

3) Each regime admits a contraction rate ρ in the sense of (3.2), the averaged dynamics is contractive as there are $C_\rho, m, \bar{\rho} > 0$:

$$\mathbb{E}^z \exp \left(-m \int_0^t \rho(Y_s) ds \right) \leq C_\rho \exp(-\bar{\rho}t).$$

4) There is a commonly accessible regime y_c , such that $\rho(y_c) > 0$ and V has at most polynomial growth in the sense of Definition 3.3.

Then Z_t has a unique invariant measure π , moreover the distribution of Z_t converges to π geometrically fast in the Wasserstein distance generated by $m(z, z') = \mathbf{1}_{y=y'} \wedge |x - x'| + \mathbf{1}_{y \neq y'}$. In particular the following bound holds with some $C, \beta > 0$,

$$m(P_t^* \delta_z, P_t^* \delta_{z'}) \leq C e^{-\beta t} (1 + V(z)).$$

3.3 Contraction on average

Given the contraction rates ρ in each regime, the contraction on average condition (3.3) can be verified using arguments similar to the ones of dissipation on average in Section 2. Yet, there is an important difference. When we construct a Lyapunov function, it suffices to consider the transitions for large $|x|$, and upper bounds suffice to hold modulo a constant, see Lemma A.2. This is no longer the case for contraction on average, and we will need the “ \lesssim ” inequalities in Section 2 to hold with “ \leq ”. As a consequence, the spectrum and comparison arguments still work with a variant, but the scaling argument no longer works.

The following theorem is the contraction version of Theorem 2.2, while the transition rates are allowed to be non-constant.

Theorem 3.6. *If there are constants $m, \bar{\rho} > 0$ and a vector a with strictly positive components such that the following holds*

$$-m\rho(y)a(y) + [\Lambda(x)a]_y \leq -\bar{\rho}a(y) \tag{3.5}$$

for all $(x, y) \in E$, then the contraction on average (3.3) also holds. In particular, if $\Lambda(x)$ is a constant irreducible transition matrix, and over its invariant measure π , $\sum \pi(y)\rho(y) > 0$, then the conditions mentioned above hold.

Proof. Consider an auxiliary scalar process $dU_t = -m\rho(Y_t)U_t dt$ with $U_0 = 1$ then clearly

$$U_t = \exp \left(-m \int_0^t \rho(Y_s) ds \right) > 0.$$

Now consider a joint function $V(z, u) = a(y)u$, which satisfies the following for any $u > 0$,

$$\mathcal{L}V(z, u) = -m\rho(y)a(y)u + [\Lambda(x)a]_y u \leq -\bar{\rho}a(y)u = -\bar{\rho}V(z, u).$$

As a consequence, $\mathbb{E}U_t \leq e^{-\bar{\rho}t} \frac{\max_y a(y)}{\min_y a(y)}$. □

Following the representation (2.10), we can see $a(y)$ in (3.5) as the potential contraction of one regime. Therefore, the comparison principle can be formulated as follow:

Proposition 3.7. *Let $(\rho, \Lambda(x))$ satisfies (3.5) with a strictly positive vector a , then $(\tilde{\rho}, \tilde{\Lambda}(x))$ satisfies the same inequality, if it is more favorable for contraction in the sense:*

$$\tilde{\rho}(y) \geq \rho(y), \quad (\lambda(x, y, y') - \tilde{\lambda}(x, y, y'))(a(y) - a(y')) \geq 0.$$

The proof is a direct verification and ignored here. On the other hand, contraction on average will require the contraction to hold homogenous inside \mathcal{R}^d , but not just for large enough $|x|$. This is probably an intrinsic requirement due to the following example:

Example 3.8. *Let $F = \{1, -1\}$, and $dX_t = -Y_t X_t dt$ to be a scalar process, while*

$$\lambda(x, 1, -1) = x^2 + 3, \quad \lambda(x, -1, 1) = 3x^2 + 1.$$

With a simple application of Theorem 2.4, it is easy to see this system has dissipation on average. Indeed, the transition rates favor contraction when $|x|$ is large. However, when $|x|$ is close to 0, the transition favors inverse contraction. As a consequence, although $\frac{1}{4}\delta_{(0,1)} + \frac{3}{4}\delta_{(0,-1)}$ is clearly an invariant measure, starting from any $x_0 \neq 0$, x_t will never reach 0. In fact, there is at least another invariant measure on \mathcal{R}^+ with density:

$$p(x, 1) = p(x, -1) = x \exp(-x^2), \quad x > 0.$$

The invariance can be verified by the Fokker Planck equation, which is the dual of \mathcal{L} . On the other hand, if we replace the ODE of X_t with $dX_t = -y_t X_t dt + dW_t$, the noise will connect the invariant measures, so the process becomes ergodic due to Theorem 3.4.

4 Geometric ergodicty through random PDMPs

4.1 Random PDMPs

Piecewise deterministic Markov processes (PDMP) are special cases of diffusions with random switching, as they require the SDEs of X_t to be ODEs, in other words, $\sigma(Z_t) \equiv 0$. In this case, assuming Y_s stays in regime y up to time t , the value of X_t is given by a diffeomorphism $X_t = \Psi_{s,t}^y X_s$ for any $s \leq t$. A PDMP can be defined, based on $\Psi_{s,t}^y$ and transition rates $\lambda(x, y, y')$. In fact, X_t is completely govern by the jumps of Y_t : if Y_t has jumps at time $\mathbf{t} = (t_1, \dots, t_n)$ with a jump sequence $\mathbf{y} = (y_1, \dots, y_n)$, then X_t is given by

$$X_t = \Psi(X_0, \mathbf{t}, \mathbf{y}, t) := \Psi_{t_n, t}^{y_n} \circ \Psi_{t_{n-1}, t_n}^{y_{n-1}} \circ \dots \circ \Psi_{0, t_1}^{y_0} X_0. \quad (4.1)$$

Moreover, the probability density of such event, that is having n jumps before time t with jump times τ_k and jumps going to y_k , is given by formula 3.10 of [43]

$$p_{n, \mathbf{t}, \mathbf{y}}^{z_0, t} dt := \mathbf{1}_{t_1 < t_2 < \dots < t_n} \exp\left(-\int_0^t \bar{\lambda}(z_s) ds\right) \prod_{i=1}^n (\lambda(z_{t_i-}, y_{t_i}) dt_i). \quad (4.2)$$

In (4.2), the X part of z_s is given by (4.1), and $z_{s-} = \lim z_r$ with r approaches s from left.

On the other hand, diffusions with random switching can be viewed as random PDMPs. As noted in Section 3.2, the solution of the SDE in a fixed regime can be written as $X_t =$

$\Psi_{s,t}^{y,\omega} X_s$, where ω denotes the realization of the Wiener process W_s , and we denote the law of ω as P_W . Therefore, if we condition on each realization ω , a diffusion with random switching is simply a PDMP. Following the nomenclature of statistical physics [44], this PDMP will be called a *quenched* process, as it is the conditioning of the original joint process Z_t , on one realization of random outcome ω . In contrast, the original process without conditioning will be called the *annealed* process. We will adopt this simple terminology.

Viewing a diffusion with random switching as a random PDMP gives us two explicit formulas. First, given the jump times \mathbf{t} and jumps \mathbf{y} , X_t is given by $\Psi_\omega(X_0, \mathbf{t}, \mathbf{y}, t)$, which is defined in (4.1) with $\Psi_{s,t}^y$ replaced by $\Psi_{s,t}^{y,\omega}$. Second, if $p_{n,\mathbf{t},\mathbf{y}}^{z_0,t,\omega}$ denotes the density (4.2) with the diffeomorphisms being $\Psi_{s,t}^{y,\omega}$, then the law of the annealed process can be recovered by averaging over P_W :

$$\mathbb{P}^{z_0}(Z_{s \leq t} \in A) = \int P_W(d\omega) \sum_{n=0}^{\infty} \sum_{\mathbf{y} \in F^n} \int_{[0,t]^n} dt p_{n,\mathbf{t},\mathbf{y}}^{z_0,t,\omega} \mathbf{1}_{z_s \leq t \in A}, \quad (4.3)$$

where z_s has its x part given by $x_s = \Psi_\omega(x_0, \mathbf{t}, \mathbf{y}, s)$ and $y_s = y_k$ if $t_k \leq s \leq t_{k+1}$. These explicit formulas will be instrumental for our proofs below.

4.2 Accessibility analysis

In both the minorization and contraction on average scenarios, we need a good regime to be commonly accessible. In this section we will discuss the consequence of this assumption and also provide a simple way to verify in Lemma 4.2. Most derivations here are relatively standard and may have a simpler version in [31, 28, 29] when the transition rates are bounded. We provide the complete proofs here to be self-contained.

Lemma 4.1. *Suppose Z_t admits a Lyapunov function V , $\Lambda(x)$ is continuous in x while $\bar{\lambda} \leq MV$ for some $M > 0$.*

- 1) *If $\mathbb{P}^z(Y_t = \tilde{y}) > 0$, then $\mathbb{P}^z(Y_{t+s} = \tilde{y}) > 0$ for any $s \geq 0$;*
- 2) *For each z and $t > 0$, there exists a neighbor of x , $O_x \subset \mathbb{R}^d$, such that*

$$\mathbb{P}^{x',y}(Y_t = \tilde{y}) \geq \frac{1}{2} \mathbb{P}^z(Y_t = \tilde{y}), \quad \forall x' \in O_x;$$

- 3) *If there is a $\tilde{y} \in F$ that is commonly accessible, then for any fixed compact set C , there exists some $t_0, m_0 > 0$ such that*

$$\mathbb{P}^z(Y_{t_0} = \tilde{y}) \geq m_0, \quad \forall z \in C.$$

Proof. Claim 1) Our condition implies $p_{n,\mathbf{t},\mathbf{y}}^{z,t,\omega} > 0$ with $y_n = \tilde{y}$ for certain $\omega, \mathbf{t}, \mathbf{y}$. Then observe that,

$$p_{n,\mathbf{t},\mathbf{y}}^{z,t+s,\omega} = p_{n,\mathbf{t},\mathbf{y}}^{z,t,\omega} \exp \left(- \int_0^s \bar{\lambda}(\Psi_{t,t+r}^{y_n,\omega} x_t, y_n) dr \right).$$

The exponential term above is nonzero for P_W -a.s. ω , because (2.3) leads to

$$\begin{aligned} \int_0^s \int P_W(d\omega) \bar{\lambda}(\Psi_{t,t+r}^{y_n, \omega} x_t, y_n) dr &\leq M \int_0^s \int_0^s P_W(d\omega) V(\Psi_{t,t+r}^{y_n, \omega} x_t, y_n) dr \\ &\leq sK_s M(V(x_t, y_n) + 1) < \infty. \end{aligned}$$

So for P_W -a.s. ω , $p_{n, \mathbf{t}, \mathbf{y}}^{z, t, \omega} > 0$ implies $p_{n, \mathbf{t}, \mathbf{y}}^{z, t+s, \omega} > 0$, and annealing (4.3) produces our claim.

Claim 2) By formula (4.2), $p_{n, \mathbf{t}, \mathbf{y}}^{z_0, \omega, t}$ depends continuously on $x_{s \leq t}$, which depends continuously on x_0 because of (4.1). Thus $p_{n, \mathbf{t}, \mathbf{y}}^{z_0, \omega, t}$ depends continuously on x_0 , and claim 2) follows by applying Fatou's lemma to the following annealing formula over any sequences $z' \rightarrow z$.

$$\mathbb{P}^{z'}(Y_t = \tilde{y}) = \int P_W(d\omega) dt \sum_{n, \mathbf{y}: y_n = \tilde{y}} p_{n, \mathbf{t}, \mathbf{y}}^{z', t, \omega}.$$

Claim 3) By claim 2) and the fact that C is compact, we can find a finite cover of C , $\{O_i\}_{i=1, \dots, n}$, and a sequence of time t_i such that

$$\mathbb{P}^{z'}(Y_{t_i} = \tilde{y}) > 0 \quad \forall z' \in O_i.$$

Let $t_0 = \max\{t_i\}$, we have $\mathbb{P}^z(Y_{t_0} = \tilde{y}) > 0$ for all $z \in C$ by claim 1). Then using the compactness again with claim 2), we can find a uniform lower bound m_0 for the transition probability. \square

The following Lemma provides an easy verification that a regime y^* is commonly accessible.

Lemma 4.2 (Burst mechanism). *Under the same conditions of Lemma 4.1, fix any $z_0 \in E$ and a sequence in F y_0, y_1, \dots, y_n such that*

$$\lambda(x_0, y_i, y_{i+1}) > 0, \quad i = 0, 1, \dots, n-1. \quad (4.4)$$

Then for any $t > 0$, $\mathbb{P}^{z_0}(Y_t = y_n) > 0$. Therefore, if there is a $y^ \in F$ such that for any $z_0 \in E$, there is a sequence $y_0, \dots, y_n = y^*$ such (4.4) holds, then y^* is commonly accessible.*

Proof. By claim 1) of Lemma 4.1, it suffices to show our claim for sufficiently small t . Since λ is continuous in x , so we can find $0 < \delta < 1$ and an $M > 0$ such that the following holds:

$$\lambda(x, y_i, y_{i+1}) > 0, \quad \bar{\lambda}(x, y_i) < M, \quad \forall \|x - x_0\| \leq \delta, i = 0, 1, \dots, n-1.$$

Then for P_W -a.s. realization of ω , because $\Psi_{s,t}^{y, \omega} x$ is continuous in s, t and x , by induction there is a sequence of measurable functions $\epsilon_k(\omega) \leq \delta, k = 0, 1, \dots, n$ such the following holds:

$$|\Psi_{s,t}^{y_{k-1}, \omega} x - x_0| < \epsilon_k(\omega), \quad \forall x : |x - x_0| < \epsilon_{k-1}(\omega), s \leq t \leq \epsilon_k(\omega),$$

where $\epsilon_n(\omega) = \delta$. Then pick any ϵ such that $P_W(\epsilon_0(\omega) > \epsilon) > 0$, and consider any fixed jump time sequence $\mathbf{t} = (t_1, t_2, \dots, t_n)$ with $t_n < \epsilon$ and the generated process $x_s = \Psi_\omega(x_0, \mathbf{t}, \mathbf{y}, s)$. with $\mathbf{y} = (y_1, \dots, y_n)$. It is easy to verify that if $\epsilon_0(\omega) > \epsilon$, then $|x_s - x_0| < \delta$ for all $s \leq \epsilon$, therefore $p_{n, \mathbf{y}, \mathbf{t}}^{z_0, \omega, \epsilon} > 0$ for these ω . This completes the proof by the annealing formula (4.3). \square

4.3 Ergodicity with a minorization regime

Due to Theorem 3.1, the proof of Theorem 3.4 is a relative standard verification of the small set condition for the full Markov semigroup P_t .

Proof of Theorem 3.4. In order to apply Theorem 3.1, it suffices for us to show the minorization condition. By the equivalence between total variation and coupling measures, this is equivalent to building a coupling of Z_t and Z'_t such that $P(Z_t = Z'_t) \geq \epsilon$, if $V(Z_0), V(Z'_0) \leq C$. Our strategy will be showing there are $t_0 > 0$ and $\delta > 0$,

$$\mathbb{P}^z(Y_{t_0} = y_h, V(Z_{t_0}) \leq C) > \delta, \quad \forall z \in U, \quad (4.5)$$

and then use the fact that Y_t is possible to stay as y_h , while the minorization in regime h can be used to build a coupling.

Since y_h is commonly accessible, by Lemma 4.1 3), there are t_0, m_0 , such that $\mathbb{P}^z(Y_{t_0} = y_h) > m_0$ for all $z \in \mathcal{R}^d \times F$. Because V is a Lyapunov function, $\mathbb{E}^z(V(Z_{t_0})) \leq e^{-\gamma t_0} C + K/\gamma$ for all C . By picking a large C , we can make

$$\mathbb{P}^z(V(Z_{t_0}) > C) \leq \frac{\mathbb{E}^z V(Z_{t_0})}{C} \leq \frac{1}{2} m_0,$$

hence (4.5) holds with $\delta = \frac{1}{2} m_0$. By Lemma A.1, there is a coupling of Z_{t_0} and Z'_{t_0} with law \mathbb{Q}^z such that

$$\mathbb{Q}^z((Z_{t_0}, Z'_{t_0}) \in A_1) > \delta, \quad A_1 := \{(z, z') : y = y' = y_h, V(z), V(z') \leq C\}.$$

By our assumption, $y = y_h$ satisfied the minorization condition. This means there are $t_1 > 0$ and $\epsilon > 0$, so that for any x, x' that $V(x, y_h), V(x', y_h) \leq C$, there is a coupling of $P_W(\omega)$ and $P_W(\omega')$, denoted by $Q_W(d\omega, d\omega')$ such that

$$\mathbb{E}_{y_h}^{x, x'}(X_{t_1} = X'_{t_1}) = \int Q_W(d\omega, d\omega') \mathbb{1}_{\Psi_{0, t_1}^{y_h, \omega} x = \Psi_{0, t_1}^{y_h, \omega'} x'} \geq \epsilon. \quad (4.6)$$

Now we extend \mathbb{Q}^z from time t_0 to $T = t_0 + t_1$ by coupling ω, ω' as Q_W after time t_0 . Then Markov property yields

$$\begin{aligned} \mathbb{Q}^z(Z_T = Z'_T) &\geq \mathbb{Q}^z(Y_{t_0 \leq s \leq T} = Y'_{t_0 \leq s \leq T} = y_h, V(Z_{t_0}), V'(Z_{t_0}) \leq C, X_T = X'_T) \\ &\geq \frac{1}{2} m_0 \inf_{(z, z') \in A_1} \int Q_W(d\omega, d\omega') p_{0, \emptyset, \emptyset}^{z, t_1, \omega} p_{0, \emptyset, \emptyset}^{z', t_1, \omega'} \mathbb{1}_{\Psi_{0, t_1}^{y_h, \omega} x = \Psi_{0, t_1}^{y_h, \omega'} x'}. \end{aligned}$$

It suffices to show the density that there are no jumps till time t_1 , which is $p_{0, \emptyset, \emptyset}^{z, t_1, \omega} p_{0, \emptyset, \emptyset}^{z', t_1, \omega'}$, is bounded from below on a set of probability more than $1 - \frac{1}{2}\epsilon$, then union bound with (4.6) will yield our claim. For this purpose, note

$$p_{0, \emptyset, \emptyset}^{(x, y_h), t_1, \omega} = \exp\left(-\int_0^{t_1} \bar{\lambda}(x_s, y_h) ds\right),$$

while by polynomial growth of V within regime y_h ,

$$\int P_W(d\omega) \int_0^{t_1} \bar{\lambda}(x_s, y_h) ds \leq M \int P_W(d\omega) \int_0^{t_1} V(x_s, y_h) ds \leq M t_1 K_{t_1} (V(x_0) + 1).$$

So by Markov's inequality, there is an N such that

$$P_W \left(p_{0,\emptyset,\emptyset}^{z,t_1,\omega} \leq \exp(-N) \right) \leq P_W \left(\int_0^{t_1} \bar{\lambda}(x_s, y_h) ds \geq N \right) \leq \frac{1}{4}\epsilon.$$

Then because Q_W is a coupling, by union bound,

$$Q_W \left(p_{0,\emptyset,\emptyset}^{z,t_1,\omega} p_{0,\emptyset,\emptyset}^{z',t_1,\omega'} \leq \exp(-2N) \right) \leq \frac{1}{2}\epsilon.$$

□

4.4 Ergodicity with contraction on average

The proof of Theorem 3.5 uses the asymptotic coupling mechanism introduced by [45, 38]. Theorem 4.8 of [38] presented below, formulates our application of this mechanism.

Theorem 4.3. *Let P_t be a Markov semigroup over a Polish space E admitting a continuous Lyapunov function V , so $\mathbb{E}V(Z_t) \leq e^{-\gamma t}\mathbb{E}V(Z_0) + K$. Suppose there exists a distance-like function $d : E \times E \mapsto [0, 1]$ and a time t such that*

1) P_t is locally contracting in d :

$$d(P_t^* \delta_z, P_t^* \delta_{z'}) \leq \frac{1}{2}d(z, z'), \quad \forall d(z, z') < 1.$$

2) *Smallness: for any two z, z' such that $V(z), V(z') \leq K$, $d(P_t^* \delta_z, P_t^* \delta_{z'}) \leq 1 - \epsilon$.*

Then P_t can have at most one invariant probability measure π . Furthermore, let $\tilde{d}(z, z') = \sqrt{d(z, z')(1 + V(z) + V(z'))}$, there exists a $t > 0$ such that $\tilde{d}(P_t^ \mu, P_t^* \nu) \leq \frac{1}{2}\tilde{d}(\mu, \nu)$ for any probability measures μ, ν on E .*

In [38], $d : E \times E \mapsto \mathcal{R}^+$ is distant-like if it is symmetric, lower semicontinuous, and $d(z, z') = 0 \Leftrightarrow z = z'$. Its associated Wasserstein-1 distance is also denoted by d for notational simplicity. In other words, for two probability measures μ and ν ,

$$d(\mu, \nu) := \inf_{\Gamma} \int d(z, z') \Gamma(dz, dz'),$$

where the infimum is taken over all coupling measures of μ and ν .

The reason that 1) is called a local contraction, is that in most applications, $d(z, z') = 1$ unless z and z' are very close. Theorem 4.3 essentially extends a local contraction to a global one.

4.4.1 Contracting distance

For the construction of a contracting distance, we have the following lemma. It is a variant of Lemma 4.13 in [38] which uses a Lyapunov function instead of a super Lyapunov function. The proof goes very similar.

Proposition 4.4. *Under the conditions of Theorem 3.5, the following distance with any positive $r \leq \min\{\frac{m}{2}, \frac{1}{2}\}$ is locally $\frac{1}{2}$ -contracting for P_T with a proper T and $\delta < 2^{-\frac{1}{r}}$:*

$$d(z, z') = 1_{y \neq y'} + 1_{y=y'} \wedge \delta^{-1} \left(\inf_{\theta: x \rightarrow x'} \int_0^1 V(\theta(s), y) |\dot{\theta}(s)| ds \right)^r. \quad (4.7)$$

Here the infimum is taken over all the C^1 paths $\theta : [0, 1] \mapsto \mathcal{R}^d$ that connects x and x' .

Before we move on to the proof of Proposition 4.4, we need two pieces of arguments. The first one indicates $d(z, z') \approx 1_{y \neq y'} + 1_{y=y'} \wedge \delta^{-1} (V(z) |x - x'|)^r$.

Lemma 4.5. *Assuming condition 1) of Theorem 3.5, fix any $y \in F$, $x, x' \in \mathbb{R}^d$ and a C^1 path $\theta : [0, 1] \mapsto \mathcal{R}^d$ that connects x and x' so that*

$$\inf_{\theta: x \rightarrow x'} \int_0^1 V(\theta(s), y) |\dot{\theta}(s)| ds \leq \frac{1}{2},$$

then

$$\frac{1}{2M} |x - x'| V(z) \leq \inf_{\theta: x \rightarrow x'} \int_0^1 V(\theta(s), y) |\dot{\theta}(s)| ds \leq 2M |x - x'| V(z).$$

Proof. Recall that we can always assume that $V \geq 1$, so our condition leads to $|x - x'| \leq \frac{1}{2}$. Because of the polynomial growth of V in x , when $|u| \leq 1$, $V(x + u, y) \geq M^{-1}V(x, y) - 1$, so

$$V(x + u, y) \geq \frac{1}{2} [(M^{-1}V(x, y) - 1) + 1] = \frac{1}{2M} V(x, y), \quad \forall |u| \leq 1.$$

Therefore for any C^1 path θ that connects x with x' , while $|x - x'| \leq \frac{1}{2}$, if it completely lies in $B_x(1) = \{x + u : |u| \leq 1\}$, then

$$\int_0^1 V(\theta(s), y) |\dot{\theta}(s)| ds \geq \frac{1}{2M} V(z) \int_0^1 |\dot{\theta}(s)| ds \geq \frac{1}{2M} V(z) |x - x'|.$$

In the other case, if θ has a part lying outside $B_x(1)$, then there is an exiting time of $B_x(1)$, $\tau = \inf\{s : \theta(s) \notin B_x(1)\}$. By definition $\theta([0, \tau])$ is a C^1 path of length at least 1, hence

$$\int_0^1 V(\theta(s), y) |\dot{\theta}(s)| ds \geq \int_0^\tau V(\theta(s), y) |\dot{\theta}(s)| ds \geq \frac{1}{2M} V(z) \geq \frac{1}{2M} V(z) |x - x'|.$$

For the other side of the bound, we only need to show it holds for $\theta(s) = x + s(x - x')$, which is the following by polynomial growth:

$$|x - x'| \int_0^1 V(\theta(s), y) ds \leq M |x - x'| (V(z) + 1).$$

□

The second lemma gives a bound on the perturbation of measures caused by perturbation on the initial condition:

Lemma 4.6. *Under the conditions of Theorem 3.5, for any fixed T there is a constant D_T such that*

$$\int P_W(d\omega) \sum_{n, \mathbf{y}} \int_{[0, T]^n} dt \|D_x p_{n, \mathbf{t}, \mathbf{y}}^{z, \omega, T}\| \leq D_T (V(z) + 1).$$

Proof. Recall that the density function is

$$p_{n, \mathbf{t}, \mathbf{y}}^{z, \omega, T} = 1_{t_1 < \dots < t_n < T} \exp \left(- \int_0^T \bar{\lambda}_s(z_s) ds \right) \prod_i^n (\lambda(z_{t_i-}, y_{t_i}) dt_i).$$

Applying Fréchet derivative and the chain rule, we have

$$\begin{aligned} & \int P_W(d\omega) \sum_{n, \mathbf{y}} \int_{[0, T]^n} dt \|D_x p_{n, \mathbf{t}, \mathbf{y}}^{z, \omega, T}\| \\ & \leq \int P_W(d\omega) \sum_{n, \mathbf{y}} \int_{[0, T]^n} dt p_{n, \mathbf{t}, \mathbf{y}}^{z, \omega, T} \int_0^T \|D_x \Psi_\omega(x, \mathbf{t}, \mathbf{y}, s) \partial_x \bar{\lambda}(x_s, y_s)\| ds \\ & \quad + \int P_W(d\omega) \sum_{n, \mathbf{y}} \int_{[0, T]^n} dt p_{n, \mathbf{t}, \mathbf{y}}^{z, \omega, T} \sum_{k=1}^n \|D_x \Psi_\omega(x, \mathbf{t}, \mathbf{y}, t_k)\| \frac{\|\partial_x \lambda(z_{t_k-}, y_{t_k})\|}{\lambda(z_{t_k-}, y_{t_k})}. \end{aligned} \quad (4.8)$$

We will bound the two parts separately in below. Since for $t_k \leq s < t_{k+1}$ and every ω

$$\|D_x \Psi_\omega(x, \mathbf{t}, \mathbf{y}, s)\| = \|D_x \Psi_{t_k, s}^{y_k, \omega} \circ \Psi_{t_{k-1}, t_k}^{y_{k-1}, \omega} \circ \dots \circ \Psi_{0, t_1}^{y_0, \omega} x_0\| \leq \exp \left(- \int_0^s \rho(y_s) ds \right) \leq \exp(M_\rho T).$$

Here $M_\rho := \max_y \{-\rho(y)\} < \infty$. Using condition 2) of Theorem 3.5, the first part of (4.8) is bounded by the following

$$M \int P_W(d\omega) \sum_{n, \mathbf{y}} \int dt p_{n, \mathbf{t}, \mathbf{y}}^{z, \omega, T} \int_0^T \exp(M_\rho T) V(z_s) ds = M \exp(M_\rho T) \mathbb{E}^z \int_0^T V(Z_s) ds,$$

which is bounded further by $K_T V(z)$ with a proper K_T since V is a Lyapunov function. The equality above holds as we recognize the probabilistic meaning of the integrals and then use that V is a Lyapunov function. For the second part of (4.8), according to condition 2) of Theorem 3.5, it is clearly bounded by

$$M \exp(M_\rho T) \int P_W(d\omega) \sum_{n, \mathbf{y}} \int_{[0, T]^n} dt p_{n, \mathbf{t}, \mathbf{y}}^{z, \omega, T} \sum_{k=1}^n \frac{V(z_{t_k-})}{\lambda(z_{t_k-}, y_{t_k})} = M \exp(M_\rho T) \mathbb{E}^z \sum_{k: \tau_k \leq T} \frac{V(z_{\tau_k-})}{\lambda(Z_{\tau_k-}, Y_{\tau_k})},$$

where τ_k are the sequential jump times. Apply formula 31.18 in [46] with $b(z', z) = \frac{V(z)}{\lambda(z, y')}$ on the quenched PMDP, and then annealing, we find

$$\mathbb{E}^z \sum_{k: \tau_k \leq T} \frac{V(z_{\tau_k-})}{\lambda(Z_{\tau_k-}, Y_{\tau_k})} = \mathbb{E}^z \int_0^T V(Z_t) dt,$$

which is bounded further by $K_T V(z)$ with a proper K_T since V is a Lyapunov function. \square

We are finally at the position to prove Proposition 4.4:

Proof of Proposition 4.4. By the definition of contracting metric, it suffices for us to show that $d(P_T^* \delta_z, P_T^* \delta_{z'}) \leq \frac{1}{2} d(z, z')$ when $d(z, z') < 1$. This implies that $y = y'$, and $|x - x'| \leq \frac{1}{2}$. Since the spaces here are Polish, by the Kantorovich-Rubinstein Theorem (Theorem 11.8.2 of [47]),

$$d(P_T^* \delta_z, P_T^* \delta_{z'}) = \sup_{\varphi} \left\{ P_T^{x,y} \varphi - P_T^{x',y} \varphi \mid \|\varphi\|_{Lip(d)} \leq 1 \right\}.$$

$\|\varphi\|_{Lip(d)}$ denotes the Lipschitz norm, in other words, for any $z, z' \in E$, $\varphi(z) - \varphi(z') \leq \|\varphi\|_{Lip(d)} d(z, z')$. Hence, to prove this lemma we only need to show for any φ , $\|\varphi\|_{Lip(d)} \leq 1$,

$$P_T^{x,y} \varphi - P_T^{x',y} \varphi \leq \frac{1}{2\delta} \left(\inf_{r:x \rightarrow x'} \int_0^1 V(r(s), y) |\dot{r}(s)| ds \right)^r. \quad (4.9)$$

However, if φ has its d -Lipschitz norm less than 1, its maximum variation is less than 1, we can replace φ by $\varphi - c$ such that $\|\varphi\|_{\infty} \leq \frac{1}{2}$, yet $\int \varphi dQ^{x,y} - \varphi dQ^{x',y}$ remains invariant. So without loss of generality, we assume $\|\varphi\|_{\infty} \leq \frac{1}{2}$.

Consider splitting $P_T^{x,y} \varphi - P_T^{x',y} \varphi$ into the difference caused by the initial condition, and the difference caused by the underlying probability measure:

$$\begin{aligned} |P_T^{x,y} \varphi - P_T^{x',y} \varphi| &= |\mathbb{E}^{x,y} \varphi(\Psi_{\omega}(x, Y_{s \leq T}, T), Y_T) - \mathbb{E}^{x',y} \varphi(\Psi_{\omega}(x', Y_{s \leq T}, T), Y_T)| \\ &\leq |\mathbb{E}^{x,y} \varphi(\Psi_{\omega}(x, Y_{s \leq T}, T), Y_T) - \mathbb{E}^{x,y} \varphi(\Psi_{\omega}(x', Y_{s \leq T}, T), Y_T)| \end{aligned} \quad (4.10)$$

$$+ |\mathbb{E}^{x,y} \varphi(\Psi_{\omega}(x', Y_{s \leq T}, T), Y_T) - \mathbb{E}^{x',y} \varphi(\Psi_{\omega}(x', Y_{s \leq T}, T), Y_T)|. \quad (4.11)$$

The quantity $\mathbb{E}^{x,y} \varphi(\Psi_{\omega}(x', Y_{s \leq T}, T), Y_T)$ has its probability initialized from point (x, y) but the stochastic flow is initialized at x' , in other words:

$$\mathbb{E}^{x,y} \varphi(\Psi_{\omega}(x', Y_{s \leq T}, T), Y_T) = \int P_W(d\omega) \sum_{n, \mathbf{y}} p_{n, \mathbf{t}, \mathbf{y}}^{z, \omega, T} \int_{[0, T]^n} dt \varphi(\Psi_{\omega}(x', \mathbf{y}, T), y_T).$$

Since φ is d -Lipschitz, so the first part is bounded as below by Lemma 4.5

$$\begin{aligned} (4.10) &\leq \mathbb{E}^z d((\Psi_{\omega}(x, Y_{s \leq T}, T), Y_T), (\Psi_{\omega}(x', Y_{s \leq T}, T), Y_T)) \\ &\leq \mathbb{E}^z 1 \wedge \delta^{-1} \left(\inf_{\theta} \int_0^1 V(\theta(s), Y_T) \|\dot{\theta}(s)\| ds \right)^r \\ &\leq \mathbb{E}^z \delta^{-1} (2M)^r V(Z_T)^r |u_T|^r, \end{aligned}$$

where $u_T = \Psi_{\omega}(x', Y_{s \leq T}, T) - X_T$. By Cauchy Schwartz,

$$\mathbb{E}^z V(Z_T)^r |u_T|^r \leq \sqrt{\mathbb{E}^z V(Z_T)^{2r}} \sqrt{\mathbb{E}^z |u_T|^{2r}}.$$

Notice $\mathbb{E}^z |u_T|^{2r} \leq [\mathbb{E}^z |u_T|^m]^{\frac{2r}{m}}$, and by contraction on average:

$$\mathbb{E} |u_T|^m \leq \mathbb{E}^z |x - x'|^m \exp \left(- \int_0^T m \rho(Y_s) ds \right) \leq C_{\rho} |x - x'|^m \exp(-\bar{\rho} T).$$

Likewise, $\mathbb{E}^z |V(Z_T)|^{2r} \leq [\mathbb{E}^z V(Z_T)]^{2r} \leq C_0 V(z)^{2r}$ for a constant C_0 . As a consequence, there is a constant C_1 such that

$$(4.10) \leq C_1 \exp(-\frac{2r}{m} \bar{\rho} T) |x - x'|^r V(z)^r,$$

which further leads to the following by Lemma 4.5 for a constant C_2 :

$$(4.10) \leq \delta^{-1} C_2 \exp(-\frac{2r}{m} \bar{\rho} T) \left(\inf_{\theta: x \rightarrow x'} \int_0^1 V(\theta(s), y) \|\dot{\theta}(s)\| ds \right)^r.$$

In the following, we will fix a T such that $C_2 \exp(-\frac{2r}{m} \bar{\rho} T) \leq \frac{1}{4}$. In order to bound (4.11), by definition:

$$\begin{aligned} (4.11) &= \left| \int P_W(d\omega) \sum_{n, \mathbf{y}} (p_{n, \mathbf{t}, \mathbf{y}}^{z, \omega, T} - p_{n, \mathbf{t}, \mathbf{y}}^{z', \omega, T}) \int_{[0, T]^n} d\mathbf{t} \varphi(\Psi_\omega(x', \mathbf{t}, \mathbf{y}, T), y_T) \right| \\ &\leq \|\varphi\|_\infty \int P_W(d\omega) \sum_{n, \mathbf{y}} \int_{[0, T]^n} d\mathbf{t} |p_{n, \mathbf{t}, \mathbf{y}}^{z, \omega, T} - p_{n, \mathbf{t}, \mathbf{y}}^{z', \omega, T}| \\ &\leq \frac{1}{2} \int P_W(d\omega) \sum_{n, \mathbf{y}} \int_{[0, T]^n} d\mathbf{t} \int_0^1 \|D_x p_{n, \mathbf{t}, \mathbf{y}}^{z_s, \omega, T}\| |\dot{\theta}(s)| ds. \end{aligned}$$

where $z' = (x', y)$ and $z_s = (\theta(s), y)$, and θ is any C^1 path from x to x' . By Lemma 4.6, there is a constant D_T such that

$$\int P_W(d\omega) \sum_{n, \mathbf{y}} \int_{[0, T]^n} d\mathbf{t} \|D_x p_{n, \mathbf{t}, \mathbf{y}}^{z_s, \omega, T}\| \leq D_T V(z_s).$$

Hence for any C^1 path θ that connects x and x' while $\int_0^1 V(z_s) |\dot{\theta}(s)| ds \leq 1$, because $r < 1$,

$$(4.11) \leq \frac{D_T}{2} \int_0^1 V(z_s) |\dot{\theta}(s)| ds \leq \frac{D_T}{2} \left(\int_0^1 V(z_s) |\dot{\theta}(s)| ds \right)^r.$$

So combine the bounds for (4.10) and (4.11), we have

$$\mathbb{E}^{x, y} \varphi(Z_T) - \mathbb{E}^{x', y} \varphi(Z_T) \leq \left(\frac{1}{4\delta} + \frac{D_T}{2} \right) \left(\int_0^1 V(z_s) |\dot{\theta}(s)| ds \right)^r,$$

it suffice to take $\delta \leq D_T^{-1}$ in (4.9) for this proposition to hold. \square

4.4.2 Small set verification

Verification of condition 2) of Theorem 4.3 is given by the following

Lemma 4.7. *Under the condition of Theorem 3.5, for any fixed strictly positive C , there are some $T, \epsilon > 0$ such that:*

$$d(P_T^* \delta_z, P_T^* \delta_{z'}) \leq 1 - \epsilon, \quad \forall z, z' : V(z), V(z') \leq C.$$

Proof. The proof is exactly the same as Theorem 3.4, and a coupling will be constructed through two steps. We first couple Y_{t_0}, Y'_{t_0} to y_c , which follows exactly the same commonly accessible argument as in the proof of Theorem 3.4. Then we keep the value of Y to be the same afterwards, until the contracting dynamics brings the X part close enough. For this part, it suffices to show that there is a t_1 such that from any two x and x' with $V(x, y_c), V(x', y_c) \leq C$, there is a coupling Q_W of $P_W(d\omega)$ and $P_W(d\omega')$, such that

$$\int Q_W(d\omega, d\omega') d((\Psi_{t_1}^{y_c, \omega} x, y_c), (\Psi_{t_1}^{y_c, \omega'} x', y_c)) \leq 1 - \epsilon. \quad (4.12)$$

Then (4.12) replaces (4.6), while the remainder of the proof of Theorem 3.4 can be used again.

By Lemma 4.5, for all ω ,

$$d((\Psi_t^{y_c, \omega} x, y_c), (\Psi_t^{y_c, \omega} x', y_c)) \leq \delta^{-1} (2M)^r \exp(-r\rho(y_c)t) V((\Psi_t^{y_c, \omega} x), y_c)^r |x - x'|^r.$$

Note that $V(x, y_c), V(x', y_c) \leq C$ implies $|x - x'|$ is bounded, so for a t large enough,

$$d((\Psi_t^{y_c, \omega} x, y_c), (\Psi_t^{y_c, \omega} x', y_c)) \leq \frac{\exp(-\frac{1}{2}r\rho(y_c)t)}{2C} V((\Psi_t^{y_c, \omega} x), y_c)^r \leq \frac{\exp(-\frac{1}{2}r\rho(y_c)t)}{2C} V((\Psi_t^{y_c, \omega} x), y_c).$$

Therefore, if we let $\omega = \omega'$, that is $Q_W(d\omega, d\omega') = P(d\omega)\delta_{\omega'=\omega}$, the left hand side of (4.12) is bounded by (3.1)

$$\frac{\exp(-\frac{1}{2}r\rho(y_c)t)}{2C} \int P_W(d\omega) V((\Psi_t^{y_c, \omega} x), y_c) \leq \exp(-\frac{1}{2}r\rho(y_c)t) K_t.$$

Since K_t has at most polynomial growth, a large $t = t_1$ would produce (4.12). \square

4.4.3 Proof of Theorem 3.5

With the conditions of Theorem 4.3 verified, it is rather elementary to show Theorem 3.5.

Proof. By the existence of Lyapunov function, and Krylov-Bogouliubov's Theorem, there exists an invariant measure π with $\mathbb{E}^\pi V = \int \pi(dz) V(z) < \infty$. Let d be as in Proposition 4.4 and \tilde{d} be defined by Theorem 4.3, then

$$\tilde{d}(P_{nt}^* \delta_z, P_{nt}^* \pi) \leq \frac{1}{2} \tilde{d}(P_{(n-1)t}^* \delta_z, P_{(n-1)t}^* \pi) \leq \dots \leq \frac{1}{2^n} \tilde{d}(\delta_z, \pi).$$

By Lemma 4.5, for some constant C_0

$$\begin{aligned} \tilde{d}(\delta_z, \pi) &\leq \int \pi(dz') \sqrt{(1 + V(z) + V(z')) (2M)^r \delta^{-1} |x - x'|^r V(z)^r} \\ &\leq C_0 + C_0 V(z)^{\frac{r+1}{2}} \int \pi(dz') |x - x'|^{\frac{r}{2}} + C_0 V(z)^{\frac{r}{2}} \int \pi(dz') |x - x'|^{\frac{r}{2}} V(z')^{\frac{1}{2}}. \end{aligned}$$

Since $V(z') \geq M^{-1} |x|^{\frac{1}{n}}$, using the bound above, we can pick $r < \frac{2}{n}$ in the beginning such that

$$\tilde{d}(\delta_z, \pi) \leq C_0 (1 + 2V(z) \mathbb{E}^\pi V).$$

On the other hand, by Lemma 4.5 and that $1 \wedge u^r \geq 1 \wedge u$ for any $u \geq 0$,

$$\tilde{d}(z, z') \geq 1_{y=y'} \wedge \frac{|x-x'|^r}{(2M)^r} V(z)^r + 1_{y \neq y'} \geq 1_{y=y'} \wedge \frac{|x-x'|}{2M} + 1_{y \neq y'} \geq \frac{1}{2M} m(z, z').$$

Combining the inequalities above, we have our claim. \square

5 Conclusion and discussion

Diffusions with random switching is a general class of stochastic processes with applications in many different areas. Such system consists of a diffusion process part X_t and a Markov jump part Y_t , and their dynamics can be fully coupled. The stability and ergodicity of such processes remain open questions if the transition rates are not bounded or Lipschitz. This paper closes this gap by developing a new analytical framework.

The first part of this paper constructs polynomial type Lyapunov functions when there is dissipation on average. These functions can be used to derive moment bounds on the diffusion part. The incorporation of potential dissipation of each regime is found to be an efficient way to capture averaged dissipation. This idea can be easily applied to the classical case where the transition rates are constants as Theorem 2.2. It also leads to a simple illustration of comparison principles, Theorem 2.9. Moreover, with a Fredholm alternative argument, we demonstrate how can the Lyapunov function be inductively constructed, as a dual process of the averaging procedure, Theorems 2.4, 2.7, assuming the transition rates have a multiscale structure.

The second part of this paper is devoted to the geometric ergodicity of diffusions with random switching, assuming a Lyapunov function exists. If there is a commonly accessible regime that satisfies the minorization condition, Theorem 3.4 proves geometric ergodicity under total variation distance. When there is contraction on average, using the asymptotic coupling framework of [37, 38], Theorem 3.6 demonstrates geometric ergodicity under a proper Wasserstein distance.

There are a few interesting ideas came to the authors when this paper is written. We share them in below to inspire further research.

1. The authors conjecture that the results here hold in a similar form if the regime process Y_t is instead a continuous stochastic process. One way to see this is taking the limit of a jump process on grid points of vanishing size. But the authors suspect that an independent mechanism can be setup for these processes without little change of the proofs. Such theory will be applicable to many nonlinear models that exhibits intermittency, for example [48].
2. The attraction or contraction rates used in this paper provides a uniform control over each component of X_t . A more general situation is that each component of X_t has a different regime based attraction rate, in other words, the attraction rate is given by a matrix. A simple example will be $dX_t = A(Y_t)X_t dt$, where A is a matrix valued function. It is known such system is very sensitive to the switching even if A has all eigenvalues of negative real parts [49]. It will be interesting if our results can be generalized to this case.

3. Theorem 3.4 can probably be generalized. [30] shows that a PDMP is geometric ergodic in total variation despite that each regime is degenerate. The proofs there only require a Hörmander type of condition on the vector fields generated by different regimes. The authors conjecture that for diffusions with random regime switching, in order to have geometric ergodicity, it suffices to have the Lie algebra generated by stochastic flows of all regimes spanning \mathcal{R}^d . But this requires a completely different set of techniques, and Assumption 3.2 should be general enough to cover most applications.

Acknowledgment The authors would like to thank R van Handel for his suggestions over the presentation of this article. This research is supported by the MURI award grant N-000-1412-10912, where A.J.M. is the principal investigator, while X.T.T. was supported as a postdoctoral fellows. This research is also partly supported by the NUS grant R-146-000-226-133, where X.T.T. is the principal investigator.

A Proofs for miscellaneous claims

Lemma A.1. *For two probability measures μ, ν on a polish space, assume that $\mu(X \in A) \geq p$ and $\nu(Y \in B) \geq p$, then there exists a coupling Γ of μ and ν such that $\Gamma(X \in A, Y \in B) \geq p$.*

Proof. This is also Lemma A.1 in [33]. □

Lemma A.2. *Suppose the generators of two processes, \mathcal{L} and \mathcal{L}' , are the same outside a compact set C . Suppose V is a Lyapunov function of the generator \mathcal{L} , so $\mathcal{L}V \leq -\gamma V + K$. Then V is also a Lyapunov function for \mathcal{L}' if $\mathcal{L}'V$ and V are bounded in C , since*

$$\mathcal{L}'V \leq -\gamma V + K + \sup_{z \in C} \{|\gamma V - K| + \mathcal{L}'V\}.$$

Lemma A.3. *Under the linear regime dissipation Assumption 2.1, with any $m \in \mathcal{R}$ and $n \geq \max\{2, 2 - m\}$, the function $f_m(x) = \frac{|x|^{m+n}}{1+|x|^n}$ is \mathcal{C}^2 , $f_m(x) \leq |x|^m$. Moreover, for any $\delta > 0$*

$$\mathcal{L}_X(y)f_m(x) \lesssim -(m\gamma(y) - \delta)f_m(x).$$

Proof. The \mathcal{C}^2 regularity of f_m and that $f_m(x) \leq |x|^m$ can be verified directly. The gradient of f_m is

$$\nabla f_m(x) = \frac{m|x|^{m+n-2}x}{1+|x|^n} + \frac{n|x|^{m+n-2}x}{(1+|x|^n)^2},$$

from which we can conclude that the Hessian $\nabla^2 f_m \lesssim M|x|^{m-2}$ with a constant M . Then from Assumption 2.1, we find

$$\mathcal{L}_X(y)f_m(x) = \frac{m|x|^{m+n-2}\langle b(x, y), x \rangle}{1+|x|^n} + O(|x|^{m+n-\epsilon}).$$

Then by Young's inequality, for any $\delta > 0$ we have $\mathcal{L}_X(y)f_m(x) \lesssim -(m\gamma(y) - \delta)f_m(x)$. □

Lemma A.4. *Based on the definition of averaging procedures and V_0 , the duality equations (2.15) hold.*

Proof. For the first duality equation, simply check

$$\begin{aligned} [\tilde{\Lambda}(x)\tilde{V}(x, \cdot)]_y &= \sum_{y' \in \tilde{F}} \tilde{\lambda}(x, y, y') [\tilde{V}(x, y') - \tilde{V}(x, y)] \\ &= \sum_{y' \in F/F'} \tilde{\lambda}(x, y, y') [\tilde{V}(x, y') - \tilde{V}(x, y)] + \sum_k \tilde{\lambda}(x, y, g_k) [\tilde{V}(x, g_k) - \tilde{V}(x, y)]. \end{aligned}$$

The first summation above is $\sum_{y' \in F/F'} \lambda(x, y, y') [V_0(x, y') - V_0(x, y)]$. Based on (2.13), the k -th term in the second summation can be written as

$$\sum_{y' \in G^c} \lambda(x, y, y') p_{F'}(y', g_k) [\tilde{V}(x, g_k) - V_0(x, y)] + \sum_{y' \in G_k} \lambda(x, y, y') [V_0(x, g_k) - V_0(x, y)].$$

If we sum the first term above over all k , it is $\sum_{y' \in G^c} \lambda(x, y, y') [V_0(x, y') - V_0(x, y)]$. Putting these back to the decomposition of $[\tilde{\Lambda}(x)\tilde{V}(x, \cdot)]_y$, we can conclude the first duality equation holds.

Likewise, the left hand of the second duality equation can be written as

$$[\tilde{\Lambda}(x)\tilde{V}(x, \cdot)]_{g_k} = \sum_{y' \in F/F'} \tilde{\lambda}(x, g_k, y') [\tilde{V}(x, y') - \tilde{V}(x, g_k)] + \sum_{j \neq k} \tilde{\lambda}(x, g_k, g_j) [\tilde{V}(x, g_j) - \tilde{V}(x, g_k)].$$

Based on (2.13), for $y' \in F/F'$, each term in the first summation is

$$\tilde{\lambda}(x, g_k, y') [\tilde{V}(x, y') - \tilde{V}(x, g_k)] = \sum_{y \in G_k} \pi_k(y) \lambda(x, y, y') [V_0(x, y') - V_0(x, y)].$$

And each term in the second summation can be written as

$$\begin{aligned} &\sum_{j \neq k} \left(\sum_{y \in G_k} \pi_k(y) \sum_{y' \in G_j} \lambda(x, y, y') + \sum_{y \in G_k} \pi_k(y) \sum_{y' \in G^c} \lambda(x, y, y') p_{F'}(y', g_j) \right) [\tilde{V}(x, g_j) - \tilde{V}(x, g_k)] \\ &= \sum_{y \in G_k} \pi_k(y) \left(\sum_{y' \in G_j, \forall j} \lambda(x, y, y') [V_0(x, y') - V_0(x, y)] + \sum_{y' \in G^c} \lambda(x, y, y') [V_0(x, y') - V_0(x, y)] \right). \end{aligned}$$

If we sum the second term on the right hand side over all k , it is $\sum_{y' \in G^c} \lambda(x, y, y') [V_0(x, y') - V_0(x, y)]$. Therefore, we can conclude the second duality equation holds. \square

References

- [1] F Malrieu. Some simple but challenging Markov processes. arXiv:1412.7516.
- [2] A J Majda and B Khouider. Stochastic and mesoscopic models for tropical convection. *Proc. Natl Acad. Sci.*, 99:1123–1128, 2002.
- [3] B Khouider, A J Majda, and M A Katsoulakis. Coarse-grained stochastic models for tropical convection and climate. *Proc. Natl. Acad. Sci.*, 100(21):11941–11946, 2003.

- [4] A J Majda, C Franzke, and B Khouider. An applied mathematics perspective on stochastic modelling for climate. *Phil. Trans. Roy. Soc.*, 336(1875):2427–2453, 2008.
- [5] Y Frenkel, A J Majda, and B Khouider. Using the stochastic multicloud model to improve tropical convective parameterization: a paradigm example. *J. Atmos. Sci.*, 69(3):1080–1105, 2012.
- [6] Y Frenkel, A J Majda, and B Khouider. Stochastic and deterministic multicloud parameterizations for tropical convection. *Climate Dynamics*, 41(5-6):1527–1551, 2013.
- [7] Q Deng, B Khouider, and A J Majda. The MJO in a coarse-resolution GCM with a stochastic multicloud parameterization. *J. Atmos. Sci.*, 72(1):55–74, 2015.
- [8] B Khouider. A coarse grained stochastic multi-type particle interacting model for tropical convection: Nearest neighbour interactions. *Comm. Math. Sci.*, 12(8):1379–1407, 2014.
- [9] S Thual, A J Majda, and S Stechmann. A stochastic skeleton model for mjo. *J. Atmos. Sci.*, 71:697–715, 2014.
- [10] S Thual, A J Majda, and S Stechmann. Assymetrical intraseasonal events in the stochastic skeleton mjo model with seasonal cycle. Accepted by *Climate Dynamics*, 2014.
- [11] M A Katsoulakis, A J Majda, and A Sopasakis. Multiscale couplings in prototype hybrid deterministic/stochastic systems: Part i, deterministic closures. *Comm. Math. Sci.*, 2(2):255–294, 2004.
- [12] M A Katsoulakis, A J Majda, and A Sopasakis. Multiscale couplings in prototype hybrid deterministic/stochastic systems: Part ii, stochastic closures. *Comm. Math. Sci.*, 3(3):453–478, 2005.
- [13] M A Katsoulakis, A J Majda, and A Sopasakis. Intermittency, metastability and coarse graining for coupled deterministic-stochastic lattice systems. *Nonlinearity*, 19:1021–1323, 2006.
- [14] M A Katsoulakis, A J Majda, and A Sopasakis. Hybrid deterministic stochastic systems with microscopic look-ahead dynamics. *Comm. Math. Sci.*, 8(2):409–437, 2010.
- [15] E Buckwar and M Riedler. An exact stochastic hybrid model of excitable membranes including spatio-temporal evolution. *J. Math. Bio.*, 63(6):1051–1093, 2011.
- [16] C Schutte, J Walter, C Hartmann, and W Huisinga. An averaging principle for fast degrees of freedom exhibiting long-term correlations. *Multiscale Model. Simul.*, 2(3):501–526, 2004.
- [17] D T Crommelin and E Vanden-Eijnden. Subgrid scale parameterization with conditional Markov chains. *J. Atmos. Sci.*, 65:2661–2675, 2008.
- [18] B Gershgorin, J Harlim, and A J Majda. Test models for improving filtering with model errors through stochastic parameter estimation. *J. Comput. Phys.*, 229:1–31, 2010.

- [19] A J Majda and J Harlim. *Filtering complex turbulent systems*. Cambridge University Press, Cambridge, UK, 2012.
- [20] W Lee and A M Stuart. Derivation and analysis of simplified filters for complex dynamical systems. arXiv:1512.03647.
- [21] B de Saporta and J Yao. Tail of a linear diffusion with Markov switching. *Ann. Appl. Probab.*, 15(1B):992–1018, 2005.
- [22] J B Bardet, H Guérin, and F Malrieu. Long time behavior of diffusions with Markov switching. *ALEA Lat. Am. J. Probab. Math. Stat.*, 7:151–170, 2010.
- [23] Y Bakhtin, T Hurth, and J C Mattingly. Regularity of invariant densities for 1D systems with random switching. *Nonlinearity*, 28(11):3755–3787, 2015.
- [24] S D Lawley, J C Mattingly, and M C Reed. Stochastic switching in infinite dimensions with applications to random parabolic PDEs. *SIAM J. Mathematical Analysis*, 47(4):3035–3063, 2015.
- [25] A J Majda and X Wang. *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*. Cambridge University Press, Cambridge, UK, 2006.
- [26] A J Majda and X T Tong. Ergodicity of truncated stochastic Navier-Stokes with deterministic forcing and dispersion. *J Nonlinear Sci*, DOI: 10.1007/s00332-016-9310-0, 2016.
- [27] X Mao and C Yuan. Asymptotic stability in distribution of stochastic differential equations with Markovian switching. *Stochastic processes and their applications*, 103:277–291, 2003.
- [28] M Benaïm, S Le Borgne, F Malrieu, and P A Zitt. Quantitative ergodicity for some switched dynamical systems. *Electron. Commun. Probab.*, 17(56):1–14, 2012.
- [29] B Cloez and M Hairer. Exponential ergodicity for Markov processes with random switching. *Bernoulli*, 21(1):505–536, 2015.
- [30] Y Bakhtin and T Hurth. Invariant densities for dynamical systems with random switching. *Nonlinearity*, 25(10):2937–2952, 2012.
- [31] M Benaïm, S Le Borgne, F Malrieu, and P A Zitt. Qualitative properties of certain piecewise deterministic Markov process. *Ann. Inst. H. Poincaré Probab. Statist.*, 51(3):1040–1075, 2015.
- [32] J Shao. Ergodicity of regime-switching diffusions in Wasserstein distances. *Stochastic Process. Appl.*, 125:739–758, 2015.
- [33] A J Majda and X T Tong. Geometric ergodicity for piecewise contracting processes with applications for tropical stochastic lattice models. *Comm. Pure Appl. Math*, 69(6):1110–1153, 2016.

- [34] G A Pavliotis and A M Stuart. *Multiscale Methods: Averaging and Homogenization*, volume 53 of *Texts in applied mathematics*. Springer, 2008.
- [35] K Pakdaman, M Thieullen, and G Wainrib. Asymptotic expansion and central limit theorem for multiscale piecewise-deterministic markov processes. *Stochastic Process. Appl.*, 122:2292–2318, 2012.
- [36] S Meyn and R Tweedie. *Markov chains and stochastic stability*. Springer-Verlag, 1993.
- [37] M Hairer and J C Mattingly. Ergodicity of the 2d navier-stokes equations with degenerate stochastic forcing. *Annals of Mathematics*, 164:993–1032, 2006.
- [38] M Hairer, J C Mattingly, and M Scheutzow. Asymptotic coupling and a general form of Harris’ theorem with applications to stochastic delay equations. *Probab. Theory Related Fields*, 149(1-2):223–259, 2011.
- [39] J R Norris. *Markov Chains*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, 1997.
- [40] G Da Prato and J Zabczyk. *Ergodicity for infinite dimensional systems*. London mathematical society lecture note series. Cambridge University Press, 2006.
- [41] J C Mattingly, A M Stuart, and D J Higham. Ergodicity for sdes and approximations: Locally lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232, 2002.
- [42] H Kunita. *Lectures on Stochastic Flows and Applications*. Springer-Verlag, 1986.
- [43] M Jacobson. *Point Process Theory and Applications Marked POint and Piecewise Deterministic Processes*. Probability and its Applications. Birkhauser, 2006.
- [44] S Tavaré and O Zeitouni. *Lectures on Probability Theory and Statistics, Ecole d’Eté de Probabilités de Saint-Flour XXXI - 2001*, volume 1837 of *Lecture Notes in Mathematics*. Springer, 2004.
- [45] M Hairer and J C Mattingly. A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs. *Electron. Commun. Probab.*, 16(23):658–738, 2011.
- [46] M H A Davis. *Markov Models and Optimization*. Monographs on Statistics and Applied Probability. Chapman & Hall, 1993.
- [47] R M Dudley. *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2004.
- [48] A J Majda and Y Lee. Conceptual dyanmical models for turbulence. *Proc. Natl. Acad. Sci.*, 111(18):6548–6533, 2014.
- [49] S D Lawley, J C Mattingly, and M C Reed. Sensitivity to switching rates in stochastically switched odes. *Communiations in Mathematical Sciences*, 12(7):1343–1352, 2014.