Approximating a Diffusion by a Finite-State Hidden Markov Model

I. Kontoyiannis* S.P. Meyn†

December 7, 2016

Abstract

For a wide class of continuous-time Markov processes evolving on an open, connected subset of $\mathbb{R}^d$, the following are shown to be equivalent:

(i) The process satisfies (a slightly weaker version of) the classical Donsker-Varadhan conditions;

(ii) The transition semigroup of the process can be approximated by a finite-state hidden Markov model, in a strong sense in terms of an associated operator norm;

(iii) The resolvent kernel of the process is 'v-separable', that is, it can be approximated arbitrarily well in operator norm by finite-rank kernels.

Under any (hence all) of the above conditions, the Markov process is shown to have a purely discrete spectrum on a naturally associated weighted L_∞ space.

Keywords: Markov process, hidden Markov model, hypoelliptic diffusion, stochastic Lyapunov function, discrete spectrum

*__Corresponding author__. Department of Informatics, Athens University of Economics and Business, Patission 76, Athens 10434, Greece. Email: `yiannis@aueb.gr`. I.K. was supported by the European Union and Greek National Funds through the Operational Program Education and Lifelong Learning of the National Strategic Reference Framework through the Research Funding Program Thales-Investing in Knowledge Society through the European Social Fund.

†Department of Electrical and Computer Engineering, University of Florida, Gainesville, USA. Email: `meyn@ece.ufl.edu`. S.P.M. was supported in part by the National Science Foundation ECS-0523620, and AFOSR grant FA9550-09-1-0190. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or AFOSR.

1 Introduction

Consider a continuous-time Markov process $\boldsymbol{\Phi} = \{\Phi(t) : t \geq 0\}$ taking values in an open, connected subset X of $\mathbb{R}^d$, equipped with its associated Borel σ-field $\mathcal{B}$. We begin by assuming that $\boldsymbol{\Phi}$ is a diffusion; that is, it is the solution of the stochastic differential equation,

$$d\Phi(t) = u(\Phi(t))dt + M(\Phi(t))dB(t), \quad t \geq 0,\ \Phi(0) = x, \tag{1}$$

where $u = (u_1, u_2, \dots, u_d)^T : \mathsf{X} \to \mathbb{R}^d$ and $M : \mathsf{X} \to \mathbb{R}^d \times \mathbb{R}^k$ are locally Lipschitz, and $\boldsymbol{B} = \{B(t) : t \geq 0\}$ is k-dimensional standard Brownian motion. [Extensions to more general Markov processes are briefly discussed in Section 1.4.] Unless explicitly stated otherwise, throughout the paper we assume that:

The strong Markov process $\boldsymbol{\Phi}$ is the unique strong solution of (1) with continuous sample paths. **(A1)**

The distribution of the process $\boldsymbol{\Phi}$ is described by the initial condition $\Phi(0) = x \in \mathsf{X}$ and the transition semigroup $\{P^t\}$: For any $t \geq 0$, $x \in \mathsf{X}$, $A \in \mathcal{B}$,

$$P^t(x, A) := \mathsf{P}_x\{\Phi(t) \in A\} := \Pr\{\Phi(t) \in A \,|\, \Phi(0) = x\}.$$

Recall that the kernel P^t acts as a linear operator on functions $f : \mathsf{X} \to \mathbb{R}$ on the right and on signed measures ν on $(\mathsf{X}, \mathcal{B})$ on the left, respectively, as,

$$P^t f\,(x) = \int f(y) P^t(x, dy), \quad \nu P^t\,(A) = \int \nu(dx) P^t(x, A), \ \ x \in \mathsf{X},\ A \in \mathcal{B},$$

whenever the above integrals exist. Also, for any signed measure ν on $(\mathsf{X}, \mathcal{B})$ and any function $f : \mathsf{X} \to \mathbb{R}$ we write $\nu(f) := \int f d\nu$, whenever the integral exists. In this paper we will constrain the domain of functions f to a Banach space defined with respect to a weighted L_∞ norm.

One of the central assumptions we make throughout the paper is the following regularity condition on the semigroup:

The transition semigroup admits a continuous density: There is a continuous function p on $(0, \infty) \times \mathsf{X} \times \mathsf{X}$ such that,

$$P^t(x, A) = \int_A p(t, x, y)\, dy\,, \qquad x \in \mathsf{X},\ A \in \mathcal{B}.$$

(A2)

Hörmander's theorem [30, Thm. 38.16] gives sufficient conditions for (A2). Explicit bounds on the density are also available; see [27] and its references.

1.1 Irreducibility, drift, and semigroup approximations

The ergodic theory of continuous-time Markov processes is often most easily addressed by translating results from the discrete-time domain. This is achieved, e.g., in [8, 24, 25, 23] through consideration of the Markov chain whose transition kernel is defined by one of the *resolvent kernels* of $\boldsymbol{\Phi}$, defined as,

$$R_\alpha := \int_0^\infty e^{-\alpha t} P^t\, dt, \quad \alpha > 0. \tag{2}$$

In the case $\alpha = 1$ we simply write $R := R_1 = \int_0^\infty e^{-t} P^t\, dt$, and call R "the" resolvent kernel of the process $\mathbf{\Phi}$.

The family of resolvent kernels $\{R_\alpha\}$ is simply the Laplace transform of the semigroup, so that each R_α admits a density under (A2). This density will not be continuous in general, so we will truncate to obtain the positive kernel,

$$\overline{R}_\alpha = \int_{t_0}^{t_1} e^{-\alpha t} P^t\, dt\,, \tag{3}$$

where $0 < t_0 < t_1 < \infty$ will be chosen so that $\overline{R}_\alpha$ is a good approximation to R_α. The approximation admits a continuous density under (A2),

$$\overline{R}_\alpha(x, A) = \int_A \bar{\xi}_\alpha(x, y)\, dy\,, \qquad x \in \mathsf{X},\ A \in \mathcal{B}, \tag{4}$$

where for each x, y,

$$\bar{\xi}_\alpha(x, y) = \int_{t_0}^{t_1} e^{-\alpha t} p(t, x, y)\, dt.$$

Proposition 1.1. *Under Assumptions* (A1) *and* (A2)*, for any* $\alpha > 0$*, the resolvent kernel* R_α *has the strong Feller property. Moreover, there exist continuous functions* $s_\alpha, n_\alpha \colon \mathsf{X} \to \mathbb{R}_+$ *that are not identically zero, and satisfy,*

$$R_\alpha(x, dy) \geq s_\alpha(x) n_\alpha(y)\, dy, \qquad x, y \in \mathsf{X}. \tag{5}$$

Proof. Condition (A2) implies the strong Feller property for the semigroup $\{P^t\}$, that is, the function $P^t f$ is continuous whenever f is measurable and bounded, for $t > 0$. It is then straightforward to show that the kernel R_α also has the strong Feller property for any $0 < \alpha < \infty$.

The existence of the functions s_α and n_α in the lower bound follows from the obvious bound $R_\alpha \geq \overline{R}_\alpha$. □

The function s_α and the positive measure defined by $\mu_\alpha(dy) = n_\alpha(y) dy$ are called *small*, and the inequality (5) is written in terms of an outer product as, $R_\alpha \geq s_\alpha \otimes \mu_\alpha$; cf. [26, 22]. Without loss of generality (through normalization) we always assume that $\mu_\alpha(\mathsf{X}) = 1$, so that μ_α defines a probability measure on $(\mathsf{X}, \mathcal{B})$.

Some of the results on ergodic theory require the following 'reachability' condition for $\mathbf{\Phi}$; it is a mild irreducibility assumption:

> There is a state $x_0 \in \mathsf{X}$ such that, for any $x \in \mathsf{X}$ and any open set O containing x_0, we have,
>
> $$P^t(x, O) > 0, \qquad \text{for all } t \geq 0 \text{ sufficiently large.}$$
>
> **(A3)**

Under (A3) we are assured of a single communicating class, since then the process is ψ-irreducible and aperiodic with $\psi(\,\cdot\,) := R(x_0, \cdot\,)$: For all $x \in \mathsf{X}$ and all $A \in \mathcal{B}$ such that $R(x_0, A) > 0$, we have,

$$P^t(x, A) > 0, \qquad \text{for all } t \text{ sufficiently large.}$$

See [24, Theorem 3.3] and Proposition 2.2 below.

Recall that the *generator* of $\mathbf{\Phi}$ is expressed, for bounded C^2 functions $f\colon \mathsf{X}\to\mathbb{R}$, as,

$$\mathcal{D}f\,(x) = \sum_i u_i(x)\frac{d}{dx_i}f\,(x) + \tfrac{1}{2}\sum_{ij}\Sigma_{ij}(x)\frac{d^2}{dx_i\,dx_j}f\,(x), \qquad x\in\mathsf{X}, \tag{6}$$

or, in more compact notation,

$$\mathcal{D} = u\cdot\nabla + \tfrac{1}{2}\mathrm{trace}\,(\Sigma\nabla^2),$$

where $\Sigma = MM^T$. Rather than restricting attention to C^2 functions, we consider the extended generator, as in our previous work [20, 24]. The function $f\colon\mathsf{X}\to\mathbb{R}$ is in the domain of $\mathcal{D}$ if there exists a function $g\colon\mathsf{X}\to\mathbb{R}$ such that the stochastic process defined by,

$$M(t) = f(\Phi(t)) - \int_0^t g(\Phi(s))\,ds, \qquad t\ge 0, \tag{7}$$

is a *local martingale*, for each initial condition $\Phi(0)$ [9, 30]. We then write $g=\mathcal{D}f$.

If $\boldsymbol{M}$ is in fact a martingale, then the following integral equation holds:

$$P^t f = f + \int_0^t P^s g\,ds, \qquad t\ge 0\,. \tag{8}$$

See Proposition 2.4 for a class of functions (f,g) solving (8).

Fleming's *nonlinear generator* [12] for the continuous-time Markov process $\mathbf{\Phi}$ is defined via,

$$\mathcal{H}(F) := e^{-F}\mathcal{D}e^{F}\,. \tag{9}$$

Its domain is the set of functions F for which $f=e^F$ is in the domain of $\mathcal{D}$. Theory surrounding multiplicative ergodic theory and large deviations based on the nonlinear generator is described, e.g., in [10, 36, 11, 20]. We say that the *Lyapunov drift criterion* (DV3) *holds with respect to the Lyapunov function* $V:\mathsf{X}\to(0,\infty]$, if there exist a function $W\colon\mathsf{X}\to[1,\infty)$, a compact set $C\subset\mathsf{X}$, and constants $\delta>0$, $b<\infty$, such that,

$$\mathcal{H}(V) \le -\delta W + b\mathbb{I}_C\,. \tag{DV3}$$

In most of the subsequent results, the following strengthened version of (DV3) is assumed:

> Condition (DV3) holds with respect to continuous functions V,W that have compact sublevel sets. **(A4)**

Recall that the sublevel sets of a function $F\colon\mathsf{X}\to\mathbb{R}_+$ are defined by,

$$C_F(r) = \{x\in\mathsf{X} : F(x)\le r\}, \qquad r\ge 0. \tag{10}$$

Note that the local Lipschitz assumption in (1) together with (DV3) imply (A1); namely, that (1) has a unique strong solution $\mathbf{\Phi}$ with continuous sample paths; see [25, Theorem 2.1] and [30, Theorem 11.2].

Conditions (A1–A4) are essentially equivalent to (but weaker than) the conditions imposed by Donsker and Varadhan in their pioneering work [5, 6, 7]. Condition (DV3) is a generalization of the drift condition of Donsker and Varadhan. Variants of this drift condition are used in [1, 36, 28, 19, 15], and (DV3) is the central assumption in [20].

One important application of (DV3) here and in [20] is in the truncation of the state space – this is how we obtain a hidden Markov model (HMM) approximation, where the approximating process eventually evolves on a compact set. Important related results have been obtained by Wu; see [35, 36, 37] and the references therein. Wu, beginning with his 1995 work [35], has developed a similar truncation technique for establishing large deviations limit theorems, as well as the existence of a spectral gap in the L_p norm, in a spirit similar to this paper and [20]. For bibliographies on these methods and other applications see [14, 15]. A significant further contribution of the present paper, in contrast to the earlier work mentioned, is the introduction of the *weighted* L_∞ *norm* for applications to large deviations theory and spectral theory. In particular, for *non-reversible* Markov processes, the theory is greatly simplified and extended by posing spectral theory within the weighted L_∞ framework.

The weighted norm is based on the Lyapunov function V from (DV3). We let $v = e^V$ and define, for any measurable function $g\colon \mathsf{X} \to \mathbb{R}$,

$$\|g\|_v := \sup\Big\{\frac{|g(x)|}{v(x)} : x \in \mathsf{X}\Big\};$$

cf. [34, 18, 17] and the discussion in [22]. The corresponding Banach space is denoted $L_\infty^v := \{g\colon \mathsf{X} \to \mathbb{R} : \|g\|_v < \infty\}$, and the induced operator norm on linear operators $K\colon L_\infty^v \to L_\infty^v$ is,

$$|\!|\!| K |\!|\!|_v := \sup\Big\{\frac{\|Kh\|_v}{\|h\|_v} : h \in L_\infty^v,\ \|h\|_v \neq 0\Big\}.$$

An analogous weighted norm is defined for signed measures ν on $(\mathsf{X}, \mathcal{B})$ via,

$$\|\nu\|_v := \sup\Big\{\frac{|\nu(h)|}{\|h\|_v} : h \in L_\infty^v,\ \|h\|_v \neq 0\Big\}.$$

The operator on L_∞^v induced by the resolvent kernel R will be shown to satisfy $|\!|\!| R |\!|\!|_v < \infty$ under (DV3) (see Proposition 2.1), and it is known that $|\!|\!| P^t |\!|\!|_v$ is uniformly bounded in t under this condition (see the proof of Theorem 6.1 of [25]).

All of the approximations in this paper are obtained with respect to $|\!|\!| \cdot |\!|\!|_v$. Our main results are all based on Theorem 1.5 below, which establishes conditions ensuring that the semigroup $\{P^t\}$ of the process $\mathbf{\Phi}$ can be approximated (in this weighted operator norm) by a semigroup written in terms of finite-rank kernels. In particular, Theorem 1.5 states that the Donsker-Varadhan condition (DV3) holds if and only if the process $\mathbf{\Phi}$ can be approximated by an HMM in operator norm.

The approximating HMM is based on a generator that is a finite-rank perturbation of the identity, of the form,

$$\mathcal{E} = \kappa\Big[-I + \mathbb{I}_{C_0} \otimes \nu_1 + \sum_{i,j=1}^N r_{ij}\, \mathbb{I}_{C_i} \otimes \nu_j\Big] \tag{11}$$

where $\{C_i : 1 \leq i \leq N\}$ is a finite collection of disjoint, precompact sets, C_0 is the complement of their union, $\mathsf{X} \backslash \cup_{1 \leq i \leq N} C_i$, and $\{\nu_i\}$ are probability measures on $(\mathsf{X}, \mathcal{B})$ with each ν_i supported on C_i. The constants κ and $\{r_{ij}\}$ are nonnegative, and the $\{r_{ij}\}$ define a transition matrix on the finite set $\{1, 2, \dots, N\}$. The approximating semigroup is expressed as the exponential family,

$$Q^t = e^{t\mathcal{E}}, \qquad t \geq 0, \tag{12}$$

where the exponential is defined via the usual power-series expansion. The family of resolvent kernels of the semigroup $\{Q^t\}$ is denoted T_α, $\alpha > 0$, where,

$$T_\alpha = \int_0^\infty e^{-\alpha t} Q^t \, dt. \tag{13}$$

The generator $\mathcal{E}$ will be constructed so that T_α approximates R_α in L_∞^v for α in a neighborhood of unity (see Proposition 3.4).

While connections between separability and condition (DV3) were previously established in [37, 20], Theorem 1.5 goes well beyond prior work. In particular, the equivalence between (DV3) and the finite-state HMM approximation in the strong sense given in the theorem cannot be foreseen based on earlier results. Although the main results of [37, 20] admit extensions to Markov models in continuous time, essential properties of a diffusion must be exploited to obtain the uniform bound (14).

Theorem 1.2. [(DV3) $\Leftrightarrow$ HMM APPROXIMATION] *For a Markov process* $\mathbf{\Phi}$ *on* X *satisfying conditions* (A1)*,* (A2) *and* (A3)*, the following are equivalent:*

(i) DONSKER-VARADHAN ASSUMPTION: *Condition* (DV3) *holds in the form given in* (A4)*.*

(ii) HMM APPROXIMATION: *There exists a continuous function* $v\colon \mathsf{X} \to [1, \infty)$ *with compact sublevel sets (possibly different from the function* v *in* (i)*), such that the following approximations hold: For each* $\varepsilon > 0$ *and* $\delta \in (0, 1)$*, there exists a semigroup* $\{Q^t\}$ *as in* (12) *with generator* $\mathcal{E}$ *of the form given in* (11) *and with an associated family of resolvent kernels* $\{T_\alpha\}$ *as in* (13)*, satisfying the following:*

  (a) Resolvent approximation: *The resolvent kernels* (2) *and* (13) *satisfy,*

  $$\|\!|R_\alpha - T_\alpha|\!\|_v \leq \varepsilon, \qquad \delta \leq \alpha \leq \delta^{-1}.$$

  (b) Semigroup approximation:

  $$\|P^t g - Q^t g\|_v \leq \varepsilon(\|g\|_v + \|\mathcal{D}^2 g\|_v), \qquad t \geq 0\,, \tag{14}$$

  for each C^4 *function* g *with compact support.*

  (c) Invariant measure approximation: *The two semigroups have unique invariant probability measures* π *and* ϖ*, satisfying,*

  $$\|\pi - \varpi\|_v \leq \varepsilon.$$

Proof. The proof is based on several results contained in Section 3:

For the implication (i) $\Rightarrow$ (ii), the function v appearing in (A4) can be chosen the same as the function v appearing in (iia)–(iic). The implication (i) $\Rightarrow$ (iia) is contained in Proposition 3.4; the implication (i) $\Rightarrow$ (iib) follows from Proposition 3.7 combined with Proposition 3.8; and the implication (i) $\Rightarrow$ (iic) is given in Corollary 3.9.

Finally, the implication (ii) $\Rightarrow$ (i) follows from Proposition 3.1: Under (ii) it follows that (A4) holds for continuous functions V_-, W_-, where $V_- \in L_\infty^V$. $\square$

We next consider the probabilistic side of this theory, and we show that a Markov process with generator of the form given in (11) admits a representation as a finite-state hidden Markov model.

1.2 Hidden Markov model approximations

A finite-state space hidden Markov model (HMM) in continuous time is defined as a pair $(\Upsilon, \boldsymbol{I})$, where $\boldsymbol{I}$ is a Markov process with finite state space X_I. The first component Υ is called the observation process; it is a stochastic process taking values in some set Y. The joint dynamics are described as follows: There is a family of probability measures $\{\nu_i\}$ on Y such that, for all measurable $A \subset \mathsf{Y}$,

$$\mathsf{P}\{\Upsilon(t) \in A \mid (\Upsilon(s), I(s)),\ s < t;\ I(t) = i\} = \nu_i(A), \quad i \in \mathsf{X}_I\,.$$

Here we explain how, under our conditions, the continuous time Markov process $\boldsymbol{\Phi}$ may be approximated by the finite-state process Υ of an appropriately constructed HMM. In fact, here the HMM will be special, in that the process Υ itself will be Markovian.

Recall that the generator $\mathcal{D}$ of $\boldsymbol{\Phi}$ will be approximated by a generator $\mathcal{E}$ of the form given in (11). Let Y denote the compact set $\mathsf{Y} := \overline{\bigcup_{i=1}^N C_i}$, and let $\boldsymbol{\Psi}$ denote the continuous-time Markov process with generator $\mathcal{E}$, and with corresponding transition semigroup $\{Q^t\}$ defined in (12). The process $\boldsymbol{\Psi}$ will define the observation process Υ in our HMM approximation.

A probabilistic description of $\boldsymbol{\Psi}$ is based on a sequence of jump times $\{\tau_k : k \geq 0\}$, with $\tau_0 := 0$. The description of τ_1 depends on the initial condition $\Psi(0) = x$: Let i denote the unique index for which $x \in C_i$. If $i \neq 0$, we construct N independent exponential random variables with respective means equal to $\{(\kappa r_{ij})^{-1} : 1 \leq j \leq N\}$, and the first jump after time $\tau_0 := 0$ is defined as the minimum of these exponential random variables. If $i = 0$, i.e., $\Psi(0) = x \in C_0$, then τ_1 is given by the value of an exponential random variable with mean $1/\kappa$. Letting j denote the index corresponding to the minimizing exponential random variable if $i \neq 0$, or taking $j = 1$ if $x \in C_0$, we define $\Psi(t) = x$ for $0 = \tau_0 \leq t < \tau_1$, and let $\Psi(\tau_1)$ be a sample from the distribution ν_j.

This procedure is continued iteratively to define the sequence of sampling times $\{\tau_k\}$ along with the jump process $\boldsymbol{\Psi}$. To see that $\boldsymbol{\Psi}$ can be viewed as an HMM we first present a simplified expression for the semigroup $\{Q^t\}$.

Proposition 1.3. *Consider the process $\boldsymbol{\Psi}$ with generator $\mathcal{E}$ as in (11) and semigroup $\{Q^t\}$ as in (12). If the initial state $\Psi(0)$ is distributed according to some probability measure $\Psi(0) \sim \mu$ of the form $\mu = \sum_{i=1}^N p_i \nu_i$, where the vector $p = (p_1, p_2, \dots, p_N) \in \mathbb{R}_+^N$ satisfies $\sum p_i = 1$, then the distribution μQ^t of $\Psi(t)$ at time $t > 0$ can be expressed as,*

$$\mu Q^t = \sum p_i(t) \nu_i, \quad t > 0, \quad \textit{where} \quad p(t) = e^{-\kappa(I-r)t} p\,,$$

and $\kappa, r = \{r_{ij}\}$ are the coefficients of the generator $\mathcal{E}$ in (11).

Proof. It suffices to prove the result with $\mu = \nu_i$ for some i; the general case follows by linearity.

The power series representation of Q^t implies that $\nu_i Q^t$ can be expressed as a convex combination of $\{\nu_j\}$ for each t,

$$\nu_i Q^t = \sum_{j=1}^N \varrho_{ij}(t) \nu_j\,, \qquad t \geq 0. \tag{15}$$

An expression for the coefficients $\{\varrho_{ij}(t)\}$ can be obtained from the differential equation,

$$\frac{d}{dt}Q^t = \mathcal{E}Q^t,$$

as follows: Writing $\nu_i\mathcal{E} = \kappa\bigl(-\nu_i + \sum_{j=1}^N r_{ij}\nu_j\bigr)$, we conclude that, for any $t \ge 0$,

$$\frac{d}{dt}\nu_i Q^t = \kappa\Bigl[-\nu_i + \sum_{j=1}^N r_{ij}\nu_j\Bigr]Q^t = \kappa\sum_{k=1}^N\Bigl[-\varrho_{ik}(t)\nu_k + \sum_{j=1}^N r_{ij}\varrho_{jk}(t)\nu_k\Bigr].$$

Therefore, the coefficients $\varrho \in \mathbb{R}^{N^2}$ appearing in (15) satisfy,

$$\frac{d}{dt}\varrho_{ik}(t) = \kappa\Bigl[-\varrho_{ik}(t) + \sum_{j=1}^N r_{ij}\varrho_{jk}(t)\Bigr].$$

Given the initial condition $\varrho_{ij}(0) = I$, the solution to this ODE is given by, $\varrho(t) = e^{-\kappa(I-r)t}$, $t \ge 0$, as required. □

For the HMM construction, let $\boldsymbol{I}$ denote a finite-state, continuous-time Markov process, with values in $\{0, 1, 2, \ldots, N\}$. Its rate matrix is denoted by $q_{ij} := \kappa r_{ij}$ for $i \ne j$, and $q_{ii} := -\sum_{j\ne i} q_{ij}$. We take $r_{01} = 1$ and $r_{0i} = 0$ for all $i \ne 1$.

Written as an $(N+1)\times(N+1)$ matrix, this becomes $q = -\kappa(I - r)$. The process $\boldsymbol{I}$ is the hidden state process; the set $\mathsf{X}_I = \{1, 2, \ldots, N\}$ will be an absorbing set for $\boldsymbol{I}$. Conditional on $\boldsymbol{I}$, we define the observed HMM process, denoted $\boldsymbol{\Upsilon} = \{\Upsilon(t)\}$, as follows. Letting $\{\tau_i\}$ denote the successive jump times of $\boldsymbol{I}$, $\Upsilon(t)$ is constant on the interval $t \in [\tau_i, \tau_{i+1})$, and satisfies for each $A \in \mathcal{B}$ and $i = 0, 1, 2, \ldots$,

$$\mathsf{P}\{\Upsilon(\tau_i) \in A \mid \Upsilon(t), t < \tau_i;\ I(t), t < \tau_i;\ I(\tau_i) = k\} = \mathsf{P}\{\Upsilon(\tau_i) \in A \mid I(\tau_i) = k\} = \nu_k(A).$$

An immediate consequence of the definitions is that $\boldsymbol{\Psi}$ can be expressed as an HMM:

Proposition 1.4. *Suppose that $\Psi(0) \sim \nu_i$ for some $i \ge 1$, and that the HMM is initialized in state i, i.e., $I(0) = i$. Then the jump process $\boldsymbol{\Psi}$ and the HMM $\boldsymbol{\Upsilon}$ are identical in law. More generally, if $\Psi(0) = x \in C_i$ and $I(0) = i$, for some $i = 0, 1, \ldots, N$, then the jump process $\boldsymbol{\Psi}$ and the HMM $\boldsymbol{\Upsilon}$ are identical in law following the first jump,*

$$\{\Psi(t) : t \ge \tau_1\} \stackrel{\text{dist}}{=} \{\Upsilon(t) : t \ge \tau_1\}.$$

1.3 Separability and the spectrum

The key property we will use to establish that a process $\boldsymbol{\Phi}$ can be approximated by an HMM as in Theorem 1.2 will be the "v-separability" of its resolvent R. Following [20] we say that a kernel K is *v-separable* with respect to some function $v : \mathsf{X} \to [1, \infty)$, if $|||K|||_v < \infty$ and, for each $\varepsilon > 0$, there exists a compact set $\mathsf{Y} \subset \mathsf{X}$ and a finite-rank, probabilistic kernel T

supported on Y, such that $\|K - T\|_v \le \varepsilon$. By 'finite-rank' we mean there are functions $\{s_i\}$, measures $\{\nu_j\}$, and nonnegative constants $\{\theta_{ij}\}$ such that,

$$T = \sum_{i,j=1}^{N} \theta_{ij} s_i \otimes \nu_j. \tag{16}$$

A kernel T is 'probabilistic' if $T(x, \mathsf{X}) = 1$ for all $x \in \mathsf{X}$.

Our next result gives an alternative characterization of the Donsker-Varadhan condition (DV3), showing that it is equivalent to v-separability of the resolvent. A similar result in discrete time appears in [37, 20]. The implication (ii) ⇒ (i) is contained in Proposition 3.1. The forward implication (i) ⇒ (ii) follows from Proposition 3.2.

Theorem 1.5. [(DV3) ⇔ v-SEPARABILITY] *For a Markov process* $\mathbf{\Phi}$ *on* X *satisfying conditions* (A1) *and* (A2)*, the following are equivalent:*

(i) DONSKER-VARADHAN ASSUMPTIONS: *Condition* (DV3) *holds in the form given in* (A4)*.*

(ii) v-SEPARABILITY: *The resolvent kernel* R *is* v*-separable, for a continuous function* v *with compact sublevel sets, possibly different from the one in* (i)*.*

The following result follows immediately from Theorem 1.5 and Proposition 3.4, combined with [20, Theorem 3.5]. Recall that the *spectrum* $\mathcal{S}(K) \subset \mathbb{C}$ of a linear operator K on L_∞^v is the set of $z \in \mathbb{C}$ such that the inverse $[Iz - K]^{-1}$ does not exist as a bounded linear operator on L_∞^v.

Theorem 1.6. [(DV3) ⇒ DISCRETE SPECTRUM] *Let* $\mathbf{\Phi}$ *be a Markov process satisfying conditions* (A1) *and* (A2)*. If* $\mathbf{\Phi}$ *also satisfies the drift condition* (DV3) *in the form given in* (A4)*, then the spectrum of the resolvent kernel is discrete in* L_∞^v*.*

1.4 Extensions

Further connections between (DV3), v-separability, multiplicative mean ergodic theorems, and large deviations for continuous-time Markov processes will be considered in subsequent work, generalizing and extending the discrete-time results of [20]. In particular, under (DV3), the process $\mathbf{\Phi}$ is "multiplicatively regular" and satisfies strong versions of the "multiplicative mean ergodic theorem." These results, in turn, can be used to deduce a large deviations principle for the empirical measures induced by $\mathbf{\Phi}$. Moreover, the rate function can be expressed in terms of the entropy rate, as in [5, 4, 20].

The technical arguments used in the proofs of all the central results here can easily be extended beyond the class of continuous-sample-path diffusions in $\mathbb{R}^d$. Although such extensions will not be pursued further in this paper, we note that the assumption (A1) can be replaced by the condition that $\mathbf{\Phi}$ is a nonexplosive Borel right process (so that it satisfies the strong Markov property and has right-continuous sample paths) on a Polish space X. Assumptions (A2) and (A3) can be maintained as stated; the conclusions of Proposition 2.2 continue to hold in this more general setting. Assumption (A4) can also be maintained without modification. The resolvent equations in Proposition 2.4 hold in this general setting, which is what is required in the converse theory that provides the implication (ii) $\Longrightarrow$ (i) in Theorem 1.2.

Finally, there are applications to consider, as well as bridges to other areas such as statistics, machine learning, and operations research [2, 3]. The approximation introduced in this paper is similar to the approximation performed in the modeling technique known as *probabilistic latent semantic analysis* (PLSA); see [16] for the basic concepts, and [13, 31] for surveys that describe connections with techniques from other fields. Given a large $m \times m$ matrix P representing associations between different objects, the goal is to find an approximating matrix T, an $m \times r$ matrix S, and an $r \times m$ matrix N such that $r \ll m$ and,

$$T = SN = \sum_{i=1}^{r} s_i n_i^T,$$

where $\{s_i : 1 \leq i \leq r\}$ denote the columns of S, and $\{n_i^T : 1 \leq i \leq r\}$ denote the rows of N. Hence, the goal is to find a transition matrix of reduced rank, exactly as in this paper. Our work provides motivation and rigorous justification for the use PLSA models, even when the state space is general, and even for Markov models evolving in continuous time, as well as motivation for the development of approximation theory for diffusions based on observed trajectories of the process.

The remainder of the paper is organized as follows. The following section develops results establishing approximations between the process $\boldsymbol{\Phi}$ and a simple jump process. This is a foundation for Section 3 that establishes similar approximations with an HMM.

2 Resolvents and Jump-Process Approximations

We begin in this section with an approximation of the process $\boldsymbol{\Phi}$ by a pure jump-process denoted $\boldsymbol{\Phi}^\kappa$, evolving on the state space X. The fixed constant $\kappa > 0$ denotes the jump rate. The jump times $\{\tau_i : i \geq 0\}$ define a Poisson process: $\tau_0 = 0$, and the increments are i.i.d. with exponential distribution and mean κ^{-1}. At the time of the ith jump we have $\Phi^\kappa(\tau_i) \sim \kappa R_\kappa(x, \cdot)$, given that $\Phi^\kappa(\tau_{i-1}) = x$. This process is Markov, with generator,

$$\mathcal{D}_\kappa := \kappa[-I + \kappa R_\kappa]. \tag{17}$$

This is the generator for the Markov process used in the proof of the Hille-Yosida theorem in [29].

Throughout this section it is assumed that $|||R_\kappa|||_v < \infty$, with v being continuous, with compact sublevel sets. Hence the generator $\mathcal{D}_\kappa$ also has finite norm. This is justified by the following proposition, whose proof may be found in the Appendix. The following drift condition is a relaxation of (DV3),

$$\mathcal{D}v \leq -v + b_v, \tag{18}$$

where b_v is a finite constant, and $v \colon \mathsf{X} \to [1, \infty)$.

Proposition 2.1. *Let* $\boldsymbol{\Phi}$ *be a Markov process satisfying* (A1)*.*

(i) *If* (A4) *holds, then there is a function* $v \colon \mathsf{X} \to [1, \infty)$ *and a finite constant* b_v *satisfying* (18)*.*

(ii) *If* (18) *holds for a function* $v\colon \mathsf{X} \to [1,\infty)$ *and a positive constant* b_v*, then the following bounds hold,*

$$\begin{aligned} \||(\alpha R_\alpha)^n\||_v &\le 1 + b_v, && \text{for all } n \ge 1,\ \alpha > 0; \\ \pi(v) &\le b_v, && \text{for any invariant probability measure } \pi. \end{aligned}$$

where b_v *is the constant in* (18).

□

We next review some background on ψ-irreducible Markov processes.

2.1 Densities, irreducibility and ergodicity

The density condition (A2) combined with the existence of a Lyapunov function as in (DV3) implies ergodicity. Recall that a Markov process $\mathbf{\Phi}$ with a unique invariant probability measure π is called v-uniformly ergodic for some function $v : \mathsf{X} \to \mathbb{R}$, if there are constants $\beta_0 > 0$, $B_0 < \infty$, such that,

$$\||P^t - 1 \otimes \pi\||_v \le e^{B_0 - \beta_0 t}, \qquad t \ge 0.$$

See [24] for basic theory of ψ-irreducible Markov processes, including definitions of small sets and aperiodicity in this general state-space setting.

Proposition 2.2. *If conditions* (A1), (A2) *and* (A3) *hold, then the Markov process* $\mathbf{\Phi}$ *is* ψ*-irreducible and aperiodic with* $\psi(\,\cdot\,) := R(x_0, \,\cdot\,)$*, and all compact sets are small. If, in addition,* (DV3) *holds, then the process is* v*-uniformly ergodic with* $v = e^V$*.*

Proof. Under (A1) and (A2) the Markov process is a T-process, since R has the strong Feller property [24]. This combined with (A3) easily implies ψ-irreducibility with $\psi(\,\cdot\,) = R(x_0, \,\cdot\,)$. Under (A3), for any set A satisfying $\psi(A) > 0$, we have $P^t(x, A) > 0$ for all $t \ge 0$ sufficiently large. The proof is similar to the proof of Proposition 6.1 of [24]. Hence the process is aperiodic. To see that all compact sets are small, we note that all compact sets are petite by [24, Theorem 4.1]. Under aperiodicity, petite sets are small; this is proved as in the discrete-time case [22, Theorem 5.5.7].

To see that $\mathbf{\Phi}$ is v-uniformly ergodic note that, under (DV3), we have,

$$\mathcal{D}v \le -\delta v + b_v^0 \mathbb{I}_C\,,$$

where $b_v^0 = b \sup_{x \in C} v(x)$. This is condition (V4) of [8], and hence the conclusion follows from the main result of [8]. □

Ergodic theory based on drift conditions such as (V4) is based in part on the following *Comparison theorem*; see [22] for the discrete-time counterpart.

Proposition 2.3. *If* $\mathcal{D}h \le -f + g$ *for nonnegative functions* (h, f, g)*, and if* h *is continuous, then*

(i) *For any* $T > 0$*,*

$$\mathsf{E}_x\Big[h(\Phi(T)) + \int_0^T f(\Phi(t))\,dt\Big] \le h(x) + \mathsf{E}_x\Big[\int_0^T g(\Phi(t))\,dt\Big]. \tag{19}$$

(ii) *For any* $\alpha > 0$,

$$\alpha R_\alpha h + R_\alpha f \le h + R_\alpha g.$$

Proof. The proof of (i) is precisely the same as in the proof of the comparison theorem in discrete time [22]. Part (ii) follows from (i) on multiplying each side of (19) by $\alpha e^{-\alpha T}$, and integrating over $T \in \mathbb{R}$. □

2.2 Resolvent equations

Recall the construction of the process $\mathbf{\Phi}^\kappa$ with generator $\mathcal{D}_\kappa$ as in (17). We denote the semigroup of $\mathbf{\Phi}^\kappa$ by $P_\kappa^t := e^{t\mathcal{D}_\kappa}$, $t \ge 0$, and its associated family of resolvent kernels by $R_{\kappa,\alpha}$:

$$R_{\kappa,\alpha} := \int_0^\infty e^{-\alpha t} P_\kappa^t \, dt, \qquad \alpha > 0. \tag{20}$$

Proposition 2.4 states the resolvent equations, and establishes some simple corollaries.

Proposition 2.4. *Suppose the process* $\mathbf{\Phi}$ *satisfies* (A1) *and* $\mathbf{\Phi}^\kappa$ *is the jump process with generator* $\mathcal{D}_\kappa$ *as in* (17)*. Then, for any positive constants* α, β *we have:*

(i) *The resolvent equation holds,*

$$R_\alpha = R_\beta + (\beta - \alpha) R_\beta R_\alpha = R_\beta + (\beta - \alpha) R_\alpha R_\beta. \tag{21}$$

(ii) *For each* $\alpha > 0$ *and any measurable function* $h \colon \mathsf{X} \to \mathbb{R}$ *for which* $R_\alpha |h|$ *is finite-valued, the function* $f = R_\alpha h$ *is in the domain of* $\mathcal{D}$*, and,*

$$\mathcal{D} R_\alpha h = \alpha R_\alpha h - h. \tag{22}$$

Moreover, with $g = \alpha R_\alpha h - h$ *the stochastic process* (7) *is a martingale, so that* (8) *holds.*

(iii) *The resolvent of* $\mathbf{\Phi}^\kappa$ *satisfies the analogous identity,*

$$\mathcal{D}_\kappa R_{\kappa,\alpha} h = \alpha R_{\kappa,\alpha} h - h, \qquad \text{if } R_{\kappa,\alpha}|h| \text{ is finite valued.} \tag{23}$$

(iv) *The generators for* $\mathbf{\Phi}$ *and* $\mathbf{\Phi}^\kappa$ *are related by,*

$$\mathcal{D}_\kappa h = \mathcal{D}[\kappa R_\kappa] h \qquad \text{if } R_\kappa |h| \text{ is finite valued;} \tag{24}$$

Proof. Part (i) is the usual resolvent equation [9]. Part (iii) follow directly from (ii), and (iv) follows from (i) and (ii).

It remains to prove the resolvent equation (22) in the strong form: (8) holds with $f = R_\alpha h$ and $g = \alpha R_\alpha h - h$. We have by Fubini's theorem,

$$P^T f = \int_0^\infty e^{-\alpha t} P^{t+T} h \, dt = e^{\alpha T} \int_T^\infty e^{-\alpha t} P^t h \, dt.$$

Suppose first that h is bounded. It follows from Assumption A2 then $P^t h$ is a continuous function of t. Hence $P^T f$ is C^1 with,

$$\frac{d}{dT} P^T f = \alpha e^{\alpha T} \int_T^\infty e^{-\alpha t} P^t h \, dt - e^{\alpha T} P^T h = P^T g.$$

The identity (8) thus holds, by the fundamental theorem of calculus.

If h is not bounded we can construct a sequence of functions $\{h_n\}$ satisfying $|h_n(x)| \le \min(|h(x)|, n)$ for each n and x, and $h_n(x) \to h(x)$ as $n \to \infty$ for each x. We then have for each n and t, with $f_n = R_\alpha h_n$ and $g_n = \alpha R_\alpha h_n - h_n$,

$$P^t f_n = f_n + \int_0^t P^s g_n \, ds.$$

Under the assumption that $R_\alpha |h|$ is finite-valued, it follows that $P^t|f|$ and $\int_0^t P^s |g| \, ds$ are finite-valued. The desired conclusion (8) thus follows by dominated convergence. □

The resolvent equation (22) implies that $[\alpha I - \mathcal{D}]$ is a left inverse of R_α for any $\alpha > 0$, in the sense that $[\alpha I - \mathcal{D}] R_\alpha f = f$ for an appropriate class of functions f. While R_α *cannot* be expressed as a true operator inverse on the space L_∞^v, it is in fact possible to obtain such a representation for $R_{\kappa,\alpha}$. This is made precise in the following.

Lemma 2.5. *Suppose the process* $\mathbf{\Phi}$ *satisfies* (A1) *and the drift condition* (18)*. Then, for any* $\alpha > 0$,

$$R_{\kappa,\alpha} = [\alpha I - \mathcal{D}_\kappa]^{-1} = [\alpha I - \kappa(\kappa R_\kappa - I)]^{-1} = \frac{\kappa}{(\kappa+\alpha)^2} \sum_{n=-1}^\infty (1 + \alpha \kappa^{-1})^{-n} (\kappa R_\kappa)^{n+1}, \qquad (25)$$

where the sum converges in L_∞^v*. Moreover,*

$$|||\alpha R_{\kappa,\alpha}|||_v \le 1 + b_v. \qquad (26)$$

Proof. For any $n \ge 0$, $\kappa > 0$, we have the bound $|||(\kappa R_\kappa)^{n+1}|||_v \le 1 + b_v$, from Proposition 2.1 (ii). The representation (23) implies that the inverse can be expressed as the power series (25), which is convergent in L_∞^v. Since $\alpha I - \mathcal{D}_\kappa$ is a left inverse of $R_{\kappa,\alpha}$, it then follows that $R_{\kappa,\alpha} = [\alpha I - \mathcal{D}_\kappa]^{-1}$.

To establish the bound (26) we apply the triangle inequality,

$$|||R_{\kappa,\alpha}|||_v \le \frac{\kappa}{(\kappa+\alpha)^2} \sum_{n=-1}^\infty (1 + \alpha\kappa^{-1})^{-n} |||\kappa R_\kappa|||_v^{n+1}.$$

Using once more the bound $|||\kappa R_\kappa|||_v^{n+1} \le 1 + b_v$, and simplifying the expression for the sum in the following bound,

$$|||R_{\kappa,\alpha}|||_v \le (1 + b_v) \frac{\kappa}{(\kappa+\alpha)^2} \left(\frac{1 + \alpha\kappa^{-1}}{1 - (1 + \alpha\kappa^{-1})^{-1}} \right),$$

we obtain the bound in (26), as claimed. □

2.3 Resolvent approximations

Under (DV3) or, more generally, under the weaker drift condition (18), we obtain the following strong approximation for the resolvent kernels:

Proposition 2.6. *Suppose the process* $\boldsymbol{\Phi}$ *satisfies* (A1) *and* $\boldsymbol{\Phi}^\kappa$ *is the jump process with generator* $\mathcal{D}_\kappa$ *defined in* (17)*. If* $\boldsymbol{\Phi}$ *satisfies the drift condition* (18)*, then, for each* $\alpha < \kappa$*:*

$$\|\!|R_{\kappa,\alpha} - R_\alpha\|\!|_v \le \frac{4}{\kappa}(1 + b_v).$$

Proof. We first obtain a power series representation for R_α not in terms of its generator, but in terms of the resolvent kernel R_κ. The resolvent equation (21) with $\beta = \kappa$ and $\alpha > 0$ arbitrary gives $[I - (\kappa - \alpha)R_\kappa]R_\alpha = R_\kappa$. Since $0 < \alpha < \kappa$ and $R_\kappa(x, \mathsf{X}) = \kappa^{-1}$ for each x, it follows that R_α can be expressed as the power series,

$$R_\alpha = [I - (1 - \alpha\kappa^{-1})\kappa R_\kappa]^{-1} R_\kappa = \frac{1}{\kappa}\sum_{n=0}^{\infty}(1 - \alpha\kappa^{-1})^n(\kappa R_\kappa)^{n+1}.$$

Proposition 2.1 (ii) gives the uniform bound, $\|\!|(\kappa R_\kappa)^{n+1}\|\!|_v \le 1 + b_v$, which implies that this sum converges in L_∞^v.

Applying Lemma 2.5, we conclude that the difference of the two resolvent kernels $R_{\kappa,\alpha}$ and R_α can be decomposed into three terms:

$$R_{\kappa,\alpha} - R_\alpha = \Big(\frac{\kappa}{(\kappa+\alpha)^2} - \frac{1}{\kappa}\Big)\sum_{n=0}^{\infty}(1 + \alpha\kappa^{-1})^{-n}(\kappa R_\kappa)^{n+1} \tag{27a}$$

$$+ \frac{1}{\kappa}\sum_{n=0}^{\infty}\Big((1 + \alpha\kappa^{-1})^{-n} - (1 - \alpha\kappa^{-1})^n\Big)(\kappa R_\kappa)^{n+1} \tag{27b}$$

$$+ \Big(\frac{\kappa}{(\kappa+\alpha)^2}(1 + \alpha\kappa^{-1})^{-n}\Big|_{n=-1}\Big) I. \tag{27c}$$

To bound the first term (27a) we apply Proposition 2.1 (ii):

$$\begin{aligned}
\frac{1}{(1+b_v)}\|\!|\text{RHS of (27a)}\|\!|_v &\le \Big|\frac{\kappa}{(\kappa+\alpha)^2} - \frac{1}{\kappa}\Big|\sum_{n=0}^{\infty}(1 + \alpha\kappa^{-1})^{-n} \\
&= \Big(\frac{1}{\kappa} - \frac{\kappa}{(\kappa+\alpha)^2}\Big)\Big(1 - (1 + \alpha\kappa^{-1})^{-1}\Big)^{-1} \\
&= \Big(\frac{2\kappa\alpha + \alpha^2}{\kappa(\kappa+\alpha)^2}\Big)\Big(\frac{\alpha}{\kappa+\alpha}\Big)^{-1} \\
&= \frac{\kappa + (\kappa+\alpha)}{\kappa(\kappa+\alpha)} = \frac{1}{\kappa+\alpha} + \frac{1}{\kappa}\,.
\end{aligned}$$

This implies the bound,

$$\|\!|\text{RHS of (27a)}\|\!|_v \le \frac{2}{\kappa}(1 + b_v).$$

The next inequality also uses the bound $\|(\kappa R_\kappa)^n\|_v \le 1 + b_v$:

$$\begin{aligned}\|\text{RHS of (27b)}\|_v &\le (1+b_v)\frac{1}{\kappa}\sum_{n=0}^{\infty}\Big((1+\alpha\kappa^{-1})^{-n} - (1-\alpha\kappa^{-1})^n\Big)\\ &= (1+b_v)\frac{1}{\kappa}\Big([1-(1+\alpha\kappa^{-1})^{-1}]^{-1} - [1-(1-\alpha\kappa^{-1})]^{-1}\Big)\\ &= \frac{1}{\kappa}(1+b_v).\end{aligned}$$

The final term (27c) is elementary:

$$\|\text{RHS of (27c)}\|_v = \frac{\kappa}{(\kappa+\alpha)^2}(1+\alpha\kappa^{-1}) = \frac{1}{\kappa+\alpha}.$$

Substituting these three bounds completes the proof. □

3 Separability

In this section we develop consequences of the separability assumption. In particular, we describe the construction of an approximating semigroup $\{Q^t\}$ with generator of the form given in (11), as described in Theorem 1.2. This is accomplished in four steps:

(i) First we note that under (DV3) the resolvent kernel R of $\boldsymbol{\Phi}$ can be truncated to a compact set.

(ii) Then we argue that, again on a compact set, R can be approximated by a finite-rank kernel T.

(iii) We next prove that the generator, $\mathcal{D}_\kappa := \kappa[-I + \kappa R_\kappa]$, of the jump process $\boldsymbol{\Phi}^\kappa$ constructed in Section 2, can be approximated by a generator $\mathcal{E}$ of the form (11),

$$\mathcal{E} = \kappa\Big[-I + \mathbb{I}_{C_0}\otimes\nu_1 + \sum_{i,j=1}^{N} r_{ij}\,\mathbb{I}_{C_i}\otimes\nu_j\Big],$$

as long as $\kappa > 0$ is chosen sufficiently large. This key result is described in Proposition 3.2.

(iv) Finally we show that the transition semigroup $\{P^t\}$ of the original process $\boldsymbol{\Phi}$ can be approximated by the semigroup $\{P^t_\kappa\}$ of the jump process $\boldsymbol{\Phi}^\kappa$ (Proposition 3.7), and that the semigroup $\{P^t_\kappa\}$ can in turn be approximated by the semigroup $\{Q^t\}$ corresponding to an HMM with a generator $\mathcal{E}$ as above (Proposition 3.8).

Again, the starting point of these results is justified by applying (DV3) to obtain the truncation described in (i). A converse is obtained in the following result. The proof is based on the resolvent equations, and is found in the Appendix.

Proposition 3.1. *Suppose that the Markov process $\boldsymbol{\Phi}$ satisfies conditions* (A1) *and* (A2)*, and that its resolvent kernel R is v-separable for some continuous function $v\colon \mathsf{X}\to[1,\infty)$ with compact sublevel sets. Then* (A4) *holds for some continuous V_-, W_- on X, and the function V_- is in L_∞^V.* □

3.1 Truncations and finite approximations

Let $\mathbf{\Phi}$ be a Markov process satisfying condition (A1), with generator $\mathcal{D}$ and associated resolvent kernels $\{R_\alpha\}$. Recall the definition of the corresponding jump process $\mathbf{\Phi}^\kappa$ in the beginning of Section 2, with generator $\mathcal{D}_\kappa$ and associated resolvents $\{R_{\kappa,\alpha}\}$.

Our result here shows that condition (DV3) implies that the generator $\mathcal{D}_\kappa$ of the jump process $\mathbf{\Phi}^\kappa$ can be approximated by a generator $\mathcal{E}$ as in (11). This result is a corollary of Proposition C.4, whose proof is given in the Appendix.

Proposition 3.2. *Suppose the Markov process $\mathbf{\Phi}$ satisfies conditions* (A1), (A2). *If* (DV3) *holds as in assumption* (A4), *then, for each $\kappa > 0$ and any $\epsilon > 0$, there exists a generator $\mathcal{E}$ of the form given in* (11), *such that all the r_{ij} are strictly positive, and the generator $\mathcal{D}_\kappa$ of the jump process $\mathbf{\Phi}^\kappa$ can be approximated in operator norm as,*

$$|\!|\!|\mathcal{D}_\kappa - \mathcal{E}|\!|\!|_v \le \varepsilon, \tag{28}$$

with $v := e^V$. □

From Proposition 3.2 we have a generator $\mathcal{E}$ of the form (11), and with $Q^t := e^{t\mathcal{E}}$, $t > 0$, being the associated transition semigroup, the corresponding resolvent kernels $\{T_\alpha\}$ are defined, as usual, in (13). Using the approximation of the generator $\mathcal{E}$ in (28), we next show that the kernels $\{T_\alpha\}$ can be expressed as operator inverses, in a way analogous to the representations obtained in Lemma 2.5 for the resolvents $\{R_{\kappa,\alpha}\}$.

Lemma 3.3. *Suppose that the assumptions of Proposition 3.2 hold, and choose $\kappa > 0$ and $\epsilon_0 > 0$ such that $|\!|\!|\mathcal{D}_\kappa - \mathcal{E}|\!|\!|_v \le \epsilon_0$. Then the resolvent obtained from the semigroup $\{Q^t\}$ can be expressed as an inverse operator on L_∞^v: For all $\alpha > (1 + b_v)\epsilon_0$,*

$$T_\alpha = [\alpha I - \mathcal{E}]^{-1},$$

where b_v is as in Proposition 2.1 (i). Moreover, for all such α we have the norm bound,

$$|\!|\!|T_\alpha|\!|\!|_v \le \frac{1 + b_v}{\alpha - (1 + b_v)\varepsilon_0}. \tag{29}$$

Proof. Note that we already have from the resolvent equation the formula $[\alpha I - \mathcal{E}]T_\alpha = I$ on L_∞^v. It remains to show that $[\alpha I - \mathcal{E}]$ admits an inverse. We can write, on some domain,

$$[\alpha I - \mathcal{E}]^{-1} = [\alpha I - \mathcal{D}_\kappa + \mathcal{D}_\kappa - \mathcal{E}]^{-1} = R_{\kappa,\alpha}[I + (\mathcal{D}_\kappa - \mathcal{E})R_{\kappa,\alpha}]^{-1}.$$

The right-hand-side admits a power series representation whenever $|\!|\!|(\mathcal{D}_\kappa - \mathcal{E})R_{\kappa,\alpha}|\!|\!|_v < 1$. In fact, under the assumptions of the Lemma, using the bound in Lemma 2.5 we have,

$$|\!|\!|(\mathcal{D}_\kappa - \mathcal{E})R_{\kappa,\alpha}|\!|\!|_v \le |\!|\!|\mathcal{D}_\kappa - \mathcal{E}|\!|\!|_v \cdot |\!|\!|R_{\kappa,\alpha}|\!|\!|_v \le \epsilon_0(1 + b_v)/\alpha < 1,$$

and the resulting bound is precisely (29). □

Our next result shows that v-separability implies that each of the resolvent kernels R_α can be approximated by the kernels $\{T_\alpha\}$ obtained from a finite-rank semigroup. Specifically, R_α will be approximated by a resolvent T_α of the form (13), where the transition semigroup $\{Q^t\}$ is that of a Markov process with generator $\mathcal{E}$ as in (11).

Proposition 3.4. *Under the assumptions of Proposition 3.2, for each $\epsilon > 0$ and $\delta \in (0,1)$, there exists a generator $\mathcal{E}$ of the form given in (11), such that the corresponding resolvent kernels $\{T_\alpha\}$ defined in (13) satisfy the following uniform bound:*

$$\|R_\alpha - T_\alpha\|_v \le \epsilon, \qquad \delta \le \alpha \le \delta^{-1}.$$

Proof. To establish the uniform bound in operator norm, first we approximate R_α by $R_{\kappa,\alpha}$. Under (DV3), Proposition 2.1 (i) implies that we can use Proposition 2.6 as follows: We fix $\kappa \ge \delta^{-1}$ such that the right-hand-side of this bound is no greater than $\frac{1}{2}\varepsilon$, giving,

$$\|R_{\kappa,\alpha} - R_\alpha\|_v \le \tfrac{1}{2}\varepsilon, \qquad \alpha \le \delta^{-1}. \tag{30}$$

We now invoke Proposition 3.2: Fix an operator $\mathcal{E}$ of the form (11) satisfying,

$$\|\mathcal{D}_\kappa - \mathcal{E}\|_v \le \varepsilon_0,$$

where $\varepsilon_0 \in (0,\varepsilon)$ is to be determined. Lemma 2.5 and Lemma 3.3 give,

$$T_\alpha = [\alpha I - \mathcal{E}]^{-1}, \quad R_{\kappa,\alpha} = [\alpha I - \mathcal{D}_\kappa]^{-1}.$$

Hence the difference can be expressed,

$$\begin{aligned} T_\alpha - R_{\kappa,\alpha} &= T_\alpha[\mathcal{E} - \mathcal{D}_\kappa]R_{\kappa,\alpha} \\ &= [T_\alpha - R_{\kappa,\alpha}][\mathcal{E} - \mathcal{D}_\kappa]R_{\kappa,\alpha} + R_{\kappa,\alpha}[\mathcal{E} - \mathcal{D}_\kappa]R_{\kappa,\alpha}, \end{aligned}$$

and applying the triangle inequality together with the sub-multiplicativity of the operator norm,

$$\|T_\alpha - R_{\kappa,\alpha}\|_v \le \|T_\alpha - R_{\kappa,\alpha}\|_v \|\mathcal{E} - \mathcal{D}_\kappa\|_v \|R_{\kappa,\alpha}\|_v + \|R_{\kappa,\alpha}\|_v^2 \|\mathcal{E} - \mathcal{D}_\kappa\|_v.$$

Lemma 2.5 gives the bound $\|\alpha R_{\kappa,\alpha}\|_v \le (1 + b_v)$, and hence for $\alpha \in [\delta, \delta^{-1}]$,

$$\|\mathcal{E} - \mathcal{D}_\kappa\|_v \|R_{\kappa,\alpha}\|_v \le \varepsilon_0(1 + b_v)/\delta.$$

Assuming that $\varepsilon_0 > 0$ is chosen so that the right-hand-side is less than one, we can substitute into the previous bound and rearrange terms to obtain,

$$\|T_\alpha - R_{\kappa,\alpha}\|_v \le \frac{\|R_{\kappa,\alpha}\|_v^2}{1 - \|\mathcal{E} - \mathcal{D}_\kappa\|_v \|R_{\kappa,\alpha}\|_v} \|\mathcal{E} - \mathcal{D}_\kappa\|_v \le \Big(\frac{(1 + b_v)^2}{1 - \varepsilon_0(1 + b_v)/\delta}\Big)\Big(\frac{\varepsilon_0}{\delta^2}\Big).$$

Choosing $\varepsilon_0 = \frac{1}{4}(1 + b_v)^{-2}\varepsilon\delta^2$ then gives,

$$\|T_\alpha - R_{\kappa,\alpha}\|_v \le \tfrac{1}{4}\varepsilon \frac{1}{(1 - \frac{1}{4}\varepsilon)} \le \tfrac{1}{2}\varepsilon, \qquad \alpha \in [\delta, \delta^{-1}].$$

This combined with (30) and the triangle inequality completes the proof. □

3.2 Ergodicity

To establish solidarity over an infinite time horizon we impose the reachability condition (A3) throughout the remainder of this section. Recall the construction of the approximating HMM process $\mathbf{\Psi}$ in Section 1.2, and the definition of v-uniform ergodicity from Section 2.1.

Proposition 3.5. *Suppose the process $\mathbf{\Phi}$ satisfies conditions* (A1) – (A4)*, so that, in particular, $\mathbf{\Phi}$ is v-uniformly ergodic with $v = e^V$ by Proposition 2.2. Then:*

(i) *For each $\kappa > 1$, the jump process $\mathbf{\Phi}^\kappa$ is v-uniformly ergodic, with $v = e^V$.*

(ii) *The HMM process $\mathbf{\Psi}$ is v-uniformly ergodic, with $v = e^V$.*

Before proceeding with the proof we prove Lyapunov bounds that are useful in later results.

Lemma 3.6. *Under the assumptions of Proposition 3.5, there exist $\delta_\circ > 0$ and $b_\circ < \infty$ such that the following bound holds for each $\kappa > 1$:*

$$\mathcal{D}_\kappa v \le -\delta_\circ v + b_\circ. \tag{31}$$

Consequently, the following bound holds for the semigroup,

$$|\!|\!| P_\kappa^t |\!|\!|_v \le 1 + b_\circ/\delta_\circ, \qquad t \ge 0. \tag{32}$$

Proof. The bound (32) follows from (31) using a version of the comparison theorem (see eqn. (31) of [8]):

$$P_\kappa^t v \le e^{-\delta_\circ t} v + b_\circ/\delta_\circ.$$

The proof of (31) begins with the bound $\mathcal{D}v \le [-\delta + \mathbb{I}_C]v$, which holds under (DV3) because $W \ge 1$ everywhere. Letting $b_v^0 = b \max_C v$ then gives,

$$\mathcal{D}v \le -\delta v + b_v^0 \mathbb{I}_C\,.$$

Applying Proposition 2.3 (ii) with $h = v$, $f = \delta v$ and $g \equiv b_v^0 \mathbb{I}_C$ implies that,

$$\kappa R_\kappa v + \delta R_\kappa v \le v + R_\kappa g \le v + \kappa^{-1} b_v^0.$$

On rearranging terms this gives,

$$\kappa R_\kappa v \le (1 + \delta\kappa^{-1})^{-1}(v + \kappa^{-1} b_v^0),$$

and thence,

$$\kappa R_\kappa v - v \le -\frac{\delta}{\delta + \kappa} v + \frac{1}{\delta + \kappa} b_v^0.$$

From the definition of the generator for the jump process we conclude that the desired bound holds,

$$\mathcal{D}_\kappa v = \kappa[\kappa R_\kappa v - v] \le -\frac{\delta\kappa}{\delta + \kappa} v + \frac{\kappa}{\delta + \kappa} b_v^0.$$

This gives (31) on choosing the worst-case over $\kappa \ge 1$:

$$\delta_\circ = \delta/(\delta + 1), \quad b_\circ = b_v^0.$$

□

Proof of Proposition 3.5. To establish (i) we first demonstrate that $\mathbf{\Phi}^\kappa$ is irreducible and aperiodic. If ψ is a maximal irreducibility measure for $\mathbf{\Phi}$, then Lemma 2.5 implies that $\psi \prec R_\kappa(x,\,\cdot\,)$ for each x. This implies that the chain with transition kernel κR_κ is ψ-irreducible and aperiodic. Irreducibility and aperiodicity for $\mathbf{\Phi}^\kappa$ is then obvious since it is a jump process with Poisson jumps, and jump distribution κR_κ.

To complete the proof of (i) we establish condition (V4) of [8]. From Lemma 3.6 we obtain,

$$\mathcal{D}_\kappa v \le -\tfrac{1}{2}\delta_\circ v + b_\circ \mathbb{I}_{C_\circ},$$

where $C_\circ = \{x : \frac{1}{2}\delta_\circ v(x) \le b_\circ\}$. The sublevel set $C_\circ$ is compact, and Proposition 2.2 implies that compact sets are small, so this implies that the jump process is v-uniformly ergodic.

Analogous arguments for $\mathbf{\Psi}$ will establish (ii): ψ-irreducibility and aperiodicity are immediate by Proposition 1.4 and the fact that all r_{ij} in the definition of $\mathcal{E}$ are strictly positive, from Proposition 3.2. To show that $\mathbf{\Psi}$ is v-uniformly ergodic simply note that, by the definition of $\mathcal{E}$,

$$\mathcal{E}v = -\kappa v + \kappa T v \le -\kappa v + \kappa b \mathbb{I}_{\mathsf{Y}},$$

where $b := \sup_{b\in\mathsf{Y}} v(x)$. Again, this is a version of condition (V4) of [8], and the conclusion follows from [8]. □

3.3 Semigroup approximations

We begin with an approximation bound between the semigroups corresponding to $\mathbf{\Phi}$ and $\mathbf{\Phi}^\kappa$.

Proposition 3.7. *Suppose that $\mathbf{\Phi}$ satisfies conditions* (A1) – (A4)*. Then there exists $b_\bullet$ depending only on $\mathbf{\Phi}$ such that,*

$$\|P^t g - P^t_\kappa g\|_v \le b_\bullet \kappa^{-1} \|\mathcal{D}^2 g\|_v, \qquad t \ge 0,\ \kappa \ge 1,$$

for any C^4 function g with compact support.

Proof. Under the assumption of the proposition, the local-martingale assumption can be strengthened to the martingale property (8). That is, for any $T > 0$,

$$\begin{aligned} \mathsf{E}_x\big[g(\Phi(T))\big] &= g(x) + \mathsf{E}_x\Big[\int_0^T \mathcal{D}g(\Phi(t))\,dt\Big] \\ \mathsf{E}_x\big[\mathcal{D}g(\Phi(T))\big] &= \mathcal{D}g(x) + \mathsf{E}_x\Big[\int_0^T \mathcal{D}^2 g(\Phi(t))\,dt\Big], \qquad x \in \mathsf{X}. \end{aligned}$$

It follows that $P^T g$ is differentiable in T, and the same is true for $P^T_\kappa g$.

Denote the difference $\varepsilon_g(t) = P^t g - P^t_\kappa g$. We have for any t,

$$\begin{aligned} \frac{d}{dt}\varepsilon_g(t) &= P^t \mathcal{D} g - P^t_\kappa \mathcal{D}_\kappa g \\ &= \mathcal{D}_\kappa[P^t g - P^t_\kappa g] + P^t[\mathcal{D} - \mathcal{D}_\kappa] g \\ &= \mathcal{D}_\kappa[P^t g - P^t_\kappa g] - \kappa^{-1} P^t[\mathcal{D}\mathcal{D}_\kappa g], \end{aligned}$$

where in the second equation we have used here the fact that the operators P^t, P^t_κ, and $\mathcal{D}_\kappa$ all commute. The final equation follows from (24) and the definition of $\mathcal{D}_\kappa$ in (17). Writing $h = \mathcal{D}\mathcal{D}_\kappa g$, this can be solved to give,

$$\varepsilon_g(t) = \varepsilon_g(0) - \kappa^{-1}\int_0^t e^{s\mathcal{D}_\kappa} P^{t-s} h\, ds.$$

Substituting $P^s_\kappa = e^{s\mathcal{D}_\kappa}$ and $\varepsilon_g(0) = 0$ simplifies this expression:

$$\varepsilon_g(t) = -\kappa^{-1}\int_0^t P^s_\kappa P^{t-s} h\, ds. \tag{33}$$

We have $\pi(h) = \pi(P^{t-s}h) = 0$, so that by Proposition 2.2 we have for some $B_0 < \infty$ and $\beta_0 > 0$,

$$\|P^{t-s}h\|_v \le e^{B_0 - \beta_0(t-s)}\|h\|_v, \qquad 0 \le s \le t$$

Consequently, for each $x \in \mathsf{X}$ and $t \ge 0$,

$$\Big|\int_0^t P^s_\kappa P^{t-s} h\,(x)\, ds\Big| \le \|h\|_v e^{B_0}\int_0^t e^{-\beta_0(t-s)} P^s_\kappa v\,(x)\, ds$$

Recalling the bound (32) on $|\!|\!|P^s_\kappa|\!|\!|_v$ and substituting into (33) gives $\|\varepsilon_g(t)\|_v \le \kappa^{-1}\beta_0^{-1}\|h\|_v e^{B_0}$. We have $h = \mathcal{D}\mathcal{D}_\kappa g$, and hence the generator relationship (24) and the generator bound in Proposition 2.1 (ii) give,

$$\|h\|_v \le |\!|\!|\kappa R_\kappa|\!|\!|_v \|\mathcal{D}^2 g\|_v \le (1 + b_v)\|\mathcal{D}^2 g\|_v.$$

Finally, substituting this into the previous bound on $\|\varepsilon_g(t)\|_v$ completes the proof. □

Similar arguments provide approximation bounds for the semigroups corresponding to $\mathbf{\Phi}^\kappa$ and $\mathbf{\Psi}$, where the latter is denoted $\{Q^t\}$ and defined in (12).

Proposition 3.8. *Suppose that* $\mathbf{\Phi}$ *satisfies conditions* (A1)–(A4). *Then there exists* $b_\bullet$ *depending only on* $\mathbf{\Phi}$ *such that for* $g \in L^v_\infty$,

$$\|P^t_\kappa g - Q^t g\|_v \le b_\bullet \varepsilon \|g\|_v.$$

Proof. The proof is similar to the proof of Proposition 3.7: We fix $g \in L^v_\infty$, and denote the error by,

$$\varepsilon_g(t) = P^t_\kappa g - Q^t g, \qquad t \ge 0.$$

The right hand side is differentiable by construction of the two semi-groups, with

$$\frac{d}{dt}\varepsilon_g(t) = \mathcal{D}_\kappa \varepsilon_g(t) + [\mathcal{D}_\kappa - \mathcal{E}]Q^t g$$

This can be solved to give,

$$\varepsilon_g(t) = \varepsilon_g(0) + \int_0^t P^s_\kappa [\mathcal{D}_\kappa - \mathcal{E}] Q^{t-s} g\, ds.$$

We have $\varepsilon_g(0) = 0$. Moreover, $[\mathcal{D}_\kappa - \mathcal{E}]1 \equiv 0$, which implies that $[\mathcal{D}_\kappa - \mathcal{E}]g = [\mathcal{D}_\kappa - \mathcal{E}][g - \varpi(g)]$. Here, ϖ denotes the unique invariant measure of the process $\boldsymbol{\Psi}$, guaranteed to exist by Proposition 3.5. Hence,

$$\|\varepsilon_g(t)\|_v \leq \|\!|\mathcal{D}_\kappa - \mathcal{E}|\!\|_v \int_0^t \|\!|P_\kappa^s|\!\|_v \|(Q^{t-s} - 1 \otimes \varpi)g\|_v \, ds.$$

Substituting the bound $\|\!|P_\kappa^s|\!\|_v \leq 1 + b_\circ/\delta_\circ$ from Lemma 3.6 gives,

$$\|\varepsilon_g(t)\|_v \leq \|\!|\mathcal{D}_\kappa - \mathcal{E}|\!\|_v (1 + b_\circ/\delta_\circ)\|g\|_v \int_0^t \|\!|Q^{t-s} - 1 \otimes \varpi|\!\|_v \, ds\,,$$

and Proposition 3.2 gives $\|\!|\mathcal{D}_\kappa - \mathcal{E}|\!\|_v \leq \varepsilon$. This establishes the result with

$$b_\bullet = (1 + b_\circ/\delta_\circ) \int_0^\infty \|\!|Q^r - 1 \otimes \varpi|\!\|_v \, dr\,,$$

which is finite, by Proposition 3.5. □

The following bound is an immediate consequence of the last Proposition.

Corollary 3.9. *Under the assumptions of Proposition 3.8, for each $\varepsilon > 0$ we can construct the approximating process $\boldsymbol{\Psi}$ described in Section 1.2 so that the the Markov processes $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ have unique invariant probability measures π and ϖ, respectively, satisfying,*

$$\|\pi - \varpi\|_v \leq \varepsilon.$$

□

Appendix

A Appendix: Proof of Proposition 2.1

The drift condition (DV3) can be expressed as follows, in terms of the function $v = e^V$:

$$\mathcal{D}v \leq (-\delta W + b\mathbb{I}_C)v.$$

By assumption, we have $\delta W(x) \geq \delta$ everywhere. Moreover, $\delta W(x) \geq 1$ on the complement of the sublevel set $C_W(\delta^{-1})$ (see (10)). This set is compact under (A4), so that the desired bound holds with,

$$b_v := b\Bigl(\sup_{x\in C} v(x)\Bigr) + (1-\min(\delta,1))\Bigl(\sup_{x\in C_F(r)} v(x)\Bigr) < \infty\,.$$

This establishes part (i).

Under (18) we can apply Proposition 2.3 (ii) with $h = v$, $f = v$ and $g \equiv b_v$ to obtain $\alpha R_\alpha v + R_\alpha v \leq v + \alpha^{-1} b_v$, or

$$\alpha R_\alpha v \leq (1+\alpha^{-1})^{-1}\bigl(v + \alpha^{-1}b_v\bigr). \tag{34}$$

Iterating this bound we obtain, for any $n \geq 1$,

$$\begin{aligned}(\alpha R_\alpha)^n v &\leq (1+\alpha^{-1})^{-n} v + \alpha^{-1} b_v \sum_{k=1}^{n} (1+\alpha^{-1})^{-k}\\ &\leq v + \alpha^{-1} b_v [1-(1+\alpha^{-1})]^{-1} = v + b_v.\end{aligned}$$

Hence $(\alpha R_\alpha)^n v \leq (1+b_v)v$, which is the first bound.

The second follows from (34) and the (discrete-time) comparison theorem of [22], which gives,

$$\pi(v) < \infty \quad \textit{and} \quad \pi(v) \leq (1+\alpha^{-1})^{-1}\bigl(\pi(v) + \alpha^{-1} b_v\bigr).$$

Rearranging terms gives $(1+\alpha^{-1})\pi(v) \leq \pi(v) + \alpha^{-1}b_v$, or $\pi(v) \leq b_v$ as claimed. ☐

B Appendix: Proof of Proposition 3.1

The following strengthening of the strong Feller property is required:

Lemma B.1. *Suppose that the kernel T has the strong Feller property and is also v-separable for some continuous function $v\colon \mathsf{X} \to [1,\infty)$. Then Tf is a continuous function for each $f \in L_\infty^v$.*

Proof. Let $\{\chi_n : n \geq 0\}$ denote a sequence of non-decreasing continuous functions satisfying $0 \leq \chi_n(x) \leq 1$ for each x. Assume that each function has compact support, and for each x,

$$\lim_{n\to\infty} \chi_n(x) = 1$$

Letting $\mathbb{I}_{\chi_n}$ denote the associated multiplication operator, denote $T_n = \mathbb{I}_{\chi_n} T \mathbb{I}_{\chi_n}$. Under v-separability it follows that $\lim_{n\to\infty} \| T - T_n \|_v = 0$.

Under the strong Feller property for T it follows that $T_n f$ is continuous for each n, and that $T_n f \to T f$ uniformly on compact sets as $n \to \infty$, for any $f \in L_\infty^v$. Continuity of Tf follows. ☐

Under the separability assumption we can find, for each $n \geq 1$, a compact set Y_n and a kernel T_n supported on Y_n satisfying $\| R - T_n \|_v \leq 2^{-n}$. We assume without loss of generality that $\mathsf{Y}_n \subseteq \mathsf{Y}_{n+1}$ for each n, and that $\bigcup_n \mathsf{Y}_n = \mathsf{X}$.

Writing $v_n = v \mathbb{I}_{\mathsf{Y}_n^c}$ we have $\| v_n \|_v = 1$ and $T_n v_n \equiv 0$. Consequently, for each $n \geq 1$,

$$Rv_n = (R - T_n)v_n + T_n v_n = (R - T_n)v_n \leq \| R - T_n \|_v v \leq 2^{-n} v, \qquad n \geq 1. \tag{35}$$

The desired solution to (DV3) is constructed as follows. First define the sequence of finite-valued functions on X,

$$u_-^m := v + \sum_{n=1}^m v_n = \Big(1 + \sum_{n=1}^m \mathbb{I}_{\mathsf{Y}_n^c}\Big) v, \qquad v_-^m = R u_-^m, \quad m \geq 1,$$

and denote $u_- = \lim_{m\to\infty} u_-^m$, $v_- = \lim_{m\to\infty} v_-^m$. Applying (35), we conclude that $v_- \in L_\infty^v$, with the explicit bound,

$$\| v_- \|_v \leq \| R \|_v + \sum_{n=1}^\infty \| R v_n \|_v \leq \| R \|_v + 1.$$

Each of the functions v_-^m is continuous since R has the strong Feller property, and the assumptions of Lemma B.1 hold with $R = T$. These functions converge to v_- uniformly on compact subsets of X, showing that v_- is continuous. We let $V_- = \log(v_-)$, which is also continuous.

It follows from Proposition 2.4 that the resolvent equation holds, $\mathcal{D} v_- = v_- - u_-$, and consequently, recalling the nonlinear generator (9),

$$\mathcal{H}(V_-) = (v_-)^{-1} \mathcal{D} v_- = 1 - u_-/v_-.$$

By construction, the function u_-/v_- has compact sublevel sets. Writing $W = \max(u_-/v_- - 1, 1)$ and $C = \{x : W(x) \leq 1\}$ then gives,

$$\mathcal{H}(V_-) \leq -W + 2\mathbb{I}_C,$$

which is a version of (DV3). The function W is not continuous. However, it has compact sublevel sets, so there exists a continuous function $W_- \colon \mathsf{X} \to [1, \infty)$ with compact sublevel sets, satisfying $W_- \leq W$ everywhere. The pair (V_-, W_-) is the desired solution to (DV3). ☐

C Proof of Proposition 3.2

Before giving the proof, we state and prove some preliminary results. The assumptions of Proposition 3.2 remain in effect throughout this subsection.

On setting $h = v = e^V$, $f = \delta W h$, and $g = b\mathbb{I}_C h$ in Proposition 2.3 we obtain the following bound:

Lemma C.1. *Under* (DV3)*, with* $v = e^V$*, we have,*

$$RI_{W_1} v \le v + b s_0,$$

where $W_1 = 1 + \delta W$*,* $s_0 = RI_C v$ *and for any function* F*,* I_F *denotes the multiplication kernel* $I_F(x, dy) = F(x)\delta_x(dy)$*.*

For each $r \ge 1$, we define the compact sets,

$$C_r = C_v(r) \cap C_W(r),$$

in the notation of equation (10). From the assumption that V and W are continuous with compact sublevel sets, we obtain,

$$\lim_{r\to\infty} \inf_{x \in C_r^c} V(x) = \lim_{r\to\infty} \inf_{x\in C_r^c} W(x) = \infty. \tag{36}$$

The above bounds on the resolvent will allow us to approximate R by a kernel supported on C_r for suitably large $r \ge 1$. To that end, we choose and fix a continuous function $W_0 \colon \mathsf{X} \to [1, \infty)$ in L_∞^W, satisfying $\|W_0^2\|_W = 1$, and whose growth at infinity is strictly slower than $W^{\frac{1}{2}}$ in the sense that,

$$\lim_{r\to\infty} \|W_0^2 \mathbb{I}_{C_W(r)^c}\|_W = 0\,. \tag{37}$$

This can be equivalently expressed,

$$\lim_{r\to\infty} \sup_{x\in\mathsf{X}} \Big[\frac{W_0(x)}{\sqrt{W(x)}} \mathbb{I}_{\{W(x)>r\}} \Big] = 0\,.$$

The weighting function is simultaneously increased to,

$$v_0 = W_0 v.$$

The following Lemma justifies truncating R to a compact set.

Lemma C.2. *Under* (DV3) *the resolvent kernel* R *satisfies* $|\!|\!| RI_W |\!|\!|_v < \infty$*, and,*

$$\lim_{r\to\infty} |\!|\!| I_{W_0}(R - I_{C_r} R I_{C_r}) I_{W_0} |\!|\!|_{v_0} = 0. \tag{38}$$

Proof. Lemma C.1 implies that $|\!|\!| RI_W |\!|\!|_v$ is finite as claimed: We have the explicit bound $|\!|\!| RI_W |\!|\!|_v \le \delta^{-1}(1 + b\|s_0\|_v)$. The limit (38) is also based on the same lemma. Starting with the identity $R - I_{C_r} R I_{C_r} = I_{C_r} R I_{C_r^c} + I_{C_r^c} R$ we obtain,

$$\begin{aligned} I_{W_0}(R - I_{C_r} R I_{C_r}) I_{W_0} v_0 &= I_{W_0}(R - I_{C_r} R I_{C_r}) I_{W_0^2} v \\ &= I_{W_0}[I_{C_r} R I_{C_r^c} I_{W_0^2}] v + I_{W_0}[I_{C_r^c} R I_{W_0}] v_0. \end{aligned} \tag{39}$$

These two terms can be bounded separately. For the first term on the right-hand-side consider the following,

$$[I_{C_r} R I_{C_r^c} I_{W_0^2}] v \le |\!|\!| I_{C_r} R I_{C_r^c} I_W |\!|\!|_v \varepsilon_r v \le |\!|\!| RI_W |\!|\!|_v \varepsilon_r v,$$

where $\varepsilon_r := \sup_{x\in C_r^c} W_0(x)W^{-\frac{1}{2}}(x)$. Multiplying both sides by W_0 then gives,

$$[I_{W_0} I_{C_r} R I_{C_r^c} I_{W_0}]v_0 \le \|RI_W\|_v \varepsilon_r v_0, \qquad r \ge 1,$$

which means that $\|I_{W_0} I_{C_r} R I_{C_r^c} I_{W_0}\|_{v_0} \le \|RI_W\|_v \varepsilon_r$ for each r.

Bounds on the second term in (39) are obtained similarly through a second truncation. Write, for any $n \ge 1$,

$$[I_{C_r^c} R I_{W_0^2}]v = [I_{C_r^c} R I_{W_0^2} \mathbb{I}_{C_n}]v + [I_{C_r^c} R I_{W_0^2} \mathbb{I}_{C_n^c}]v.$$

Arguing as above we have $\|I_{C_r^c} R I_{W_0^2} \mathbb{I}_{C_n^c}\|_v \le \|RI_W\|_v \varepsilon_n$. Moreover, $W_0^2 v \le Wv \le n^2$ on C_n, which gives,

$$[I_{C_r^c} R I_{W_0^2}]v \le n^2 I_{C_r^c} + \|RI_W\|_v \varepsilon_n v.$$

Multiplying both sides of this equation by W_0 gives,

$$[I_{W_0} I_{C_r^c} R I_{W_0}]v_0 \le n^2 I_{C_r^c} W_0 + \|RI_W\|_v \varepsilon_n v_0,$$

so that

$$\|I_{W_0} I_{C_r^c} R I_{W_0}\|_{v_0} \le n^2 \|I_{C_r^c} W_0\|_{v_0} + \|RI_W\|_v \varepsilon_n. \tag{40}$$

And also,

$$\|I_{C_r^c} W_0\|_{v_0} = \sup_{x\in C_r^c} \frac{W_0(x)}{v_0(x)} = \sup_{x\in C_r^c} \frac{1}{v(x)} \le \frac{1}{r}.$$

This combined with (36) implies that (40) can be made arbitrarily small by choosing large n and then large r. □

Lemma C.3. *Under* (A1) *and* (A2)*, for each $r \ge 1$ and $\varepsilon > 0$, there exists $t_0 > 0$ and $t_1 < \infty$ in the definition* (3) *such that,*

$$\|I_{W_0} I_{C_r} (R - \overline{R}) I_{C_r} I_{W_0}\|_{v_0} \le \varepsilon,$$

where $\overline{R} = \overline{R}_1$.

Proof. Since W_0 and v_0 are bounded on C_r, we can apply the bound,

$$\|I_{W_0} I_{C_r} (R - \overline{R}) I_{C_r} I_{W_0}\|_{v_0} \le \big(\sup_{x\in C_r} W_0(x)^2 v_0(x)\big) \|I_{C_r} (R - \overline{R}) I_{C_r}\|_1$$

Hence it is sufficient to prove the result with $W_0 = v_0 = 1$.

We have by definition of $\overline{R}$,

$$\|I_{C_r} (R - \overline{R}) I_{C_r}\|_1 = \sup_{x\in C_r} \int_{t\in[t_0,t_1]^c} e^{-t} P^t(x, C_r)\, dt.$$

The right hand side is bounded by $t_0 + e^{-t_1}$, which can be made arbitrarily small by choice of $t_0 > 0$ and $t_1 < \infty$. □

Proposition 3.2 will be seen as a corollary to the following more general bound:

Proposition C.4. *For any $\varepsilon > 0$ there exists a finite-rank kernel T satisfying:*

$$\|I_{W_0}[R - T]I_{W_0}\|_{v_0} \le \varepsilon.$$

The kernel can be taken of the form,

$$T = \sum_{ij} r_{ij} \mathbb{I}_{C_i} \otimes \nu_j \tag{41}$$

where $\{C_i : 1 \le i \le N\}$ is a finite collection of disjoint, open, precompact sets, $\{r_{ij}\}$ are non-negative constants, and $\{\nu_i\}$ are probability measures on $(\mathsf{X}, \mathcal{B})$ with each ν_i supported on C_i.

Proof. Lemma C.2 and Lemma C.3 imply that for any $\varepsilon > 0$ we can find $r_0 \ge 1$ such that,

$$\|I_{W_0}(R - I_{C_{r_0}} \overline{R} I_{C_{r_0}})I_{W_0}\|_{v_0} \le \varepsilon/2.$$

With this value of r_0 fixed, note that (A2) implies that for any $\varepsilon_0 > 0$ we can construct a kernel $T(x, dy) = t(x, y)dy$ of the form given in (16) such that $|t(x, y) - \overline{\xi}(x, y)| \le \varepsilon_0$ for $(x, y) \in C_{r_0} \times C_{r_0}$ (see definition of $\overline{\xi}$ above Proposition 1.1). In particular, the functions $\{s_i\}$ and the densities of the ν_i can be taken as indicator functions, so that this is simply the approximation of the continuous function $\overline{\xi}(\,\cdot\,,\,\cdot\,)$ by simple functions. Consequently,

$$\begin{aligned}
\|I_{W_0}[R - T]I_{W_0}\|_{v_0} &\le \varepsilon/2 + \|I_{W_0} I_{C_{r_0}}[R - T] I_{C_{r_0}} I_{W_0}\|_{v_0} \\
&\le \varepsilon/2 + \varepsilon_0 \sup_{x \in C_{r_0}} W_0(x) \sup_{x \in C_{r_0}} \bigl(v(x) W_0^2(x)\bigr)\, \mu^{\text{Leb}}(C_{r_0}) \\
&\le \varepsilon/2 + \varepsilon_0 r_0^4\, \mu^{\text{Leb}}(C_{r_0}),
\end{aligned}$$

where $\mu^{\text{Leb}}(C_{r_0})$ denotes the Lebesgue measure of the bounded set C_{r_0}. The right-hand-side is bounded by ε on choosing $\varepsilon_0 = [r_0^4\, \mu^{\text{Leb}}(C_{r_0})]^{-1}(\varepsilon/2)$. □

Proof of Proposition 3.2. Since Proposition C.4 was proved for an arbitrary function W_0 satisfying (37), we can take W_0 equal to a constant, say $w \ge 1$. First consider the case $\kappa = 1$. There, applying Proposition C.4 with $\epsilon/2$ instead of ϵ, we obtain a finite-rank kernel of the form (41). Letting $\mathcal{E}_0 = \kappa[-I + T]$,

$$\|\mathcal{D}_\kappa - \mathcal{E}_0\|_v = \|\kappa[\kappa R_\kappa - T]\|_v = \|R - T\|_v = \frac{1}{w}\|I_{W_0}[R - T]I_{W_0}\|_{v_0} \le \epsilon/2.$$

Now we define $\mathcal{E} = \mathcal{E}_0 + \mathbb{I}_{C_0} \otimes \nu_1$, with $C_0 = \mathsf{X} \setminus \cup_{1 \le i \le N} C_i$ and ν_1 a probability measure supported on C_1. We have $\mathcal{E}1 \equiv 0$ as required, and the following bound holds:

$$\|\mathcal{D}_\kappa - \mathcal{E}\|_v \le \|\mathcal{D}_\kappa - \mathcal{E}_0\|_v + \nu_1(v)\Bigl(\sup_{x \in C_0} \frac{1}{v(x)}\Bigr).$$

Recall that $C_0^c = \bigcup_{i \ge 1} C_i$. If the $\{C_i : i \ge 1\}$ are constructed so that $C_0^c \subset C_v(r)$, then the right hand side is bounded by $\nu_1(v) r^{-1}$. For $r > 0$ sufficiently large, this is less than ε, as required.

For a fixed, general κ we consider the scaled process $\{Z(t) := \Phi(t/\kappa) : t \ge 0\}$ and note it satisfies exactly the same assumptions as $\{\Phi(t)\}$. Also, κR_κ, is the resolvent kernel for $\{Z(t)\}$ (corresponding to the parameter $\alpha = 1$) so that, as before by Proposition C.4, we obtain the required bound. □

References

[1] S. Balaji and S.P. Meyn. Multiplicative ergodicity and large deviations for an irreducible Markov chain. *Stochastic Process. Appl.*, 90(1):123–144, 2000.

[2] A. Bušić, I. Vliegen, and A. Scheller-Wolf. Comparing Markov chains: Aggregation and precedence relations applied to sets of states, with applications to assemble-to-order systems. 37(2):259–287, 2012.

[3] K. Deng, P. Mehta, and S. Meyn. Optimal Kullback-Leibler aggregation via spectral theory of Markov chains. 56(12):2793 –2808, Dec. 2011.

[4] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Springer-Verlag, New York, second edition, 1998.

[5] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. I. II. *Comm. Pure Appl. Math.*, 28:1–47; ibid. 28:279–301, 1975.

[6] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.*, 29(4):389–461, 1976.

[7] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.*, 36(2):183–212, 1983.

[8] D. Down, S.P. Meyn, and R.L. Tweedie. Exponential and uniform ergodicity of Markov processes. *Ann. Probab.*, 23(4):1671–1691, 1995.

[9] S.N. Ethier and T.G. Kurtz. *Markov Processes : Characterization and Convergence.* John Wiley & Sons, New York, 1986.

[10] J. Feng. Martingale problems for large deviations of Markov processes. *Stochastic Process. Appl.*, 81:165–212, 1999.

[11] J. Feng and T.G. Kurtz. *Large deviations for stochastic processes*, volume 131 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2006.

[12] W.H. Fleming. Exit probabilities and optimal stochastic control. *App. Math. Optim.*, 4:329–346, 1978.

[13] E. Gaussier and C. Goutte. Relation between PLSA and NMF and implications. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 601–602, New York, NY, USA, 2005. ACM.

[14] F.Z. Gong and L.M. Wu. Spectral gap of positive operators and applications. *J. Math. Pures Appl.*, 85:151–191, 2006.

[15] A. Guillin, C. Léonard, L. Wu, and N. Yao. Transportation-information inequalities for Markov processes. 144(3):669–695, July 2009.

[16] T. Hofmann. Unsupervised learning by Probabilistic Latent Semantic Analysis. *Mach. Learn.*, 42(1-2):177–196, 2001.

[17] N.V. Kartashov. Criteria for uniform ergodicity and strong stability of Markov chains with a common phase space. *Theor. Probability Appl.*, 30:71–89, 1985.

[18] N.V. Kartashov. Inequalities in theorems of ergodicity and stability for Markov chains with a common phase space. *Theor. Probability Appl.*, 30:247–259, 1985.

[19] I. Kontoyiannis and S.P. Meyn. Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.*, 13:304–362, February 2003.

[20] I. Kontoyiannis and S.P. Meyn. Large deviation asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab.*, 10(3):61–123, 2005.

[21] C. Lobry. Contrôlabilité des systèmes non linéaires. *SIAM J. Control*, 8:573–605, 1970.

[22] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, London, 2nd edition, 2009. Published in the Cambridge Mathematical Library. 1993 edition online: `http://black.csl.uiuc.edu/~meyn/pages/book.html`.

[23] S.P. Meyn and R.L. Tweedie. Generalized resolvents and Harris recurrence of Markov processes. *Contemporary Mathematics*, 149:227–250, 1993.

[24] S.P. Meyn and R.L. Tweedie. Stability of Markovian processes II: Continuous time processes and sampled chains. *Ann. Appl. Probab.*, 25:487–517, 1993.

[25] S.P. Meyn and R.L. Tweedie. Stability of Markovian processes III: Foster-Lyapunov criteria for continuous time processes. *Ann. Appl. Probab.*, 25:518–548, 1993.

[26] E. Nummelin. *General Irreducible Markov Chains and Nonnegative Operators*. Cambridge University Press, Cambridge, 1984.

[27] S. Polidoro, C. Cinti, and S. Menozzi. Two-sided bounds for degenerate processes with densities supported in subsets of $\mathbb{R}^n$. *arXiv preprint arXiv:1203.4918*, 2012.

[28] L. Rey-Bellet and L. E. Thomas. Fluctuations of the entropy production in anharmonic chains. *Ann. Inst. Henri Poincaré*, 3(3):483–502, 2002.

[29] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 1*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Foundations, Reprint of the second (1994) edition.

[30] L.C.G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge University Press, Cambridge, 2000.

[31] M. Shashanka, B. Raj, and P. Smaragdis. Probabilistic Latent Variable Models as nonnegative factorizations. *Computational Intelligence and Neuroscience*, pages 1–8, 2008.

[32] D.W. Stroock and S.R. Varadhan. On the support of diffusion processes with applications to the strong maximum principle. In *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, pages 333–368. University of California Press, 1972.

[33] H.J. Sussmann and V. Jurdjevic. Controllability of nonlinear systems. *J. Differential Equations*, 12:95–116, 1972.

[34] A.F. Veinott Jr. Discrete dynamic programming with sensitive discount optimality criteria. *Ann. Math. Statist.*, 40(5):1635–1660, 1969.

[35] L.M. Wu. Large deviations for Markov processes under superboundedness. *C. R. Acad. Sci Paris Série I*, 324:777–782, 1995.

[36] L.M. Wu. Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Process. Appl.*, 91(2):205–238, 2001.

[37] L. Wu. Essential spectral radius for Markov semigroups. I. Discrete time case. *Prob. Theory Related Fields*, 128(2):255–321, 2004.