Construction of Lyapunov Functions for Hybrid Piecewise-Deterministic Markov Processes

Alexandre R. Mesquita
Center for Control, Dynamical Systems and Computation
University of California
Santa Barbara, CA 93106-9560 USA
mesquita@umail.ucsb.edu

João P. Hespanha
Center for Control, Dynamical Systems and Computation
University of California
Santa Barbara, CA 93106-9560 USA
hespanha@ece.ucsb.edu

ABSTRACT

We present a method for the construction of Foster-Lyapunov functions to prove ergodicity of hybrid piecewise-deterministic Markov processes. Our method searches for the Lyapunov function that maximizes a measure of the rate of convergence that appears in the theory of large deviations. We provide conditions under which such a Lyapunov function is the limit of an iterative procedure. Analytical solutions are often possible as exemplified with a problem for which a Lyapunov function was not known before: the problem of controlling the probability density of a swarm of robotic agents.

1. INTRODUCTION

Markov processes with jumps, such as hybrid piecewise-deterministic Markov processes, offer a significant challenge to the construction of stability proofs due to the intricate nonlocal interactions in the state space that are introduced by jumps. Foster-Lyapunov stability criteria for such processes involve solving partial integro-differential inequalities, what is typically difficult to do numerically. This paper shows that Lyapunov functions for such processes can be obtained from a convex optimization problem arises in the theory of large deviations [7, 5]. In particular, we show that the Lyapunov function that maximizes a specific measure of convergence often used in large deviations theory satisfies a nonlinear integral equation without differential terms in the unknown. In addition, we also show that this equation can be solved using an iterative procedure that converges exponentially fast to its solution. It turns out that, for some problems, the iterative method to find the Lyapunov equation can be bypassed altogether since one can find closed-form solutions to the integral equation.

An additional feature of our constructive technique is that it provides a parametrization of Lyapunov functions so that one can find different Lyapunov functions by prescribing different weights to points in the space.

HSCC 2010 Stockholm, Sweden

We illustrate the use of our method by finding a Lyapunov function for a process for which a Lyapunov function was not previously known. This process consists of a hybrid Markov Chain Monte Carlo (MCMC) approach in which a vehicle is induced to perform a random walk with some prespecified stationary distribution. This process is inspired by bacterial chemotaxis and we refer to it as Optimotaxis. Stability was proven in a previous work [14] without the use of Lyapunov techniques.

The key advantage in using the Lyapunov-based techniques proposed here to prove stability of a Markov process is that one obtains information about the rate of convergence of the process to the steady-state. In particular, the method used in this paper to construct Lyapunov functions provide conditions under which the law of the process in Optimotaxis converges exponentially fast to the steady-state distribution. Information about the speed of convergence is important to (1) design processes with fast convergence rates and (2) estimate how long one needs to wait to be sufficiently close to the steady-state distribution.

The results in this paper have two limitations. First, the optimization problem that we solve to find the Lyapunov function, may have a solution that is not a Lyapunov functions. This will certainly be true if we apply this procedure to a process that does not converge to the invariant distribution. However, conceivably even for a process for which the law of the process converges exponentially to the steady-state distribution, it could happen that the solution to our optimization is not a Lyapunov function. In practice, this means that one needs to verify if the function obtained by the procedure proposed in this paper is indeed a Lyapunov function, which generally is a trivial procedure.

A second limitation of the results in this paper is that our current proof that the iterative procedure proposed to find the Lyapunov function required the assumption that the jump kernel of the process be irreducible and have finite rank. While irreducibility is not very limiting, the finite rank requirement limits the way in which the reset maps can depend on the continuous part of the state, can be a problem for more general stochastic hybrid systems. In our Optimotaxis example, the above assumptions on the jump kernel are not fully satisfied, yet we are able to find a Lyapunov function using our method.

In Section 2, we define concepts and provide some key results in the theory of Markov processes and piecewise-deterministic Markov processes. In Section 3 we establish some fundamental connections between Lyapunov functions and the concept of rate functions for general Markov processes. These results are specialized to piecewise-deterministic Markov processes in Section 4. In Section 5, we apply our method to the problem of controlling the probability density of a swarm of robotic agents.

2. PRELIMINARIES

In the next sections we first derive our results for a general Markov process and then we discuss how these results specialize to piecewise-deterministic processes. We consider a continuous-time Markov process $\{\mathbf{\Phi}(t)\}_{t\in\mathbb{R}^+}$ taking values in a space $\mathcal{Y}$ equipped with a Borel σ-algebra $\mathcal{B}$. Here $\mathcal{Y}$ is assumed to be a locally compact separable metric space. In this paper we will consistently use boldface symbols to denote random variables. The process is described using the *transition semigroup* $\{P^t\}_{t\in\mathbb{R}^+}$, where each P^t is an operator from $B(\mathcal{Y})$ to $B(\mathcal{Y})$, which is the space of bounded and measurable real functions on $(\mathcal{Y}, \mathcal{B})$ equipped with the norm of the supremum. In particular, for $h \in B(\mathcal{Y})$,

$$P^t h(y) := \mathrm{E}_y\{h(\mathbf{\Phi}(t))\} \ ,$$

where E_y denotes expected value given the initial condition $\mathbf{\Phi}(0) = y$. When (2) is applied to the indicator function $1_A(\cdot)$ of the set $A \in \mathcal{B}$, we use the notation $P^t(y, A) := P^t 1_A(y)$. We say that a probability measure π is an *invariant probability measure* for P^t if, for every $A \in \mathcal{B}$,

$$\pi(A) = \int_{\mathcal{Y}} P^t(y, A) d\pi(y) \ \ \forall t \geq 0 \ .$$

When P^t admits an invariant probability measure, we say that $\mathbf{\Phi}$ is *positive*.

We recapitulate some Foster-Lyapunov criteria for the stability of Markov processes from [6, 17]. For a nontrivial σ-finite measure ψ, we say that $\mathbf{\Phi}$ is ψ*-irreducible* if, for every $A \in \mathcal{B}$, $\psi(A) > 0$ implies $\mathrm{E}\{\eta_A \mid \mathbf{\Phi}(0) = y\} > 0$, $\forall y \in \mathcal{Y}$, where η_A is the occupancy time, defined by $\eta_A := \int_0^\infty 1_A\{\mathbf{\Phi}(t)\}\ dt$. We shall assume that ψ is a maximal irreducibility measure, i.e, any other irreducibility measure is absolutely continuous with respect to ψ.

A nonempty set C is called *petite* if there exists a nontrivial measure ν on $\mathcal{B}$ and a sampling distribution a on $(0, \infty)$ satisfying

$$\int P^t(y, \cdot)\ a(dt) \geq \nu(\cdot), \ y \in C \ .$$

We say that a ψ-irreducible process $\mathbf{\Phi}$ is *aperiodic* if, for some petite set C such that $\psi(C) > 0$, there exists a T such that $P^t(y, C) > 0$, for all $t \geq T$ and $y \in C$.

2.1 Exponential Ergodicity

Next, we recall some results involving the exponential ergodicity of Markov process. To this purpose, we the define *extended generator* $\mathcal{L}$ of $\mathbf{\Phi}$ as in [6]. The domain $D(\mathcal{L})$ is defined as the set of functions $h : \mathcal{Y} \to \mathbb{R}$ such that there exists a measurable function $h : \mathcal{Y} \to \mathbb{R}$ satisfying

$$\mathrm{E}_y\{h(\mathbf{\Phi}(t))\} = h(y) + \mathrm{E}_y\left\{\int_0^t g(\mathbf{\Phi}(s))\ ds\right\}$$

and

$$\int_0^t \mathrm{E}_y\{|g(\mathbf{\Phi}(s))|\}\ ds < \infty \ .$$

The value of the generator is then denoted $\mathcal{L}h := g$. This is an extension of the *infinitesimal generator* $\hat{\mathcal{L}}$, which is defined as $\hat{\mathcal{L}}h := \lim_{t\to 0} t^{-1}(P^t h - h)$ for $h \in B(\mathcal{Y})$.

Consider a candidate Lyapunov function $V : \mathcal{Y} \to [1, \infty]$. For a measure μ on $\mathcal{B}$, we define the norm

$$\|\mu\|_V = \sup_{|h|\leq V} \left|\int h\ d\mu\right| \ .$$

We say that $\mathbf{\Phi}(t)$ is V*-exponentially ergodic* if there exists $V : \mathcal{Y} \to [1, \infty]$, finite for at least one $y \in \mathcal{Y}$, an invariant probability measure π, and constants $B_0, b_0 > 0$ such that

$$\|P^t(y, \cdot) - \pi\|_V \leq B_0 V(x) e^{-b_0 t}, \ \forall y \in \mathcal{Y} \ .$$

If $\mathbf{\Phi}$ is V-exponentially ergodic with $V(y)$ finite for all $y \in \mathcal{Y}$, then we say that $\mathbf{\Phi}$ is *exponentially ergodic.*

We define *the continuous drift condition*:

(CD). For constants $c > 0$, $b < \infty$, a petite set C, the function $V : \mathcal{Y} \to [1, \infty]$ verifies

$$\mathcal{L}V \leq -cV + b1_C \ .$$

Recall that compact sets are petite for a wide class of Markov processes, e.g., irreducible T-processes[16].

Theorem 1 ([6]) *Suppose $\mathbf{\Phi}$ is ψ-irreducible and aperiodic. Then, the condition (CD) satisfied for some function V is equivalent to V-exponential ergodicity.*

Because the Lyapunov function V above is possibly unbounded and we want to be able to compare these functions, we need to define another metric space other than $B(\mathcal{Y})$. The approach adopted in [18] consists of using a weighted supremum norm. For $V : \mathcal{Y} \to [1, \infty]$ we define the space L_∞^V as the space of measurable functions bounded on the norm

$$\|h\|_V = \sup_{y\in\mathcal{Y}} \frac{|h(y)|}{V(y)} \ .$$

A second approach to we adopt in this paper is to consider the space $L(\pi_0)$ of measurable functions on $\mathcal{B}$ with norm $\|h\|_{\pi_0} := \int |h|\ d\pi_0$, where pi_0 is some probability measure to be specified later.

2.2 Nonlinear Generators

A useful object in the theory of large deviations is Fleming's *nonlinear generator* [7] that we define as:

$$\mathcal{H}(G) := e^{-G}\mathcal{L}e^{G}$$

with domain $D(\mathcal{H})$. For simplicity, we assume that $D(\mathcal{L})$ is an algebra and that $h \in D(\mathcal{L})$ implies $e^h \in D(\mathcal{L})$. Thus, we identify $D(\mathcal{H}) = D(\mathcal{L})$. See [7] for a thorough characterization of the domain $D(\mathcal{H})$. Based on [12], we define a related semigroup $\{\mathcal{H}^t\}$ as

$$\mathcal{H}^t(G) = \log e^{-G} P^t e^G \ .$$

for G measurable. From our definition, we have that $\mathcal{H}(G) = \lim_{t\to 0} t^{-1}\mathcal{H}^t(G)$ whenever the limit exists. It turns out that the above condition for exponential ergodicity can be written equivalently in terms of $\mathcal{H}^t$ [12]:

(CD1). The Markov process $\mathbf{\Phi}$ is ψ-irreducible, aperiodic and there exist functions $V : \mathcal{Y} \to (0, \infty]$ and $W : \mathcal{Y} \to [1, \infty]$, a petite set C and constants $\delta,\ b < \infty$ such that

$$\mathcal{H}^t V \le -\delta W + b 1_C$$

for some (and then all) $t > 0$ and all $y \in \mathcal{Y}$ such that $V(y) < \infty$.

We say that condition $(CD1+)$ holds if (CD) holds and if, for all initial conditions, $\mathbf{\Phi}$ admits a continuous probability density for some time $t_0 > 0$ (this a relaxed version of condition (DV3+) in [12]).

When (CD1) holds with an unbounded function W, we take a fixed function $W_0 : \mathcal{Y} \to [1, \infty)$ whose growth at infinity is strictly slower than W, i.e.,

$$\lim_{r\to\infty} \sup_{y\in\mathcal{Y}} \left[\frac{W_0(y)}{W(y)} 1_{\{W(y)>r\}} \right] = 0 \ .$$

We denote by $\mathcal{M}_1^{W_0}$ the set of measures with finite norm $\|\cdot\|_{W_0}$.

2.3 Piecewise-Deterministic Markov Processes

Our definition of Piecewise-Deterministic Markov Processes (PDP) follows closely the framework introduced in [4] and extended in [11]. In a PDP, state trajectories are right continuous with only finitely many discontinuities (*jumps*) on a finite time interval. The continuous evolution of the process is described by a deterministic flow whereas the jumps occur at randomly distributed times and have stochastic lengths.

We consider state variables $x \in \mathcal{X} \subset \mathbb{R}^d$ and $v \in \mathcal{V}$ so that $\mathcal{Y} = \mathcal{X} \times \mathcal{V}$, where $\mathcal{V}$ is a compact set. During flows, the *continuous state* $\mathbf{x}(t)$ evolves according to the vector field $f(x, v)$, whereas the *discrete state* $\mathbf{v}(t)$ remains constant and only changes with jumps. For a fixed $v \in \mathcal{V}$, we denote by $\varphi_t(x, v)$ the continuous flow at time t defined by the vector field $f(\cdot, v)$ and starting at x at time 0. The conditional probability that a jump occurs between the time instants t and s, $0 < s < t$, given $\mathbf{x}(s)$ and $\mathbf{v}(s)$, is

$$1 - \exp\left(- \int_s^t \lambda(\varphi_{\tau-s}(\mathbf{x}(s), \mathbf{v}(s)), \mathbf{v}(s)) d\tau \right) \ , \qquad (1)$$

where the nonnegative function $\lambda(x, v)$ is called the *jump rate* at $(x, v) \in \mathcal{X} \times \mathcal{V}$. At each jump, the overall state $\boldsymbol{\xi} := (\mathbf{x}, \mathbf{v})$ assumes a new value distributed according to the *jump kernel* Q. Namely, if a jump occurs at time t_k, then

$$\Pr(\boldsymbol{\xi}(t_k) \in A \mid \boldsymbol{\xi}^-(t_k)) = \xi) = Q(\xi, A)$$

for $A \in \mathcal{B}$, where the superscript minus indicates the left limit of a processes.

This PDP model is captured by several stochastic hybrid system models that appeared in the literature, including our stochastic hybrid models discussed in [9], or the hybrid models initially proposed in [10] by Hu, Lygeros and co-workers and further expanded in a series of subsequent papers [3]. Fig. 1 depicts a schematic representation of our PDP.

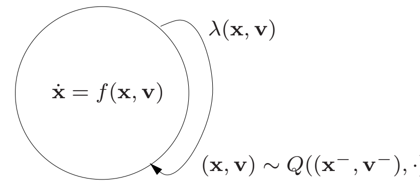

Figure 1: Hybrid automaton for the PDP

With some abuse of notation, we use the symbol Q to denote the operator given by $Qh(y) = \int h(\xi) Q(y, d\xi)$ for $h \in B(\mathcal{Y})$. Under a continuity property on λ and Q [11], the generator $\mathcal{L}$ for the PDP has form

$$\mathcal{L}h = f \cdot \nabla h + \lambda Q h - \lambda h$$

for $h \in D(\mathcal{L})$ continuously differentiable (see [11] for the expression of the full generator), where ∇ denotes the gradient with respect to x.

We define $p(x, v, t)$ as the joint probability density of the state (x, v) at time t. Here it is important to explicit the structure of the parameter space $\mathcal{V}$. We consider $\mathcal{V}$ to be a compact subset of a locally compact separable metric space equipped with a Borel measure ν such that $\nu(\mathcal{V}) = 1$. Note that, as opposed to [4], we do not require V to be countable. This more general setting for PDP's is supported by the theory developed in [11]. Denoting by m the Lebesgue measure in $\mathcal{X}$, we have that $\int_{\mathcal{X}\times\mathcal{V}} p(x, v, t)\ dm \times d\nu = 1,\ \forall t \ge 0$. We denote by $L^1(m \times \nu)$ the space of real functions integrable with respect to $m \times \nu$. As in [2], we assume the existence of a kernel Q^* such that

$$m(dx)Q(x, dy) = m(dy)Q^*(y, dx) \ .$$

In this case, the adjoint $\mathcal{L}^*$ of $\mathcal{L}$ restricted to $L^1(m \times \nu)$ is given by

$$\mathcal{L}^* p = \nabla \cdot fp + Q^*(\lambda p) - \lambda p$$

for $p \in D(\mathcal{L}^*) \subset L^1(m \times \nu)$.

3. LYAPUNOV FUNCTIONS AND RATE FUNCTIONS

In this section we discuss how one can obtain Lyapunov functions by solving relatively simple convex optimization problems. Such Lyapunov functions are related with a notion of convergence rate that appears in the theory of large deviations of Markov processes [7]. For a probability measure μ on $\mathcal{B}$, we define the *rate function*

$$I(\mu) := \sup \left\{ \int -\frac{\mathcal{L}u}{u}\ d\mu : u \in D(\mathcal{L}), u \ge 1 \right\} \ . \qquad (2)$$

Intuitively, we can think of $-\frac{\mathcal{L}u}{u}$ evaluated at $y \in \mathcal{Y}$ as the rate of convergence of the function u at the point y and, therefore, $I(\mu)$ would correspond to the fastest μ-weighted average rate of convergence achievable for some function u in the domain of the generator. Rate functions have a fundamental role in the study of the probability of rare events in the context of large deviations theory. Our objective is to construct a Lyapunov function u by solving the maximization problem posed in (2).

Throughout this section we assume that (CD1+) holds with an unbounded function W. In view of [7], we can rewrite the rate function in terms of an optimization problem that will be shown to be convex:

$$I(\mu) = \sup\left\{\langle \mu, -\mathcal{H}(G)\rangle : G \in D(\mathcal{H}) \cap L_\infty^{W_0}\right\} \tag{3}$$

where $\langle \cdot,\cdot\rangle$ denotes integration. We can likewise define the rate functions $I^t(\mu)$

$$I^t(\mu) = \sup\left\{\langle \mu, -\mathcal{H}^t(G)\rangle : G \in L_\infty^{W_0}\right\} \ . \tag{4}$$

Two simple facts from [12, Prop. 4.9] highlight the relation between the choice of the measure μ and the rate function. The first fact is that the rate function is 0 if μ is an invariant measure for $\mathbf{\Phi}$. In this case the supremum is attained by any constant function. The second fact is that the rate function is infinity for a choice of μ that is not absolutely continuous with respect to the invariant measure π.

The next result provides a good indication that an optimizer G will be a Lyapunov function satisfying (CD1). Once the optimizer is found, all one needs to do is to check if $\bar{C}$ is indeed petite, which is a typically a simple verification.

Proposition 1 *Suppose that (CD1+) holds with an unbounded function W so that $\mathbf{\Phi}$ is exponentially ergodic. In addition, suppose that $\mu \in \mathcal{M}_1^{W_0}$ is not an invariant measure. Then, a function G that attains the supremum in (4) for a fixed $t > 0$ satisfies*

$$\mathcal{H}^t(G) \le -\bar{\delta}\bar{W} + \bar{b}1_{\bar{C}}$$

for constants $\bar{\delta}$, $\bar{b} > 0$, a function $\bar{W} \in L_\infty^{W_0}$ with $W \ge 1$ and a set $\bar{C}$ such that $\mu(\bar{C}) < 1$.

Proof. This theorem follows from Proposition 4.9 in [12]. The proposition states that there exists F bounded such that $\langle \mu, \mathcal{H}^t(F)\rangle < 0$ when μ is not an invariant measure. This implies that $I^t(\mu) > 0$. Therefore, if G attains the supremum in (4) we must have $\langle \mu, \mathcal{H}^t(G)\rangle < 0$. The proposition then follows, given that $\langle \mu, \bar{W}\rangle < 1$. □

3.1 Characterization of the Optimizers

To find the function u that attains the supremum in (3), we need to introduce the notion of twisted semigroups. For a strictly positive function $g \in B(\mathcal{Y})$, we define the *twisted semigroup* $\{P_g^t\}$ as

$$P_g^t h = \frac{P^t(gh)}{P^t g} \ .$$

The corresponding generator is given by

$$\mathcal{L}_g h = \frac{\mathcal{L}(gh) - h\mathcal{L}g}{g}$$

with domain $D(\mathcal{L}_g) = D(\mathcal{L})$ under our assumption that $D(\mathcal{L})$ is an algebra, which is typically the case in applications. Since P_g^t is a positive semigroup with $P_g^t 1 = 1$, it is a semigroup of Markov operators. It can be shown that it defines the same Markov process $\mathbf{\Phi}$ as P^t under an exponential change of measure [20].

The following theorem provides a necessary and sufficient condition for a function to attain the supremum in (3) are given in terms of the twisted semigroup and generator.

Theorem 2 *i. The optimization problem in (3) is a convex optimization problem.*

ii. If $G \in D(\mathcal{H})$ attains the supremum in (3) and F is such that $\mathcal{L}F = 0$, then $G + F$ also attains the supremum in (3).

iii. $G \in L_\infty^{W_0}$ attains the supremum in the definition of $I^t(\mu)$ for all $t > 0$ if and only if μ is an invariant measure under the twisted operator P_g^t, where $g = e^G$, i.e.,

$$\langle \mu, P_g^t h\rangle = \langle \mu, h\rangle$$

for all $h \in B(\mathcal{Y})$.

iv. $G \in D(\mathcal{H}) \cap L_\infty^{W_0}$ attains the supremum in the definition of (3) if and only if

$$\langle \mu, \mathcal{L}_g h\rangle = 0 \tag{5}$$

for all $h \in D(\mathcal{L})$.

Let $\mathcal{L}^*$ be the adjoint of $\mathcal{L}$ acting on $\mathcal{M}(\mathcal{Y})$, the space of finite signed measures on $\mathcal{B}$ with the total variation norm. We can expand (5) in terms of $\mathcal{L}^*$ to obtain a sufficient condition for G to be a maximizer of $I(\mu)$:

$$g\mathcal{L}^*\left(\frac{\mu}{g}\right) = \mu\frac{\mathcal{L}g}{g} \tag{6}$$

The above condition is also necessary when $D(\mathcal{L})$ separates $\mathcal{M}(\mathcal{Y})$. Conditions for this are given in [4].

We now specialize this result for piecewise deterministic Markov processes and therefore take $\mathcal{Y} = \mathcal{X} \times \mathcal{V}$ as in Section 2.3 and assume that the generator $\mathcal{L}$ and its adjoint $\mathcal{L}^*$, acting on $L^1(m \times \nu)$, have the form

$$\mathcal{L}h = f \cdot \nabla h + Kh - \lambda h$$

for $h \in D(\mathcal{L})$; and

$$\mathcal{L}^* p = \nabla \cdot fp + K^* p - \lambda p$$

for $p \in D(\mathcal{L}^*) \subset L^1(m \times \nu)$. Here f is a continuous differentiable function that characterizes the process continuous drift, $\lambda \in B(\mathcal{Y})$ is the jump intensity, and K is a bounded operator that, for the time being, may include both jump and diffusion terms. The following corollary is obtained by expanding (6) for $\mathcal{L}$ of the above form.

Corollary 1 *A sufficient condition for a differentiable function u to attain the supremum in (3) for $d\mu = p\ dm$, $p \in D(\mathcal{L}^*)$, is*

$$uK^*\left(\frac{p}{u}\right) - \nabla \cdot fp - \frac{p}{u}Ku = 0 \ . \tag{7}$$

From Theorem 2, we know that if U is such that $u = e^U$ is a solutions to (7), then $U + F$ is also a solution for any harmonic function F ($\mathcal{L}F = 0$). This is always true for F =constant. One can also verify that $U + F$ is also a solution if $KF = F$.

3.2 Proof of Theorem 2

The two parts of Theorem 2 follows from the next proposition.

Proposition 2 *$\mathcal{H}$ and $\mathcal{H}^t$ are pointwise convex:*

$$\mathcal{H}(\theta F + (1-\theta)G) \le \theta\mathcal{H}(F) + (1-\theta)\mathcal{H}(G)$$

for $F, G \in D(\mathcal{H})$ and $\theta \in (0,1)$. Moreover, equality holds if and only if $F - G$ is a harmonic function for P^t, i.e., $\mathcal{L}(F-G) = 0$.

Proof. Boundedness was proven in [12, Prop. 4.4]. By Hölder's inequality:

$$P^t(e^{\theta F+(1-\theta)G}) \le (P^t e^F)^\theta (P^t e^G)^{1-\theta} \ .$$

Convexity of $\mathcal{H}^t$ then follows by taking the logarithm and rearranging the terms. Since $\lim_{t\to 0} t^{-1}\mathcal{H}^t(F)$ exists for $F \in D(\mathcal{H})$, the convexity of $\mathcal{H}$ follows from the convexity of $\mathcal{H}^t$. Still by Hölder's inequality, equality holds if and only if $F(y) - G(y)$ is constant $P^t(y,\cdot)$-a.e. This is the same as $F - G$ being a harmonic function for P^t, i.e., $P^t(F-G) = F - G$. □

In the next proposition, we see that the Fréchet derivatives of $\mathcal{H}^t$ and $\mathcal{H}$ can be written in terms of the twisted semigroup and its generator respectively. The boundedness property was proven in [12, Prop. 4.4], the other conclusions follow from elementary calculus of variations.

Proposition 3 *Suppose (CD1+) holds with an unbounded function W. Then $\mathcal{H}^t : L_\infty^{W_0} \to L_\infty^V$ and $\mathcal{H} : L_\infty^{W_0} \cap D(\mathcal{H}) \to L_\infty^V$ are smooth and admit the following Taylor expansions in terms of Fréchet derivatives:*

$$\mathcal{H}^t(F_0 + aF) = \mathcal{H}^t(F_0) + a(P^t_{f_0} - I)(F) + \frac{1}{2}a^2\left[P^t_{f_0}(F^2) - (P^t_{f_0}F)^2\right] + O(a^3)$$

for $F_0, F \in L_\infty^{W_0}$, $f_0 = e^{F_0}$ and $a \in \mathbb{R}$; and

$$\mathcal{H}(F_0 + aF) = \mathcal{H}(F_0) + a\mathcal{L}_{f_0}F + \frac{1}{2}a^2\left[\mathcal{L}_{f_0}(F^2) - 2F\mathcal{L}_{f_0}F\right] + O(a^3)$$

for $F_0, F \in D(\mathcal{H}) \cap L_\infty^{W_0}$, $f_0 = e^{F_0}$ and $a \in \mathbb{R}$.

From Proposition 3, a local maxima of $\langle \mu, -\mathcal{H}^t(G)\rangle$ must satisfy $\langle \mu, (P^t_g - I)h\rangle = 0$ for all $h \in B(\mathcal{Y})$. The convexity of $\mathcal{H}^t$ implies that $\langle \mu, -\mathcal{H}^t(G)\rangle$ is concave and, therefore, any local maximum is also global. Aperiodicity of the process grants that the optimizer is the same for all $t > 0$. This proves the third part of the theorem. The fourth part follows with a similar reasoning.

4. COMPUTATION OF LYAPUNOV FUNCTIONS

The main result in this section lies in the observation that solving (7) for u can be done with relative ease for a significant number of cases.

A first case to be considered is when the process is deterministic. In this case, the solution is trivial and u cannot be used as a Lyapunov function. In fact, since $K = 0$ and $K^* = 0$, (7) either has no solution ($\nabla \cdot fp \neq 0$) or it is solved by any u ($\nabla \cdot fp = 0$). This is consistent with the fact that the ratio $\mathcal{L}u/u = f \cdot \nabla \ln u$ can be arbitrarily increased by making $u = u_0^k$, for some Lyapunov function u_0, and increasing the power k.

For nondeterministic processes, however, nontrivial results can be obtained. In the case of a pure diffusion, for example, (7) leads to an advection equation which can be solved explicitly in the scalar case. When K involves both a jump and a diffusion term, then one would obtain a partial integro-differential equation from (7), which is typically difficult to solve. However, the complexity of solving (7) is considerably reduced when the operator K is a probability kernel, which is the case for piecewise-deterministic processes. In particular, when the operator K has finite rank, solving (7) reduces to a finite-dimensional problem. This is the case of the example in the next section, where an analytical solution to (7) is provided.

For a fixed $\bar{U} \in L^\infty(m)$, let $\bar{u} = e^{\bar{U}}$ and let u be the positive solution of the quadratic equation

$$u^2 K^*\left(\frac{p}{\bar{u}}\right) - u\nabla \cdot fp - pK\bar{u} = 0 \ . \tag{8}$$

We denote by T the map $\bar{U} \mapsto \ln u$.

Define the *normalizing* projection $N : L(\pi_0) \to L(\pi_0)$ as $N := I - \Pi_0$, where $\Pi_0 U = \int U \, d\pi_0$. When applied to U, N has the same effect as normalizing e^U. A fixed point U of NT gives a solution $u = e^U$ for (7). We show in the next theorem that the iteration of the map NT is an effective computational procedure to solve (7).

Assumption 1 *i. Suppose the kernel K has finite rank.*

ii. Suppose the kernel K is irreducible, i. e., $K - \lambda I$ is the generator of an irreducible process as defined above.

Theorem 3 *Suppose that $p > 0$ and that there exists $\bar{V} \in L(\pi_0)$ such that $e^{\bar{V}}$ solves (7). Then, under Assumption 1, the sequence $\{U_n\}$ defined recursively as*

$$U_{n+1} = NTU_n, \ n \ge 0$$

converges exponentially in $L(\pi_0)$ to $\bar{V}+c$ for any $U_0 \in L(\pi_0)$, where c is a constant.

The proof of this result is given in the next subsection. This proof suggests that a generalization for irreducible compact kernels K may be possible [see Remark 1].

The conclusions of Theorem 3 would still hold for a different definition of the map T, where $\bar{u} = e^{\bar{U}}$ is mapped to $u = \ln U$

that solves the equation

$$u^2 K^* \left(\frac{p}{\bar{u}}\right) - \bar{u}\nabla \cdot fp - pK\bar{u} = 0 \ .$$

4.1 Proof of Theorem 3

Our proof makes use of Banach fixed-point theorem [1]. To apply this theorem we need to verify that $NT : L(\pi_0) \to L(\pi_0)$ is a contractive map on some closed invariant set $C \subset L_\infty$, that is, $\|NT(U) - NT(V)\|_{\pi_0} \leq \gamma \|U - V\|_{\pi_0}$ for every $U, V \in C$ and some $\gamma < 1$. To this purpose, we show that the Gâteaux derivative $dNT(U, \delta)$ of NT at U in the direction δ satisfies $\|dNT(U, \cdot)\|_{\pi_0} \leq 1$ for all $U \in L(\pi_0)$. This implies that NT is nonexpansive (see [22] for the relation between derivatives and the Lipschitz constant) and that the closed balls $C_r = \{U \in L(\pi_0) : \|U - \bar{V}\|_{\pi_0} < r\}$ are invariant under NT for all $r > 0$ given that $\bar{V} \in L(\pi_0)$ is a fixed point of NT. Finally, we will show that $\|dNT(U, \cdot)\|_{\pi_0} < \gamma < 1$ for all $U \in C$ and the result of the theorem follows from Banach fixed-point theorem.

Next we compute the derivative dT and show that it is a contraction in $L(\pi_0)$. We can rewrite (8) as

$$\alpha(\bar{u})u - \beta - u^{-1}\gamma(\bar{u}) = 0 \ . \tag{9}$$

for $\alpha(\bar{u}) = K^*(p/\bar{u})$, $\beta = \nabla \cdot fp$ and $\gamma = pK\bar{u}$. Let $d\mathcal{A}(\bar{u}, \delta)$ denote the derivative of the operator $\mathcal{A}$ at $\bar{u}$ in the direction δ. We first compute the pointwise derivative and then make the point that it is indeed a Gâteaux derivative. Recall that $u = e^{T\bar{U}}$ and take the derivative in (9) with respect to $\bar{U}$ to obtain

$$(u\alpha(\bar{u}) + u^{-1}\gamma(\bar{u}))dT(\bar{U}, \delta) + ud\alpha(\bar{u}, \bar{u}\delta) - u^{-1}d\gamma(\bar{u}, \bar{u}\delta) = 0 \ .$$

Because $p > 0$, we have that $u\alpha + u^{-1}\gamma > 0$. We can then solve for dT to obtain

$$dT(\bar{U}, \delta) = -\frac{ud\alpha(\bar{u}, \bar{u}\delta) - u^{-1}d\gamma(\bar{u}, \bar{u}\delta)}{u\alpha(\bar{u}) + u^{-1}\gamma(\bar{u})} \ .$$

Expanding α and γ, we have

$$dT(\bar{U}, \delta) = \frac{uK^*(p\delta/\bar{u}) + u^{-1}pK(\bar{u}\delta)}{uK^*(p/\bar{u}) + u^{-1}pK(\bar{u})} \ . \tag{10}$$

Note that $dT(\bar{U}, 1) = 1$ and that dT is a positive operator since K and K^* are. Therefore, dT is a probability kernel in δ and $\|dT\| \leq 1$, where $\|\cdot\|$ denotes the norm of the supremum. Since $B(\mathcal{Y})$ is dense in $L(\pi_0)$, we conclude that $\|dT\|_{\pi_0} \leq 1$ as well. This shows that dT is well defined as Gâteaux derivative and that T is nonexpansive. Since N is a projection, we have immediately that $\|dNT\|_{\pi_0} \leq 1$ and NT is nonexpansive as desired.

To show that NT is contractive, we want to decompose dT as

$$dT(U, \delta) = \Pi_1(U)\delta + M(U)\delta \tag{11}$$

where Π_1 is a projection on the space of constant functions and $\|M(U)\|_{\pi_0} < \kappa < 1$ for all U. In this case, using the fact that $\Pi_0\Pi_1 = \Pi_1$, we have that

$$dNT(U, \cdot) = (I - \Pi_0)(\Pi_1(U) + M(U)) = NM(U)$$

so that $\|dNT(U, \cdot)\|_{\pi_0} < \kappa < 1$ for every $U \in C$.

In order to show that such a decomposition is possible we use the irreducibility of K and the fact that it has finite rank. Note from (10) that dT can be regarded as the convex combination of two probability kernels:

$$dT(\bar{U}, \delta) = \lambda_1 \frac{K^*(p\delta/\bar{u})}{K^*(p/\bar{u})} + \lambda_2 \frac{K(\bar{u}\delta)}{K\bar{u}}$$

where $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$. Moreover, $\lambda_2 = 0$ if and only if $\lambda = 0$. This implies that dT can be regarded as transition operator for a Markov chain that executes the jumps corresponding to K with some positive probability and the reverse jumps corresponding to K^* with some probability. More importantly, dT inherits irreducibility and aperiodicity from K. Thus the decomposition in (11) is simply an application of Perron-Frobenius theorem [19] for the case of a finite rank K. In addition, the existence of a uniform bound $\gamma < 1$ is possible as follows. Let $C = TC_r$. Then, C is invariant under T and compact since T is compact [compactness of T follows from compactness of K and continuity of T]. Hence, there exists a constant γ such that $\|M(U)\|_{\pi_0} < 1$ $\forall U \in C$.

Finally, given any initial condition U_0, we can find and appropriate set C such that $NTU_0 \in C$ and apply the Banach fixed-point theorem to conclude our result.

Remark 1 *A generalization of this proof may be possible for compact irreducible kernels. Indeed, the decomposition (11) is possible for compact kernels. It is just not clear under what conditions $M(U)$ will be contractive. Another possible generalization is for ergodic kernels K; in this case, a different proof that uses the Lyapunov functions for K may be possible.]*

5. EXAMPLE: LYAPUNOV FUNCTIONS FOR OPTIMOTAXIS

In this section we apply our results to the process we introduced in [14]. In that paper, convergence to a stationary probability density was proven with a method that does not provide information on the rate of convergence.

Optimotaxis was introduced in [14] as a solution to an in loco optimization problem with point measurements only. This problem was extended in [15], where it was posed as the problem of controlling the probability density of a PDP by selecting the jump intensity λ and the jump kernel Q as a function of an output. Applications were provided in the area of mobile robotics, where the method can be used to solve problems such as search, deployment and monitoring.

In this example we consider a simple instance of the controlled process obtained in [14]. The process represents vehicles moving with position $\mathbf{x} \in \mathcal{X} = \mathbb{R}^d$ and velocity $\mathbf{v} \in \mathcal{V} = \mathbb{S}^d$, the unit sphere in $\mathbb{R}^d$. The measure ν is the Lebesgue measure on the sphere modulo a normalization factor. In this case we have $f = v$. In our output feedback formulation, the controller can only observe the output function $q(x)$, which represents measurements of some physical signal taken at position x. Our objective is to make the probability density of the vehicles' position to converge to the output function $q(x)$ and then have an external observer that can measure the vehicles position to collect information about $q(x)$, much like in MCMC methods [8].

The jump intensity is chosen such that $\lambda(x, v) = \eta(x) - v \cdot$

$\nabla \ln q(x)$, where the function η is a design parameter that must be chosen such that λ is nonnegative. Note that for such a η to exist it is necessary that $\ln g$ be locally Lipschitz.

The jump kernel is such that x does not change and v has a jump distribution that is uniform on $\mathcal{V}$. More precisely,

$$Qh(x, v) = \int_{\mathcal{V}} h(x, v)\nu(dv)$$

for $h \in B(\mathcal{Y})$. We have shown that a process with these characteristics has indeed an invariant density $q(x)$ [14]. As discussed in [14], this controller can be implemented using just the information from the output $q(x)$.

Finding Lyapunov functions for this process is not trivial due to the intricate relationship between the continuous state x and the discrete mode v. To illustrate this, we consider the Metropolis-Hastings algorithm, which is a classic MCMC algorithm. Optimotaxis and Metropolis-Hastings are similar in the sense that the probabilities to reject a point in Metropolis-Hastings and the probability to reject a velocity v in Optimotaxis are essentially the same. The main difference is that, because in Metropolis-Hastings the state represents a variable in a computer, the controller can look at a point and reject it without moving the state to that point, which in turn is not possible if the state represents the position of a physical vehicle. In [21], it is shown that $q^{-1/2}$ is a Lyapunov function for the Metropolis-Hastings algorithm in terms of the goal distribution q [indeed, this is also a common Lyapunov function for diffusions]. However, this fails to be a Lyapunov function for Optimotaxis, which would be naturally expected since the dependence on v is not taken into account.

Using the method developed in the previous sections, we were able to find a Lyapunov function for Optimotaxis which turns out to be a nontrivial modification of the Lyapunov function for the random walk generated by the Metropolis-Hastings algorithm. This Lyapunov function is $u = \lambda^{-1/2}q^{-1/2}$. With such a u we conclude exponential ergodicity of the PDP in the following theorem. For some $\epsilon > 0$, let η be a constant such that

$$\max_{v \in \mathcal{V}} \|\nabla \ln q(x)\| + \epsilon \le \eta < \infty \ . \tag{12}$$

Note that a constant η is not necessary for our result, but it will simplify our proof. The next assumption characterizes distributions with exponential decaying tails.

Assumption 2 *1. $\|\nabla \ln q\|$ is bounded*

2. $\liminf_{\|x\| \to \infty} \|\nabla \ln q\| > 0$

3. The Hessian $H_{xx} \ln g$ converges to 0 as $\|x\| \to \infty$.

Theorem 4 *Suppose that the output function q satisfies Assumption 2. Then, for η as in (12), $u = \sqrt{\lambda q}$ is a Lyapunov function for the PDP and the PDP is exponentially ergodic:*

$$\|P^t((x, v), \cdot) - \pi\|_u \le B_0 u e^{-b_0 t}$$

for some positive constants B_0 and b_0 and any initial condition $(x, v) \in \mathcal{X} \times \mathcal{V}$, where $d\pi = qdm$.

In the next section, we will describe in detail the process that led to construction of this Lyapunov function, but for now we prove this result using Theorem 1.

Proof. Because $\lambda > \epsilon$ and v is restarted uniformly after jumps, we have that, for any $A,\ C \in \mathcal{B}$ such that $m(A) > 0$ and C is compact, there exists a time $T < \infty$ such that the probability of reaching A from C is positive for $t \ge T$. This shows that the process is m-irreducible and aperiodic with compact sets as petite sets. From (19) in the next subsection we have

$$\mathcal{L}u \le -\frac{c_0}{2}u + b1_C \ ,$$

for $u = q^{-1/2}\sqrt{\lambda}$, a compact set C and positive constants b and c_0. We can then apply Theorem 1 to conclude the result. □

The condition of q having an exponentially decaying tail in Theorem 4 was proven to be necessary and sufficient for exponential ergodicity of the Metropolis algorithm in the one-dimensional case [13]. Hence, it comes as no surprise the fact that we were not able to find a Lyapunov function to prove exponential convergence when q has a polynomially decaying tail. In the next subsection, we describe in detail how we applied our method to construct a Lyapunov function for Optimotaxis.

5.1 Constructing a Lyapunov function

Although $K = \lambda Q$ is not a compact operator in $L^1(m \times \nu)$, it is a finite rank operator in $L^1(\nu)$ for every fixed $x \in \mathcal{X}$. In fact, if we regard Q as an operator in $L^1(\nu)$ for a fixed x, its range is spanned by the constant function. Moreover, the operator Q has the property that $Q(\alpha(x)h) = \alpha(x)Qh$ for any $h \in B(\mathcal{Y})$ and any α independent of v. That implies that the set of solutions to (7) is invariant under multiplication by a function of x only.

We can rearrange (7) to obtain

$$u = \frac{\nabla \cdot pf + \sqrt{(\nabla \cdot pf)^2 + 4\lambda p \int u \ d\nu \int \frac{\lambda p}{u} \ d\nu}}{2\int \frac{\lambda p}{u} \ d\nu} \ .$$

This implies that there exists funcitons r and s such that u satisfies the following structure

$$u = r(x)\left(\nabla \cdot pf + \sqrt{(\nabla \cdot pf)^2 + \lambda ps(x)}\right) \ . \tag{13}$$

For some positive matrix $\Sigma \in \mathbb{R}^{d \times d}$, take $p(x, v)$ to be $(2\pi)^{-d/2}|\sigma\Sigma|^{-1/2}\exp(-x'\sigma\Sigma^{-1}x/2)$ and let σ tend to infinity. The u that results from the limit is equivalent to that we would obtain with $p = 1$, but in this case p would not be integrable. Although our theory has no need to restrict μ to be a probability measure, we avoid this path due to its more complicated interpretation. The resulting u is of the form

$$u = r(x)\sqrt{\lambda s(x)} \ .$$

Recalling that the set of solutions to (7) is invariant under multiplication by a function of x only, we have

$$u = \gamma(x)\sqrt{\lambda} \ . \tag{14}$$

for any $\gamma(x)$ such that $u \in D(\mathcal{L})$.

This u can be interpreted as the function that maximizes the rate of convergence with equal weight for every (x, v). Since the set of functions u of the form (14) is so large, it is not hard to see that not all elements of this set are Foster-Lyapunov functions for the PDP. Thus, we have applied the results of the previous section fully and have not arrived to a Foster-Lyapunov function. However, we have arrived for a structure for Foster-Lyapunov functions without which we were not able to find Lyapunov functions in the past. Next, we show how we can choose $\gamma(x)$ so that u is a Foster-Lyapunov function for the PDP.

Since the considered PDP is m-irreducible and compact sets are petite, we only need to analyze the behavior of $\mathcal{L}u$ as $\|x\|$ goes to infinity. From the definition of the generator, we have

$$\frac{\mathcal{L}u}{u} = \frac{1}{2} v \cdot \ln \lambda + v \cdot \ln \gamma - \lambda + \sqrt{\lambda} \int \sqrt{\lambda} \, d\nu \ . \tag{15}$$

Define the auxiliary funcitons $\alpha := v \cdot \nabla \ln q$ and $\beta := \int \sqrt{\lambda} \, d\nu = \int \sqrt{\eta - \alpha} \, d\nu$. We can rewrite (15) as

$$\frac{\mathcal{L}u}{u} = -\frac{1}{2} \frac{v' \ \mathrm{H}_{xx} \ln q \ v}{\lambda} + v \cdot \nabla \ln \gamma + \alpha + \beta \sqrt{\eta - \alpha} - \eta \ ,$$

where $'$ denotes the transpose. Notice that when calculating $I(\mu)$ the term that depends on γ vanishes upon integration in ν:

$$\int \frac{\mathcal{L}u}{u} \, d\nu = -\frac{1}{2} \mathrm{tr} \left(\mathrm{H}_{xx} \ln q \int \frac{v v'}{\lambda} \, d\nu \right) + \beta^2 - \eta \ . \tag{16}$$

Let $\gamma(x) = q(x)^{-k}$ for some constant $k > 0$. Then, we can rewrite

$$\frac{\mathcal{L}u}{u} = -\frac{1}{2} \frac{v' \ \mathrm{H}_{xx} \ln q \ v}{\lambda} - k\alpha + \alpha + \beta \sqrt{\eta - \alpha} - \eta \ .$$

We split the right-hand side into two parts and analyze them separately:

$$A := -k\alpha + \alpha + \beta \sqrt{\eta - \alpha} - \eta$$
$$B := -\frac{1}{2} \frac{v' \ \mathrm{H}_{xx} \ln q \ v}{\lambda} \ .$$

Maximizing A on $\alpha \in [-\eta, \eta]$, we have the worst-case bound

$$A \leq -\eta k + \frac{\beta^2}{4(1-k)} \ . \tag{17}$$

We can find the roots of the right-hand side of (17) as a function of k to conclude that $A \leq 0$ for

$$\frac{1 - \sqrt{1 - \beta^2/\eta}}{2} \leq k \leq \frac{1 + \sqrt{1 - \beta^2/\eta}}{2} \ .$$

In special, $A \leq 0$ holds independently of β if and only if $k = 1/2$. When $\gamma = q^{-1/2}$, we have

$$A \leq \frac{\beta^2 - \eta}{2} \leq 0 \ . \tag{18}$$

By Jensen's inequality, we have that $\beta \leq \sqrt{\eta}$ with equality if and only if λ does not depend on v. This analysis provides the valuable intuition that the term A in the convergence rate is taking into account how inhomogeneous the jump rate is in v. Thus, for $\gamma = q^{-1/2}$, we have $A \leq 0$ with equality if and only if $\nabla \ln g = 0$.

One can prove that the bound on A is minimized for η as small as possible. Thus, an important design principle that follows from from our analysis is that η must be chosen as small as possible in order to minimize the bound on A and therefore maximize the convergence rate.

To analyze the interplay between A and B, we make a distinction between two typical cases: a) when g has an exponential tail, e.g., $q = \exp(-c\|x\|)$; and b) when q has a polynomial tail, e.g., $q = \|x\|^{-c}$ for $\|x\|$ larger than some r.

5.1.1 Invariant density with exponential tail

In this case, $\liminf_{\|x\| \to \infty} \|\nabla \ln q\| > 0$ and, therefore, there is a positive constant c_0 such that $\beta^2 < \eta - c_0$ for x large. On the other hand, $H_{xx} \ln q$ is bounded by a constant times $\|x\|^{-1}$ for x large. Thus, A dominates B and we can use the bound in (18) to conclude

$$\limsup_{\|x\| \to \infty} \frac{\mathcal{L}u}{u} \leq -c_0/2 \ , \tag{19}$$

for $u = q^{-1/2} \sqrt{\lambda}$.

5.1.2 Invariant density with polynomial tail

Both A and B decay with rate $\|x\|^{-2}$ in this case and there are instances in which B dominates A as $\|x\| \to \infty$. Hence, the analysis must be case by case. There are even instances where a positive value for the rate function may not be attained. For example, consider (16) for x large and q with a tail of order $\|x\|^{-c}$. Then, for x large, one verifies that

$$\begin{aligned} \int \frac{\mathcal{L}u}{u} \, d\nu &\approx -\frac{1}{2cd\eta} \mathrm{tr} \left(\int v v' \, d\nu \right) \left(\mathrm{tr}(H_{xx} \ln q) + \frac{1}{2} \|\nabla \ln q\|^2 \right) \\ &= -\frac{1}{2d\bar{\eta}} \mathrm{tr} \left(\int v v' \, d\nu \right) \left(2 - d + \frac{c}{2} \right) \frac{1}{\|x\|^2} \end{aligned} \tag{20}$$

where d is the dimension of the Euclidian space $\mathcal{X}$ and where $\eta = c\bar{\eta}$ (note that the choice of η must change depending on c). Interestingly, if the dimension d is larger than $c/2 + 2$, a positive value for the rate function may not be possible according to (20). This illustrates a situation where the result of Proposition 1 does not hold. This is so because of two potential reasons: p does not verify the conditions of the proposition; the process is not exponentially ergodic. Yet, one may be able to use the optimizer u to prove (non-exponential) ergodicity.

Because $\beta \to 0$ as $\|x\| \to \infty$, a Lyapunov function $u = g^{-k} \sqrt{\lambda}$ maintains A nonpositive for x large only if $k = 2$. Thus, if we are interested in using this u to verify condition (CD) for a Lyapunov stability proof when q has a tail of order $\|x\|^{-c}$, we need $c \geq 4$ since $\mathcal{L}u$ is of the order $\|x\|^{c/2-2}$. This is consistent with the fact that q is not a valid probability density when $c \leq 1$.

6. CONCLUSIONS

We have presented a method for the construction of Foster-Lyapunov functions for PDPs based on the maximization of a certain notion of rate of convergence. Under appropriate conditions, solutions can be found by inspection or using an exponentially convergent numerical iterative procedure. Some open questions are whether one can give conditions that guarantee that the maximizing function is a Foster-Lyapunov function and how one can select μ in order to obtain Lyapunov functions and/or simplify calculations.

7. ACKNOWLEDGEMENTS

The authors are indebted to professor Sean Meyn for suggestions regarding the literature on large deviations. The authors acknowledge the support of the Inst. for Collaborative Biotechnologies through grant DAAD19-03-D-0004 from the U.S. Army Research Office. A. R. Mesquita was partially funded by CAPES (Brazil) grant BEX 2316/05-6.

8. REFERENCES

[1] K. Atkinson and W. Han. *Theoretical numerical analysis: A functional analysis framework*. Springer London, Limited, 2009.

[2] J. Bect. A unifying formulation of the fokker-planck-kolmogorov equation for general stochastic hybrid systems, 2008.

[3] M. Bujorianu and J. Lygeros. General stochastic hybrid systems: Modelling and optimal control. In *43rd IEEE Conference on Decision and Control, 2004. CDC*, volume 2, 2004.

[4] M. Davis. *Markov models and optimization*. Monographs on statistics and applied probability. Chapman & Hall, London, UK, 1993.

[5] J. Deuschel and D. Stroock. *Large deviations*. Academic Press, 1989.

[6] D. Down, S. Meyn, and R. Tweedie. Exponential and uniform ergodicity of Markov processes. *The Annals of Probability*, pages 1671–1691, 1995.

[7] J. Feng and T. Kurtz. *Large deviations for stochastic processes*. American Mathematical Society, 2006.

[8] W. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, 1996.

[9] J. Hespanha. Modeling and analysis of stochastic hybrid systems. *IEE Proc — Control Theory & Applications,* Special Issue on Hybrid Systems, 153(5):520–535, 2007.

[10] J. Hu, J. Lygeros, and S. Sastry. Towards a theory of stochastic hybrid systems. In N. Lynch and B. Krogh, editors, *Hybrid Systems: Computation and Control*, volume 1790 of *LNCS*, pages 160–173. Springer, 2000.

[11] M. Jacobsen. *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes*. Birkhauser, 2006.

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

[13] K. Mengersen and R. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Annals of Statistics*, 24(1):101–121, 1996.

[14] A. R. Mesquita, J. P. Hespanha, and K. J. Åström. Optimotaxis: A stochastic multi-agent optimization procedure with point measurements. In M. Egerstedt and B. Mishra, editors, *Hybrid Systems: Computation and Control*, number 4981, pages 358–371. Springer-Verlag, Berlin, Mar. 2008. Available at `http://www.ece.ucsb.edu/~hespanha/published`.

[15] A. R. Mesqutia and J. P. Hespanha. Jump Control of Probability Densities with Applications to Autonomous Vehicle Motion. 2009. Submitted to IEEE Transactions on Automatic Control. Available at `http://www.ece.ucsb.edu/~hespanha/published`.

[16] S. Meyn and R. Tweedie. Stability of Markovian processes II: Continuous-time processes and sampled chains. *Advances in Applied Probability*, pages 487–517, 1993.

[17] S. Meyn and R. Tweedie. Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, pages 518–548, 1993.

[18] S. Meyn and R. Tweedie. *Markov chains and stochastic stability*. Springer, 1996.

[19] E. Nummelin. *General irreducible Markov chains and non-negative operators*. Cambridge University Press, 2004.

[20] Z. Palmowski and T. Rolski. A technique for exponential change of measure for Markov processes. *Bernoulli*, pages 767–785, 2002.

[21] G. Roberts and R. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110, 1996.

[22] A. Siddiqi. *Applied functional analysis: numerical methods, wavelet methods, and image processing*. CRC, 2004.