Construction of Lyapunov Functions for Piecewise-Deterministic Markov Processes

Alexandre R. Mesquita and João P. Hespanha

***Abstract*— The purpose of this contribution is twofold: 1) to present for the first time a Lyapunov function that proves exponential ergodicity of a process studied by the authors in [1], where the problem of controlling the probability density of a swarm of robotic agents was solved; 2) to introduce alongside the method used to construct this Lyapunov function, which is of interest in its own since it may be applicable to a wide class of piecewise-deterministc 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. Analytical solutions are often possible as shown by examples.**

I. Introduction

Markov processes with jumps, such as 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. Lyapunov stability criteria for such processes involve solving partial integro-differential inequalities, which is typically difficult to do numerically. This paper shows that Lyapunov functions for such processes can be obtained from the minimization of a convex functional that arises in the theory of large deviations [2], [3]. 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. This equation is considerably simpler than integro-differential inequalities and we present cases in which it can be solved in closed form.

Our method was developed in the attempt to find Lyapunov functions for a process studied by the authors in [1]. 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 the mentioned previous work without the use of Lyapunov techniques.

The key advantage in using the Lyapunov-based techniques 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 provides 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) estimate how long one needs to wait to be sufficiently close to the steady-state distribution and (2) design processes with fast convergence rates. This second point was exploited in Section III-B.

The results in this paper have the limitation that, 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 here is indeed a Lyapunov function, which is often a simple procedure.

In Section II, we define concepts and recall some key results in the theory of Markov processes and piecewise-deterministic Markov processes. Our main result is given in Section III, where we characterize the exponential ergodicity in Optimotaxis. In Section IV we establish connections between Lyapunov functions and the concept of rate functions. In Section V, we apply our method to prove ergodicity of a scalar linear transport process.

II. Preliminaries

A. Piecewise-Deterministic Markov Processes

Our paper follows closely the definition of piecewise-deterministic Markov Processes (PDP) introduced in [4] and extended in [5]. We consider the PDP $\boldsymbol{\xi}(t) := (\mathbf{x}(t), \mathbf{v}(t))$ evolving on $\mathcal{Y} = \mathcal{X} \times \mathcal{V}$, where $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{V}$ is a compact set. We denote by $\mathcal{B}$ the corresponding Borel σ-field and by $B(\mathcal{Y})$ the set of bounded $\mathcal{B}$-measurable functions. For brevity, we define the PDP in terms of its extended generator [4], [5]:

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

for $h \in D(\mathcal{L})$ [for the sake of a simpler exposition, we judiciously omit the fact that the extended generator takes also path-differentiable functions in its domain; we remark, however, that our derivations are valid on the full domain of the generator]. Here ∇ denotes the gradient with respect to x, $f : \mathcal{Y} \to X$ is the vector field that describes the deterministic flow, the *jump rate* $\lambda : \mathcal{Y} \to \mathbb{R}^+$ defines the infinitesimal probability λdt of a jump occurring in the interval $[0, dt]$, and the *jump kernel* Q is such taht $Qh(y) := \int h(\xi) Q(y, d\xi)$ for $h \in B(\mathcal{Y})$, where, given that a jump occurs at $t = \tau$, $Q(\xi, A) := \Pr(\boldsymbol{\xi}(\tau) \in A \mid \boldsymbol{\xi}(\tau^-) = \xi)$ for $A \in \mathcal{B}$.

This PDP model is captured by several stochastic hybrid system models that appeared in the literature, including [6], [7]. Fig. 1 depicts a schematic representation of our PDP.

The authors are with the Center for Control, Dynamical Systems and Computation, University of California, Santa Barbara, CA 93106

This material is based upon work supported by the NSF grant CNS-0720842 and by the Institute for Collaborative Biotechnologies through grant W911NF-09-D-0001 from the U.S. ARO. A. R. Mesquita was partially funded by CAPES (Brazil) grant BEX 2316/05-6.

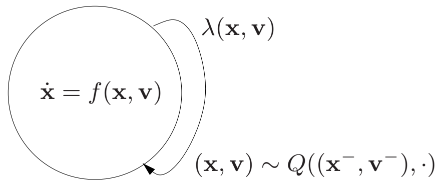

Fig. 1. Stochastic hybrid automaton for the PDP

Let m denote the Lebesgue measure in $\mathcal{X}$. We assume $\mathcal{V}$ to be a compact subset of a locally compact separable metric space equipped with a Borel measure ν such that $\nu(\mathcal{V}) = 1$. We denote by $L^1(m \times \nu)$ the space of real functions integrable with respect to $m \times \nu$. As in [8], we assume the existence of a kernel Q^* such that

$$m \times \nu(d\xi_1)Q(\xi_1, dx_2) = m \times \nu(d\xi_2)d\nu Q^*(\xi_2, d\xi_1) \ .$$

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

B. Exponential Ergodicity

Next, we briefly recall the drift condition for exponential ergodicity of Markov process. The reader is referred to [9] for a more complete definition of the main concepts. For a Markov process $\mathbf{\Phi}(t)$, let $P^t(y, A) = \Pr(\mathbf{\Phi}(t) \in A | \mathbf{\Phi}(0) = y)$, for $A \in \mathcal{B}$, and let $\mathcal{L}$ denote its extended generator.

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

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

where $\|\mu\|_V = \sup_{|h| \leq V} \left| \int h \ d\mu \right|$ for a measure μ on $\mathcal{B}$.

We define *the continuous drift condition*

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

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

We note that for the examples considered in this paper, all compact sets are petite.

Theorem 1 ([9]) *Suppose $\mathbf{\Phi}$ is ψ-irreducible and aperiodic. Then, V-exponential ergodicity is equivalent to the existence of a function V that satisfies the condition (CD).*

III. Lyapunov Functions for Optimotaxis

In this section we give a Lyapunov proof for exponential ergodicity of the process generated by the Optimotaxis algorithm. Optimotaxis was introduced in [1] as a solution to an in loco optimization problem with point measurements only. This problem was extended in [10], 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 paper we consider a simple instance of the controlled process obtained in [1]. 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.

The jump intensity is chosen such that $\lambda(x, v) = \eta - v \cdot \nabla \ln q(x)$, where the constant η is a design parameter that must be chosen such that λ is nonnegative. 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 $q(x)$ [1]. As discussed in [1], this controller can be implemented using just the information from the output $q(x)$.

Finding Lyapunov functions for this process is difficult 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 [11], it is shown that $q^{-1/2}$ is a Lyapunov function for the Metropolis-Hastings algorithm in terms of the goal distribution q. However, it is is easy to see that no function that is independent of v can satisfy the condition (CD) for Optimotaxis.

Using the method developed in the following 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

$$\|\nabla \ln q(x)\| + \epsilon \leq \eta < \infty \ . \tag{1}$$

We 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 1 *1) $\|\nabla \ln q\|$ is bounded*

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

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

Theorem 2 *Suppose that the output function q satisfies Assumption 1. Then, for η as in (1), $u = \sqrt{\lambda/q}$ is a Lyapunov function for the PDP and the PDP is u-exponentially ergodic:*

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

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

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 $t \geq T$. This shows that the process is m-irreducible and aperiodic with compact sets as petite sets. From (6) in the next subsection we have

$$\mathcal{L}u \leq -\frac{c_0}{2} u + b 1_C \ ,$$

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

The process that led to the construction of this Lyapunov function is described in the following sections.

The condition of q having an exponentially decaying tail in Theorem 2 was proven to be necessary and sufficient for the exponential ergodicity of the Metropolis algorithm in the one-dimensional case [12]. Hence, it comes as no surprise that we were not able to find a Lyapunov function to prove exponential convergence when q has a polynomially decaying tail.

A. Constructing a Lyapunov function

Our starting point is the candidate Lyapunov function

$$u = \gamma(x)\sqrt{\lambda} \ , \tag{2}$$

where $\gamma(x)$ is some uniformly positive function to be identified. How we arrived to this candidate Lyapunov function is the theme of the next sections.

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

$$\frac{\mathcal{L}u}{u} = \frac{1}{2} v \cdot \ln\lambda + v \cdot \ln\gamma - \lambda + \sqrt{\lambda}\int \sqrt{\lambda}\, d\nu \ . \tag{3}$$

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 (3) 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. 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

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

We can find the roots of the right-hand side of (4) 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

$$A \leq \frac{\beta^2 - \eta}{2} \leq 0 \ , \tag{5}$$

where the last inequality follows from Jensen's inequality. In addition, equality holds if and only if $\nabla \ln q = 0$. This analysis provides the valuable intuition that the term A in the convergence rate is taking into account how much information is provided by the gradient of $\ln q$.

One can also prove that the bound on A is minimized for η as small as possible. Thus, a design guideline that follows 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 q has an exponential tail, e.g., $q = \exp(-c\|x\|)$; and b) when q has a polynomial tail, e.g., $q = \|x\|^{-c}$ for $\|x\|$ large.

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 (5) to conclude

$$\limsup_{\|x\|\to\infty} \frac{\mathcal{L}u}{u} \leq -c_0/2 \ , \tag{6}$$

for $u = \sqrt{\lambda/q}$.

2) Invariant density with polynomial tail: Both A and B decay proportionally to $\|x\|^{-2}$ in this case. As a consequence, our candidate Lyapunov function cannot be used to prove exponential ergodicity. Yet, it may be used to prove (non-exponential) ergodicity (see [13]). Results here depend on the specific invariant density q. 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 in a Lyapunov stability proof when q has a tail of order $\|x\|^{-c}$, we must have $c \geq 4$ 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 $c \leq 1$.

B. Consequences for the design

The analysis in the previous subsection suggests how we can change the algorithm of Optimotaxis to improve convergence properties. We do this by increasing the nominal velocity. To this purpose, we redefine the vector field and the jump rate to be $f = v\rho$, for some scalar nominal velocity ρ, and $\lambda = \rho(\eta - v \cdot \nabla \ln q\rho)$. All other parameters are kept the same. Under the framework of [10], one can verify that q remains the unique stationary density of the process as long as $\ln \rho$ is some Lipschitz smooth function of x and independent of v.

When ρ is a constant, we have that our previous Lyapunov function $u = \sqrt{q(\eta - v \cdot \nabla \ln q)}$ has its convergence rate $\mathcal{L}u/u$ in (3) multiplied by ρ. Provided the original process is exponentially ergodic, one can improve the convergence rate in the tail of u arbitrarily by increasing ρ. This, however, could even worsen the convergence rate to steady state since it only takes into account the behavior on the tails. Following our analysis in the previous subsection, the term A would become more negative, but the term B would become necessarily more positive on the neighborhoods of the maxima of q. Intuitively, to have good convergence one wants vehicles to move slowly in the neighborhoods of the maxima of q and to scape quickly from the regions where q is small. This motivates the use of an output-modulated nominal velocity $\rho(q)$ with the properties just mentioned. For this process, one can use the method in the next section with $p = 1/\rho$ to find the Lyapunov function $u = \sqrt{\lambda q/\rho}$ that proves exponential ergodicity provided $\|\nabla \ln q\rho\|$ is uniformly positive for $\|x\|$ large. Simulation results for this new design are illustrated in Fig. 2, where we see that an output-dependent ρ in the range $[23, 40]$ improves speed of convergence with respect to $\rho = 25$ while ultimately providing better convergence than $\rho = 50$ [after vehicles approach the maxima of q].

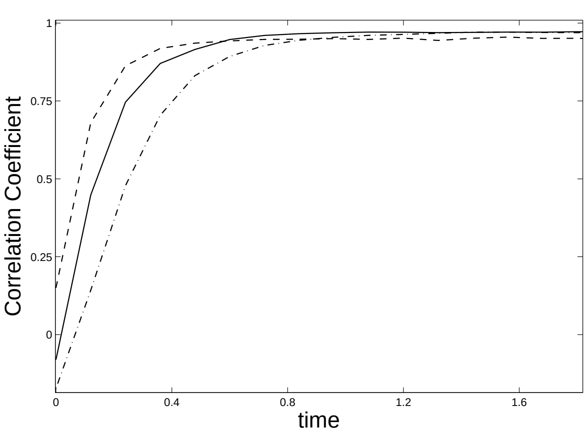

Fig. 2. Coefficient of correlation between the probability density of vehicles at x and the output function $q(x) = 0.4e^{-\|x\|} + 0.6e^{-\|x-[1.5\ \ -1.5]'\|}$ for $\rho = 25$ (dot-dashed), $\rho = 50$ (dashed) and $\rho = 40\tanh([23 + 1/q^2]/40)$ (solid). For further simulation details, see [1].

IV. Lyapunov Functions and Rate Functions

In this section we discuss how one can obtain Lyapunov functions for PDPs 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 [2]. For a probability measure μ on $\mathcal{B}$, we define the *rate function*

$$I(\mu) := \sup\left\{ \int -\frac{\mathcal{L}u}{u}\, d\mu : u \in D(\mathcal{L}), u \geq 1 \right\} \quad . \qquad (7)$$

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. However, it is not common to solve the maximization posed in (7) explicitly. Our objective is to construct a Lyapunov function u by solving this maximization problem.

A converse result is given in the following proposition, which is proven in [14]. We say that a function W_0 grows strictly slower than W if $\limsup_{\|x\|\to\infty} W_0(x)/W(x) = 0$. We denote this relation by $W_0 \ll W$.

Proposition 1 *Suppose that the V-exponentially ergodic process $\mathbf{\Phi}(t)$ satisfies (CD) with W unbounded off petite sets. Then, given $u_0 \in D(\mathcal{L})$ satisfying $1 \leq u_0 \leq V$ and the growth condition $-u_0^{-1}\mathcal{L}u_0 \ll W$, there exists a unique probability measure μ such that u_0 attains the supremum in (7). In particular, this is true for Lyapunov functions satisfying the growth condition.*

In view of [3], we can rewrite the rate function in terms of an optimization problem that will be shown to be convex:

$$I(\mu) = \sup\left\{ \langle \mu, -e^{-U}\mathcal{L}e^{U} \rangle : e^{U} \in D(\mathcal{L}) \right\} \qquad (8)$$

where $\langle \cdot, \cdot \rangle$ denotes integration. From this point on it is convenient to assume that $D(\mathcal{L})$ is an algebra and that $h \in D(\mathcal{L})$ implies $e^h \in D(\mathcal{L})$.

Along with convexity, another important property of the functional $\langle \mu, -e^{-U}\mathcal{L}e^{U} \rangle$ is that it depends affinely of the differential operator $f \cdot \nabla$. As a consequence, this operator does not appear in the first variation optimality condition (9) or in the second variation of $\langle \mu, -e^{-U}\mathcal{L}e^{U} \rangle$. This allows one to exploit the compactness properties typically present in jump kernels.

Some useful facts about $I(\mu)$ are established in [3] for the discrete-time case. From [3, Prop. 4.9], we have that $I(\mu) \geq 0$ and $I(\mu) = 0$ if and only if μ is an invariant measure for $\mathbf{\Phi}$, and in this case the supremum is attained by any constant function. Also from [3, Prop. 4.6] we have that $I(\mu) < \infty$ only if μ is absolutely continuous with respect to the invariant measure for $\mathbf{\Phi}$.

In the following theorem we provide a sufficient condition for a function to attain the supremum in (8). Define $K = \lambda Q$.

Theorem 3 *i. The optimization problem in (8) is a convex optimization problem in* U.

ii. A sufficient condition for $u = e^U \in D(\mathcal{L})$, $U \geq 0$, *to attain the supremum in (7) for* $d\mu = p\ dm$, $p \in D(\mathcal{L}^*)$, *is*

$$uK^*\left(\frac{p}{u}\right) - \nabla \cdot fp - \frac{p}{u}Ku = 0 \quad m\text{-a.e.} \tag{9}$$

iii. Suppose e^U *is a solution to (9). Then,* $e^G \in D(\mathcal{L})$ *is also a solution if and only if* $G(y) - U(y) = G(z) - U(z)$ $\mu(dz)K(z, dy)$*-a.e.*

Roughly speaking, Theorem 3 (iii) implies that the set of solutions to (9) is invariant under multiplication by harmonic functions of Q (i.e., functions h such that $Qh = h$). In particular, this is always true for multiplication by a constant. A proof for the theorem is provided in [14].

A. Computation of Optimizers

The main result in this section lies in the observation that solving (9) for u can be done with relative ease for a significant number of PDPs. This is true because (9) has no differential terms in u, which makes it possible to exploit the compactness properties of K. In particular, when K and K^* have finite rank, solving (9) reduces to a finite-dimensional problem. A wide class that satisfies this property is given by a generalization of the Markov Jump Linear systems in [15], where one may allow Markov transitions to depend on the continuous state. This is also the case of our Optimotaxis example, where a closed form solution to (9) is provided.

To see why this is possible, note that we can rewrite (9) using the fact that u solves an implicit quadratic equation:

$$u = \frac{\nabla \cdot pf + \sqrt{(\nabla \cdot pf)^2 + 4pK(u)K^*(p/u)}}{2K^*(p/u)} \ . \tag{10}$$

When K and K^* have finite rank, this expression defines a finite dimensional manifold where u lies. Therefore, even when it is not possible to solve (10) explicitly, this equation gives us structure to make good guesses for candidate Lyapunov functions.

One can also attempt to solve (10) via an iterative procedure as follows: given u_n, replace u on the the right hand side of (10) with u_n and define u_{n+1} to be that value modulo some normalization. The normalization is necessary since the class of solutions to (9) is invariant under multiplication by a constant. Under reasonable conditions, this iteration is verified to converge for finite rank jump kernels.

Finally, it is important to remark that this is the point where our approach takes advantage of the specialized setting of PDPs. If, for example, one was to consider a purely deterministic process, the solution set to (9) would be trivial (either empty or the whole space of functions) and u cannot be used as a Lyapunov function. On the other hand, nontrivial results can be obtained in the nondeterministic case. But, when one considers a general process that includes both jumps and diffusion, the solution to (9) is typically difficult, since one would be dealing with a partial integro-differential equation.

B. A Candidate Lyapunov Function for Optimotaxis

Although $K = \lambda Q$ is not a compact operator in $B(\mathcal{Y} = \mathcal{X} \times \mathcal{V})$, it is a finite rank operator in $B(\mathcal{V})$ for every fixed $x \in \mathcal{X}$. In fact, if we regard Q as an operator in $B(\mathcal{V})$ 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. This implies that the set of solutions to (9) is invariant under multiplication by a function of x only.

We obtain from (10)

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

This implies that there exist functions r and s such that u satisfies the following structure

$$u = r(x)\left(v \cdot \nabla p + \sqrt{(v \cdot \nabla p)^2 + \lambda p s(x)}\right) \ .$$

Let p be a multivariable normal distribution and let its covariance tend to infinity. The u that results from the limit is equivalent to the one 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 limit satisfies

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

Recalling that the set of solutions to (9) is invariant under multiplication by a function of x only, we have

$$u = \gamma(x)\sqrt{\lambda} \ . \tag{11}$$

for any $\gamma(x)$ such that $u \in D(\mathcal{L})$, $u \geq 1$.

This u can be interpreted as the function that maximizes the rate of convergence with equal weight for every (x, v). It is clear that not all elements of the form (11) are Lyapunov functions for the PDP. However, we have arrived to a structure for Lyapunov functions without which we were not able to find Lyapunov functions in the past.

V. Example: A Scalar Linear Transport Process

In this section we show how the method proposed above can be applied to a process different from Optimotaxis. The process we consider in this example appears in many fields ranging from neutron transport phenomena to biology (see [16] and references therein). This well studied process provides a simple example of how our method can be used to find Lyapunov functions in closed form. This process describes a particle with position $\mathbf{x} \in \mathcal{X} = \mathbb{R}$ moving with velocity $\mathbf{v} \in \mathcal{V} = \{-1, +1\}$. Velocity jumps occur with intensity $\lambda_v(x) \geq \epsilon > 0$ and with $Q = \delta_{\{-v\}}$, where δ denotes the Dirac mass. A process so defined is aperiodic and irreducible, with compact sets being petite.

From (10), we can derive an optimizer of the form

$$u_v(x) = \frac{v \cdot \nabla p_v + \sqrt{(v \cdot \nabla p_v)^2 + 4p_v\lambda_v u_{-v}\lambda_{-v}p_{-v}/u_{-v}}}{2p_{-v}\lambda_{-v}/u_{-v}} \ .$$

As above, take the limit as $p \to 1$. The resulting limit satisfies

$$u_v(x) = u_{-v}\sqrt{\frac{\lambda_v}{\lambda_{-v}}} \ .$$

Therefore, the class of minimizers is characterized by functions of the form

$$u_v(x) = \gamma(x)\sqrt{\lambda_v}$$

where the function γ only depends on x . To construct a Lyapunov function, we need now to select γ properly. To this purpose, we evaluate

$$\frac{\mathcal{L}u_v}{u_v} = v(\ln\gamma)' + \frac{v}{2}(\ln\lambda_v)' + \sqrt{\lambda_v\lambda_{-v}} - \lambda_v$$

where $'$ denotes derivative with respect to x. In order to make the right hand side negative for both values of v, we must choose γ so that

$$-\frac{1}{2}(\ln\lambda_{-1})' + \sqrt{\lambda_1\lambda_{-1}} - \lambda_{-1} < (\ln\gamma)' < -\frac{1}{2}(\ln\lambda_1)' - \sqrt{\lambda_1\lambda_{-1}} + \lambda_1 \ .$$

Assuming that the inequalities hold, we choose $(\ln\gamma)'$ to be the mean of the two bounds:

$$(\ln\gamma)' = -\frac{1}{2}(\ln\sqrt{\lambda_1\lambda_{-1}})' + \frac{1}{2}\lambda_1 - \frac{1}{2}\lambda_{-1} \ .$$

This leads to a consistent choice of $\gamma \geq 1$ as long as $\lambda_1 \geq \lambda_{-1}$ in the positive tail and $\lambda_1 \leq \lambda_{-1}$ in the negative tail. For this choice of γ we have

$$\frac{\mathcal{L}u_v}{u_v} = \frac{1}{2}\left(\ln\frac{\sqrt{\lambda_1}}{\sqrt{\lambda_{-1}}}\right)' + \sqrt{\lambda_1\lambda_{-1}} - \frac{\lambda_1+\lambda_{-1}}{2} \ . \quad (12)$$

For x large enough, the derivative term in the right hand side of (12) is either dominated by λ_1 or negative. Therefore, exponential ergodicity depends solely on the difference between the geometric and the arithmetic averages of the jump intensities, which is uniformly negative provided $|\lambda_1 - \lambda_{-1}| > 0$ uniformly outside a compact set.

In summary, we have constructed a Lyapunov function that predicts exponential ergodicity for the transport process given that there exists $\epsilon > 0$ such that $\mathrm{sgn}(x)(\lambda_1 - \lambda_{-1}) > \epsilon$ for $|x|$ large enough. If this condition holds with ϵ depending on x but decaying more slowly than $1/|x|$, one can show that $\mathcal{L}u_v < 0$ uniformly off a compact set, which is a known condition for (non-exponential) ergodicity (see [13]). These results are consistent with the expression for the stationary probability density $q_v(x)$ for this process [16]:

$$q_v(x) + q_{-v}(x) = c\exp\left(\int_0^x \lambda_{-1} - \lambda_1 \ dz\right) \ .$$

This density is integrable if and only if $\mathrm{sgn}(x)(\lambda_1 - \lambda_{-1}) > 0$ decays more slowly than $1/|x|$, which shows that our candidate Lyapunov function provides nonconservative conditions for ergodicity. Although this is not the focus of our contribution, it is worth mentioning that, to the best of our knowledge, there is no similar Lyapunov approach to this problem in the literature. Finally we note that, despite the similarities with Optimotaxis, this example shows that our technique can construct Lyapunov functions for processes with jump kernel qualitatively different.

VI. Conclusions

We have presented a method for the construction of Lyapunov functions for PDPs based on the maximization of a certain notion of rate of convergence. This method allowed us to construct a Lyapunov function to prove exponential ergodicity for the Optimotaxis algorithm. Some open questions are how one can select μ such that the maximizer function is guaranteed to be a Lyapunov function; and to find general conditions under which the iterative procedure suggested in Section IV-A converges to the solution of the optimization problem.

VII. Acknowledgements

The authors are indebted to professor Sean Meyn for suggestions regarding the literature on large deviations.

References

[1] A. R. Mesquita, J. P. Hespanha, and K. J. Åström, "Optimotaxis: A stochastic multi-agent optimization procedure with point measurements," in *Hybrid Systems: Computation and Control*, M. Egerstedt and B. Mishra, Eds. Berlin: Springer-Verlag, Mar. 2008, no. 4981, pp. 358–371.

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

[3] I. Kontoyiannis and S. Meyn, "Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes," *Electron. J. Probab*, vol. 10, no. 3, pp. 61–123, 2005.

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

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

[6] J. Hespanha, "Modeling and analysis of stochastic hybrid systems," *IEE Proc — Control Theory & Applications*, Special Issue on Hybrid Systems, vol. 153, no. 5, pp. 520–535, 2007.

[7] J. Hu, J. Lygeros, and S. Sastry, "Towards a theory of stochastic hybrid systems," in *Hybrid Systems: Computation and Control*, ser. LNCS, N. Lynch and B. Krogh, Eds., vol. 1790. Springer, 2000, pp. 160–173.

[8] J. Bect, "A unifying formulation of the fokker-planck-kolmogorov equation for general stochastic hybrid systems," 2008. [Online]. Available: arXiv:0801.3725v2[math.PR]

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

[10] 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.

[11] G. Roberts and R. Tweedie, "Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms," *Biometrika*, vol. 83, no. 1, pp. 95–110, 1996.

[12] K. Mengersen and R. Tweedie, "Rates of convergence of the Hastings and Metropolis algorithms," *Annals of Statistics*, vol. 24, no. 1, pp. 101–121, 1996.

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

[14] A. R. Mesquita and J. P. Hespanha, "Construction of Lyapunov Functions for Piecewise-Deterministic Markov Processes," University of California, Santa Barbara, Tech. Rep., Mar. 2010.

[15] O. do Valle Costa, M. Fragoso, and R. Marques, *Discrete-time Markov jump linear systems*. Springer Verlag, 2005.

[16] H. Othmer and T. Hillen, "The diffusion limit of transport equations II: Chemotaxis equations," *SIAM Journal on Applied Mathematics*, vol. 62, no. 4, pp. 1222–1250, 2002.