Fastest Mixing Markov Chain on a Compact Manifold

Shiba Biswal, Karthik Elamvazhuthi, and Spring Berman

***Abstract*— In this paper, we address the problem of optimizing the convergence rate of a discrete-time Markov chain, which evolves on a compact smooth connected manifold without boundary, to a specified target stationary distribution. This problem has been previously solved for a discrete-time Markov chain on a finite graph that converges to the uniform distribution. In contrast to this previous work, we consider arbitrary positive target measures that are supported on the entire state space of the system and are absolutely continuous with respect to the Riemannian volume. Similar to the earlier work, we pose the optimization problem in terms of maximizing the spectral gap of the operator that pushes forward measures, also known as the forward operator. Prior to formulating the optimization problem, we prove the existence of a Kolmogorov forward operator that can stabilize the class of measures that we consider. In addition, we prove the existence of an optimal solution to our problem. Lastly, we develop a numerical scheme for solving the optimization problem and validate our approach on a simulated system that evolves on a torus in $\mathbb{R}^3$.**

I. Introduction

In this paper, we consider discrete-time Markov chains (DTMC) that evolve on compact, smooth, connected manifolds without boundary. We focus on stabilizing and optimizing the convergence rate of a DTMC to target probability measures that are positive almost everywhere on the manifold and that are absolutely continuous with respect to the Riemannian volume, with L^∞ Radon-Nikodym derivatives that are known as *densities* in simple terms. Our primary motivation stems from applications in multi-agent control systems; specifically, the problem of distributing an ensemble of identity-independent agents on a state space of choice.

While there are numerous well-established methods for control of multi-agent systems [4], [21], [23], many of these control approaches do not scale well to very large agent populations. When all agents follow the same control laws and these control laws are independent of agents' identities, an alternative approach is to apply control techniques to a fluid approximation of the swarm in the form of a mean-field model [8], [9]. This approximation is justified by modeling each agent's dynamics as a DTMC, and then the mean-field behavior of the population is determined by the Kolmogorov forward equation corresponding to the DTMC. A standard assumption in the literature on multi-agent control is that the state space of the agents is Euclidean. However, the state spaces of many mechanical systems are naturally represented as manifolds [22]. Some works have extended multi-agent control approaches on Euclidean spaces to manifolds; for example, consensus and coverage strategies on manifolds are presented in [25] and [1], respectively. In this work, we consider the problem addressed in our recent paper [2], the stabilization of a class of discrete-time nonlinear systems that describe agents evolving on a compact subset of a Euclidean space, for agent state spaces that are manifolds. As we will see, most of our results in [2] carry straightforwardly to our results here. Our approach of analyzing the stability of a dynamical system from a measure-theoretic point of view is quite classical [18], and it is also extensively used in the context of mean-field games [12], optimal transport theory [29], and mean-field control [11]. We give a literature review of significant works that have influenced research on the stabilization and optimization of Markov processes in [2].

It is known that a DTMC admits a stationary distribution under certain conditions of irreducibility, recurrency, and aperiodicity. If it is feasible to make a desired distribution invariant by choosing appropriate transition probabilities for the DTMC, then it is possible to compute optimized transition probabilities that maximize the rate of convergence to the invariant distribution. One of the earliest works on this problem is [3], which optimizes the transition probabilities of a DTMC on a finite graph to stabilize the uniform distribution. Our present paper generalizes this optimization problem to continuous state spaces and non-uniform target measures.

The problems addressed in this paper are posed in Section III. We identify the types of target measures that can be stabilized by DTMCs in Section IV. Instead of working with arbitrary probability measures, we will specifically consider those with L^2 density functions so that we can conduct our analysis on a Hilbert space. In Section IV, we construct a *forward operator*, an analogue of the transition probability matrix, which has a specified target L^2 density as its fixed point. We formulate an optimization problem that maximizes the Markov chain's convergence rate in Section V, prove the existence of an optimal solution in Section VI, and solve the problem numerically in Section VII. In Section VIII, we apply our optimization approach to an example system and confirm through simulations that the system converges to a given target measure.

II. Notation

We first present notation that will be used throughout the paper. We define $\bar{\mathbb{R}}_+ := [0, \infty)$ and $\mathbb{R}_+ := (0, \infty)$. Similarly, we define $\bar{\mathbb{Z}}_+$ as the set of all non-negative integers and $\mathbb{Z}_+$ as the set of all positive integers.

This work was supported by ONR Young Investigator Award N00014-16-1-2605.

Shiba Biswal, Karthik Elamvazhuthi, and Spring Berman are with the School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, 85287 USA {sbiswal, karthikevaz, Spring.Berman}@asu.edu.

We denote the state space by $\mathcal{M}$, a d-dimensional smooth, compact, connected manifold without boundary. Let $T_x\mathcal{M}$ denote the tangent space of the manifold at $x \in \mathcal{M}$. We assume that $\mathcal{M}$ is equipped with a *bi-invariant Riemannian metric* of $g : T\mathcal{M} \times T\mathcal{M} \to \bar{\mathbb{R}}_+$, where $T\mathcal{M} = \cup_{x\in\mathcal{M}} T_x\mathcal{M}$ denotes the tangent bundle of the manifold $\mathcal{M}$. In particular, the natural measure associated with the Riemannian manifold, known as the Riemannian volume, will be denoted by m_g. Let $d_g : \mathcal{M} \times \mathcal{M} \to \bar{\mathbb{R}}_+$ denote the Riemannian distance on $\mathcal{M} \times \mathcal{M}$. For $x \in \mathcal{M}$ and $h > 0$, let $B_g(x, h) = \{y \in \mathcal{M};\ d_g(x, y) \leq h\}$ denote the ball of radius h centered at x. We denote the space of probability measures on $\mathcal{M}$ by $\mathcal{P}(\mathcal{M})$.

Let $(\mathcal{X}, \mathcal{N}, m)$ be a measure space, where $\mathcal{N}$ is the sigma algebra and m is a measure. We define $L^p(\mathcal{X}, m)$, where $p \in [1, \infty)$, as the space $\{f : \mathcal{X} \to \mathbb{R};\ f$ is measurable and $\|f\|_p < \infty\}$, where $\|f\|_p = (\int |f|^p dm)^{1/p}$. We also define $L^\infty(\mathcal{X}, m) = \{f : \mathcal{X} \to \mathbb{R};\ f$ is measurable and $\|f\|_\infty < \infty\}$, where $\|f\|_\infty = \operatorname{ess\,sup}_{x\in\mathcal{X}} |f(x)|$. For topological spaces $\mathcal{X}, \mathcal{Y}$, if $T : \mathcal{X} \to \mathcal{Y}$ is an operator, it will be understood that $\|T\|$ stands for the operator norm, defined as $\sup_x \frac{\|Tx\|_\mathcal{Y}}{\|x\|_\mathcal{X}}$.

A *transition kernel* or *Markov kernel* (or simply *kernel*) is a map $Q : \mathcal{X} \times \mathcal{N} \to [0, 1]$, where $Q(\cdot, E)$ is a measurable function on $\mathcal{X}$ for each fixed $E \in \mathcal{N}$ and $Q(x, \cdot)$ is a measure on $\mathcal{X}$ for each fixed $x \in \mathcal{X}$. Furthermore, for ν on $\mathcal{P}(\mathcal{X})$, the transition kernel Q induces an operator $T : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ defined as:

$$T\nu(E) = \int_\mathcal{X} Q(x, E)\ d\nu(x), \quad E \in \mathcal{N}. \tag{1}$$

Similarly, Q can be used to define an operator on $L^2(\mathcal{X}, m)$. Suppose that ν is absolutely continuous with respect to m, denoted as $\nu \ll m$, and the Radon-Nikodym derivative of ν with respect to m, $d\nu/dm$, is given by $d\nu/dm = f_\nu \in L^2(\mathcal{X}, m)$. In simple terms, f_ν is called the *density* of ν. Then Q induces an operator $T^* : L^2(\mathcal{X}, m) \to L^2(\mathcal{X}, m)$, the adjoint of T, defined as

$$T^* f(x) = \int_\mathcal{X} f(y) Q(x, dm(y)), \quad x \in \mathcal{X},\ f \in L^2(\mathcal{X}, m). \tag{2}$$

We say that Q is *regular* if there exists a function $q \in L^\infty(\mathcal{X} \times \mathcal{X}, m \times m)$ such that for each $x \in \mathcal{X}$, the measure $Q(x, \cdot)$ is absolutely continuous with respect to m and $Q(x, dy) = q(x, y)dm$. The function $q : \mathcal{X} \times \mathcal{X} \to \bar{\mathbb{R}}_+$ will be called the *kernel function* of the transition kernel Q.

The Borel sigma algebra on a measure space $\mathcal{X}$ will be denoted as $\mathcal{B}(\mathcal{X})$.

The spectrum $\sigma(T)$ of a continuous linear operator T on the Banach space $\mathcal{X}$ is the non-void compact set of complex numbers λ for which $T - \lambda I$ does not have a continuous inverse on $\mathcal{X}$. The spectral radius of T will be denoted by $r(T) := \sup\{|\lambda| : \lambda \in \sigma(T)\}$.

A linear operator T on a real ordered vector space $\mathcal{X}$ is said to be *positive*, denoted by $T > 0$, if for $x \in \mathcal{X}$, $x \geq 0$ implies that $Tx \geq 0$.

Consider the Hilbert space of real-valued square integrable functions $L^2(\mathcal{X}, m)$. The dual space of this space is itself. For $x, y \in L^2(\mathcal{X}, m)$, $\langle x, y\rangle = \int xydm$ defines an inner product on $L^2(\mathcal{X}, m)$. The *weak topology* on $L^2(\mathcal{X}, m)$, denoted as w, is the topology defined by the family of seminorms $\{p_{x^*} : x^* \in L^2(\mathcal{X}, m)\}$, where $p_{x^*}(x) = |\langle x, x^*\rangle|$. The *weak* topology*, denoted as w^*, is defined on the dual space $L^2(\mathcal{X}, m)$ by the family of seminorms $\{p_x : x \in L^2(\mathcal{X}, m)\}$, where $p_x(x^*) = |\langle x, x^*\rangle|$. The *weak operator topology* (WOT) on $\mathbb{B}(L^2(\mathcal{X}, m))$, the space of linear bounded operators that map $L^2(\mathcal{X}, m)$ to $L^2(\mathcal{X}, m)$, is defined by seminorms $\{p_{x,y} : x, y \in L^2(\mathcal{X}, m)\}$, where $p_{x,y}(T) = |\langle Tx, y\rangle|$ for $T \in \mathbb{B}(L^2(\mathcal{X}, m))$. Convergence in this topology is as follows:

$$(T_i)_i \overset{WOT}{\longrightarrow} T \iff \langle T_i x, y\rangle \to \langle Tx, y\rangle,\ \forall x, y \in L^2(\mathcal{X}, m).$$

III. Problem Formulation

We begin by stating our assumptions. We consider measures in $\mathcal{P}(\mathcal{M})$ that have square integrable Radon-Nikodym derivatives with respect to m_g; this assumption gives us the advantage of working on a Hilbert space, $L^2(\mathcal{M}, m_g)$, which significantly simplifies the analysis. Consider the following discrete-time flow on the space of probability densities $L^2(\mathcal{M}, m_g)$:

$$\begin{aligned} f_{n+1} &= Pf_n, \quad n = 0, 1, 2, ... \\ f_0 &\in L^2(\mathcal{M}, m_g), \end{aligned} \tag{3}$$

where $P : L^2(\mathcal{M}, m_g) \to L^2(\mathcal{M}, m_g)$ is the induced forward operator. To define P, let $K : \mathcal{M} \times \mathcal{B}(\mathcal{M}) \to [0, 1]$ be a transition kernel. To ensure that P preserves probability densities, we impose the following property on K:

$$\int_\mathcal{M} K(x, \mathcal{M}) dm_g(x) = 1, \quad \text{for } m_g\text{-a.e. } x \in \mathcal{M}. \tag{4}$$

Using system (3), we define a discrete-time Markov chain (DTMC) $\Phi = \{\Phi_0, \Phi_1, \ldots\}$ on $\mathcal{M}$ that describes an agent's dynamics on the state space. The Markov chain induces a probability measure $\mathbb{P}$ on $\mathcal{M}^\infty$, defined as follows: $\mathbb{P}(E)$ is the probability of the event $\{\Phi \in E\}$, where $E \in \bigvee_{i=0}^\infty \mathcal{B}(\mathcal{M}_i)$ (the product sigma algebra) with $\mathcal{M}_i = \mathcal{M}$ for each $i \in \bar{\mathbb{Z}}_+$. For every $n \in \mathbb{Z}_+$, we say that the random variable Φ_n is distributed according to μ_n, the measure corresponding to f_n, if $\mathbb{P}(\Phi_n \in E) = \mu_n(E)$. Suppose that Φ_n is the current agent state and is distributed according to μ_n. Then the Markov property implies that Φ_{n+1} is distributed according to μ_{n+1}, where the density f_{n+1} corresponding to μ_{n+1} is given by (3).

The action of P on a function $f \in L^2(\mathcal{M}, m_g)$ can be represented as follows: for $E \in \mathcal{B}(\mathcal{M})$,

$$\int_E (Pf)(x) dm_g(x) = \int_\mathcal{M} K(x, E) f(x) dm_g(x). \tag{5}$$

If K is regular, then we can obtain an explicit expression for P, rather than defining P through (5). Defining $k : \mathcal{M} \times \mathcal{M} \to \mathbb{R}_+$ as the kernel function of K, we have that $k \in$

$L^\infty(\mathcal{M}\times\mathcal{M}, m_g\times m_g)$. From (5), we obtain the following: for $y\in\mathcal{M}$ and $f\in L^2(\mathcal{M}, m_g)$,

$$Pf(y)=\int_{\mathcal{M}} k(x,y)f(x)dm_g(x). \tag{6}$$

Operators of this form are called *integral operators* [6]. The function k is called the *kernel of the integral operator*.

We will first consider the problem of stabilizing system (3) to a target density.

Problem III.1. *Given a target density $f_d\in L^\infty(\mathcal{M}, m_g)$, determine whether there exists a transition kernel $K:\mathcal{M}\times\mathcal{B}(\mathcal{M})\to[0,1]$ such that* (3) *satisfies $\lim_{n\to\infty}P^n f_0\to f_d$ for all initial densities $f_0\in L^2(\mathcal{M}, m_g)$, where the forward operator P is defined in* (5).

This problem will be addressed in Section IV. When the state space $\mathcal{M}$ is a Lie group, we can in fact show the existence of a regular transition kernel K, in which case P is defined in (6).

Given that there exists such a transition kernel, we then address the problem of choosing the transition kernel that optimizes the convergence rate (mixing rate) of system (3) to the target density. DTMCs that converge exponentially fast to their stationary distributions are called *geometrically ergodic*. A Markov chain is geometrically ergodic if the forward operator that operates on the densities of the process has a spectral gap in L^2; the converse is only true for reversible Markov chains [24]. Therefore, the convergence rate is characterized by the L^2 spectral gap. Toward this goal, we will prove the existence of a spectral gap for P. Further, we will prove in the next section that 1 is the unique largest eigenvalue of P, which implies that P is stochastic, as in the case of DTMCs that evolve on a discrete state space. Let $\lambda_2(P)$ be the eigenvalue of P with the second-largest modulus.

Problem III.2. ***(Optimization of convergence rate)*** *Let $\mathcal{K}$ be the set of all Markov kernels defined on $\mathcal{M}\times\mathcal{B}(\mathcal{M})\to[0,1]$ that each correspond to a well-defined bounded operator on $L^2(\mathcal{M}, m_g)$. Given a target density $f_d\in L^\infty(\mathcal{M}, m_g)$, determine whether the following optimization problem admits a solution:*

$$\min_{\mathcal{K}}|\lambda_2(P)|$$

subject to the constraint $Pf_d=f_d$, where P is the forward operator (5).

IV. Existence of a Solution to Problem III.1

Let a target density f_d be given that is strictly positive almost everywhere on $\mathcal{M}$ and satisfies $f_d, \frac{1}{f_d}\in L^\infty(\mathcal{M}, m_g)$. In this section, we will prove the existence of an operator P that has f_d as its fixed point, i.e., $Pf_d=f_d$.

We will first address the problem for the case where the state space is $\mathcal{M}$, a compact manifold as defined in Section II. We define a regular kernel K associated with a kernel function k as

$$K(x, dm_g(y))=k(x,y)dm_g(y)=\frac{\chi_{B_g(x,h)}}{m_g(B_g(x,h))}dm_g(y) \tag{7}$$

for all $x,y\in\mathcal{M}$, where $\chi_{(\cdot)}$ is the characteristic function. This transition kernel induces the Markov chain known as the "ball walk" [19]. Let $S:L^2(\mathcal{M}, m_g)\to L^2(\mathcal{M}, m_g)$ be the operator defined by this transition kernel as per (6),

$$Sf(y)=\int_{\mathcal{M}} k(x,y)f(x)dm_g(x),\ \ f\in L^2(\mathcal{M}, m_g).$$

The resulting Markov chain has the invariant measure $m_g(B_g(x,h))dm_g(y)$, i.e., $Sf_\pi = f_\pi$ with $f_\pi(x) = m_g(B_g(x,h))$ for all $x\in\mathcal{M}$. Note that in the case when $\mathcal{M}$ is a Lie group, S is a self-adjoint operator.

For Markov chains on discrete state spaces, the types of stabilizable measures can be characterized using the classical *Perron-Frobenius theorem* [14]. In order to extend these results to continuous state spaces, we require an appropriate generalization of the Perron-Frobenius theorem for infinite-dimensional vector spaces. This generalization has motivated the theory of *Banach lattices* and *positive operators* [26], which has now been developed to the point where the classical theorems of Perron-Frobenius are known to hold under very general conditions. Here, we present several definitions from this theory that will be used to characterize S. A Banach lattice is a Banach space with an order defined on it. Let $\mathcal{X}$ be a Banach lattice. A linear subspace $\mathcal{I}$ of a Banach lattice is a *lattice ideal* if the following condition holds: if $|g|\le|h|$ pointwise and $h\in\mathcal{I}$, then $g\in\mathcal{I}$. A positive linear operator $T:\mathcal{X}\to\mathcal{X}$ is said to be *irreducible* if there exists no closed T-invariant lattice ideal distinct to 0 and $\mathcal{X}$. A positive operator T is called *primitive* if the spectral radius $r(T)$ is the only eigenvalue on the spectral circle, defined as the set $\{\lambda\in\mathbb{C}:|\lambda|=r(T)\}$. Primitivity of P implies *aperiodicity* and irreducibility of the associated Markov chain; see [26].

We will now establish some fundamental spectral properties of the operator S. For small $h>0$, these results were demonstrated in [19] using the theory of pseudodifferential operators, along with precise quantitative estimates of $\lambda_2(S)$. Here we sketch an alternative proof, using basic functional analytic principles, that S has a spectral gap for arbitrary $h>0$. The following generalized Perron-Frobenius theorem will be used to establish the simplicity of the eigenvalue 1.

Theorem IV.1. ***(Jentzsch-Perron)****[14] Let T be a linear operator on a Banach lattice $\mathcal{X}$. Suppose that T is positive and compact. If T is irreducible, then $r(T)$ is a positive eigenvalue of algebraic multiplicity one.*

We can now state the following theorem.

Theorem IV.2. *The operator S is compact and $Sf_\pi=f_\pi$, where $f_\pi\in L^\infty(\mathcal{M}, m_g)$ is given by*

$$f_\pi(x)=Cm_g(B_g(x,h)) \tag{8}$$

for all $x\in\mathcal{M}$. Here, C is a normalizing constant such that $\int_{\mathcal{M}} f_\pi(x)dm_g(x)=1$. Moreover, $r(S)=1$ is a simple eigenvalue of the operator S, and $|\lambda_2(S)|<1$.

Proof. The operator S is a compact operator since it is an integral operator with an essentially bounded integral kernel

k ([6], Proposition II.4.7). Let π denote a measure that is absolutely continuous with respect to m_g with density f_π. In addition, let $M_{f_\pi} : L^2(\mathcal{M}, m_g) \to L^2(\mathcal{M}, \pi)$ be a *multiplication operator*, defined as $M_{f_\pi} g = f_\pi g$. Since $f_\pi \in L^\infty(\mathcal{M}, m_g)$, M_{f_π} is bounded and well-defined ([6], Theorem II.1.5). Consider the operator $\hat{S} : L^2(\mathcal{M}, m_g) \to L^2(\mathcal{M}, \pi)$ that is given by

$$\hat{S} = M_{f_\pi}^{-1} S M_{f_\pi}. \tag{9}$$

The operator $\hat{S}$ is an integral operator with integral kernel $\hat{q}$, defined as

$$\hat{q} = \frac{k(x,y) f_\pi(x)}{f_\pi(y)}. \tag{10}$$

Then it follows from the proof of Proposition IV.5 in [2] that $\hat{S}$ is a contraction, and hence $r(S) = r(\hat{S}) = 1$.

To establish that $|\lambda_2(S)| < 1$, we consider the set $U_0^x = \{x\}$ for each $x \in \mathcal{M}$ and inductively define the sets $U_m^x = \cup_{y \in U_{m-1}^x} B_g(y, h)$ for each $m \in \mathbb{Z}_+$. Since the manifold $\mathcal{M}$ is compact, there exists $n \in \mathbb{Z}_+$, independent of $x \in \mathcal{M}$, such that $\mathcal{M} = U_n^x$. From this, it follows that if $f \in L^2(\mathcal{M}, m_g)$ is a non-zero non-negative function, then $S^n f$ is positive almost everywhere on $\mathcal{M}$. Hence, from Theorem 6.1 of [14], it follows that the operator S is primitive, and the only eigenvalue of S with modulus 1 is $r(S)$. This last observation can also be concluded from the fact that $\hat{S}$ is self-adjoint. Using the fact that $S^n f$ is positive almost everywhere on $\mathcal{M}$ if $f \in L^2(\mathcal{M}, m_g)$ is a non-zero non-negative function, we can conclude that S is irreducible. We can now invoke Theorem IV.1 to conclude that the eigenvalue 1 is simple, and hence that $|\lambda_2(S)| < 1$. □

In the case that $\mathcal{M}$ is a *Lie Group* [20], [17], the invariant measure characterized by f_π can be described explicitly under a particular condition on the metric g. Let $x \cdot y$ denote the *right-translation* of $x \in \mathcal{M}$ by y. Similarly, $y \cdot x$ denotes the *left-translation* of x by y. If the metric g is *bi-invariant*, then the distance d_g is invariant under translations, i.e., $d_g(x \cdot y, y \cdot z) = d_g(x, y) = d_g(z \cdot x, z \cdot y)$ for all $x, y, z \in \mathcal{M}$. In this case, the invariant measure coincides with the Riemannian volume m_g, or more specifically, the *Haar volume*. Due to the bi-invariance of the metric, it follows that for each $x, y \in \mathcal{M}$, $B_g(x \cdot y, h) = B_g(x, h) \cdot y := \{z \cdot y,\ z \in B_g(x, h)\}$. Similarly, for each $x, y \in \mathcal{M}$, $B_g(y \cdot x, h) = y \cdot B_g(x, h) := \{y \cdot z,\ z \in B_g(x, h)\}$. Hence, we have that $m_g(B_g(x, h)) = m_g(B_g(e, h)) = x^{-1} \cdot m_g(B_g(x, h)) = m_g(B_g(x, h)) \cdot x^{-1}$ for all $x \in \mathcal{M}$, where $e \in \mathcal{M}$ is the unique identity element of $\mathcal{M}$. Therefore, Theorem IV.2 can be rewritten for Lie groups as follows.

Theorem IV.3. *Let $\mathcal{M}$ be a Lie Group such that the metric g is bi-invariant. Then $S : L^2(\mathcal{M}, m_g) \to L^2(\mathcal{M}, m_g)$ is a compact operator, and $S\mathbf{1} = \mathbf{1}$. Moreover, S has a spectral gap, and hence $|\lambda_2(S)| < 1$.*

Our goal is to construct an operator P that has f_d as its fixed point. Toward this end, we define a multiplication operator $D : L^2(\mathcal{M}, m_g) \to L^2(\mathcal{M}, m_g)$ as $D(g) = \frac{g f_\pi}{f_d}$. Since $\frac{f_\pi}{f_d} \in L^\infty(\mathcal{M}, m_g)$ (note that f_d is bounded from below), D is well-defined and bounded. We define P as

$$P = (S - I)\varepsilon D + I, \quad 0 < \varepsilon << 1, \tag{11}$$

where I is the identity operator on $L^2(\mathcal{M}, m_g)$.

Remark IV.4. *For ε small enough, P is a positive operator.*

We note that since the identity operator I is not compact, P is not compact, and therefore it cannot be represented as an integral operator (6) with an L^2 kernel. Instead, we will show that P can be represented as (5) with a Markov kernel $Q : \mathcal{M} \times \mathcal{B}(\mathcal{M}) \to [0, 1]$. Unlike the kernel K in (7), Q is not regular. From (11), we can write this kernel as follows:

$$Q(x, E) = \int_E k(x, y) a(x) dy + (1 - a(x)) \delta_x(E), x \in \mathcal{M}, \tag{12}$$

where $E \in \mathcal{B}(\mathcal{M})$, $a(x) = \frac{\varepsilon f_\pi(x)}{f_d(x)}$, and $\delta_{(\cdot)}$ is the Dirac measure. This can be easily confirmed to be a Markov transition kernel. Next, we establish properties of the spectrum of the new operator P. The proof of the result closely follows the proofs of Theorems V.11 and V.12 in our paper [2], and hence we will omit the proof here.

Proposition IV.5. *The operator P defined in* (11) *satisfies $P^*\mathbf{1} = \mathbf{1}$, $Pf_d = f_d$. The eigenvalue* 1 *is algebraically simple, isolated (i.e., is not a limit point), and coincides with the spectral radius of P. Furthermore, for ε small enough, and with f_π, f_d bounded from below, P is primitive.*

The construction of P concludes our discussion on the existence of an operator that has a unique fixed point at f_d. We note that such an operator P is not necessarily unique. Next, we move on to optimizing over all such operators in order to maximize the convergence rate of system (3) to f_d.

V. Formulation of the Optimization Problem

In this section, we present a solution to a relaxed version of Problem III.2. The reason for this relaxation will be explained shortly. In the previous section, we proved the existence of an operator P with the following properties: P has a spectral gap, the desired density f_d is its unique eigenvector, and P makes f_d an asymptotically stable equilibrium point for the system (3). In this section, we investigate whether we can pose an optimization problem to search for such an operator P such that the system (3) converges exponentially fast to the equilibrium f_d. The spectral gap of P will determine the rate of convergence of system (3); the larger the gap, the faster the convergence [24]. Recall our assumptions that f_d is in $L^\infty(\mathcal{M}, m_g)$ and is a.e. strictly positive on $\mathcal{M}$. Let μ_d be a measure that is absolutely continuous with respect to m_g with density f_d.

Instead of constructing P, we will pose this optimization problem for $\hat{P} = M_{f_d}^{-1} P M_{f_d}$, defined as in (9), which has the same spectrum as P. The advantage here is that $\hat{P}$ is doubly stochastic, which simplifies the formulation of the optimization problem. We know that given an operator T on a Hilbert space $\mathcal{H}$, for all $\lambda \in \sigma(T)$, we have that $|\lambda(T)| \leq$

$\|T\|$. Unless the operator is self-adjoint or normal, there is no convex formula, that we know of, to characterize the moduli of the eigenvalues. Since we are not searching for a self-adjoint or normal operator $\hat{P}$, the second largest eigenvalue modulus of $\hat{P}$ can only be bounded from above. We obtain this bound by restricting $\hat{P}$ to the subspace obtained after removing the eigenspace $\mathrm{span}(\mathbf{1})$ corresponding to its largest eigenvalue 1:

$$\lambda_2(\hat{P}) = \lambda_1(\hat{P} \circ \mathrm{Proj}_{\mathbf{1}^\perp}) \leq \|\hat{P} \circ \mathrm{Proj}_{\mathbf{1}^\perp}\|_2, \tag{13}$$

where $\mathrm{Proj}_{(.)}$ is the *projection operator* onto a subspace, and $\|\cdot\|_2$ denotes the $L^2(\mathcal{M}, \mu_d)$ norm. The optimization objective is then to minimize the right-hand side of the equation above, knowing that it will be an upper bound for the moduli of all eigenvalues of $\hat{P}$. This is the relaxation that we mentioned at the beginning of the section.

The projection of an arbitrary vector $v \in L^2(\mathcal{M}, \mu_d)$ onto the eigenspace $\mathbf{1}$ is $\mathrm{Proj}_{\mathbf{1}}(v) = \frac{\langle v, \mathbf{1}\rangle}{\|\mathbf{1}\|_2^2}\mathbf{1}$, and the projection of v onto $\mathbf{1}^\perp$ is $\mathrm{Proj}_{\mathbf{1}^\perp}(v) = (I - \mathrm{Proj}_{\mathbf{1}})(v)$. Therefore, we calculate that

$$\left(\hat{P} \circ \mathrm{Proj}_{\mathbf{1}^\perp}\right) v = \hat{P}\left(v - \frac{\langle v, \mathbf{1}\rangle}{\|\mathbf{1}\|_2^2}\mathbf{1}\right) = \hat{P}v - \frac{\langle v, \mathbf{1}\rangle}{\|\mathbf{1}\|_2^2}\mathbf{1}.$$

We now formulate the optimization problem. The optimization variable is the transition kernel function K in the definition (5) of P. The relationship between $\hat{P}$ and P is enforced as constraint (15) in the optimization problem, defined below.

$$\min_K \quad \left\|\hat{P}(K) \circ \mathrm{Proj}_{\mathbf{1}^\perp}\right\| \tag{14}$$

subject to

$$\hat{P} = M_{f_d}^{-1} P M_{f_d}, \tag{15}$$

$$K(x, E) \geq 0 \quad \forall x \in \mathcal{M},\ E \in \mathcal{B}(\mathcal{M}), \tag{16}$$

$$\int_{\mathcal{M}} K(x, \mathcal{M}) dx = 1 \quad \forall x \in \mathcal{M}, \tag{17}$$

$$\int_{\mathcal{M}} f_d(y) K(x, dy) = f_d(x) \quad \forall x \in \mathcal{M}, \tag{18}$$

$$K(x, \mathcal{M} \backslash B_g(x, r)) = 0, \quad \forall x \in \mathcal{M}. \tag{19}$$

The constraints (16), (17) ensure that K is indeed a transition kernel. Constraint (18) ensures that f_d is the stationary distribution of P. Equation (19) imposes a localization constraint on the corresponding Markov chain; that is, starting from any point $x \in \mathcal{M}$, the probability of choosing a point lying outside a ball of specified radius r is zero. This constraint captures physical limitations on an agent's motion as it traverses the state space, which restrict the agent to moving a distance bounded by r in a single time step.

We will discuss the convexity of the optimization problem in the next section.

VI. Optimal Solution

In this section, we show that an optimal solution to the optimization problem (14)-(19) exists. In order to show this, we must prove that the set of decision variables, which will be defined shortly, is compact in some topology and that the objective function (14) is continuous on this set with respect to the chosen topology.

We begin with a definition. Operators that are described by expression (2), where the kernel is not necessarily regular, are called *pseudo-integral operators* [27]; integral operators form a subset. Suppose that $(\mathcal{X}, \mathcal{N}, \mu)$ is a finite Borel measure space.

Definition VI.1. *A bounded linear operator $T : L^2(\mathcal{X}, \mu) \to L^2(\mathcal{X}, \mu)$, where μ is a kernel, is called a pseudo-integral operator if T is given by the expression*

$$(Tf)(x) = \int f(y)\mu(x, dy),\ m\text{-a.e.} \tag{20}$$

for every $f \in L^2(\mathcal{X}, \mu)$.

Remark VI.2. *In fact, the kernel is uniquely determined by the operator in the sense that if $\nu(x, dy)$ satisfies* (20)*, then $\mu(x, \cdot) = \nu(x, \cdot)$ for m-almost every x.*

The following result ([27], Theorem 3.1) will be used in our upcoming discussion.

Theorem VI.3. *Let T be defined as in* (20)*. Then T is a pseudo-integral operator with a positive kernel if and only if T is a positive operator.*

The decision variable in the optimization problem is the transition kernel K. However, in view of the remark and theorem above, we shall instead define a set of operators that satisfy the optimization constraints and identify the transition kernels with these operators. This formulation will significantly simplify our analysis. To begin, we define the following set.

$$\begin{aligned}\mathscr{P} = \Big\{ & P : L^2(\mathcal{M}, m_g) \to L^2(\mathcal{M}, m_g) \text{ is a pseudo-} \\ & \text{integral operator with a kernel } K : \mathcal{M} \times \mathcal{B}(\mathcal{M}) \to \\ & [0, 1],\ P\mathbf{1} = \mathbf{1},\ P^* f_d = f_d \ \text{ for } f_d \in L^2(\mathcal{M}, m_g), \\ & Pf(x) = \int_{B_g(x,r)} f(y) K(x, dm_g(y)) \ \text{ for } x \in \mathcal{M} \Big\}. \end{aligned} \tag{21}$$

Here, the constraint $Pf(\cdot) = \int_{B_g(\cdot,r)} fK(\cdot, dm_g)$ is equivalent to the condition $K(x, \mathcal{M} \backslash B_g(x, r)) = 0$ in (19). For sufficiency, we obtain this condition by choosing f to be the constant function $\mathbf{1}$. The necessary direction is straightforward to prove. Note that the set $\mathscr{P}$ is defined in terms of operators that are of the form (2) due to the statement of Theorem VI.3, which involves operators of the form (20).

Proposition VI.4. *The set $\mathscr{P}$ is closed in the WOT topology.*

Proof. Let $(P_i)_i$ be a sequence in $\mathscr{P}$, and suppose that $(P_i)_i$ converges to P in WOT. We will show that $P \in \mathscr{P}$. WOT convergence implies that $\langle P_i f, g\rangle \xrightarrow{i \to \infty} \langle Pf, g\rangle$ for all $f, g \in L^2(\mathcal{M}, m_g)$. In particular, take $f = \mathbf{1}$. Since $P_i\mathbf{1} = \mathbf{1}$ for all i, we have that $\langle P\mathbf{1}, g\rangle = \langle \mathbf{1}, g\rangle$ for all $g \neq 0$, which implies that $P\mathbf{1} = \mathbf{1}$. Similarly, $P^* f_d = f_d$. We now show that P is a positive operator. Suppose that $f, g \in L^2(\mathcal{M}, m_g)$ are positive functions. Then, $P_i f$ is non-negative for every i.

Since $\bar{\mathbb{R}}_+$ is closed, the limit $\langle Pf, g\rangle$ must be non-negative, which implies that P must be a positive operator. From Theorem VI.3 and the condition $P\mathbf{1} = \mathbf{1}$, we conclude that P must be a pseudo-integral operator with a kernel K taking values in $[0, 1]$. We now consider the last constraint in the set (21). Let $f, g \in L^2(\mathcal{M}, m_g)$ be positive functions. Again, from the definition of the WOT topology, we have that $\lim_i \langle (P_i - P)f, g\rangle = 0$; that is, $\lim_i \int_{\mathcal{M}} (P_i - P)fg dm_g = 0$. Since P_i, P are positive operators, this implies that $(P_i - P)fg \to 0$ in L^1. Therefore, there exists a subsequence such that $\lim_j ((P_i)_j - P)fg = 0$ a.e. [10]. Since g is positive, this implies that $\lim_j ((P_i)_j - P)f = 0$ a.e. Finally, for all i and fixed $x \in \mathcal{M}$, $(P_i f)x$ is non-zero over the set $B_g(x, r)$, and since f is positive, we conclude that the limit $Pf(x)$ must be zero everywhere outside the ball $B_g(x, r)$. Hence, $P \in \mathscr{P}$. □

Remark VI.5. *We note that $\mathscr{P}$ is a set of operators of the form* (2)*, whereas we are interested in the adjoints of these operators, which are of the form* (1)*. Therefore, we require that $\mathscr{P}^* := \{T^*, T \in \mathscr{P}\}$ be closed in the WOT topology, which follows from the fact that the map $P \to P^*$ is WOT continuous.*

Since the optimization problem is a minimization problem, it is sufficient for us to prove that the objective function is only lower-semicontinuous, rather than continuous. We prove this in the following proposition.

Proposition VI.6. *The map $P \mapsto \|M_{f_d}^{-1} P M_{f_d} \circ \mathrm{Proj}_{\mathbf{1}^1}\|$ is lower-semicontinuous on $\mathscr{P}$ and convex.*

Proof. It is clear that the map $P \to M_{f_d}^{-1} P M_{f_d} \circ \mathrm{Proj}_{\mathbf{1}^\perp}$ is continuous. Further, by Problem 109 in [15], the operator norm is weak* lower-semicontinuous on the dual space $L^2(\mathcal{M}, m_g)$. We observe that the objective function is a composition of a lower-semicontinuous function and a continuous function, and is therefore a lower-semicontinuous function on $\mathscr{P}$. Convexity follows from the fact that the objective function is defined as a norm function. □

We can now state the following result, which proves the existence of an optimal solution.

Theorem VI.7. *The optimization problem* (14)-(19) *has an optimal solution.*

Proof. We know that the unit ball in $\mathcal{B}(\mathcal{H})$ is compact in WOT ([16], Theorem 5.1.3). In addition, Theorem IV.2 guarantees the existence of an operator P that satisfies the constraints of the optimization problem, and is therefore an element of $\mathscr{P}$ and is bounded. The optimization algorithm will hence generate bounded operators with norms that do not exceed the norm of P. Accordingly, these operators will form a bounded subset of $\mathscr{P}$, which we will refer to as $\mathscr{P}'$. Since $\mathscr{P}$ is closed in WOT by Proposition VI.4, we conclude that $\mathscr{P}'$ is closed and bounded in WOT, and is therefore compact in WOT. By Theorem VI.3, we can identify a positive kernel K with each pseudo-integral operator P. Therefore, the set of kernels K that satisfy the constraints of the optimization problem is compact in the topology induced by this bijective identification. We denote this set by $\mathcal{K}'$.

Finally, in view of Proposition VI.6, the infimum of the map $K \mapsto \|M_{f_d}^{-1} P(K) M_{f_d} \circ \mathrm{Proj}_{\mathbf{1}^1}\|$ over the set $\mathcal{K}'$ can indeed be achieved; that is, there exists an optimal (minimal) solution. □

A. Special Case

In the case where $\mathcal{M}$ is a Lie group and f_d corresponds to $\mathbf{1}$, the uniform distribution, we know that there exists a regular transition kernel K, defined in (7). Then, the optimization problem (14)-(19) can be posed in terms of the kernel function k. In this case, i.e. when $\mathcal{M}$ is a Lie group and $f_d = \mathbf{1}$, we can expect an exact minimization of the second largest eigenvalue modulus of $\hat{P}$. This is because the inequality in (13) is in fact an equality when the operator $\hat{P}$ is self-adjoint, and we have shown that there exists at least one solution to the optimization problem, the operator S constructed in Section IV, that is self-adjoint. In this subsection, we outline a proof of the existence of the optimal solution in this case.

Let c be a positive constant and r be the radius in constraint (19). In contrast to our approach for the general case, in which we identified the kernels K (the decision variables) with a set of operators P, here we can directly define a set of L^∞ kernel functions, denoted by $\mathcal{K}$, that satisfy the constraints of the optimization problem:

$$\begin{aligned}\mathcal{K} = \Big\{ & k \in L^\infty(\mathcal{M} \times \mathcal{M}, m_g \times m_g) : \\ & 0 \le \|k\|_\infty \le c, \quad \textstyle\int_{\mathcal{M}} k(x, z) dm_g(z) = \mathbf{1}, \\ & \textstyle\int_{\mathcal{M}} k(z, y) dm_g(z) = \mathbf{1}, \\ & k(x, y) = 0 \text{ if } d_g(x, y) > r \quad \forall x, y \in \mathcal{M} \Big\}. \end{aligned} \tag{22}$$

Now, $\mathcal{K}$ is the set of decision variables.

Proposition VI.8. *$\mathcal{K}$ is compact in the weak* topology and is convex.*

Proof. Since $\mathcal{M}$ has finite measure, we have that $\mathcal{K} \subseteq L^\infty(\mathcal{M} \times \mathcal{M}, m \times m) \subseteq L^2(\mathcal{M} \times \mathcal{M}, m \times m)$. First, we will show that $\mathcal{K}$ is closed and bounded in the topology induced by the $\|\cdot\|_2$ norm. Let $(k_i)_i \in \mathcal{K}$ be such that $k_i \xrightarrow{\|\cdot\|_2} \bar{k}$. That is, $\int |k_i - \bar{k}|^2 \to 0$. We will show that $\bar{k} \in \mathcal{K}$. It is straightforward to show that the limit $\bar{k}$ must satisfy $0 \le \|\bar{k}\|_\infty \le c$. Next, we observe that $\left|\int (k_i - \bar{k})\right| \le \int |k_i - \bar{k}| \le \|k_i - k\|_2$ (by Holder's inequality). Therefore, we have that $\left|\int (k_i - \bar{k})\right| \to \mathbf{1}$, which implies that $\int \bar{k} = \mathbf{1}$. We now consider the last constraint in (22). Note that $k_i(x, y) = 0$ for all $x, y \in \mathcal{M}$ such that $d_g(x, y) > r$. Since $k_i \to \bar{k}$ in $\|\cdot\|_2$, there exists a subsequence $(k_{ij})_j$ that converges to $\bar{k}$ m_g-a.e. It then follows that $\bar{k}(x, y) = 0$. Hence, $\mathcal{K}$ is closed in the $\|\cdot\|_2$ norm. The boundedness of $\mathcal{K}$ follows trivially from the condition $\|k(\cdot, \cdot)\|_\infty \le c$.

The convexity of $\mathcal{K}$ follows from brief algebraic computations which show that the constraints in (22) are convex. An application of *Mazur's theorem* ([7], Proposition 12.2)

proves that $\mathcal{K}$ is closed in the weak topology. On the real-valued function space L^2, this implies that $\mathcal{K}$ is in fact closed in the weak* topology. Finally, we obtain our result by applying *Alaoglu's theorem* ([5], Corollary 3.15), which states that a set that is weak* closed and bounded is also weak* compact. □

In the proposition below, we will prove lower-semicontinuity of the map considered in Proposition VI.6.

Proposition VI.9. *The map* $k \mapsto \|M_{f_d}^{-1} P(k) M_{f_d} \circ \mathrm{Proj}_{\mathbf{1}^1}\|$ *is weakly lower-semicontinuous on* $\mathcal{K}$ *and convex.*

Proof. Let $(k_i)_i \in \mathcal{K}$ be such that $(k_i)_i \to k \in \mathcal{K}$ in the weak* topology. Let $P_i, P \in \mathbb{B}(L^2(\mathcal{M}, m_g))$ be the corresponding operators defined by k_i and k, respectively. Consider the WOT topology on $\mathbb{B}(L^2(\mathcal{M}, m_g))$. We will show that $P_i \to P$ in WOT. Convergence in WOT entails showing that $\langle P_i f, g\rangle \overset{i\to\infty}{\longrightarrow} \langle Pf, g\rangle$ for all $f, g \in L^2(\mathcal{M}, m_g)$, which implies that:

$$\int_{\mathcal{M}}\int_{\mathcal{M}} k_i(x,y) f(x) g(y) dm_g(x) dm_g(y) \overset{i\to\infty}{\longrightarrow} \int_{\mathcal{M}}\int_{\mathcal{M}} k(x,y) f(x) g(y) dm_g(x) dm_g(y).$$

The *tensor product* for functions $f, h \in L^2(\mathcal{M}, m_g)$ is denoted by $f \otimes h : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$, which is defined as $f \otimes h(x,y) := f(x)h(y)$. By exercise 1.4.25 of [28], $f \otimes h \in L^2(\mathcal{M} \times \mathcal{M}, m_g \times m_g)$. Therefore, the equation above can also be written as

$$\int_{\mathcal{M}\times\mathcal{M}} k_i(z)(f \otimes g)(z) dz \overset{i\to\infty}{\longrightarrow} \int_{\mathcal{M}\times\mathcal{M}} k(z)(f \otimes g)(z) dz.$$

This is exactly the definition of convergence of $(k_i)_i$ to k in weak*. Therefore, we have that $(P_i)_i \to P$ in WOT. The rest of the proof is similar to the proof of Proposition VI.6. □

In conclusion, the objective function is weak* lower-semicontinuous on the compact set $\mathcal{K}$, and therefore the infimum of the objective function can indeed be achieved.

VII. Numerical Optimization

In this section, we present a numerical approach to solving the optimization problem (14)-(19). As stated in Section II, we assume that the state space $\mathcal{M}$ is a compact smooth connected manifold, without boundary, of dimension d. The subset $\mathcal{M}$ is partitioned into $N \in \bar{\mathbb{Z}}_+$ sets, $\widetilde{\mathcal{M}} = \{\mathscr{M}_1, \ldots, \mathscr{M}_N\}$, where $\mathcal{M} = \cup_{i=1}^N \mathscr{M}_i$ and the sets $\mathscr{M}_i$ have intersections of zero Riemannian volume. We define an equivalent of the transition kernel K for this discretized state space. Let $\tilde{k}_{ij}$ be the probability of jumping to $\mathscr{M}_j$, given that the system state is in $\mathscr{M}_i$. This probability is given by,

$$\tilde{k}_{ij} = \int_{\mathscr{M}_i} K(x, \mathscr{M}_j) dx.$$

We define $\mathbf{K}$ as the matrix $[\tilde{k}_{ij}]_{i,j\in\mathcal{I}}$, where $\mathcal{I} = \{1, \ldots, N\}$. We use $\mathbf{K}$ to construct an approximating Markov chain on the finite state space $\mathcal{I}$. Let $\mathcal{G} = (\mathcal{I}, \mathcal{E})$ be a graph defined on $\mathcal{I}$ with edge set $\mathcal{E} = \{(i,j) : i, j \in \mathcal{I},\ \tilde{k}_{ij} > 0\}$, which specifies the transitions of the Markov chain. An edge (i,j) is in the edge set $\mathcal{E}$ if the distance between the centers of $\mathscr{M}_i$ and $\mathscr{M}_j$ does not exceed r, as per the constraint (19).

Let $\mu \in \mathcal{P}(\widetilde{\mathcal{M}})$ and $\mathbf{P} \in \mathbb{M}(\mathbb{R}^N)$, the space of real-valued matrices. Then $\mathbf{P}$ defined below is equivalent to the operator defined in (5):

$$(\mathbf{P}\mu)(j) = \sum_{i\in\mathcal{I}} \tilde{k}_{ij}\mu(i), \quad j \in \mathcal{I}. \tag{23}$$

Let $\mu_d \in \mathcal{P}(\widetilde{\mathcal{M}})$ be a desired distribution that is positive on $\widetilde{\mathcal{M}}$, and define a diagonal matrix $\mathbf{M}_d = \mathrm{diag}(\mu_d)$.

We can now formulate a finite-dimensional quadratic program that is equivalent to optimization problem (14)-(19) as follows:

$$\min_{\mathbf{K}} \left\| \widehat{\mathbf{P}} - \frac{\mathbf{1}\mathbf{1}^T}{N} \right\| \tag{24}$$

subject to

$$\widehat{\mathbf{P}} = \mathbf{M}_d^{-1}\mathbf{P}\mathbf{M}_d, \tag{25}$$

$$(\mathbf{P}\mu)(j) = \sum_{i\in\mathcal{I}} \tilde{k}_{ij}\mu(i) \quad \forall j \in \mathcal{I},\ \forall \mu \in \mathcal{P}(\widetilde{\mathcal{M}}), \tag{26}$$

$$\tilde{k}_{ij} \geq 0 \quad \forall i, j \in \mathcal{I}, \tag{27}$$

$$\mathbf{K}\mathbf{1} = \mathbf{1}, \tag{28}$$

$$\mu_d \mathbf{K} = \mu_d, \tag{29}$$

$$\tilde{k}_{ij} = 0 \quad \forall (i,j) \notin \mathcal{E}. \tag{30}$$

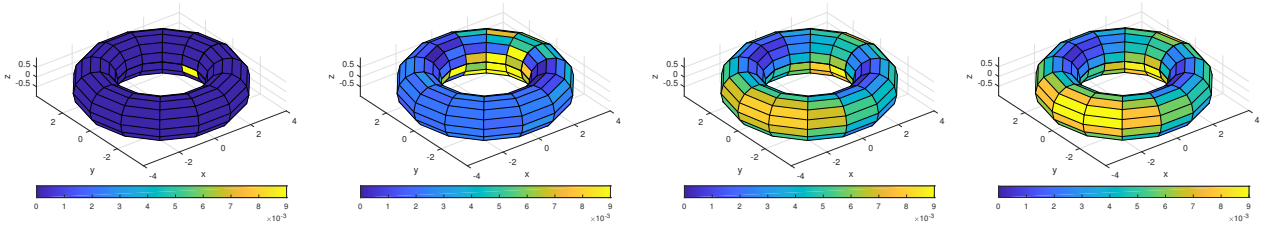
Note that $\mathbf{1}$ in (24) is a vector in $\mathbb{R}^N$. The constraint in (25) ensures that the matrix $\widehat{\mathbf{P}}$ is bistochastic. The constraint (30) is equivalent to (19). Observe that this optimization problem is convex and is similar to the optimization problem solved in [3].

VIII. Simulation Results

In this section, we apply our numerical optimization procedure to an example system evolving on a torus in $\mathbb{R}^3$. For graphs of modest sizes (e.g., 10^3 edges), the convex optimization problem (24)-(30) can be solved using the semidefinite programming (SDP) techniques described in [3], which can be implemented with the MATLAB package CVX [13]. For much larger graphs (e.g., 10^5 edges), the problem can be solved using a subgradient method [3]. Since our focus is on the theoretical formulation of the optimization problem rather than its efficient solution, we make use of CVX to compute an approximate solution of our example. We recall that, as described in Section V, we are solving a relaxation of the optimization problem in Problem III.2.

The state space $\mathcal{M}$ in our example is the torus in $\mathbb{R}^2$ embedded in $\mathbb{R}^3$. This state space is discretized into a 15×15 grid ($N = 225$). We define the initial and target measures on the discretized space as shown in Figs. 1a and 1d, respectively. The value of r in constraint (19) is chosen to be 0.2, a number greater than the partition size $1/15$. We solve the optimization problem (24)-(30) to obtain a transition probability matrix $\mathbf{K}$. Defining $\mathbf{P}$ from the resulting $\mathbf{K}$, we simulate the following version of system (3):

$$\mu_{n+1} = \mathbf{P}\mu_n. \tag{31}$$



(a) Initial measure μ_0 at time $n = 0$ (b) Measure at time $n = 10$ (c) Measure at time $n = 35$ (d) Target measure μ_d

Fig. 1. Simulation of the system (31) at different times n.

Figures 1a-c show snapshots of the simulation of system (31) at three different times. It is evident from the time evolution of the snapshots that the simulated measure μ_n converges asymptotically to the target measure μ_d. To quantify this degree of convergence, we computed the 2-norm error metric $\gamma_n = ||\mu_n - \mu_d||_2$ at the times of the snapshots. The corresponding values of γ_n for $n = 0$, 10, and 35 are 0.7611, 0.0824, and 6.7382×10^{-4}, which clearly shows that μ_n is tending toward μ_d.

IX. Conclusion

In this paper, we presented an approach to optimizing the convergence rate of a discrete-time Markov chain that evolves on a compact, smooth, connected manifold without boundary to a target probability measure. The target measure must satisfy certain properties; namely, it must be absolutely continuous with respect to the Riemannian volume with L^∞ derivatives and positive almost everywhere on the manifold. We proved the existence of a solution to this problem, and we explicitly constructed an operator that stabilizes any measure that satisfies these properties. We also proved the existence of an optimal solution and presented a numerical approach to solving the optimization problem. A possible direction for future work is to extend these results to nonlinear dynamical systems evolving on manifolds, as we did in our paper [2] for nonlinear discrete-time systems evolving on $\mathbb{R}^d$.

References

[1] Subhrajit Bhattacharya, Robert Ghrist, and Vijay Kumar. Multi-robot coverage and exploration on riemannian manifolds with boundaries. *The International Journal of Robotics Research*, 33(1):113–137, 2014.

[2] Shiba Biswal, Karthik Elamvazhuthi, and Spring Berman. Stabilization of nonlinear discrete-time systems to target measures using stochastic feedback laws, 2019. Submitted to *IEEE Transactions on Automatic Control*. Preprint available at: https://www.researchgate.net/publication/330103622.

[3] Stephen Boyd, Persi Diaconis, and Lin Xiao. Fastest mixing Markov chain on a graph. *SIAM Review*, 46(4):667–689, 2004.

[4] Francesco Bullo, Jorge Cortes, and Sonia Martinez. *Distributed control of robotic networks: a mathematical approach to motion coordination algorithms*, volume 27. Princeton University Press, 2009.

[5] Francis Clarke. *Functional analysis, calculus of variations and optimal control*, volume 264. Springer Science & Business Media, 2013.

[6] John B Conway. *A course in functional analysis*, volume 96. Springer Science & Business Media, 2013.

[7] Emmanuele DiBenedetto. *Real analysis*. Springer, 2002.

[8] Karthik Elamvazhuthi and Spring Berman. Nonlinear generalizations of diffusion-based coverage by robotic swarms. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 1341–1346, Dec. 2018.

[9] Karthik Elamvazhuthi, Matthias Kawski, Shiba Biswal, Vaibhav Deshmukh, and Spring Berman. Mean-field controllability and decentralized stabilization of Markov chains. In *2017 IEEE Conference on Decision and Control (CDC)*, pages 3131–3137, Dec. 2017.

[10] Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013.

[11] Massimo Fornasier, Benedetto Piccoli, and Francesco Rossi. Mean-field sparse optimal control. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2028):20130400, 2014.

[12] Diogo A Gomes, Joana Mohr, and Rafael Rigao Souza. Discrete time, finite state space mean field games. *Journal de mathématiques pures et appliquées*, 93(3):308–328, 2010.

[13] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

[14] JJ Grobler. Spectral theory in Banach lattices. In *Operator theory in function spaces and Banach lattices*, pages 133–172. Springer, 1995.

[15] Paul Richard Halmos. *A Hilbert space problem book*, volume 19. Springer Science & Business Media, 2012.

[16] Richard V Kadison and John R Ringrose. *Fundamentals of the theory of operator algebras*, volume 1. American Mathematical Society, 1997.

[17] Alexander A Kirillov. *An introduction to Lie groups and Lie algebras*, volume 113. Cambridge University Press, 2008.

[18] Andrzej Lasota and Michael C Mackey. *Chaos, fractals, and noise: stochastic aspects of dynamics*, volume 97. Springer Science & Business Media, 2013.

[19] Gilles Lebeau, Laurent Michel, et al. Semi-classical analysis of a random walk on a manifold. *The Annals of Probability*, 38(1):277–315, 2010.

[20] John M Lee. *Introduction to smooth manifolds*. Springer, 2001.

[21] Frank L Lewis, Hongwei Zhang, Kristian Hengster-Movric, and Abhijit Das. *Cooperative control of multi-agent systems: optimal and adaptive design approaches*. Springer Science & Business Media, 2013.

[22] Jerrold E Marsden and Tudor S Ratiu. *Introduction to mechanics and symmetry: a basic exposition of classical mechanical systems*, volume 17. Springer Science & Business Media, 2013.

[23] Mehran Mesbahi and Magnus Egerstedt. *Graph theoretic methods in multiagent networks*, volume 33. Princeton University Press, 2010.

[24] Gareth O Roberts and Richard L Tweedie. Geometric L^2 and L^1 convergence are equivalent for reversible Markov chains. *Journal of Applied Probability*, 38(A):37–41, 2001.

[25] Alain Sarlette and Rodolphe Sepulchre. Consensus optimization on manifolds. *SIAM Journal on Control and Optimization*, 48(1):56–76, 2009.

[26] Helmut H Schaefer. Banach lattices. In *Banach Lattices and Positive Operators*, pages 46–153. Springer, 1974.

[27] AR Sourour. Pseudo-integral operators. *Transactions of the American Mathematical Society*, 253:339–363, 1979.

[28] Terence Tao. *An Epsilon of Room, I: Real Analysis: pages from year three of a mathematical blog*. American Mathematical Society, Providence, RI, 2010.

[29] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Society, 2003.