NONSTANDARD REPRESENTATION OF THE DIRICHLET FORM

ROBERT M. ANDERSON, HAOSUI DUANMU, AND AARON SMITH

Abstract. The Dirichlet form is a generalization of the Laplacian, heavily used in the study of many diffusion-like processes. In this paper we present a nonstandard representation theorem for the Dirichlet form, showing that the usual Dirichlet form can be well-approximated by a hyperfinite sum. One of the main motivations for such a result is to provide a tool for directly translating results about Dirichlet forms on finite or countable state spaces to results on more general state spaces, without having to translate the details of the proofs. As an application, we prove a generalization of a well-known comparison theorem for Markov chains on finite state spaces, and also relate our results to previous generalization attempts.

1. Introduction

The Dirichlet form, introduced in Beurling and Deny [BD58], can be used in place of the usual Laplacian in situations where the “usual” Laplacian may not be convenient or make sense at all. Among other applications, it is an important tool for defining diffusive processes on various complicated spaces (see e.g. an early paper on fractals Kusuoka [Kus89] and an introductory paper on infinite-dimensional processes Schmuland [Sch99]), potential theory (see e.g. Albeverio and Schachermayer [AS03] for the application of Dirichlet forms and Doob [Doo12] for the classical theory) and for comparing processes (see versions of such results in Davies [Dav89] and Levin, Peres, and Wilmer [LPW09a]). See e.g. Ma and Röckner [MR12] and Fukushima [Fuk96] for broad introductions to Dirichlet forms and their uses.

The main result of this paper, Theorem 4.3, develops a nonstandard representation theorem for Dirichlet forms. We give a very informal summary here. Recall that the Dirichlet form of a transition kernel $g$ with stationary measure $\pi$ on state space $X$ can be applied to functions $f \in L^2(\pi)$ via the formula:

$$\mathcal{E}_g(f,f) = \frac{1}{2} \int_{x \in X} \int_{y \in X} [f(x) - f(y)]^2 g(x,1,dy)\pi(dx).$$ (1.1)

Theorem 4.3 says that, under appropriate conditions, this Dirichlet form can be well-approximated by an appropriate hyperfinite sum. With notation to be fixed later in the paper, the main conclusion of the theorem is written

$$\mathcal{E}_g(f,f) \approx \frac{1}{2} \sum_{s,t \in \mathcal{S}_X} [F(s) - F(t)]^2 H_s(\{t\})\Pi(\{s\}),$$ (1.2)
where $S_X$ is a set meant to approximate $X$, $F$ is a function meant to approximate $f$, $H_s$ and $\pi$ are measures meant to approximate $g(s, 1, \cdot)$ and $\pi$ respectively, and the symbol “≈”, which is defined precisely in Section 2, means two quantities equal up to an infinitesimal.

The immediate motivation for a result such as (1.2) is that the hyperfinite sum over $S_X$ behaves a great deal like a more-familiar finite sum. This allows one to directly translate certain results about processes on discrete spaces to results about processes on more general spaces, without worrying about translating each step in the associated proof. Such direct translation can lead to substantially simpler and shorter proofs, and sometimes allows one to avoid assumptions, e.g. by allowing one to bypass differentiability assumptions that would be needed in translating the proof steps but which are not needed to translate the theorem statement.

As an illustration of Theorem 4.3 and how it may be applied, we give a simple translation of the well-known comparison theorem for Markov chains, first proved in Diaconis and Saloff-Coste [DSC93]. Previous generalizations of this result have been obtained by purely standard results (see Yuen [Yue00]). However, our version offers several important improvements - for example, we remove various differentiability conditions (which in fact often fail in practice) and are able to obtain substantially sharper estimates in some basic but important test cases. See Section 6 for a brief application and discussion, and our forthcoming companion paper focused on these applications for more details.

1.1. Nonstandard Analysis and Transferring Program. We view the main contribution of this paper as the development of a nonstandard version of the classical Dirichlet form. This paper is a small part of an ongoing effort to provide nonstandard analogues to a variety of important objects in probability theory and statistics: [DRW18], [ADS18], [ADS19a], [ADS19b], [DRS17] and [DR18]. The main motivation is the rough observation that many interesting theorems in probability have the following properties:

- The theorem is initially proved on a discrete (or finite-dimensional) space, and
- The theorem statement does not seem to rely heavily on the space being discrete or finite-dimensional, but
- Several steps in the most natural proof do seem to rely heavily on the space being discrete or finite-dimensional.

For some examples, see e.g. [And76], [Kei84], [DRW18], [ADS18], [DRS17] and [DR18]. For results that have this form, it is natural to try to directly translate the theorem statement without needing to go through the details of translating the full proof. This idea of translation is at the heart of nonstandard analysis, where it is formalized in the notion of “transfer.” This basic idea has let us and others make progress on several problems that otherwise appear difficult, including:

- In Duannu, Rosenthal, and Weiss [DRW18], we prove the Markov chain ergodic theorem for a large class of continuous time general Markov processes, generalizing the well-known Markov chain ergodic theorem for discrete Markov processes.
In Duanmu and Roy [DR18], we show that a decision procedure is extended admissible if and only if it has infinitesimal excess Bayes risk under a non-standard prior distribution. This result holds under complete generality and is a generalization of existing complete class theorems.

In Anderson, Duanmu, and Smith [ADS18], we show that mixing time and hitting time are asymptotically equivalent for general Markov processes under moderate regularity condition, generalizing the same result for finite Markov processes in Peres and Sousi [PS15].

In Duanmu, Roy, and Smith [DRS17], we show that matching priors for specific families of credible sets exist on compact metric spaces, extending a result for finite spaces in Müller and Norets [MN16].

As discussed in the introduction, this paper includes another example of the power of this approach: we provide a translation of Levin, Peres, and Wilmer [LPW09b, Thm. 13.23] that is in many ways more powerful than the extension in previously published work Yuen [Yue00] and Yuen [Yue02].

2. Introduction to Nonstandard Analysis

We briefly introduce the setting and notation from nonstandard analysis. For those who are not familiar with nonstandard analysis, Duanmu, Rosenthal, and Weiss [DRW18] and Duanmu and Roy [DR18] provide introduction tailored to statisticians and probabilists. Arkeryd, Cutland, and Henson [ACH97], Cutland et al. [Cut+95], and Wolff and Loeb [WL00] provide thorough introductions.

We use $^*$ to denote the nonstandard extension map taking elements, sets, functions, relations, etc., to their nonstandard counterparts. In particular, $^*\mathbb{R}$ and $^*\mathbb{N}$ denote the nonstandard extensions of the reals and natural numbers, respectively. An element $r \in ^*\mathbb{R}$ is infinite if $|r| > n$ for every $n \in \mathbb{N}$ and is finite otherwise. An element $r \in ^*\mathbb{R}$ with $r > 0$ is infinitesimal if $r^{-1}$ is infinite. For $r, s \in ^*\mathbb{R}$, we use the notation $r \approx s$ as shorthand for the statement “$|r - s|$ is infinitesimal,” and similarly we use use $r \preceq s$ as shorthand for the statement “either $r \geq s$ or $r \approx s$.”

Given a topological space $(X, T)$, the monad of a point $x \in X$ is the set $\bigcap_{U \in T : x \in U} ^* U$. An element $x \in ^* X$ is near-standard if it is in the monad of some $y \in X$. We say $y$ is the standard part of $x$ and write $y = st(x)$. Note that such $y$ is unique. We use NS($^* X$) to denote the collection of near-standard elements of $^* X$ and we say NS($^* X$) is the near-standard part of $^* X$. The standard part map $st$ is a function from NS($^* X$) to $X$, taking near-standard elements to their standard parts. In both cases, the notation elides the underlying space $Y$ and the topology $T$, because the space and topology will always be clear from context. For a metric space $(X, d)$, two elements $x, y \in ^* X$ are infinitely close if $^* d(x, y) \approx 0$. An element $x \in ^* X$ is near-standard if and only if it is infinitely close to some $y \in X$. An element $x \in ^* X$ is finite if there exists $y \in X$ such that $^* d(x, y) < \infty$ and is infinite otherwise.

Let $X$ be a topological space endowed with Borel $\sigma$-algebra $\mathcal{B}[X]$. An internal probability measure $\mu$ on $^*((X, \mathcal{B}[X]))$ is an internal function from $^*\mathcal{B}[X] \to ^*[0, 1]$ such that

1. $\mu(\emptyset) = 0$;
2. $\mu(^* X) = 1$; and
3. $\mu$ is hyperfinitely additive.
The Loeb space of the internal probability space \( (X, \mathcal{B}[X], \mu) \) is a countably additive probability space \( (X, \mathcal{B}[X], \overline{\mu}) \) such that

\[
\overline{\mathcal{B}[X]} = \{ A \subset X \mid (\forall \epsilon > 0)(\exists A_\epsilon, A_\epsilon \in \mathcal{B}[X])(A_\epsilon \subset A \land \mu(A_\epsilon \setminus A_0) < \epsilon) \}
\]  

and

\[
\overline{\mu}(A) = \sup\{\text{st}(\mu(A_\epsilon))|A_\epsilon \subset A, A_\epsilon \in \mathcal{B}[X]\} = \inf\{\text{st}(\mu(A_0))|A_0 \supset A, A_0 \in \mathcal{B}[X]\}.
\]  

Every standard model is closely connected to its nonstandard extension via the transfer principle, which asserts that a first order statement is true in the standard model if and only if it is true in the nonstandard model. Finally, given a cardinal number \( \kappa \), a nonstandard model is called \( \kappa \)-saturated if the following condition holds: let \( \mathcal{F} \) be a family of internal sets, if \( \mathcal{F} \) has cardinality less than \( \kappa \) and \( \mathcal{F} \) has the finite intersection property, then the total intersection of \( \mathcal{F} \) is non-empty. In this paper, we assume our nonstandard model is as saturated as we need (see e.g. [ACH97, Thm. 1.7.3] for the existence of \( \kappa \)-saturated nonstandard models for any uncountable cardinal \( \kappa \)).

3. Hyperfinite Markov Processes

We start this section by giving an overview of hyperfinite Markov processes developed in [DRW18], [ADS18] and [ADS19b]. Intuitively, hyperfinite Markov processes behave like finite Markov processes but can be used to represent general Markov processes under moderate conditions. For the remainder of this paper, we assume that a probability space \( X \) is always endowed with Borel \( \sigma \)-algebra \( \mathcal{B}[X] \) unless otherwise mentioned.

**Definition 3.1.** A general hyperfinite Markov chain on \( *X \) is characterized by the following four ingredients:

1. A hyperfinite state space \( S \subset *X \).
2. A hyperfinite time line \( T = *\mathbb{N} \).
3. A set \( \{v_i : i \in S\} \) of non-negative hyperreals such that \( \sum_{i \in S} v_i = 1 \).
4. A set \( \{G_{ij} : i, j \in S\} \) consisting of non-negative hyperreals with \( \sum_{j \in S} G_{ij} = 1 \) for each \( i \in S \).

The state space \( S \) naturally inherits the \(*\)-metric of \( *X \). For every \( i, j \in S \), \( G_{ij} \) refers to the internal probability of going from \( i \) to \( j \). The following theorem shows the existence of hyperfinite Markov process.

**Theorem 3.2** ([DRW18, Thm. 7.2]). Given a non-empty hyperfinite state space \( S \subset *X \), \( \{v_i : i \in S\} \) and \( \{G_{ij} : i, j \in S\} \) as in Definition 3.1. Then there exists a \(*\)-additive probability triple \((\Omega, \mathcal{A}, P)\) with an internal stochastic process \( \{X_t\}_{t \in T} \) defined on \((\Omega, \mathcal{A}, P)\) such that

\[
P(X_0 = i_0, X_1 = i_1, ... X_t = i_t) = v_{i_0} G_{i_0 i_1} ... G_{i_{t-1} i_t}
\]

for all \( t \in *\mathbb{N} \) and \( i_0, ..., i_t \in S \).

The proof of Theorem 3.2 essentially follows from transferring the existence theorem for finite Markov processes. For every \( i \in S \) and every internal \( A \subset S \), define \( G_i(A) = \sum_{j \in A} G_{ij} \). For \( n \in T \), define \( G_i^{(n+1)}(A) = \sum_{j \in S} G_j^{(n)}(A) G_{ij} \). It is easy to see that \( G_i^{(n)}(\cdot) \) is an internal probability measure on \( S \) for every \( i \in S \) and \( n \in T \).
We call \( \{ G_i^{(n)}(\cdot) \} \) the \( n \)-step internal transition kernel of the underlying hyperfinite Markov process. Just like standard Markov process, a hyperfinite Markov process is characterized by its 1-step internal transition kernel.

As in [DRW18], [ADS18] and [ADS19b], we shall construct hyperfinite Markov processes from standard Markov processes. Let \( \{ g(x, 1, \cdot) \}_{x \in X} \) be the transition kernel of a Markov process on \( X \). We now define the hyperfinite representation of the state space \( X \).

**Definition 3.3.** Let \((X, d)\) be a compact metric space. Let \( \delta \in \ast \mathbb{R}_{>0} \) be an infinitesimal. A \( \delta \)-hyperfinite representation of \( X \) is a tuple \((S, \{ B(s) \}_{s \in S})\) such that

1. \( S \) is a hyperfinite subset of \( \ast X \).
2. \( s \in B(s) \in \ast B(X) \) for every \( s \in S \) and \( \bigcup_{s \in S} B(s) = \ast X \).
3. For every \( s \in S \), the diameter of \( B(s) \) is no greater than \( \delta \).
4. \( B(s_1) \cap B(s_2) = \emptyset \) for every \( s_1 \neq s_2 \in S \).

The set \( S \) is called the base set of the hyperfinite representation. For every \( x \in \ast X \), we use \( s_x \) to denote the unique element in \( S \) such that \( x \in B(s_x) \).

We say \((S, \{ B(s) \}_{s \in S})\) is a hyperfinite representation of \( X \) if \((S, \{ B(s) \}_{s \in S})\) is a \( \delta \)-hyperfinite representation of \( X \) for some infinitesimal \( \delta \). Hyperfinite representations always exist.

**Theorem 3.4** ([ADS18, Theorem 3.3]). Let \( X \) be a compact metric space. Then, for every positive infinitesimal \( \delta \), there exists a \( \delta \)-hyperfinite representation \( (S_\delta, \{ B(s) \}_{s \in S_\delta}) \) of \( \ast X \).

We fix such a hyperfinite representation of \( \ast X \) and denoted it by \( S \) in the remainder of this section. We now define a hyperfinite Markov process transition kernel on \( S \) from \( \{ g(x, 1, \cdot) \}_{x \in X} \). For \( i, j \in S \), define \( G_{ij} = \ast g(i, 1, B(j)) \). For every internal set \( A \subset S \), define \( G_i(A) = \sum_{j \in A} G_{ij} \). It is straightforward to check that \( \{ G_{ij} \}_{i, j \in S} \) defines the one-step internal transition kernel for some hyperfinite Markov process. In order to establish a connection between \( \{ G_{ij} \}_{i, j \in S} \) and \( \{ g(x, 1, \cdot) \}_{x \in X} \), we impose the following nonstandard condition on \( \{ \ast g(x, 1, \cdot) \}_{x \in \ast X} \).

**Condition SSF.** Let \((S, \{ B(s) \}_{s \in S})\) be a hyperfinite representation of \( X \). The transition kernel \( \{ g(x, 1, \cdot) \}_{x \in X} \) satisfies the \( S \)-strong Feller property if, for every \( x \in \ast X \) and every internal \( A \subset S \), we have

\[
\| \ast g(x, 1, \bigcup_{a \in A} B(a)) - \ast g(s_x, 1, \bigcup_{a \in A} B(a)) \| \approx 0,
\]

where \( s_x \) is the unique point in \( S \) with \( x \in B(s_x) \).

For two (internal) probability measures \( P_1 \) and \( P_2 \), we use \( \| P_1 - P_2 \| \) to denote the \( \ast \)total variational distance between \( P_1 \) and \( P_2 \). (SSF) is a consequence of the following classical condition on \( \{ g(x, 1, \cdot) \}_{x \in X} \).

**Condition DSF.** The transition kernel \( \{ g(x, 1, \cdot) \}_{x \in X} \) satisfies the strong Feller property if for every \( x \in X \) and every \( \epsilon > 0 \) there exists \( \delta > 0 \) such that

\[
(\forall y \in X)(|y - x| < \delta \implies \| g(x, 1, A) - g(y, 1, A) \| < \epsilon).
\]

Suppose \( \{ g(x, 1, \cdot) \}_{x \in X} \) is a reversible transition kernel with stationary distribution \( \pi \). Define \( \Pi \) on \((S, \{ B(s) \}_{s \in S})\) by letting \( \Pi(\{a\}) = \ast \pi(B(a)) \) for every \( a \in S \).
The internal transition kernel \( \{G_i(\cdot)\}_{i \in S} \) may not be *reversible with respect to \( \Pi \). In the following theorem, we construct an internal transition kernel \( \{H_i(\cdot)\}_{i \in S} \), which is infinitesimally close to \( \{G_i(\cdot)\}_{i \in S} \) and is *reversible with respect to \( \Pi \).

**Theorem 3.5** ([ADS19b, Thm. 3.7]). Let \( \{g(x,1,\cdot)\}_{x \in X} \) be a transition kernel on \( X \). Suppose \( \{g(x,1,\cdot)\}_{x \in X} \) is reversible with stationary measure \( \pi \) and satisfies (SSF). For every \( i,j \in S \), define

\[
H_i(\{j\}) = \frac{\int_{B(i)} \ast g(x,1,B(j)) \ast \pi(dx)}{\ast \pi(B(i))} \tag{3.4}
\]

if \( \ast \pi(B(i)) \neq 0 \). Define \( H_{ij} = G_{ij} = \ast g(i,1,B(j)) \) if \( \ast \pi(B(i)) = 0 \).

1. \( \{H_i(\cdot)\}_{i \in S} \) defines an one-step internal transition kernel.
2. \( \Pi \) is a *stationary distribution for \( \{H_i(\cdot)\}_{i \in S} \).
3. \( \{H_i(\cdot)\}_{i \in S} \) is *reversible with respect to \( \Pi \).
4. \( \max_{i \in S} \| g_i(t) \cdot - H_i(\cdot) \| \approx 0 \) for every \( t \in \mathbb{N} \).

We shall use notations \( \{g(x,1,\cdot)\}_{x \in X}, \{G_i(\cdot)\}_{i \in S} \) and \( \{H_i(\cdot)\}_{i \in S} \) as they defined in this section for the rest of the paper.

### 4. Hyperfinite Representation of Dirichlet Form

In this section, we study the relationship between Dirichlet forms for general discrete-time Markov processes and Dirichlet forms for hyperfinite Markov processes.

Let \( \{g(x,1,\cdot)\}_{x \in X} \) denote the transition kernel of a Markov process and let \( \pi \) denote its stationary distribution. the *Dirichlet form* associated with \( g \) of a function \( f : X \to \mathbb{R} \) is defined to be

\[
\mathcal{E}(f,f) = \frac{1}{2} \int_{x \in X} \int_{y \in X} [f(x) - f(y)]^2 g(x,1,dy) \pi(dx). \tag{4.1}
\]

When the state space \( X \) is finite, the integral in the above equation becomes summation.

For the remainder of this section, assume that the state space \( X \) is compact. Let \( (S,\{B(s)\}_{s \in S}) \) denote a hyperfinite representation of \( X \). Let \( \{g(x,1,\cdot)\}_{x \in X} \) be a transition kernel on \( X \) with stationary distribution \( \pi \). We define \( \{G_i(\cdot)\}_{i \in S}, \{H_i(\cdot)\}_{i \in S} \) and \( \Pi \) as in Section 3. We now use hyperfinite Dirichlet forms to approximate standard Dirichlet forms. We first quote the following well-known result in nonstandard analysis.

**Theorem 4.1** ([Cut+95, Thm. 5.3]). Let \( (X,\mathcal{B}(X),\mu) \) be a Radon probability space with Borel \( \sigma \)-algebra and let \( f : X \to \mathbb{R} \) be measurable. Then \( \ast f \) is a lifting of \( f \) with respect to \( \ast \mu \), i.e.

\[
\ast f(x) \approx f(st(x)) \tag{4.2}
\]

for \( \ast \mu \)-almost all \( x \in \ast X \). Consequently, there is a set \( Y \subset \ast X \) with \( \ast \mu(Y) = 1 \) such that for all \( y_1,y_2 \in Y \)

\[
(y_1 \approx y_2) \implies (\ast f(y_1) \approx \ast f(y_2)). \tag{4.3}
\]

Before proving the main result of this section, we quote the following useful lemma.
Lemma 4.2 ([DRW18, Lemma. 7.24]). Let $P_1$ and $P_2$ be two internal probability measures on a hyperfinite set $S$. Then
\[ \| P_1(\cdot) - P_2(\cdot) \| \approx \sup_{f: S \to [0,1]} \sum_{i \in S} P_1(\{i\})f(i) - \sum_{i \in S} P_2(\{i\})f(i), \] where $\| P_1(\cdot) - P_2(\cdot) \| = \sup_{A \in \mathcal{S}(S)}|P_1(A) - P_2(A)|$ and the function $f$ ranges over all internal functions.

Our main result in this section is:

Theorem 4.3. Suppose $X$ is a compact metric space. Let $\{U_1, U_2, \ldots, U_K\} \subset \mathcal{B}[X]$ be a hyperfinite partition of $*X$ such that the diameter of $U_j$ is infinitesimal for each $j \leq K$. Then there exists a hyperfinite representation $(S_X, \{B(s) : s \in S_X\})$ of $X$ such that

1. $S_X$ has internal cardinality $K$ and $\{B(s) : s \in S_X\} = \{U_j : j \leq K\}$.
2. Suppose (SSF) holds, then for every continuous function $f : X \to \mathbb{R}$, we have
\[ \epsilon(f, f) \approx \frac{1}{2} \sum_{s \in S_X} |F(s) - F(t)|^2H_s(\{t\})\Pi(\{s\}) \] for every $s \in S_X$.

where $F : S_X \to \mathbb{R}$ is defined as $F(s) = *f(s)$ for every $s \in S_X$.

Proof. Let $\{U_j : j \leq K\}$ be a partition of $*X$ of *Borel sets with infinitesimal radius and let $k \in \mathbb{N}$. Pick $n \in \mathbb{N}$ and let $f_1, \ldots, f_n$ be $n$ continuous functions from $X$ to $\mathbb{R}$. Let
\[ a = \max \{|f_i(x) - f_i(y)|^2 : x, y \in X, 1 \leq i \leq n\} \in \mathbb{R}. \]

For $1 \leq i \leq n$, let $r_i(x) = \int_{y \in X} |f_i(x) - f_i(y)|^2 g(x, 1, dy)$. As $r_i$ is a measurable function from $X \to \mathbb{R}$, by Theorem 4.1, there exists a Loeb measurable set $M_i$ with $\mu(M_i) = 1$ such that $*r_i(x) = r_i(\text{st}(x))$ for $x \in M_i$. Let $M = \bigcap_{i=1}^n M_i$. Note that $\mu(M) = 1$. Thus, $*r_i(y) = r_i(\text{st}(y))$ for all $1 \leq k \leq n$ and all $y \in M$.

Let $Y \subset M$ be a *Borel set such that $*\mu(Y) > 1 - \frac{1}{2n}$. For $1 \leq j \leq K$, if $U_j \cap Y = \emptyset$, we pick an arbitrary element $s_j \in U_j$. If $U_j \cap Y \neq \emptyset$, we pick an arbitrary element $s_j \in U_j \cap Y$. By the internal definition principle, we know that $S = \{s_j : j \leq K\}$ is a hyperfinite subset of $*X$. For every $s \in S$, we use $B(s)$ to denote the unique element in $\{U_j : j \leq K\}$ which contains $s$. It is easy to see that $(S, \{B(s) : s \in S\})$ is a hyperfinite representation of $*X$. Moreover, $S$ has internal cardinality $K$ and $\{B(s) : s \in S\} = \{U_j : j \leq K\}$. For every $1 \leq i \leq n$, let $F_i : S \to \mathbb{R}$ be $F_i(s) = *f_i(s)$. For every $s \in S$ and every $1 \leq i \leq n$, let
\[ R_i(s) = \sum_{t \in S} \frac{|F_i(s) - F_i(t)|^2}{H_s(\{t\})}. \]

Let $Z = \{s \in S : B(s) \cap Y \neq \emptyset\}$. It is straightforward to see that $\Pi(Z) = *\mu(\bigcup_{s \in Z} B(s)) \geq *\mu(Y) > 1 - \frac{1}{2n}$.

Claim 4.4. For every $z \in Z$ and every $1 \leq i \leq n$, we have $|R_i(z) - *r_i(z)| \leq \frac{2}{n}$ and $*r_i(z) \approx r_i(\text{st}(z))$.

Proof. Pick some $1 \leq i \leq n$ and pick some $z \in Z$. By construction, we have $z \in Y$. By Theorem 4.1, we have $*r_i(z) \approx r_i(\text{st}(z))$. Let $Y \supset Y$ be a *Borel subset of $*X$.
such that \( g(z, 1, Y_i) > 1 - \frac{1}{ka} \). We have

\[ *r_i(z) = \int_{x \in X} [(f_i(z) - f_i(x))^2] g(z, 1, dx) \]

\[ = \int_{x \in Y_i} [(f_i(z) - f_i(x))^2] g(z, 1, dx) + \int_{x \in X \setminus Y_i} [(f_i(z) - f_i(x))^2] g(z, 1, dx). \]

(4.8)

Then we have

(4.9)

Note that \( \int_{x \in X \setminus Y_i} [(f_i(z) - f_i(x))^2] g(z, 1, dx) < \frac{1}{k} \). Let

\[ Z_i = \{ s \in S : B(s) \cap Y_i \neq \emptyset \}. \]

(4.10)

Note that \( G_z(Z_i) > 1 - \frac{1}{ka} \) and \( a \) in Eq. (4.6) is an upper bound of \( [F_i(z) - F_i(s)]^2 \).

Then we have

\[ \sum_{s \in S} [F_i(z) - F_i(s)]^2 G_z({s}) \]

\[ = \sum_{s \in Z_i} [F_i(z) - F_i(s)]^2 G_z({s}) + \sum_{s \in S \setminus Z_i} [F_i(z) - F_i(s)]^2 G_z({s}). \]

(4.11)

Note that \( \sum_{s \in S \setminus Z_i} [F_i(z) - F_i(s)]^2 G_z({s}) < \frac{1}{k} \).

By the continuity of \( f_i \), we have

\[ \int_{x \in Y_i} [(f_i(z) - f_i(x))^2] g(z, 1, dx) \]

\[ = \sum_{s \in Z_i} \int_{x \in B(s) \cap Y_i} [(f_i(z) - f_i(x))^2] g(z, 1, dx) \]

\[ \approx \sum_{s \in Z_i} \int_{x \in B(s) \cap Y_i} [(f_i(z) - f_i(s))^2] g(z, 1, dx) \]

\[ = \sum_{s \in Z_i} [(F_i(z) - F_i(s))^2] g(z, 1, B(s) \cap Y_i). \]

(4.12)

(4.13)

(4.14)

(4.15)

As \( \sum_{s \in Z_i} [(F_i(z) - F_i(s))^2] G_z({s}) - \sum_{s \in Z_i} [(F_i(z) - F_i(s))^2] g(z, 1, B(s) \cap Y_i) < \frac{1}{k} \),

we have \( \|r_i(z) - \sum_{s \in S} [(F_i(z) - F_i(s))^2] G_z({s})\| < \frac{2}{k} \).

By Theorem 3.5 and Lemma 4.2, we have

\[ \sum_{s \in S} [F_i(z) - F_i(s)]^2 G_z({s}) = a \sum_{s \in S} \frac{[F_i(z) - F_i(s)]^2}{a} G_z({s}) \approx \sum_{s \in S} [F_i(z) - F_i(s)]^2 H_z({s}). \]

(4.16)

(4.17)

Hence, we have \( |r_i(z) - *r_i(z)| \leq \frac{2}{k} \) for every \( z \in Z \) and every \( 1 \leq i \leq n \). \( \square \)

For \( 1 \leq i \leq n \), let \( \mathcal{R}_S(F_i, F_i) = \frac{1}{2} \sum_{s \in S} R_i(s) \Pi({s}). \)

Claim 4.5. For \( 1 \leq i \leq n \), we have \( |\mathcal{R}_S(f_i, f_i) - \mathcal{R}_S(F_i, F_i)| < \frac{2}{k} \).
Proof. Pick some $1 \leq i \leq n$. By definition, we have

$$2\mathcal{R}_S(f_i, f_i) = \int_{x \in X} r_i(x)\pi(dx) = \int_{x \in \ast X} \ast r_i(x)\pi(dx) = \int_{x \in \ast X} \ast r_i(x)\pi(dx) + \int_{x \in \ast X \setminus Y} \ast r_i(x)\pi(dx).$$

Note that $\int_{x \in \ast X \setminus Y} \ast r_i(x)\pi(dx) < \frac{1}{k}$.

On the other hand, we have

$$2\mathcal{R}_S(F_i, F_i) = \sum_{s \in S} R_i(s)\Pi(\{s\}) = \sum_{s \in Z} R_i(s)\Pi(\{s\}) + \sum_{s \in S \setminus Z} R_i(s)\Pi(\{s\}).$$

Note that $\sum_{s \in S \setminus Z} R_i(s)\Pi(\{s\}) < \frac{1}{k}$.

We now compare terms $\int_{x \in Y} \ast r_i(x)\pi(dx)$ and $\sum_{s \in Z} R_i(s)\Pi(\{s\})$. Note that $\bigcup_{s \in Z} B(s) \supset Y$ and $\ast r_i$ is S-continuous in $Y$. By Claim 4.4, we have

$$|\int_{x \in Y} \ast r_i(x)\pi(dx) - \sum_{s \in Z} R_i(s)\ast \pi(B(s) \cap Y)| < \frac{2}{k}.$$  

(4.23)

Note that

$$\sum_{s \in Z} R_i(s)\Pi(\{s\}) - \sum_{s \in Z} R_i(s)\ast \pi(B(s) \cap Y) = \sum_{s \in Z} R_i(s)[\ast \pi(B(s)) - \ast \pi(B(s) \cap Y)] < \frac{1}{k}.$$  

(4.25)

Thus, we have $|\mathcal{R}_S(f_i, f_i) - \mathcal{R}_S(F_i, F_i)| \leq \frac{4}{k} < \frac{5}{k}$.

Let $\mathcal{C}$ denote the collection of all continuous functions from $X$ to $\mathbb{R}$ and let $\phi^\phi_f(S)$ be the conjunction of the following formulas:

1. $(S, \{U_j : j \leq K\})$ is a hyperfinite representation of $\ast X$.
2. $|\ast \mathcal{R}_S(f, f) - \frac{1}{2} \sum_{s, t \in S} \mathcal{R}_S(F(s) - F(t))(\ast \Pi(\{s\}))| < \frac{5}{k}$ where $F(s) = \ast f(s)$ for every $s \in S$.

The family $\mathcal{B} = \{\phi^\phi_f : f \in \mathcal{C}, k \in \mathbb{N}\}$ is finitely satisfiable. By saturation, there exists $S_X$ such that every formula in $\mathcal{B}$ is satisfied simultaneously. Such $(S_X, \{U_j : j \leq K\})$ is the desired hyperfinite representation. □
The hyperfinite Dirichlet form with $H$ of an internal function $F: S_X \to {^*\mathbb{R}}$ is defined to be
\[
\mathcal{F}_H(F, F) = \frac{1}{2} \sum_{s,t \in S_X} |F(s) - F(t)|^2 H_s(\{t\}) \Pi(\{s\}).
\] (4.29)

The result in this section shows that, under moderate conditions, we can use hyperfinite Dirichlet form to approximate the standard Dirichlet form.

5. Nonstandard Characterization of Density Functions

In this section, we study density functions of transition kernels and stationary distributions via studying their corresponding hyperfinite counterparts. The materials presented in this section are closely related to [Zim05]. We start by introducing the notion of S-integrability from nonstandard analysis.

**Definition 5.1.** Let $(\Omega, A, P)$ be an internal probability space and let $F: \Omega \to {^*\mathbb{R}}$ be an internally integrable function such that $\text{st}(F)$ exists $\mathcal{T}$-almost surely. Then $F$ is S-integrable with respect to $P$ if $\text{st}(F)$ is $\mathcal{T}$-integrable, and $\int F \, dP \approx \int \text{st}(F) \, d\mathcal{T}$.

The following theorem provides several ways to verify the S-integrability of an internal function $F$.

**Theorem 5.2** ([ACH97, Theorem 4.6.2]). Suppose $(\Omega, A, P)$ is an internal probability space, and $F: \Omega \to {^*\mathbb{R}}$ is an internally integrable function such that $\text{st}(F)$ exists $\mathcal{T}$-almost surely. Then the following are equivalent:

1. $\text{st}(\int |F| \, dP)$ exists and it equals to $\lim_{n \to \infty} \text{st}(\int |F_n| \, dP)$ where for $n \in \mathbb{N}$, $F_n = \min\{F, n\}$ when $F \geq 0$ and $F_n = \max\{F, -n\}$ when $F \leq 0$.
2. For every infinite $K > 0$, $\int_{|F| > K} |F| \, dP \approx 0$.
3. $\text{st}(\int |F| \, dP)$ exists, and for every $B$ with $P(B) \approx 0$, we have $\int_B |F| \, dP \approx 0$.
4. $F$ is S-integrable with respect to $P$.

For the rest of this section, let $Y$ be a compact metric space and let $\{A_i : 1 \leq i \leq T\}$ be a hyperfinite partition of $^*Y$ such that the diameter for $A_i$ is infinitesimal for every $i \leq T$.

**Definition 5.3.** Let $P, Q$ be two probability measures on $(Y, \mathcal{B}[Y])$. $Q$ dominates $P$ if for each $\epsilon > 0$, there exists $\delta > 0$ such that $P(C) < \epsilon$ for all $C \in \mathcal{B}[Y]$ with $Q(C) < \delta$. Let $\mathcal{P}$ be a family of probability measures on $(Y, \mathcal{B}[Y])$. Then $\mathcal{P}$ is uniformly dominated by $Q$ if, for every $\epsilon > 0$, there exists $\delta > 0$ such that $P(C) < \epsilon$ for all $P \in \mathcal{P}$ and all $C \in \mathcal{B}[Y]$ with $Q(C) < \delta$.

Using the nonstandard characterization of domination, $Q$ dominates $P$ if and only if $^*Q(A) \approx 0$ implies $^*P(A) \approx 0$ for $A \in ^*\mathcal{B}[Y]$. We now mimic proofs in [Zim05] to give an useful nonstandard characterization of density functions.

**Lemma 5.4** ([Zim05, Lemma 2.1]). Let $P, Q$ be two probability measures on $(Y, \mathcal{B}[Y])$. Suppose $P$ is dominated by $Q$. Then the function $\phi: ^*Y \to {^*\mathbb{R}}$ defined by
$$
\phi(x) = \sum_{i=1}^{T} \frac{^*P(A_i)}{^*Q(A_i)} I_{A_i}(x)
$$
(5.1)
is S-integrable with respect to $Q$. 

Proof. It is straightforward to verify that \( \int_Y \phi(x)^*Q(dx) = 1 \). As \( \phi(x) \geq 0 \), we know that \( \phi(x) \in \text{NS}(\mathbb{R}) \) for \( \overline{Q} \)-almost all \( x \in X \). Pick a set \( A \in \mathcal{B}[Y] \) such that \( ^*Q(A) \approx 0 \). By domination, we know that \( ^*P(A) \approx 0 \). Thus, we have

\[
\int_A \phi(x)^*Q(dx) = \sum_{i=1}^{T} ^*P(A \cap A_i) \leq ^*P(A) \approx 0.
\] 

(5.2)

Thus, by Theorem 5.2, \( \phi \) is \( S \)-integrable with respect to \( Q \). \( \square \)

We now show that \( \phi \) defined in Lemma 5.4 is infinitesimally close to the standard Radon-Nikodym derivative.

**Theorem 5.5.** Let \( P, Q \) be two probability measures on \((Y, \mathcal{B}[Y])\). Suppose \( P \) is dominated by \( Q \). Let \( f \) denote the density function of \( P \) with respect to \( Q \). Suppose \( f \) is continuous at \( y_0 \in Y \). Let \( \phi \) be the internal function defined in Eq. (5.1). Then \( \phi(x) \approx f(y) \) for \( x, y \in \text{st}(Y) \) with \( x \approx y \approx y_0 \).

**Proof.** Pick \( A_i \) such that every point in \( A_i \) is infinitely close to \( y_0 \). Note that \( \phi \) is constant on \( A_i \). By the transfer principle, we know that \( ^*P(A_i) = \int_{A_i} \phi(x)^*Q(dx) \) for every \( 1 \leq i \leq T \). We also know that \( ^*P(A_i) = \int_{A_i} \phi(x)^*Q(dx) \). So we have \( \int_{A_i} [f(x) - \phi(x)]^*Q(dx) = 0 \). As \( \phi \) is constant on \( A_i \) and \( f \) is continuous at \( y_0 \), we know that \( f(x) \approx \phi(x) \) for all \( x \in A_i \). By the continuity of \( f \) at \( y_0 \) again, we have \( f(x) \approx \phi(y) \) for all \( x, y \in \text{st}(Y) \) with \( x \approx y \approx y_0 \). \( \square \)

Theorem 5.5 implies that \( \phi(x) = f(\text{st}(x)) \) provided that the density function \( f \) is continuous at \( \text{st}(x) \). In fact, by Lemma 5.4 and Theorem 5.5, it can be shown that \( \text{st}(\phi(x)) \in \text{st}(Y) \) is a density function with respect to \( Q \) ([Zim05, Thm. 3.6]) provided that \( \text{st}(\phi) \) is constant on monads.

### 5.1. Density Functions For Markov Processes

Let \( \{g(x, 1, \cdot)\}_{x \in X} \) be a Markov transition kernel with stationary distribution \( \pi \) on compact metric space \( X \) endowed with Borel algebra \( \mathcal{B}[X] \). Let \( \{S_X, \{B(s) : s \in S_X\}\} \) be a hyperfinite representation of \( X \). We use \( G, H, \Pi \) to denote the corresponding hyperfinite objects as defined in Section 3.

**Assumption 1.** There exists a referencing measure \( \lambda \) on \((X, \mathcal{B}[X])\) such that \( \pi \) is dominated by \( \lambda \) and the family \( \{g(x, 1, \cdot)\}_{x \in X} \) is dominated by \( \lambda \).

We use \( f(x, \cdot) \) to denote a density function of \( g(x, 1, \cdot) \) with respect to \( \lambda \) and use \( k(\cdot) \) to denote the density function of \( \pi \) with respect to \( \lambda \).

**Theorem 5.6.** Suppose Assumption 1 holds. Suppose \( k \) is continuous at \( x_1 \in X \) and \( f \) is jointly continuous at \((x_1, x_2)\) for some \( x_2 \in X \). Then, for \( i, j \in S_X \) such that \( i \approx x_1 \) and \( j \approx x_2 \), we have \( \frac{\Pi(i)}{\pi(B(i))} \approx ^*k(i) \) and \( \frac{G(i)}{\pi(B(i))} \approx ^*f(i, j) \approx \frac{H(i)}{\pi(B(i))} \).

**Proof.** Clearly \( \{B(s) : s \in S_X\} \) is a hyperfinite partition of \( ^*X \) such that each \( B(s) \) has infinitesimal diameter. As \( k(\cdot) \) is continuous at \( x_1 \), by Theorem 5.5, we have

\[
\frac{^*\pi(B(i))}{^*\lambda(B(i))} = \frac{\Pi(i)}{\pi(B(i))} \approx ^*k(i)
\]

(5.3)

for all \( i \approx x_1 \).
Pick \(s, j \in S\) such that \(s \approx x_1\) and \(j \approx x_2\). By the transfer principle, we know that

\[
G_\ast(\{j\}) = g(s, 1, B(j)) = \int_{B(j)} f(s, x) \lambda(dx). \tag{5.4}
\]

On the other hand, we also know that

\[
G_\ast(\{j\}) = g(s, 1, B(j)) = \int_{B(j)} \frac{G_\ast(\{j\})}{\lambda(B(j))} \lambda(dx). \tag{5.5}
\]

As \(f\) is jointly continuous at \((x_1, x_2)\), we know that \(f(s, x) \approx f(s, y) \approx f(st(s), st(j))\) for all \(x, y \in B(j)\). This implies that \(\frac{G_\ast(\{j\})}{\lambda(B(j))} \approx f(s, j)\).

If \(\pi(B(j)) = 0\), then

\[
H_\ast(\{j\}) = G_\ast(\{j\}) \approx f(s, j). \tag{5.6}
\]

If \(\pi(B(j)) \neq 0\), by the joint continuity of \(f\) at \((x_1, x_2)\), we have

\[
H_\ast(\{j\}) = \int_{B(s)} \frac{\pi(x, 1, B(j))}{\lambda(B(j))} \pi(dx) \tag{5.7}
\]

\[
\approx \int_{B(s)} \frac{f(x, j)}{\pi(B(s))} \pi(dx) \approx f(s, j). \tag{5.8}
\]

Thus, we have the desired result.

\[\Box\]

### 6. Comparison Theorem for General Markov Processes

Our main motivation for constructing a hyperfinite representation of the Dirichlet form is to allow us to directly translate theorems about transition kernels on finite-dimensional spaces to theorems on more general state space. As an illustration, we provide a simple translation of a “comparison” theorem for Markov chains on finite state spaces (see the original Theorem 6.1 and our generalization Theorem 6.10). We also give a quick explanation, including an example, as to why direct translation does not work well for this problem. These bounds will be more fully developed in our companion paper focused on comparison.

#### 6.1. Original Comparison Bound.

We introduce comparison methods, following the same presentation as in [LPW09b]. For a finite reversible transition kernel \(\{P(x, y)\}_{x, y \in \Omega}\) with stationary distribution \(\pi\), let \(E = \{(x, y) : P(x, y) > 0\}\). An \(E\)-path from \(x\) to \(y\) is a sequence \(\Gamma = (e_1, e_2, \ldots, e_n)\) of edges in \(E\) such that \(e_1 = (x, x_1), e_2 = (x_1, x_2), \ldots, e_n = (x_{n-1}, y)\) for some vertices \(x_1, x_2, \ldots, x_{n-1} \in \Omega\). We use \(|\Gamma|\) to denote the length of a path \(\Gamma\) and let \(Q(x, y) = \pi(x)P(x, y)\) for \(x, y \in \Omega\).

Let \(P\) and \(\tilde{P}\) be two finite reversible transition kernels with stationary distribution \(\pi\) and \(\tilde{\pi}\), respectively. Suppose we can fix an \(E\)-path from \(x\) to \(y\) for every \((x, y) \in \tilde{E}\) and denote it by \(\gamma_{xy}\). Given such a choice of path, we define the congestion ratio by

\[
B = \max_{e \in E} \left( \frac{1}{Q(e)} \sum_{(x, y) : e \in \gamma_{xy}} \tilde{Q}(x, y)|\gamma_{xy}| \right). \tag{6.1}
\]
Theorem 6.1 ([LPW09b, Thm. 13.23]). Let $P$ and $\tilde{P}$ be finite reversible transition kernels with stationary distributions $\pi$ and $\bar{\pi}$ respectively. If $B$ is the congestion ratio for a choice of $E$-paths, as defined in Eq. (6.14), then
\[ E_P(f, f) \leq B E_P(f, f). \] (6.2)

The aim of this section is to extend Theorem 6.1 to a large class of general Markov Processes.

6.2. Non-Example: Barbell Graph. Before giving our new theorem, we give a quick example that illustrates why directly mimicking the form of the theorem for finite Markov chains will not work well. To understand the basic issue, note that we can view the bound in Theorem 6.1 as coming from three terms: the length $|\gamma_{xy}|$ of a path, the relative probability $\frac{\partial \pi(x, y)}{\partial \pi(x)}$ of transitions along the path, and the congestion $|\{(x, y) : e \in \gamma_{xy}\}|$. The length makes sense as written for general state spaces, and the relative probability can easily be “translated” to e.g. the Radon-Nikodym derivative. However, it is not clear how to translate the notion of congestion, which measures how many paths go exactly through an edge. The problem is that it is often simple to very slightly deform paths in continuous state spaces so that the congestion remains extremely small (or even 1), even though a very large number of paths are extremely close to an edge.

We give a concrete example. Denote by $K_n$ the usual complete graph on $n$ vertices and let $G_n = (V_n, E_n)$ be the “barbell” graph obtained by taking two copies $R_n, L_n$ of $K_n$ and adding a single edge; denote the endpoints of this new edge by $r_n \in R_n$ and $\ell_n \in L_n$. Denote by $\{X_t\}$ the usual simple random walk on this graph, with kernel $A_n$. We study the following natural collection of paths:

1. For $x, y \in L_n$ (or $x, y \in R_n$), simply take the edge from $x$ to $y$.
2. For $x \in L_n$ and $y \notin L_n$, take the edge $(x, \ell_n)$ then the edge $(\ell_n, r_n)$ and then the edge $(r_n, y)$. At the end, delete any self-edges that may appear.
3. For $x \in R_n$ and $y \notin R_n$, do the obvious analogue to path (2).

It is straightforward to check that:

1. All paths are of length at most 3,
2. All transition probabilities are either $n^{-1}$ or $(n - 1)^{-1}$ (so always roughly $n^{-1}$),
3. The stationary measure at any point is either $\frac{n}{2(n - 1)^2 + n}$ or $\frac{n - 1}{2(n - 1)^2 + n}$ (so again roughly $n^{-1}$), and
4. The edge $(r_n, \ell_n)$ appears in roughly $n^2$ paths, which is (vastly) more than any other edge.

Putting together these estimates and comparing to the kernel that simply takes i.i.d. samples from the stationary measure, Theorem 6.1 tells us that the relaxation time of $A_n$ is $O(n^2)$ (and so the mixing time is $O(n^2 \log(n))$). In fact it is straightforward to check that both the relaxation and mixing times are $\Theta(n^2)$. Sketching the relevant arguments: the matching lower bound on the relaxation time can be obtained from Cheeger’s inequality; upper and lower bounds on the mixing time can be obtained by checking that it takes $\Theta(n^2)$ steps to travel from $L_n \setminus \{\ell_n\}$ to $R_n \setminus \{r_n\}$ on average (and, to obtain the upper bound, a straightforward coupling argument).

We now consider a continuous analogue. Define $S_n = \cup_{v \in V_n} I_v$, where $I_v$ are disjoint copies of $[0, 1]$. For $v \in V_n$ and $y \in I_v \subset S_n$, write $V(y) = v$ for the vertex
We note that, ignoring a finite number of points, $\phi$ will consider this process on $S_n$ with the obvious analogues to the last two equations when $y \in V_n \setminus \{\ell_n, r_n\}$. Let $I_{\ell_n} \equiv \{0, 1\}$. In particular, this means it has the same mixing time and relaxation times as $\{X_t\}$.

Attempting to confirm this with a naive version of Theorem 6.1 runs into an immediate problem. Roughly speaking, the problem comes from the fact that on continuous spaces we can “split” the heavily-used vertices $\ell_n, r_n$ into many pieces, and this allows paths travelling between these vertices to avoid each other. Making this split precise requires some additional notation; the following somewhat-heavy notation is reused in Section 6.4.

Let $N_{L,n} : L_n \setminus \{\ell_n\} \mapsto \{1, 2, \ldots, n-1\}$ be an arbitrary ordering of the vertices in $L_n \setminus \{\ell_n\}$, and similarly let $N_{R,n}$ be an ordering of $R_n \setminus \{r_n\}$. For $u \in L_n \setminus \{\ell_n\}$, let $I_{\ell_n}(u)$ be a copy of $[0, 1]$ and for $t \in [0, 1]$ let $I_{\ell_n}(u, t) = t \in I_{\ell_n}(u)$. Let $I_{r_n}$ be the obvious analogous construction on $R_n$. Define

$$S_n' = \left( \bigcup_{v \in V_n \setminus \{\ell_n, r_n\}} I_{\ell_n} \right) \cup \left( \bigcup_{v \in L_n \setminus \{\ell_n\}} I_{\ell_n}(v) \right) \cup \left( \bigcup_{v \in R_n \setminus \{r_n\}} I_{r_n}(v) \right),$$

that is, we take $S_n$ and split the interval $I_{\ell_n}$ into $n-1$ intervals $\cup_{v \in L_n \setminus \{\ell_n\}} I_{\ell_n}(v)$, and make the analogous split for $I_{r_n}$. Define the map $\phi : S_n' \mapsto S_n$ by the following equations:

$$\phi(y) = y, \quad V(y) \in V_n \setminus \{\ell_n, r_n\},$$

$$V(\phi(y)) = \ell_n, \quad y \in \bigcup_{v \in L_n \setminus \{\ell_n\}} I_{\ell_n}(v),$$

$$|\phi(y)| = \frac{|y|}{n-1} + \frac{N_n(v) - 1}{n-1}, \quad y \in L_n(v),$$

with the obvious analogues to the last two equations when $y \in \bigcup_{v \in R_n \setminus \{r_n\}} I_{r_n}(v)$. We note that, ignoring a finite number of points, $\phi$ is a bijection between $S_n$ and $S_n'$; ignoring events of measure 0, $Y'_t \equiv \phi^{-1}(Y_t)$ defines a Markov chain with the same relaxation and mixing times as $\{Y_t\}$. For the remainder of this section we will consider this process on $S_n'$, and write $Q_n'$ for its kernel.

We now consider the following path from $x \in I_x$ to $y \in I_y$:

1. If $V(x), V(y) \in L_n$ (or $V(x), V(y) \in R_n$), simply take the edge from $x'$ to $y'$.
2. For $V(x) \in L_n$ and $V(y) \notin L_n$, take the edge from $x' \in I_x$ to $x'' \equiv I(\ell_n, |x'|) \in I_{\ell_n}$, then the edge from $x''$ to $y'' \equiv I(r_n, |y'|)$, and then the edge from $y''$ to $y'$. At the end, delete any self-edges that may appear.
3. For $V(x) \in R_n$ and $V(y) \notin R_n$, do the obvious analogue to path (2).

We again try to analyze these paths. All of the earlier comments apply, except that no edges are used more than $O(1)$ times! This occurs because we have managed to “clone” $\ell_n, r_n$ into effectively $(n-1)$ vertices, resulting in effectively $(n-1)^2$ edges between them; this exactly matches the $\Theta(n^2)$ congestion of the original discrete path. Thus, applying the formula in Theorem 6.1 naively would give a relaxation time that is $O(1)$, which is not correct.
As discussed at the start of the section, the problem is that the formula of Theorem 6.1 counts how many paths exactly reuse an edge, while the current continuous example merely compresses edges without exactly reusing them. Earlier work Yuen [Yue00; Yue02] adjusted for this compression by associating a Jacobian to its paths. Unfortunately, this approach seems to be quite restrictive. The obvious problem is that this approach requires paths to be sufficiently smooth. A less-obvious problem is that the method often gives very poor bounds. This happens even when (as in our barbell example) the continuous chain of interest has a nearly-identical continuous analogue for which comparison inequalities give nearly-sharp answers; see the original papers [Yue00; Yue02] for examples of this phenomenon.

In section 6.4, we analyze \( Q_n \) using our new theorem (and the same path) and obtain the correct answer.

6.3. Hyperfinite Approximation of Congestion Ratio. In this section, we will consider Markov chains on a state space of the form \( X = \bigcup_{i=1}^{k} X_i \), where \( X_i \) is of the form \([0,1]^{k_i}\) and the notation “\( \bigcup \)” is used to emphasize the fact that these are disjoint copies of what may be the same space; we use the word “component” to refer to such a copy, so that e.g. \( X_1 \) is one component of \( X \). We now define a metric \( d \) on \( X \) as following:

1. For every \( i \in \{1,2,\ldots,I\} \) and every \( x,y \in X_i \), define \( d(x,y) \) to be the standard Euclidean distance between \( x \) and \( y \).
2. For \( x \in X_{i_1} \) and \( y \in X_{i_2} \) where \( i_1 \neq i_2 \), define \( d(x,y) = 1 + \max_{i \in I} \sqrt{k_i} \).

It is easy to verify that \( d \) is a metric on \( X \) with the following property:

\[
\inf_{i \neq j \in I} \inf_{x \in X_i, y \in X_j} d(x,y) > \sup_{i \in I} \sup_{x,y \in X_i} d(x,y). \tag{6.4}
\]

In other words, every point in one “part” \( X_i \) is closer to every other point in the same part than it is to any point in \( X \) in any other part \( X_j \). We will always use the completion of the Borel \( \sigma \)-algebra on \( X \) associated with \( d \).

We start by constructing a specific hyperfinite representation \( S \) of \( X \). Pick an infinite \( K \in \mathbb{N} \) and let \( \delta t = \frac{1}{K} \). Define \( S = \{0, \delta t, 2\delta t, \ldots, 1-\delta t\}, S_i = S^{k_i} \), and \( S_X = \bigcup_{i=1}^{I} S_i \). For every \( 1 \leq i \leq I \) and \( s \in S_i \), \( B_i(s) \subset S_i \) is defined to be the small rectangle with length equals to \( \delta t \) on each side and containing \( s \) as its left lower point. Let \( \lambda_i \) denote the Lebesgue measure on \( X_i \). Then the hyperfinite representations \( (S_i, \{B_i(s)\}_{s \in S_i}) \) are uniform in the sense that \( *\lambda_i(B_i(s_1)) = *\lambda_i(B_i(s_2)) \) for \( s_1, s_2 \in S_i \).

We work with two transition kernels \( \{g(x,1,\cdot)\}_{x \in X} \) and \( \{\tilde{g}(x,1,\cdot)\}_{x \in X} \) on \( X \) with stationary distributions \( \pi \) and \( \tilde{\pi} \), respectively. Denote by \( \lambda \) the natural extension of the Lebesgue measure to \( X \) defined by the following formula: for Lebesgue-measurable \( A_1 \subset X_1, \ldots, A_I \subset X_I \), set

\[
\lambda(A_1 \cup \ldots \cup A_I) = \lambda_1(A_1) + \ldots + \lambda_I(A_I), \tag{6.5}
\]

where \( \lambda_i \) is the usual Lebesgue measure on \( X_i \). We assume that \( \{g(x,1,\cdot)\}_{x \in X}, \{\tilde{g}(x,1,\cdot)\}_{x \in X}, \pi, \tilde{\pi} \) are dominated by \( \lambda \). Let \( k(\cdot), \tilde{k}(\cdot) \) denote the density functions for \( \pi, \tilde{\pi} \), respectively. Let \( f(x,\cdot), \tilde{f}(x,\cdot) \) denote the density functions for \( g(x,1,\cdot), \tilde{g}(x,1,\cdot) \), respectively. We now extend notions of path and edge from finite Markov processes to general Markov processes.
Let \( E = \{(x, y) : f(x, y) > 0\} \). An \( E \)-path from \( x \) to \( y \) is a sequence \((e_1, e_2, \ldots, e_n)\) of edges in \( E \) such that \( e_1 = (x, x_1), e_2 = (x_1, x_2), \ldots, e_n = (x_{n-1}, y) \) for some vertices \( x_1, x_2, \ldots, x_{n-1} \in X \). Let \( Q(x, y) = k(x)f(x, y) \) for \( x, y \in \Omega \). \( \tilde{E} \) and \( \tilde{Q} \) are defined similarly. For \((x, y) \in X^2\), we assume that there exists an \( E \)-path connecting \( x \) and \( y \). We fix such an \( E \)-path and denote it by \( \gamma_{xy} \). Let \( \Theta = \{\gamma_{xy} : (x, y) \in X^2\} \) be the collection of all such paths. We are, of course, really interested in \( \tilde{\Theta} = \{\gamma_{xy} : (x, y) \in \tilde{E}\} \). Note that we can view \( \gamma \) as a function from \( X^2 \) to \( \Theta \). The nonstandard counterparts of \( E, Q, \tilde{E}, \tilde{Q}, \Theta, \tilde{\Theta} \) are defined as nonstandard extensions of these objects. By the transfer principle, \( \ast \gamma \) is a function from \( \ast X^2 \) to \( \ast \Theta \). To associate a nonstandard edge (path) with a standard edge (path), we need to define the distance between edges and paths. Abusing notation slightly, we write:

**Definition 6.2.** Let \( e_1 = (x_1, y_1) \) and \( e_2 = (x_2, y_2) \) be two edges in \( E \) or \( \tilde{E} \). The distance between \( e_1, e_2 \) is defined to be \( W(e_1, e_2) = \max\{d(x_1, x_2), d(y_1, y_2)\} \).

Let \( \gamma_{a_1 b_1} \) and \( \gamma_{a_2 b_2} \) be two paths in \( E \). The distance between \( \gamma_{a_1 b_1} \) and \( \gamma_{a_2 b_2} \) is defined to be \( W(\gamma_{a_1 b_1}, \gamma_{a_2 b_2}) = \max\{d(a_1, a_2), d(b_1, b_2)\} \) - that is, on paths \( W \) is a pseudometric on paths which only pays attention to the endpoints.

The distance between nonstandard edges and paths can be defined similarly. Two nonstandard edges (paths) are infinitely close to each other if and only if both their start and end points are infinitely close to each other. We write \( e_1 \approx e_2 \) (respectively \( \gamma_1 \approx \gamma_2 \)) if \( \ast W(e_1, e_2) \approx 0 \) (respectively \( \ast W(\gamma_1, \gamma_2) \approx 0 \)) for two nonstandard edges (paths) \( e_1, e_2, \gamma_1, \gamma_2 \).

We shall use \( \{G_i(\cdot)\}_{i \in S_X}, \{G_i(\cdot)\}_{i \in S_X}, \{H_i(\cdot)\}_{i \in S_X}, \{\tilde{H}_i(\cdot)\}_{i \in S_X}, \Pi(\cdot) \) and \( \tilde{\Pi}(\cdot) \) to denote various corresponding hyperfinite objects defined in Section 3. We now define the hyperfinite counterparts of \( E, Q, \tilde{E}, \tilde{Q}, F \). Let \( \mathcal{E} = \{(i, j) : H_i(\{j\}) > 0\} \). Note that \( \mathcal{E} \) is a subset of \( S_X \times S_X \). An \( \mathcal{E} \)-path from \( i \) to \( j \) is a sequence \((e_1, e_2, \ldots, e_n)\) of edges in \( \mathcal{E} \) such that \( e_1 = (i, i_1), e_2 = (i_1, i_2), \ldots, e_n = (i_{n-1}, j) \) for some vertices \( i_1, i_2, \ldots, i_{n-1} \in S_X \). Let \( \mathcal{Q}(i, j) = \Pi(\{i\})H_i(\{j\}) \) for \( i, j \in \Omega \). \( \tilde{\mathcal{E}} \) and \( \tilde{\mathcal{Q}} \) are defined similarly. We defer the definition of the collection of hyperfinite paths (hyperfinite counterpart of \( \mathcal{E} \)) to later part of the section. We now introduce the following assumption. To be precise, let \( \mathbb{P}_i \) denote the projection from \( X^2 \) to the \( i \)-th coordinate for \( i \in \{1, 2\} \).

**Assumption 2.** There exist open sets \( \mathcal{O} \subset X^2 \) such that, the density functions \( f(x, y), k(x) \) are bounded away from 0 for all \((x, y) \in \mathcal{O} \) and all \( x \in \mathbb{P}_1(\mathcal{O}) \).

In Section 5.1, Theorem 5.5 implies that division between two hyperfinite probabilities can be approximated by using the open sets and nonstandard extensions of the density functions. We assume that \( \{g(x, 1, \cdot)\}_{x \in X}, \{g(x, 1, \cdot)\}_{x \in X}, \pi, \tilde{\pi} \) are continuous on \( \mathcal{O} \).

**Assumption 3.** Let \( \mathcal{O} \) be the same open set as in Assumption 2. The density function \( f(\cdot, \cdot) \) is jointly continuous on \( \mathcal{O} \) and \( k(\cdot) \) is continuous on \( \mathbb{P}_1(\mathcal{O}) \). The density function \( f(\cdot, \cdot) \) is jointly continuous on \( X^2 \) and \( k(\cdot) \) is continuous on \( X \).

Assumption 2 and Assumption 3 imply the following result.

**Lemma 6.3.** Suppose Assumption 2 and Assumption 3 hold for \( \mathcal{O} \). Then, for \( i_1, i_2, j_1, j_2 \in S_X \) such that \((i_2, j_2) \in \text{st}^{-1}(\mathcal{O}) \), we have

\[
\frac{\tilde{\Pi}(\{i_1\})\tilde{H}_{i_1}(\{j_1\})}{\Pi(\{i_2\})\Pi_{i_2}(\{j_2\})} \approx \frac{\ast k(i_1) \ast f(i_1, j_1)}{\ast k(i_2) \ast f(i_2, j_2)} \approx \frac{\tilde{\Pi}(\{i_1\})G_{i_1}(\{j_1\})}{\Pi(\{i_2\})G_{i_2}(\{j_2\})}.
\]
Proof. As \((i_2, j_2) \in \text{st}^{-1}(\mathcal{O})\), we have \((i, j) \in \ast \mathcal{O}\) and \(i \in \ast \mathcal{P}_1(\ast \mathcal{O})\) for every \((i, j)\) with \((i, j) \approx (i_2, j_2)\). Thus, by Assumption 2 and Theorem 5.6, we know that these fractions are well-defined. Moreover, by Theorem 5.6 again, we have

\[
\frac{\tilde{\Pi}(\{i_1\}) \tilde{H}_{i_1}(\{j_1\})}{\Pi((i_2) \tilde{H}_i(\{j_2\})} = \frac{\tilde{\Pi}(\{i_1\}) \tilde{H}_{i_1}(\{j_1\})}{\Pi((i_2) \tilde{H}_i(\{j_2\})} \frac{\lambda(B(i_1)) \lambda(B(j_1))}{\lambda(B(i_2)) \lambda(B(j_2))} \quad (6.7)
\]

\[
\approx \frac{\ast k(i_1) \ast f(i_1, j_1)}{k(i_1) \ast f(i_2, j_2)} \quad (6.8)
\]

\[
\approx \frac{\tilde{\Pi}(\{i_1\}) \tilde{G}_{i_1}(\{j_1\})}{\Pi((i_2) \tilde{G}_{i_2}(\{j_2\})} \quad (6.9)
\]

We also have the following result from Assumption 2 and Assumption 3.

**Lemma 6.4.** Suppose Assumption 2 holds for \(\mathcal{O}\). Then \((i, j)\) is an element of \(\ast \mathcal{E}\) for every \(i, j \in S_X\) such that \((i, j) \in \ast \mathcal{O}\).

Proof. Suppose \((i, j) \in \ast \mathcal{O}\). By the transfer of Assumption 2, \(\ast f(i, j) > 0\) hence \((i, j) \in \ast \mathcal{E}\). By the construction of the hyperfinite representation, there exist \(A, B \in \ast \mathcal{B}[X]\) such that

1. \(A \subset B(i)\) and \(i \in A\)
2. \(B \subset B(j)\) and \(j \in B\)
3. \(A \times B \subset \ast \mathcal{O}\)
4. \(\ast \lambda(A) > 0\) and \(\ast \lambda(B) > 0\).

Then, by the construction of \(H\) (see Theorem 3.5), we conclude that \(H_i(\{j\}) > 0\) hence \((i, j) \in \mathcal{E}\). □

As we mentioned at the beginning of the section, we fix an \(E\)-path for every \((x, y) \in X^2\). Let \(\mathcal{D} = \{(x, x) : x \in X\}\) denote the collection of diagonal points of \(X^2\). Our choices of paths have been arbitrary so far. We impose the following assumption which ensures that our choices of paths consist of certain edges.

**Assumption 4.** Let \(\mathcal{O}\) be the same open set as in Assumption 2 and let \(\epsilon_0\) be a positive real number. Let

\[
\tilde{E}_{\epsilon_0} = \{(x, y) \in X^2 \setminus \mathcal{D} : \inf_{(a, b) \in \tilde{E}} W((x, y), (a, b)) \leq \epsilon_0\}. \quad (6.10)
\]

There exists a \(\tilde{\mathcal{C}} \subset \mathcal{O}\) such that \(\tilde{\mathcal{C}}\) is closed under the subspace topology of \(X^2 \setminus \mathcal{D}\) and, for every \((x, y) \in \tilde{E}_{\epsilon_0}\), there exists an \(E\)-path \(\gamma_{xy}\) consisting of edges from \(\tilde{\mathcal{C}}\).

Note that \(\text{st}^{-1} (\tilde{\mathcal{C}}) \subset \mathcal{O}\) since \(\tilde{\mathcal{C}}\) is a closed subset of \(\mathcal{O}\). With Assumption 4 holds, for \((x, y) \in \tilde{E}_{\epsilon_0}\), we let the path connecting \(x\) and \(y\) consists of edges from \(\tilde{\mathcal{C}}\). In order to generalize Theorem 6.1 to general Markov processes, we need to make sure that every edge is not contained in too many paths. We impose the following \(\epsilon\)-separated property.

**Assumption 5.** Fix \(\epsilon > 0\). For every edge \(e \in E \setminus \mathcal{D}\), let

\[
S_e = \{\gamma_{xy} \in \Theta : e \in \gamma_{xy}\} \quad (6.11)
\]
be the collection of all paths that contains the edge $e$. Then for any pair $(x_i, y_i), (x_j, y_j)$ such that $\gamma_{x_i y_i}, \gamma_{x_j y_j} \in S_e$, the following statement holds:

\[
((x_i = x_j) \lor (d(x_i, x_j) > \epsilon)) \land ((y_i = y_j) \lor (d(y_i, y_j) > \epsilon)).
\]  

(6.12)

Fix some $\epsilon > 0$. Assumption 5 then implies that, for each standard edge $e \in \mathbb{D}$, there is only finitely many paths from $\Theta$ contain it. It is also true that every nonstandard edge $v \not\in \ast \mathbb{D}$ is contained in finitely many elements of $\ast \Theta$. We assume that there is an uniform bound on the length of all paths.

**Assumption 6.** There exists $R > 0$ such that $|\gamma| \leq R$ for every $\gamma \in \Theta$.

We now impose the following regularity condition on the length of the paths in $F$.

**Assumption 7.** There exists $M > 0$ with the following property: for every $\gamma \in F$, there exists $\delta > 0$ such that, for every $\gamma' \in F$ with $d(\gamma, \gamma') < \delta$, we have $||\gamma| - |\gamma'|| \leq M$.

Assumption 7 implies the following result:

**Lemma 6.5.** Suppose Assumption 7 holds. Let $\ast \gamma_{x_1 y_1}$ and $\ast \gamma_{x_2 y_2}$ be two nonstandard paths in $\ast F$. If $x_1 \approx x_2$ and $y_1 \approx y_2$ (that is, $\ast W(\ast \gamma_{x_1 y_1}, \ast \gamma_{x_2 y_2}) \approx 0$), then $||\ast \gamma_{x_1 y_1}| - |\ast \gamma_{x_2 y_2}|| \leq M$.

We now define the collection of hyperfinite paths which we shall be primarily working with for the rest of the section. For $(i, i) \in \hat{E} \cap \ast \mathbb{D}$, the hyperfinite path $\Gamma_{ii}$ is simply $(i, i)$ which has length 0. For $(i, j) \in \hat{E} \setminus \ast \mathbb{D}$, by the transfer principle, there exists a nonstandard path $\ast \gamma_{ij} \in \ast \Theta$ connecting $i$ and $j$. By Assumption 6, we know that $|\ast \gamma_{ij}| = k$ for some natural number $k \leq R$. Thus, $\ast \gamma_{ij}$ is a sequence of edges in $\ast E$ and we can denote this sequence by $(i, i_1), (i_1, i_2), \ldots, (i_{k-1}, j)$. Note that, as $\hat{H}((\{j\}) > 0$, by the construction of $\hat{H}$ and the fact that $i \not= j$, we know that $(i, j) \in \ast E_{\Theta}$. A hyperfinite path $\Gamma_{ij}$ from $i$ to $j$ is defined by the sequence

\[
(i, s_{i_1}), (s_{i_1}, s_{i_2}), \ldots, (s_{i_{k-1}}, j)
\]

(6.13)

where, for every $1 \leq n \leq k-1$, $s_{in}$ is the unique element in $S_X$ such that $i_n \in B(s_{in})$.

We now verify that $\Gamma_{ij}$ is a well-defined hyperfinite path from $i$ to $j$.

**Lemma 6.6.** Suppose Assumption 2 holds for $\mathcal{O}$, Assumption 4 holds for $\mathcal{C}$. Let $(a, b) \in \{(i, s_{i(1)}), (s_{i(1)}, s_{i(2)}), \ldots, (s_{i(k-1)}, j)\}$. Then, $H(a, b) > 0$.

**Proof.** Let $(a, b) \in \{(i, s_{i_1}), (s_{i_1}, s_{i_2}), \ldots, (s_{i_{k-1}}, j)\}$. By Assumption 4 and the fact that $(i, j) \in \ast \hat{E}_\Theta$, we know that $(a, b) \in \ast \mathcal{C} \subset \ast \mathcal{O}$. By Lemma 6.4, we know that $H_a(\{b\}) > 0$. \hfill $\square$

Thus, $(a, b)$ is an element of $\mathcal{E}$ so $\Gamma_{ij}$ is a well-defined hyperfinite path. Let $\Upsilon = \{\Gamma_{ij} : (i, j) \in \hat{E}\}$ be a fixed collection of hyperfinite paths. By construction, we have $|\Gamma_{ij}| \leq |\ast \gamma_{ij}|$ for $(i, j) \in S_X \times S_X$.

Suppose Assumption 5 (\(\epsilon\)-separated property) holds for some $\epsilon > 0$. Let $v = (s, t) \in \ast E \setminus \ast \mathbb{D}$ for $s, t \in S_X$. By Assumption 5, $v$ is contained in finitely many nonstandard paths (elements of $\ast \Theta$). However, $v$ may be contained in infinitely many hyperfinite paths (elements of $\Upsilon$). This is because, for a nonstandard path $\ast \gamma_{ij}$ with $i, j \in S_X$, if $\ast \gamma_{ij}$ contains a nonstandard edge $e = (x, y)$ where $x \in B(s)$ and $y \in B(t)$, then the hyperfinite path $\Gamma_{ij}$ contains $v$. So we impose the following regularity condition on $\Theta$. 

Assumption 8. There exist $K \in \mathbb{R}_{>0}$ and a continuous function $m : E \setminus \mathbb{D} \to [0, K]$ with the following property: for every $\epsilon > 0$ and standard edges $e, e' \in E \setminus \mathbb{D}$, if $W(e, e') < \epsilon$ and $e \in \gamma \in \Theta$, there exists $\gamma' \in \Theta$ such that $W(\gamma, \gamma') < m(e)\epsilon$ and $e' \in \gamma'$.

By the transfer principle, $\ast m$ is an internal function from $\ast E \setminus \ast \mathbb{D}$ to $\ast \mathbb{R}$. Assumption 8 also implies that $\ast m(e) \approx \ast m(e')$ for $e, e' \in \ast E \setminus \ast \mathbb{D}$ such that $\ast W(e, e') \approx 0$ and $\text{st}(e) \not\in \mathbb{D}$. We have the following result.

Lemma 6.7. Suppose Assumption 8 holds. Let $v = (i, j) \in \ast E \setminus \ast \mathbb{D}$ be a hyperfinite edge that is contained in $n$ nonstandard paths (elements of $\ast \Theta$). Then $v$ is contained in at most $2K'n$ many hyperfinite paths. Suppose that $\text{st}(v) = (\text{st}(i), \text{st}(j)) \in E \setminus \mathbb{D}$. Then $v$ is contained in at most $2(\lceil m(\text{st}(v)) \rceil + 1) \times n$ many hyperfinite paths.

Proof. Recall that each element in $\{B(s) : s \in S_X\}$ is a rectangle with side length $\delta t$. Let $A = \{\gamma_k : k \leq n\}$ denote the collection of nonstandard paths that contain $v$ and let $\Gamma$ be a hyperfinite path that contains $v$. By the construction of hyperfinite path, there exists a nonstandard path $\ast \gamma'$ that corresponds with $\Gamma$. Note that $\ast \gamma'$ has the same start and end points as $\Gamma$ and $\ast \gamma'$ contains some nonstandard edge $e = (x, y) \in \ast E \setminus \ast \mathbb{D}$ where $x \in B(i)$ and $y \in B(j)$. As $\ast W(e, v) < \delta t$, by Assumption 8, there exists $k_0 \leq n$ such that $\ast W(\ast \gamma', \ast \gamma_{k_0}) < \ast m(v)\delta t$. By Assumption 8 and the hypothesis of the theorem, we know that $v$ is contained in at most $2K'n$ many hyperfinite paths. If $\text{st}(v) = (\text{st}(i), \text{st}(j)) \in E \setminus \mathbb{D}$, then $v$ is contained in at most $2(\lceil m(\text{st}(v)) \rceil + 1) \times n$ many hyperfinite paths. □

Lemma 6.8. Suppose Assumption 2 holds for $\mathcal{O}$. Suppose Assumption 5 holds for some $\epsilon > 0$ and Assumption 8 holds. Let $C$ be a subset of $\mathcal{O}$ such that $\ast^{-1}(C) \subset \ast \mathcal{O}$. Let $v = (i, j) \in (S_X \times S_X) \cap \ast^{-1}(C)$ be contained in $\Gamma_{st} \in \Upsilon$ with $i \neq j$. Then $e = (\text{st}(i), \text{st}(j))$ is an element of $E \setminus \mathbb{D}$ and it is contained in $\gamma_{st(\text{st}(i))} \in \Theta$.

Proof. By Lemma 6.4, we have $v \in \ast E \setminus \ast \mathbb{D}$. Note that $e \in C \subset \mathcal{O}$ and $e \not\in \mathbb{D}$. By Assumption 2, we know that $f(\text{st}(i), \text{st}(j)) > 0$ so $e \in E \setminus \mathbb{D}$. Let $\ast \gamma_{st}$ denote the nonstandard path that associates with $\Gamma_{st}$. Then $\ast \gamma_{st}$ contains $v' = (x, y)$ where $x \in B(i)$ and $y \in B(j)$. As $\ast W(v', e) \approx 0$, by Assumption 8, we have $\ast m(v') \approx \ast m(e) = m(e)$ and there exists a nonstandard path $\ast \gamma$ such that $\ast W(\ast \gamma_{st}, \ast \gamma) \approx 0$ and $e \in \ast \gamma'$. By Assumption 5, $e$ is only contained in finitely many paths. By the transfer principle, it must be contained in $\gamma_{st(\text{st}(i))}$. □

Before we present our main result, we introduce one more assumption which asserts that edges with small length can only be contained in paths with close starting and ending points. For $e = (a, b) \in E$, the length of $e$, which we denote by $\|e\|$, is defined to be $\|e\| = d(a, b)$.

Assumption 9. There exists a function $L : \mathbb{R} \to \mathbb{R}$ such that

- $L(x) > 0$ for all $x > 0$,
- $\ast L(x) \approx 0$ if and only if $x \approx 0$,
- For every $\gamma_{ab} \in \Theta$ and every $e \in \gamma_{ab}$, we have $\|e\| \geq L(d(a, b))$.

Note that, for every continuous function $L$ such that $L(0) = 0$ and $L(x) > 0$ for $x > 0$, we know that $\ast L(x) \approx 0$ if and only if $x \approx 0$.

Lemma 6.9. Suppose Assumption 9 holds. Let $v = (s, t)$ be an element of $\mathcal{E}$ such that $s \approx t$. If $\Gamma_{ij} \in \Upsilon$ contains $v$, then $i \approx j$. 


Proof. Let $\Gamma_{ij} \in \Upsilon$ contain $v$. Then the nonstandard path $^{*}\gamma_{ij}$ contains $e = (x, y) \in ^{*}E$ for $x \in B(s)$ and $y \in B(t)$. By Assumption 9 and the transfer principle, we have $|e| \geq ^{*}L(|i - j|)$. As $|e| \approx 0$, we know that $^*L(|i - j|) \approx 0$. By Assumption 9 again, we can conclude that $i \approx j$. \hfill \Box

We are now at the place to establish the main result of this section.

**Theorem 6.10.** Let $\{g(x, 1, \cdot)\}_{x \in X}$ and $\{\tilde{g}(x, 1, \cdot)\}_{x \in X}$ be two reversible transition kernels satisfying (SSF) with stationary distributions $\pi$ and $\tilde{\pi}$, respectively. Let $k(\cdot), \tilde{k}(\cdot)$ denote the density functions for $\pi, \tilde{\pi}$, respectively and let $f(x, \cdot), \tilde{f}(x, \cdot)$ denote the density functions for $g(x, 1, \cdot), \tilde{g}(x, 1, \cdot)$, respectively. Suppose Assumption 2, Assumption 3 hold for some $O$, Assumption 4 holds for some $C$ and $e_0 > 0$, Assumption 5 holds for some $\epsilon > 0$, Assumption 6 holds for some $R > 0$, Assumption 7 holds for some $M > 0$, Assumption 8 holds for some continuous function $m : E \to \mathbb{R}$, and Assumption 9 holds for some $L : \mathbb{R} \to \mathbb{R}$. Let

$$B = \sup_{e \in E} \left( \frac{2([m(e)] + 1)}{Q(e)} \sum_{(x, y) : e \in \gamma_{xy}} \tilde{Q}(x, y)(|\gamma_{xy}| + M) \right). \quad (6.14)$$

Then, for every continuous function $h : X \to \mathbb{R}$, we have

$$\delta_{\tilde{g}}(h, h) \leq B \delta_{g}(h, h). \quad (6.15)$$

Proof. Let $\{H_i(\cdot)\}_{i \in S_X}$ and $\{\tilde{H}_i(\cdot)\}_{i \in S_X}$ be two hyperfinite transition kernels associated with $\{g(x, 1, \cdot)\}_{x \in X}$ and $\{\tilde{g}(x, 1, \cdot)\}_{x \in X}$ as defined in Theorem 3.5. Let $\Pi(\cdot)$ and $\tilde{\Pi}(\cdot)$ be $^{*}$stationary distributions for $\{H_i(\cdot)\}_{i \in S_X}$ and $\{\tilde{H}_i(\cdot)\}_{i \in S_X}$, respectively. By Theorem 3.5, we know that $\{H_i(\cdot)\}_{i \in S_X}$ is $^{*}$reversible with respect to $\Pi$ and $\{\tilde{H}_i(\cdot)\}_{i \in S_X}$ is $^{*}$reversible with respect to $\tilde{\Pi}$. Let

$$B = \max_{v \in \mathbb{E}} \left( \frac{1}{Q(v)} \sum_{(s, t) : v \in \Gamma_{st}} \tilde{Q}(s, t)|\Gamma_{st}| \right) \quad (6.16)$$

be the hyperfinite congestion ratio. By the transfer principle, for every internal function $F : S_X \to ^{*}\mathbb{R}$, we have

$$\mathcal{F}_H(F, F) \leq B \mathcal{F}_\tilde{H}(F, F). \quad (6.17)$$

Pick a continuous function $h_0 : X \to \mathbb{R}$. Define an internal function $F_0 : S_X \to ^{*}\mathbb{R}$ by letting $F_0(s) = ^{*}h_0(s)$ for every $s \in S_X$. By Theorem 4.3, we have $\delta_{\tilde{g}}(h_0, h_0) \approx \mathcal{F}_H(F_0, F_0)$ and $\delta_{g}(h_0, h_0) \approx \mathcal{F}_\tilde{H}(F_0, F_0)$. Thus, to finish the proof, it is sufficient to show that $B \leq B$.

Pick $v = (i, j) \in \mathcal{E}$. Suppose $(B(i) \times B(j)) \cap ^{*}C = \emptyset$. Let $\Gamma_{ab} \in \Upsilon$. As $^*\tilde{H}_a((b)) > 0$, we conclude that $(a, b) \in ^*E_{eq}$. By Assumption 4 and the transfer principle, we know that the nonstandard path $^{*}\gamma_{ab}$ is formed by edges from $^*C$. As $(B(i) \times B(j)) \cap ^{*}C = \emptyset$, we know that $v \notin \Gamma_{ab}$. Thus, we can conclude that

$$\frac{1}{Q(v)} \sum_{(s, t) : v \in \Gamma_{st}} \tilde{Q}(s, t)|\Gamma_{st}| = 0. \quad (6.18)$$

We now suppose that $(B(i) \times B(j)) \cap ^{*}C \neq \emptyset$ but $i \approx j$. By Assumption 5, $v$ is contained in finitely many nonstandard paths. By Lemma 6.7, $v$ is contained in finitely many hyperfinite paths. By Lemma 6.9, any hyperfinite path contains $v$.
has infinitesimal length. Hence, we have
\[
\frac{1}{Q(v)} \sum_{(s,t):v \in \Gamma_{st}} \tilde{Q}(s,t) |\Gamma_{st}| = 0
\] (6.19)
in this case.

We now suppose that \( v \in \text{st}^{-1}(\mathcal{C}) \subset \text{st}^{-1}(\mathcal{O}) \subset \ast \mathcal{O} \). This immediately implies that \( i \neq j \). By Assumption 2, we know that \( \text{st}(\ast f(i,j)) > 0 \) and \( \text{st}(\ast k(i)) > 0 \). Let \( \Gamma_{ab} \) be a hyperfinite path that contains \( v \). Suppose \( \ast f(a,b) \approx 0 \). Then, by Lemma 6.3, we know that \( \frac{\tilde{Q}(a,b)}{Q(v)} \approx 0 \). By Assumption 5 and Lemma 6.7, \( v \) is contained in finitely many hyperfinite paths. Thus, we have
\[
\frac{1}{Q(v)} \sum_{(s,t):v \in \Gamma_{st}} \tilde{Q}(s,t) |\Gamma_{st}| \approx \frac{1}{Q(v)} \sum_{v \in \Gamma_{st} \subset T_v} \tilde{Q}(s,t) |\Gamma_{st}|
\] (6.20)
where \( T_v = \{ \Gamma_{st} : \text{st}(\ast f(s,t)) > 0 \wedge v \in \Gamma_{st} \} \).

By Lemma 6.8, \( e = (\text{st}(i),\text{st}(j)) \) is an element of \( E \setminus D \). It follows from Assumption 5 that \( e \) is contained in finitely many elements of \( F \). Let
\[
T_e = \{ \gamma : e \in \gamma \} = \{ \gamma_{x_1,y_1}, \gamma_{x_2,y_2}, \ldots, \gamma_{x_n,y_n} \}
\] (6.21)
for some \( n \in \mathbb{N} \). By Lemma 6.8 again, we have \( T_v = \bigcup_{i=1}^{n} A_i \) where \( A_i = T_v \cap \{ \Gamma \in \mathcal{Y} : \ast W(\Gamma,\gamma_{x_i,y_i}) \approx 0 \} \). By Lemma 6.7, for every \( 1 \leq i \leq n \), we have \( |A_i| \leq [2m(e) + 1] \). By Lemma 6.3, Assumption 2 and Assumption 3, we have
\[
\frac{1}{Q(v)} \sum_{v \in \Gamma_{st} \subset T_v} \tilde{Q}(s,t) |\Gamma_{st}| \approx \sum_{v \in \Gamma_{st} \subset T_v} \frac{\tilde{Q}(\text{st}(s),\text{st}(t))}{Q(e)} |\Gamma_{st}|.
\] (6.22)
By Lemma 6.5, we have \( |\Gamma_{st}| \leq |\gamma_{\text{st}(s),\text{st}(t)}| + M \). Thus, combining with Eq. (6.20), we have
\[
\sum_{(s,t):v \in \Gamma_{st}} \frac{\tilde{Q}(\text{st}(s),\text{st}(t))}{Q(e)} |\Gamma_{st}| \leq \frac{2([m(e)] + 1)}{Q(e)} \sum_{(x,y):x \in \gamma_{xy}} \tilde{Q}(x,y)(|\gamma_{xy}| + M).
\] (6.23)
Hence, we can conclude that \( B \lesssim B \), proving the desired result. \( \square \)

6.4. **Example: Finishing the Barbell Graph.** We continue from Section 6.2, using the same notation and paths. We will compare the kernel \( Q'_n \) to the kernel that takes i.i.d. samples from the stationary measure of \( Q'_n \). Let \( d \) be as in Section 6.3.

We now quickly check the assumptions of Theorem 6.10:

- Both transition kernels clearly satisfy (DSF) (indeed, in the statement of the condition one can take \( \delta = 0.1 \) for all \( \epsilon \geq 0 \), and thus they also satisfy the weaker condition (SSF).
- Assumption 2 and Assumption 3 hold with \( \mathcal{O} = X^2 \).
- Assumption 4 holds with \( \mathcal{C} \) consisting of all edges of length at most 1, and \( \epsilon_0 = 1 \).
- Assumption 5 holds with any \( 0 < \epsilon < 1 \).
- Assumption 6 holds with \( R = 3 \).
- Assumption 7 holds trivially with \( M = R - 1 = 2 \).
- Assumption 8 holds for any \( 0 < \epsilon < 1 \) with \( m(\epsilon) \equiv (n + 1) \).
- Assumption 9 holds with \( L(x) \equiv x \).
Applying these bounds to the formula $B$ appearing in Theorem 6.10, we find $B = \Theta(n^2)$; this gives the correct dependence on $n$ for the relaxation time.

References

[ACH97] L. O. Arkeryd, N. J. Cutland, and C. W. Henson, eds. Nonstandard analysis. Vol. 493. NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences. Theory and applications. Kluwer Academic Publishers Group, Dordrecht, 1997, pp. xiv+366.

[ADS18] R. M. Anderson, H. Duanmu, and A. Smith. “Mixing Times and Hitting Times for General Markov Processes”, submitted. 2018.

[ADS19a] R. M. Anderson, H. Duanmu, and A. Smith. “Mixing Times and Average Mixing Times for General Markov Processes” (2019). Submitted.

[ADS19b] R. M. Anderson, H. Duanmu, and A. Smith. “Mixing Times and Hitting Times for Gibbs Samplers and other Non-Feller Processes” submitted. 2019.

[And76] R. M. Anderson. “A non-standard representation for Brownian motion and Itô integration”. Israel J. Math. 25.1-2 (1976), pp. 15–46. ISSN: 0021-2172.

[AS03] S. Albeverio and W. Schachermayer. Lectures on probability theory and statistics: ecole d’été de probabilités de Saint-Flour XXX-2000. Springer Science & Business Media, 2003.

[BD58] A. Beurling and J. Deny. “Espaces de Dirichlet. I. Le cas élémentaire”. Acta Mathematica 99.1 (1958), pp. 203–224.

[Cut+95] N. J. Cutland, V. Neves, F. Oliveira, and J. Sousa-Pinto, eds. Developments in nonstandard mathematics. Vol. 336. Pitman Research Notes in Mathematics Series. Papers from the International Colloquium (CIMNS94) held in memory of Abraham Robinson at the University of Aveiro, Aveiro, July 18–22, 1994. Longman, Harlow, 1995, pp. x+260.

[Dav89] E. B. Davies. Heat Kernels and Spectral Theory. Cambridge Tracts in Mathematics. Cambridge University Press, 1989.

[Doo12] J. L. Doob. Classical potential theory and its probabilistic counterpart: Advanced problems. Vol. 262. Springer Science & Business Media, 2012.

[DR18] H. Duanmu and D. M. Roy. “On extended admissible procedures and their nonstandard Bayes risk”. The Annals of Statistics, under revision. 2018.

[DRS17] H. Duanmu, D. M. Roy, and A. Smith. “Existence of matching priors on compact spaces yielding confidence intervals”. In preparation. 2017.

[DRW18] H. Duanmu, J. Rosenthal, and W. Weiss. “Ergodicity of Markov processes via non-standard analysis”. Memoirs of the American Mathematical Society, to appear. 2018.

[DSC93] P. Diaconis and L. Saloff-Coste. “Comparison theorems for reversible Markov chains”. The Annals of Applied Probability (1993), pp. 696–730.

[Fuk96] M Fukushima. “Dirichlet forms and symmetric Markov processes”. Bull. Amer. Math. Soc 33 (1996), pp. 87–92.

[Kei84] H. J. Keisler. “An infinitesimal approach to stochastic analysis”. Mem. Amer. Math. Soc. 48.297 (1984), pp. x+184.
[Kus89] S. Kusuoka. “Dirichlet forms on fractals and products of random matrices”. *Publications of the Research Institute for Mathematical Sciences* 25.4 (1989), pp. 659–680.

[LPW09a] D. Levin, Y. Peres, and E. Wilmer. *Markov Chains and Mixing Times*. Providence, Rhode Island: American Mathematical Society, 2009, pp. xi+371. ISBN: 978-0-8218-4739-8.

[LPW09b] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. With a chapter by James G. Propp and David B. Wilson. American Mathematical Society, Providence, RI, 2009, pp. xviii+371. ISBN: 978-0-8218-4739-8.

[MN16] U. K. Müller and A. Norets. “Coverage Inducing Priors in Nonstandard Inference Problems”. *Journal of the American Statistical Association* 111.515 (2016), pp. 1233–1241.

[MR12] Z.-M. Ma and M. Röckner. *Introduction to the theory of (non-symmetric) Dirichlet forms*. Springer Science & Business Media, 2012.

[PS15] Y. Peres and P. Sousi. “Mixing times are hitting times of large sets”. *J. Theoret. Probab.* 28.2 (2015), pp. 488–519. ISSN: 0894-9840.

[Sch99] B. Schmuland. “Dirichlet forms: some infinite-dimensional examples”. *Canad. J. Statist.* 27.4 (1999), pp. 683–700. ISSN: 0319-5724.

[WL00] M. Wolf and P. A. Loeb, eds. *Nonstandard analysis for the working mathematician*. Vol. 510. Mathematics and its Applications. Kluwer Academic Publishers, Dordrecht, 2000, pp. xiv+311.

[Yue00] W. K. Yuen. “Applications of geometric bounds to the convergence rate of Markov chains on $\mathbb{R}^n$”. *Stochastic processes and their applications* 87.1 (2000), pp. 1–23.

[Yue02] W. K. Yuen. “Generalization of discrete-time geometric bounds to convergence rate of Markov processes on $\mathbb{R}^n$”. *Stochastic models* 18.2 (2002), pp. 301–331.

[Zim05] G. B. Zimmer. “A unifying Radon-Nikodym theorem through nonstandard hulls”. *Illinois J. Math.* 49.3 (2005), pp. 873–883. ISSN: 0019-2082.

University of California, Berkeley, Department of Economics
University of California, Berkeley, Department of Economics
University of Ottawa, Department of Mathematics and Statistics