INSTITUTE OF MATHEMATICS, POLISH ACADEMY OF SCIENCES

DISSERTATIONES MATHEMATICAE

514

THOMAS KAIJSER

On convergence in distribution of the Markov chain generated by the filter kernel induced by a fully dominated Hidden Markov Model

WARSZAWA 2016

Thomas Kaijser
Department of Mathematics
Linköping University
SE-581 83 Linköping, Sweden
E-mail: thomas.kaijser@liu.se

Published by the Institute of Mathematics, Polish Academy of Sciences
Typeset using TEX at the Institute
Printed and bound in Poland by HermanDruK, Warszawa
Nakład 200 egz.

Abstracted/Indexed in: Mathematical Reviews, Zentralblatt MATH, Science Citation Index Expanded, Journal Citation Reports/Science Edition, Google Science, Scopus, EBSCO Discovery Service.

Available online at **http://journals.impan.pl**

DOI: 10.4064/dm739-9-2015

ISSN 0012-3862

Contents

Abstract

Consider a Hidden Markov Model (HMM) such that both the state space and the observation space are complete, separable, metric spaces and for which both the transition probability function (tr.pr.f.) determining the hidden Markov chain of the HMM and the tr.pr.f. determining the observation sequence of the HMM have densities. Such HMMs are called fully dominated. In this paper we consider a subclass of fully dominated HMMs which we call regular.

A fully dominated, regular HMM induces a tr.pr.f. on the set of probability density functions on the state space which we call the *filter kernel* induced by the HMM and which can be interpreted as the Markov kernel associated to the sequence of conditional state distributions.

We show that if the underlying hidden Markov chain of the fully dominated, regular HMM is strongly ergodic and a certain coupling condition is fulfilled, then, in the limit, the distribution of the conditional distribution becomes independent of the initial distribution of the hidden Markov chain and, if also the hidden Markov chain is uniformly ergodic, then the distributions tend towards a limit distribution.

In the last part of the paper, we present some more explicit conditions, implying that the coupling condition mentioned above is satisfied.

Acknowledgements. I want to thank Sten Kaijser, Fuzhou Gong, Lucáš Malý and Svante Janson for valuable discussions and support. I also thank the referee for valuable comments.

2010 *Mathematics Subject Classification*: Primary 60J05; Secondary 60F05.

Key words and phrases: hidden Markov models, filtering processes, Markov chains on nonlocally compact spaces, Kantorovich distance, barycenter.

Received 4 February 2015; revised 24 September 2015.

Published online 21 March 2016.

1. Introduction

1.1. The aim of the paper. Let $S = \{1, \ldots, d\}$ be a finite set and P a transition probability matrix (tr.pr.m.) on S. Let $A = \{1, \ldots, k\}$ be another finite set and let R be a tr.pr.m. from S to A. The set $\{S, P, A, R\}$ is often called a *Hidden Markov Model* (*HMM*); later we shall also refer to such a set as a "classical HMM". We call S the *state space*, P the *state transition probability matrix*, A the *observation space* and R the *observation matrix*.

Associated to a classical HMM and an initial distribution p_0 on S there are two stochastic processes: the hidden Markov chain and the observation sequence. A simple way to obtain these is as follows: First define a tr.pr.m. M on the product space $S \times A$ simply by

$$(M)_{(i,a),(j,b)} = (P)_{i,j}(R)_{j,b}, \quad i, j \in S,\ a, b \in A.$$

REMARK 1.1. Here and throughout, if M is a matrix, we let $(M)_{i,j}$ denote its i, jth entry. Similarly, if x is a vector, we let $(x)_i$ denote the ith component.

Next, let q_0 denote an arbitrary probability vector on A, let $\tilde{p}_0 = p_0 \otimes q_0$ be the product measure of p_0 and q_0, and let $\{(X_n, Y_n),\ n = 0, 1, 2, \ldots\}$ denote the bivariate Markov chain generated by the initial distribution $\tilde{p}_0$ and the tr.pr.m. M. It is easily seen that the sequence $\{X_n,\ n = 0, 1, 2, \ldots\}$ is a Markov chain itself with tr.pr.m. P and initial distribution p_0. It is usually called the *hidden Markov chain*. The sequence $\{Y_n,\ n = 1, 2, \ldots\}$ is usually called the *observation sequence*.

Associated to a classical HMM there is a third important stochastic process defined as follows. For each $i \in S$ and every integer $n \geq 1$ define

$$Z_{n,i} = \Pr[X_n = i \mid Y_1, \ldots, Y_n], \tag{1.1}$$

$$Z_n = (Z_{n,i},\ 1 \leq i \leq d). \tag{1.2}$$

Clearly Z_n is a random probability vector on the finite set S. The sequence $\{Z_n,\ n = 1, 2, \ldots\}$ is often called either the *sequence of conditional state distributions* or the *filtering process*.

It is well-known that $\{Z_n,\ n = 1, 2, \ldots\}$ is also a Markov chain. Let $\mathbf{P}$ denote the transition probability function (tr.pr.f.) for this chain. We shall call $\mathbf{P}$ the *filter kernel* induced by the HMM $\{S, P, A, R\}$.

A natural question is under which conditions the Markov chain generated by $\mathbf{P}$ is ergodic, in the sense that it is an aperiodic Markov chain such that its distributions tend to a unique limit distribution, which is independent of the initial distribution p_0. One might conjecture that if the tr.pr.f. P of the hidden Markov chain is both irreducible and aperiodic, then the answer to the above question is affirmative, but this is not true; some extra condition is needed.

Next we introduce some notation. For $x \in \mathbb{R}^d$, let $\|x\|$ denote the l_1-norm, let $K = \{x = (x_1, \dots, x_d) : x_i \geq 0, \|x\| = 1\}$, for $x, y \in K$ define $\delta(x, y) = \|x - y\|$, let $\mathcal{O}$ denote the topology on K induced by δ, let $\mathcal{E}$ denote the σ-algebra on K induced by δ, let $C[K]$ denote the set of continuous functions on K with respect to $\mathcal{O}$, and let $\mathcal{P}(K, \mathcal{E})$ denote the set of probability measures on $(K, \mathcal{E})$.

To each $a \in A$ we now associate a $d \times d$ matrix $M(a)$ by

$$(M(a))_{i,j} = (P)_{i,j}(R)_{j,a}. \tag{1.3}$$

Following [30], we call $M(a)$ the *stepping matrix* determined by $a \in A$.

It is now easy to describe the tr.pr.f. $\mathbf{P} : K \times \mathcal{E} \to [0, 1]$ associated to the sequence of conditional state distributions. We have

$$\mathbf{P}(x, E) = \sum_{a \in A(x,E)} \|xM(a)\|, \qquad x \in K,\ E \in \mathcal{E}, \tag{1.4}$$

where

$$A(x, E) = \{a : \|xM(a)\| > 0,\ xM(a)/\|xM(a)\| \in E\}. \tag{1.5}$$

Next, let $\mathcal{K}$ be the set of matrices defined by

$$\mathcal{K} = \{cM(a_1) \cdots M(a_n) : n = 1, 2, \dots,\ a_1, a_2, \dots \in A,\ c \in \mathbb{R},\ c > 0\}.$$

The following condition, which we shall call *Condition KR* or the *rank one condition*, was introduced in [30]:

$$\textit{The closure of } \mathcal{K} \textit{ contains a rank one matrix.} \tag{1.6}$$

The following result from 2006 is due to F. Kochman and J. Reeds [30].

THEOREM 1.2. *Let $\mathcal{H} = \{S, P, A, R\}$ be a classical HMM, let p_0 be a probability on S, let $\{X_n,\ n = 0, 1, 2, \dots\}$ and $\{Y_n,\ n = 1, 2, \dots\}$ denote the hidden Markov chain and the observation sequence induced by $\mathcal{H}$ and p_0, let $\{Z_n,\ n = 1, 2, \dots\}$ denote the sequence of conditional state distributions and, for $n = 1, 2, \dots$, let μ_n denote the probability distribution of Z_n. Then, if the state tr.pr.m. P is an irreducible, aperiodic tr.pr.m. and Condition KR is satisfied, there exists a probability measure $\mu \in \mathcal{P}(K, \mathcal{E})$, independent of the initial distribution p_0, such that*

$$\lim_{n \to \infty} \int_K u(x)\, \mu_n(dx) = \int u(x)\, \mu(dx), \qquad \forall u \in C[K].$$

In the paper [26] by Kaijser from 2011, Theorem 1.2 was generalised to the case when both the sets S and A are denumerable.

The main aim of this paper is to generalise the result of Kochman and Reeds one step further, namely to the case when the HMM under consideration is such that both the state space and the observation space are complete, separable, metric spaces and the HMM is a so-called fully dominated HMM (see [7, Section 2.2]), which loosely speaking means that the tr.pr.fs involved in the definition of the HMM have densities.

1.2. Some earlier results. Let $\{X_n,\, n = 0, 1, 2, \ldots\}$ be a stationary, aperiodic, irreducible Markov chain with finite state space $S = \{1, \ldots, d\}$, tr.pr.m. P and stationary distribution p_0. Let $g : S \to A$ be a surjective mapping from S to another space A, and for $n = 0, 1, 2, \ldots$ define $Y_n = g(X_n)$. (The function g is sometimes called a *lumping function.*) Let $\|\cdot\|$, K, $\mathcal{E}$ and $\mathcal{P}(K, \mathcal{E})$ be defined as above. If we define the tr.pr.m. R from S to A by

$$(R)_{i,a} = 1 \quad \text{if } g(i) = a, \tag{1.7}$$

$$(R)_{i,a} = 0 \quad \text{if } g(i) \neq a, \tag{1.8}$$

then $\mathcal{H} = \{S, P, A, R\}$ constitutes a HMM. We call a HMM $\{S, P, A, R\}$ for which R is defined by (1.7) and (1.8) a *HMM determined by the lumping function* g, and say that the observation matrix R is determined by the function g.

In the classical paper [5] from 1957, D. Blackwell showed that the entropy rate $H_R(Y)$ for the stationary sequence $\{Y_n,\, n = 0, 1, 2, \ldots\}$ can be expressed as

$$H_R(Y) = \frac{-1}{\ln 2} \sum_{a \in A} \int_K \|xM(a)\| \ln(\|xM(a)\|)\, \mu(dx), \tag{1.9}$$

where $M(a)$, $a \in A$, denotes the stepping matrix determined by a, which in this case is defined by

$$(M(a))_{i,j} = (P)_{i,j} \text{ if } g(j) = a \quad \text{and} \quad (M(a))_{i,j} = 0 \text{ otherwise,}$$

and where μ is an invariant measure (a so-called Blackwell measure) for the tr.pr.f. $\mathbf{P} : K \times \mathcal{E} \to [0, 1]$ defined by (1.4) and (1.5).

In [5] Blackwell raised the question whether the tr.pr.f. $\mathbf{P}$ defined by (1.4) and (1.5) has a unique invariant probability measure; he proved the uniqueness if the tr.pr.m. P has "rows which are nearly identical and no element which is very small". Blackwell also made the conjecture that there is a unique invariant measure if P is indecomposable.

In the paper [21] from 1975 the following condition was introduced for the special case when the HMM is determined by a lumping function.

Condition A. There exists a finite sequence $a_1, \ldots, a_N$ of elements in A such that if we set

$$M = M(a_1) \cdots M(a_N),$$

where $M(a_i)$, $i = 1, \ldots, N$, are defined by (1.3), then M is a nonzero matrix, and if $(M)_{i_1,j_1}, (M)_{i_2,j_2} > 0$ then also $(M)_{i_1,j_2}, (M)_{i_2,j_1} > 0$.

In [21, Theorem A] it was proved that if Condition A is satisfied, then the tr.pr.f. $\mathbf{P}$ has a unique invariant probability measure μ (say), and furthermore if $\{Z_{n,p},\, n = 1, 2, \ldots\}$ denotes the Markov chain generated by the tr.pr.f. $\mathbf{P}$ and the initial distribution p, and $\mu_{n,p}$, $n = 1, 2, \ldots$, denotes the distribution of $Z_{n,p}$, then $\{\mu_{n,p},\, n = 1, 2, \ldots\}$ converges in distribution towards the unique invariant measure μ for all initial distributions p.

In [21] a simple counterexample to Blackwell's conjecture was also presented.

Condition A was originally formulated for HMMs determined by a lumping function, and Theorem A of [21] was formulated for such a HMM. This may at first seem to be a severe restriction but, as was pointed out in [21], it is not so because of an observation due to L. Baum and T. Petrie [3]. For, let $\mathcal{H} = \{S, P, A, R\}$ be an arbitrary (classical) HMM. We can then define another enlarged HMM as follows. Set $S' = S \times A$ and $A' = A$, define the tr.pr.m. P' on S' by

$$(P')_{(i,a),(j,b)} = (P)_{i,j}(R)_{j,b},$$

define $g' : S' \to A'$ by $g'(i, a) = a$, let R' denote the observation matrix determined by g' and set $\mathcal{H}' = \{S', P', A', R'\}$. It is then easy to see that if Condition A holds for the HMM $\mathcal{H}$, then it also holds for $\mathcal{H}'$. Furthermore, if μ' is a unique invariant measure for the filter kernel $\mathbf{P}'$ induced by the HMM $\{S', P', A', R'\}$, and we let μ denote the marginal distribution of μ' on S, then μ is a unique invariant measure for the filter kernel $\mathbf{P}$ induced by $\{S, P, A, R\}$.

In the paper [15] from 1996, A. Goldsmith and P. Varaiya used Theorem A of [21] in order to obtain limit formulas for the capacity and the mutual information for finite-state Markov channels.

In the paper [30] from 2006, F. Kochman and J. Reeds showed that Condition A of [21] implies that their “rank one condition” holds and they gave a simplified proof of the conclusions of Theorem A in [21]. A few years later, in the paper [8] from 2010, P. Chigansky and R. van Handel proved that the “rank one condition” of Kochman and Reeds is also a necessary condition in order for a classical HMM to have a unique invariant probability measure for its filter kernel.

2. The main theorem

In this chapter we formulate the main theorem of the present paper. We begin by introducing some further notation and concepts, most of them standard and well-known. We then state the main theorem, and we end the chapter by giving an outline of the contents of the rest of the paper (see Section 2.9).

2.1. Basic notation. If $(X, \mathcal{X})$ is a given measurable set and ϕ is a metric on X, then we always assume implicitly that there is a topology on X which is determined by the metric ϕ, and that the σ-algebra $\mathcal{X}$ is the Borel field induced by this topology. We call such a space a *metric space*, and denote it $(X, \mathcal{X}, \phi)$ or simply $(X, \mathcal{X})$.

Next, let $(X, \mathcal{X})$ be a given measurable space. We let $\mathcal{P}(X, \mathcal{X})$ denote the set of probabilities on $(X, \mathcal{X})$, we let $\mathcal{Q}(X, \mathcal{X})$ denote the set of finite, nonnegative measures on $(X, \mathcal{X})$ and let $\mathcal{Q}^\infty(X, \mathcal{X})$ denote the set of σ-finite, positive measures on $(X, \mathcal{X})$. If $\mu, \nu \in \mathcal{Q}(X, \mathcal{X})$, we let $\delta_{TV}(\mu, \nu)$ denote the total variation between μ and ν defined by

$$\delta_{TV}(\mu, \nu) = \sup\{\mu(F) - \nu(F) : F \in \mathcal{X}\} + \sup\{\nu(F) - \mu(F) : F \in \mathcal{X}\}.$$

We shall also often use the notation $\|\mu - \nu\|$ instead of $\delta_{TV}(\mu, \nu)$. If $\nu \in \mathcal{Q}(X, \mathcal{X})$, we write $\|\nu\| = \nu(X)$. We always assume implicitly that the topology on $\mathcal{Q}(X, \mathcal{X})$ is the topology generated by the total variation metric δ_{TV}.

We let $B_u[X]$ denote the set of real, $\mathcal{X}$-measurable functions on X, and $B[X]$ the set of real, bounded, $\mathcal{X}$-measurable functions on X. We may write $B[X, \mathcal{X}]$ instead of $B[X]$. If $u \in B[X]$, we set $\|u\| = \sup\{|u(x)| : x \in X\}$, $\operatorname{osc}(u) = \sup\{u(x) - u(y) : x, y \in X\}$, and if $u \in B[X]$ and $A \subset X$, we set $\mathrm{osc}_A(u) = \sup\{u(x) - u(y) : x, y \in A\}$. If $u \in B_u[X]$ and $\nu \in \mathcal{P}(X, \mathcal{X})$ then, when convenient, we write $\int_X u(x)\,\nu(dx) = \langle u, \nu\rangle$ if the integral exists. If $\lambda \in \mathcal{Q}^\infty(X, \mathcal{X})$, $\nu \in \mathcal{Q}(X, \mathcal{X})$ and there exists $f \in B_u[X]$ such that

$$\nu(F) = \int_F f(x)\,\lambda(dx), \qquad \forall F \in \mathcal{X},$$

then we write $\nu \in \mathcal{Q}_\lambda(X, \mathcal{X})$ and call f a *representative* of ν. If also $\|\nu\| = 1$, we write $\nu \in \mathcal{P}_\lambda(X, \mathcal{X})$.

If $\mu, \nu \in \mathcal{Q}_\lambda(X, \mathcal{X})$, f is a representative of μ, and g is a representative of ν, we define $\mu \wedge \nu$ by $\mu \wedge \nu(F) = \int_F \min\{f(x), g(x)\}\,\lambda(dx)$, $F \in \mathcal{X}$, and we define $\mu \vee \nu$ by $\mu \vee \nu(F) = \int_F \max\{f(x), g(x)\}\,\lambda(dx)$, $F \in \mathcal{X}$.

If $\mu, \nu \in \mathcal{P}(X, \mathcal{X})$ and $u \in B[X]$, it is well-known that

$$\left|\int u(x)\,\mu(dx) - \int u(x)\,\nu(dx)\right| \le \mathrm{osc}(u)\|\mu - \nu\|/2. \tag{2.1}$$

Next, if $Q : X \times \mathcal{X} \to [0, \infty)$ is such that $Q(x, \cdot) \in \mathcal{Q}(X, \mathcal{X})$ for all $x \in X$ and $Q(\cdot, F)$ is $\mathcal{X}$-measurable for all $F \in \mathcal{X}$, we call Q a *transition kernel.* If $Q(x, X) = 1$ for all $x \in K$, then clearly $Q : X \times \mathcal{X} \to [0, 1]$ is a tr.pr.f. If $Q : X \times \mathcal{X} \to [0, 1]$ is a tr.pr.f. on $(X, \mathcal{X})$, then we define $Q^n : X \times \mathcal{X} \to [0, 1]$ recursively by $Q^1 = Q$ and

$$Q^{n+1}(x, F) = \int_X Q(x, dx')Q^n(x', F), \qquad n = 2, 3, \ldots.$$

We call the mapping $T : B[X] \to B[X]$ defined by $Tu(x) = \int_X u(y)\,Q(x, dy)$ the *transition operator* associated to the tr.pr.f. Q. The tr.pr.f. Q also induces a map $\check{Q} : \mathcal{P}(X, \mathcal{X}) \to \mathcal{P}(X, \mathcal{X})$ by $\check{Q}(\mu)(F) = \int_X Q(x, F)\,\mu(dx)$. We usually write $\check{Q}(\mu) = \mu Q$. As is well-known (see [37, Section 1.2]),

$$\langle u, \mu Q\rangle = \langle Tu, \mu\rangle. \tag{2.2}$$

If $Q : X \times \mathcal{X} \to [0, 1]$ is a tr.pr.f. on $(X, \mathcal{X})$, $\lambda \in \mathcal{Q}^\infty(X, \mathcal{X})$ and $q : X \times X \to [0, \infty)$ is a measurable function such that

$$Q(x, F) = \int_F q(x, y)\,\lambda(dy), \qquad \forall x \in X, \forall F \in \mathcal{F},$$

then we call $q : X \times X \to [0, \infty)$ a *probability density kernel* on $\{(X, \mathcal{X}), \lambda\}$.

Next, let $(X, \mathcal{X}, \phi)$ be a metric space. We let $C[X]$ denote the set of real, bounded, continuous functions on X. If $u \in C[X]$, we define

$$\gamma(u) = \sup\left\{\frac{u(x_1) - u(x_2)}{\phi(x_1, x_2)} : x_1 \ne x_2\right\},$$
$$\mathrm{Lip}[X] = \{u \in C[X] : \gamma(u) < \infty\}, \qquad \mathrm{Lip}_1[X] = \{u \in \mathrm{Lip}[X] : \gamma(u) \le 1\}.$$

If T denotes the transition operator associated to the tr.pr.f. Q and if

$$u \in C[X] \;\Rightarrow\; Tu \in C[K],$$

then Q is called *Feller continuous*; if there exists a constant C such that

$$u \in \mathrm{Lip}[X] \;\Rightarrow\; \gamma(Tu) \le C\gamma(u),$$

then we call Q *Lipschitz continuous* and call the constant C a *bounding constant*; if furthermore there exists a constant C such that

$$\gamma(T^n u) \le C\gamma(u), \qquad n = 1, 2, \ldots, \qquad \forall u \in \mathrm{Lip}[K], \tag{2.3}$$

then we call Q *Lipschitz equicontinuous* and again call C a *bounding constant.*

Let again $(X, \mathcal{X})$ be a measurable set and let $Q : X \times \mathcal{X} \to [0, 1]$ be a tr.pr.f. on $(X, \mathcal{X})$. If there exists a probability measure $\pi \in \mathcal{P}(X, \mathcal{X})$ such that

$$\lim_{n\to\infty} \delta_{TV}(Q^n(x, \cdot), \pi) = 0, \qquad \forall x \in X,$$

then we call the tr.pr.f. Q *strongly ergodic*, and we call π the *limit measure*; if furthermore

$$\lim_{n\to\infty} \sup_{x\in X} \delta_{TV}(Q^n(x, \cdot), \pi) = 0,$$

then we call the tr.pr.f. Q *uniformly ergodic.* If $(X, \mathcal{X}, \phi)$ is a metric space and

$$\lim_{n\to\infty} \sup\{\langle u, Q^n(x, \cdot)\rangle - \langle u, Q^n(y, \cdot)\rangle : u \in \mathrm{Lip}_1[K]\,\} = 0, \qquad \forall x, y \in X, \tag{2.4}$$

then we call the tr.pr.f. Q *weakly contracting*; if furthermore there exists a probability measure π such that

$$\lim_{n\to\infty} \int_X u(y)\, Q^n(x, dy) = \int_X u(y)\, \pi(dy), \qquad \forall x \in X, \forall u \in C[X],$$

then we call Q *weakly ergodic.*

If $(X_1, \mathcal{X}_1, \phi_1)$ and $(X_2, \mathcal{X}_2, \phi_2)$ are metric spaces, then $(X_1 \times X_2, \mathcal{X}_1 \otimes \mathcal{X}_2)$ is a metric space with metric given by $\phi_3((x_1, x_2), (y_1, y_2)) = \phi_1(x_1, y_1) + \phi_2(x_2, y_2)$. Furthermore, if $\mu_1 \in \mathcal{Q}(X_1, \mathcal{X}_1)$ and $\mu_2 \in \mathcal{Q}(X_2, \mathcal{X}_2)$ then $\mu_1 \otimes \mu_2$ denotes the product measure on $(X_1 \times X_2, \mathcal{X}_1 \otimes \mathcal{X}_2)$.

If $\tilde{\mu}$ is a probability measure on the product space $(X_1 \times X_2, \mathcal{X}_1 \otimes \mathcal{X}_2)$ of two measurable spaces $(X_1, \mathcal{X}_1)$ and $(X_2, \mathcal{X}_2)$, recall that the measure $\tilde{\mu}$ is determined if it is defined for all sets in $\mathcal{X}_1 \otimes \mathcal{X}_2$ of the form $E_1 \times E_2$, where $E_1 \in \mathcal{X}_1$ and $E_2 \in \mathcal{X}_2$ (the so-called rectangular sets).

If $(X, \mathcal{X})$ is a measurable space, $Q_1 : X \times \mathcal{X} \to [0, \infty)$ and $Q_2 : X \times \mathcal{X} \to [0, \infty)$ are transition functions such that both $\sup\{Q_1(x, X) : x \in X\} < \infty$ and $\sup\{Q_2(x, X) : x \in X\} < \infty$, then we define $Q_1 Q_2 : X \times \mathcal{X} \to [0, \infty)$ by

$$Q_1 Q_2(x, F) = \int_X Q_1(x, dx') Q_2(x', F), \qquad F \in \mathcal{X}.$$

If $(X, \mathcal{X})$ is a measurable space and $(X_n, \mathcal{X}_n) = (X, \mathcal{X})$, $n = 2, 3, \ldots$, we set $X^2 = X \times X$, $\mathcal{X}^2 = \mathcal{X} \otimes \mathcal{X}$ and define X^n and $\mathcal{X}^n$ recursively for $n = 2, 3, \ldots$ by $X^{n+1} = X^n \times X$ and $\mathcal{X}^n \otimes \mathcal{X}$. We denote a generic element in X^n by x^n or $(x_1, \ldots, x_n)$.

If $(X, \mathcal{X})$ is a measurable space and $\mu \in \mathcal{Q}^\infty(X, \mathcal{X})$, we let $\mu^n \in \mathcal{Q}^\infty(X^n, \mathcal{X}^n)$ denote the nth product measure on $(X^n, \mathcal{X}^n)$ defined recursively by $\mu^2 = \mu \otimes \mu$ and $\mu^{n+1} = \mu^n \otimes \mu$.

Next, we shall recall the concept of coupling. Let μ be a probability measure on the measurable space $(X_1, \mathcal{X}_1)$ and let ν be a probability measure on the measurable space

$(X_2, \mathcal{X}_2)$. If $\tilde{\mu}$ is a probability measure on the product space $(X_1 \times X_2, \mathcal{X}_1 \otimes \mathcal{X}_2)$ such that

$$\tilde{\mu}(F \times X_2) = \mu(F), \quad \forall F \in X_1,$$
$$\tilde{\mu}(X_1 \times F) = \nu(F), \quad \forall F \in X_2,$$

then we call $\tilde{\mu}$ a *coupling* of μ and ν. We denote the set of all couplings of μ and ν by $\tilde{\mathcal{P}}(\mu, \nu, X_1 \times X_2)$.

We end this section with a simple and useful inequality for differences between normalised vectors in a normed vector space. The lemma follows easily by first adding and subtracting the term $y/\|x\|$ on the left hand side and then using the triangle inequality.

LEMMA 2.1. *Let x and y belong to a normed vector space and suppose that $\|x\|, \|y\| > 0$. Then*

$$\left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\| \leq \frac{2\|x-y\|}{\|x\|}.$$

2.2. Hidden Markov models on general state spaces. Let $(S, \mathcal{F})$ and $(A, \mathcal{A})$ be measurable spaces and let $(S \times A, \mathcal{F} \otimes \mathcal{A})$ be the product space. Let $\xi \in \mathcal{P}(S \times A, \mathcal{F} \otimes \mathcal{A})$ and let $\Lambda : S \times A \times \mathcal{F} \otimes \mathcal{A} \to [0, 1]$ be a tr.pr.f. on $(S \times A, \mathcal{F} \otimes \mathcal{A})$. The Markov chain generated by the tr.pr.f. Λ and the initial distribution ξ is called a *bivariate Markov chain.* We denote a bivariate Markov chain $\{(X_{n,\xi}, Y_{n,\xi}),\ n = 0, 1, 2, \ldots\}$.

A Hidden Markov Model (HMM), as described in the classical paper [36], consists of a finite state space S, a finite observation space A, a tr.pr.m. P on S, a tr.pr.m. R from S to A, and an initial distribution p_0. In the more modern literature (see e.g. [7]) one allows both the state space S and the observation space A to be measurable spaces, $(S, \mathcal{F})$ and $(A, \mathcal{A})$ say, and then the tr.pr.ms P and R must be replaced by tr.pr.fs.

Our definition of a HMM is slightly more general than the one given in [7, Section 2.2], and will be based on a tr.pr.f. from the state space to the product of the state space and the observation space.

DEFINITION 2.2. Let $(S, \mathcal{F})$ and $(A, \mathcal{A})$ be measurable spaces, $M : S \times (\mathcal{F} \otimes \mathcal{A}) \to [0, 1]$ be a tr.pr.f. from $(S, \mathcal{F})$ to $(S \times A, \mathcal{F} \otimes \mathcal{A})$, and define the tr.pr.f. $P : S \times \mathcal{F} \to [0, 1]$ by $P(s, F) = M(s, F \times A)$. Then we call

$$\mathcal{H} = \{(S, \mathcal{F}), P, (A, \mathcal{A}), M\} \tag{2.5}$$

a *Hidden Markov Model* (*HMM*). We call $(S, \mathcal{F})$ the *state space*, $(A, \mathcal{A})$ the *observation space*, M the *Hidden Markov Model kernel* of $\mathcal{H}$ (the *HMM-kernel*), and P the *Markov kernel* of $\mathcal{H}$.

In case the tr.pr.f. $M : S \times (\mathcal{F} \otimes \mathcal{A}) \to [0, 1]$ is determined by composing a tr.pr.f. $P : S \times \mathcal{F} \to [0, 1]$ on $(S, \mathcal{F})$ with a tr.pr.f. $R : S \times \mathcal{A} \to [0, 1]$ from $(S, \mathcal{F})$ to $(A, \mathcal{A})$ in such a way that

$$M(s, F \times B) = \int_F P(s, dt) R(t, B),$$

then we call $\mathcal{H}$ an *ordinary HMM* and write $\{(S, \mathcal{F}), P, (A, \mathcal{A}), R\}$ instead of $\{(S, \mathcal{F}), P, (A, \mathcal{A}), M\}$.

REMARK 2.3. Since the tr.pr.f. P is determined by M, we could have excluded P in the expression of the right hand side of (2.5). We have included it for the sake of clarity.

Associated to a HMM there are two stochastic processes defined as follows.

Definition 2.4. Let $\mathcal{H} = \{(S, \mathcal{F}), P, (A, \mathcal{A}), M\}$ be a HMM, and define $\Lambda : S \times A \times (\mathcal{F} \otimes \mathcal{A}) \to [0, 1])$ by

$$\Lambda(s, a, F \times B) = M(s, F \times B), \qquad \forall s \in S, \forall a \in A, \forall F \in \mathcal{F}, \forall B \in \mathcal{A}.$$

Let $x \in \mathcal{P}(S, \mathcal{F})$, $\alpha \in \mathcal{P}(A, \mathcal{A})$, $\xi = x \otimes \alpha$, and let $\{(X_{n,\xi}, Y_{n,\xi}),\ n = 0, 1, 2, \ldots\}$ denote the bivariate Markov chain generated by Λ and ξ.

We call $\{X_{n,\xi},\ n = 0, 1, 2, \ldots\}$ the *hidden Markov chain* generated by $\mathcal{H}$, and write $\{X_{n,x},\ n = 0, 1, 2, \ldots\}$ instead of $\{X_{n,\xi},\ n = 0, 1, 2, \ldots\}$ since the first component is independent of the initial distribution $\alpha \in \mathcal{P}(A, \mathcal{A})$; we call $\{Y_{n,\xi},\ n = 1, 2, \ldots\}$ the *observation sequence* generated by $\mathcal{H}$ and write $\{Y_{n,x},\ n = 1, 2, \ldots\}$ instead of $\{Y_{n,\xi},\ n = 1, 2, \ldots\}$ since also the second component is independent of the initial distribution α if $n \geq 1$.

Remark 2.5. Our way to define the hidden Markov chain and the observation sequence of a HMM is similar to the way these sequences are defined in [7].

Remark 2.6. It is for the sake of convenience that we start the observation sequence with $n = 1$ instead of $n = 0$.

2.3. Fully dominated hidden Markov models. Let $\mathcal{H} = \{(S, \mathcal{F}), P, (A, \mathcal{A}), M\}$ be a HMM. Now suppose that

(1) the state space $(S, \mathcal{F})$ is a complete, separable, metric space $(S, \mathcal{F}, \delta_0\}$,
(2) there exists a positive σ-finite measure λ on $(S, \mathcal{F})$,
(3) the observation space $(A, \mathcal{A})$ is a complete, separable, metric space $(A, \mathcal{A}, \varrho)$,
(4) there is a σ-finite positive measure τ on $(A, \mathcal{A})$, and
(5) $m : S \times S \times A \to [0, \infty)$ is an $\mathcal{F} \otimes \mathcal{F} \otimes \mathcal{A}$-measurable function such that

$$M(s, F \times B) = \int_F \int_B m(s, t, a)\, \lambda(dt)\, \tau(da), \qquad \forall s \in S, \forall F \in \mathcal{F}, \forall B \in \mathcal{A}.$$

Then, following [7], we call $\mathcal{H}$ a *fully dominated HMM*. We denote a fully dominated HMM by

$$\{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\} \tag{2.6}$$

where $p : S \times S \to [0, \infty)$ is defined by

$$p(s, t) = \int_A m(s, t, a)\, \tau(da).$$

We call λ and τ *base measures*, m the *probability density kernel* of $\mathcal{H}$, the tr.pr.f. $P : S \times \mathcal{F} \to [0, 1]$, defined by $P(s, F) = \int_F p(s, t)\, \lambda(dt)$, the *Markov kernel* determined by (p, λ), and $M : S \times (\mathcal{F} \otimes \mathcal{A}) \to [0, 1]$ the *HMM-kernel* determined by (m, λ, τ).

In case there exists a measurable function $r : S \times A \to [0, \infty)$ such that the probability density kernel $m : S \times S \times A \to [0, \infty)$ can be factorised as

$$m(s, t, a) = p(s, t) r(t, a),$$

we call $\mathcal{H}$ an *ordinary, fully dominated HMM* and write $\{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (r, \tau)\}$ instead of $\{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$.

If S is denumerable, we always assume that the associated σ-algebra $\mathcal{F}$ is the power set of S, δ_0 is the discrete metric and λ is the counting measure. Similarly, if A is denumerable, we always assume that the σ-algebra $\mathcal{A}$ is the power set of A, ϱ is the discrete metric and τ is the counting measure. Therefore, if both the state space and the observation space are denumerable, we will denote such a HMM by $\{S, P, A, M\}$.

Furthermore, if the Markov kernel P of $\mathcal{H}$ is strongly ergodic [uniformly ergodic] with limit measure π we call $\mathcal{H}$ a *strongly ergodic* [*uniformly ergodic*], *fully dominated HMM with limit measure* π.

Next, set $K = P_\lambda(S, \mathcal{F})$, let the metric $\delta : K \times K \to [0, 2]$ be defined by

$$\delta(x, y) = \delta_{TV}(x, y) = \int |f(s) - g(s)|\, \lambda(ds),$$

where, in the last expression, f and g are representatives of x and y respectively; further let $\mathcal{E}$ denote the σ-algebra on K generated by the metric δ.

We shall now introduce a subset of the set of fully dominated HMMs as follows. Let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau))\}$ be a fully dominated HMM and let P be the tr.pr.f. determined by (p, λ). For each $a \in A$ we define $M_a : \mathcal{Q}_\lambda(S, \mathcal{F}) \to \mathcal{Q}_\lambda(S, \mathcal{F})$ by

$$M_a(x)(F) = \int_{s \in S} \int_{t \in F} m(s, t, a)\, x(ds)\, \lambda(dt). \tag{2.7}$$

We call $M_a : \mathcal{Q}_\lambda(S, \mathcal{F}) \to \mathcal{Q}_\lambda(S, \mathcal{F})$ the *stepping function* determined by $a \in A$. We shall usually write xM_a instead of $M_a(x)$.

Further let $\overline{M} : \mathcal{Q}_\lambda(S, \mathcal{F}) \times A \to \mathcal{Q}_\lambda(S, \mathcal{F})$ be defined by

$$\overline{M}(x, a) = xM_a. \tag{2.8}$$

We call $\overline{M} : \mathcal{Q}_\lambda(S, \mathcal{F}) \times A \to \mathcal{Q}_\lambda(S, \mathcal{F})$ the *stepping kernel* of $\mathcal{H}$.

Definition 2.7. If $\overline{M} : \mathcal{Q}_\lambda(S, \mathcal{F}) \times A \to \mathcal{Q}_\lambda(S, \mathcal{F})$ defined by (2.7) and (2.8) is continuous, then we call $\mathcal{H}$ a *fully dominated*, *regular HMM*.

A trivial, but yet important, example of a fully dominated, regular HMM is a HMM $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ for which the observation space A is denumerable, the metric ϱ is the discrete metric and the measure τ is the counting measure, since in this case, if we let f and g be representatives of x and y in $\mathcal{Q}_\lambda(S, \mathcal{F})$, then for all $a \in A$ we have

$$\begin{aligned}\|xM_a - yM_a\| &\le \int_S \int_S |f(s) - g(s)| m(s, t, a)\, \lambda(ds)\, \lambda(dt)\\ &\le \int_S \int_S |f(s) - g(s)| p(s, t)\, \lambda(ds)\, \lambda(dt) \le \int_S \int_S |f(s) - g(s)|\, \lambda(ds) = \|x - y\|.\end{aligned}$$

For a less trivial example see Example 9.6 in Chapter 9.

2.4. The filter kernel. Now consider a fully dominated, regular HMM $\{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$. Let $M_a : \mathcal{Q}_\lambda(S, \mathcal{F}) \to \mathcal{Q}_\lambda(S, \mathcal{F})$ and $\overline{M} : \mathcal{Q}_\lambda(S, \mathcal{F}) \times A \to \mathcal{Q}_\lambda(S, \mathcal{F})$ be defined by (2.7) and (2.8) respectively. Further, for each $x \in K$ and $E \in \mathcal{E}$, set

$$A_x^+ = \{a : \|xM_a\| > 0\}, \tag{2.9}$$

and let

$$A(x,E) = \{a : \|xM_a\| > 0,\, xM_a/\|xM_a\| \in E\}. \tag{2.10}$$

We now define $\mathbf{P} : K \times \mathcal{E} \to [0,1]$ as follows:

$$\mathbf{P}(x,E) = \int_{A(x,E)} \|xM_a\|\, \tau(da). \tag{2.11}$$

That $\mathbf{P} : K \times \mathcal{E} \to [0,1]$ is a tr.pr.f. is easily proved. First we note that if $E = \bigcup_{i=1}^{\infty} E_i$ where $E_i \cap E_j = \emptyset$, $i \neq j$, then

$$A(x,E) = \bigcup_{i=1}^{\infty} A(x,E_i),$$

from which it easily follows that $\mathbf{P}(x,\cdot)$ is a probability measure on $(K,\mathcal{E})$ for each $x \in K$. That $\mathbf{P}(\cdot,E)$ is an $\mathcal{E}$-measurable function for each open set E follows easily from Lemma 2.1 and the continuity of $\overline{M} : \mathcal{Q}_\lambda(S,\mathcal{F}) \times A \to \mathcal{Q}_\lambda(S,\mathcal{F})$, and since it is easily proved that $\mathcal{G} = \{E \in K : \mathbf{P}(\cdot,E) \text{ is measurable}\}$ is a σ-algebra and therefore $\mathcal{E} \subset \mathcal{G}$, it follows that $\mathbf{P}(\cdot,E)$ is a $\mathcal{E}$ -measurable function for each $E \in \mathcal{E}$. Hence $\mathbf{P} : K \times \mathcal{E} \to [0,1]$, defined by (2.11) and (2.10), is a tr.pr.f. on $(K,\mathcal{E})$. We call $\mathbf{P}$ the *filter kernel induced by* $\{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$.

Remark 2.8. The main reason that we introduce the set of fully dominated *and regular* HMMs is that we need this extra property in order to be able to prove that $\mathbf{P}(\cdot,E) : K \to K$, where $\mathbf{P}$ is defined by (2.11) and (2.10), is a measurable function for all $E \in \mathcal{E}$. Since it seems to us that in most concrete examples of fully dominated HMMs the regularity condition is satisfied, we do not consider this extra assumption to be a severe restriction. A nice and useful property is that the set of fully dominated and regular HMMs is closed under an operation we call composition (see Section 3.1).

Remark 2.9. Note that if S and A are finite sets, δ_0 and ϱ are the discrete measures, and λ and τ are the counting measures, then the definition of $\mathbf{P}$, as given by (2.11) and (2.10), agrees with the definition given by (1.4) and (1.5).

Remark 2.10. Whenever we consider a fully dominated and regular HMM

$$\mathcal{H} = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$$

we always implicitly let K denote the set $\mathcal{P}_\lambda(S,\mathcal{F})$, $\mathcal{E}$ the σ-algebra on K generated by the total variation metric on K, and $\mathbf{P} : K \times \mathcal{E} \to [0,1]$ the filter kernel defined by (2.11) and (2.10).

We shall end this section by writing down the expression for the transition operator $\mathbf{T} : B[K] \to B[K]$ associated to the filter kernel $\mathbf{P}$. We have

$$\mathbf{T}u(x) = \int_{A_x^+} u\left(\frac{xM_a}{\|xM_a\|}\right) \|xM_a\|\, \tau(da), \tag{2.12}$$

where A_x^+ is defined by (2.9).

2.5. Condition E. In this section we shall formulate a condition which in Theorem 2.13 below will replace the rank one condition (Condition KR) of Theorem 1.2.

We first recall the well-known concept of barycenter (for some basic facts about this notion see e.g. the book [9] by G. Choquet).

Let $(S, \mathcal{F}, \delta_0)$ be a metric space, let $\lambda \in \mathcal{Q}^\infty(S, \mathcal{F})$, set $K = \mathcal{P}_\lambda(S, \mathcal{F})$, let $\mathcal{E}$ be the σ-algebra on K determined by the total variation metric on K and let $\mu \in \mathcal{P}(K, \mathcal{E})$. The *barycenter* of μ, which we denote by $\bar{b}(\mu)$, is a probability measure in K defined by

$$\bar{b}(\mu)(F) = \int_K \int_F x(ds)\, \mu(dx), \qquad F \in \mathcal{F}.$$

That $\bar{b}(\mu) : \mathcal{F} \to [0, 1]$ is a probability in K is easily verified. Furthermore, if $u \in B[S]$ and $\mu \in \mathcal{P}(K, \mathcal{E})$ has barycenter π, it follows easily from the definition that

$$\int_K \langle u, x \rangle\, \mu(dx) = \langle u, \bar{b}(\mu) \rangle = \langle u, \pi \rangle. \tag{2.13}$$

If $x \in K$, we let $\mathcal{P}(K|x)$ denote the set of probability measures in $\mathcal{P}(K, \mathcal{E})$ with barycenter x.

We now introduce Condition E.

Definition 2.11. Let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a strongly ergodic, fully dominated, regular HMM with invariant measure π, and let $\mathbf{P}$ be the induced Markov kernel. We define *Condition E* as follows: For every $\rho > 0$, there exist an integer N and a number α such that, for any measures μ and ν in $\mathcal{P}(K|\pi)$, there exists a coupling $\tilde{\mu}_N$ of $\mu\mathbf{P}^N$ and $\nu\mathbf{P}^N$ such that if we set $D_\rho = \{(x, y) \in K \times K : \delta_{TV}(x, y) < \rho\}$, then

$$\tilde{\mu}_N(D_\rho) \geq \alpha.$$

Remark 2.12. The important point of Condition E is that the number α does not depend on the choice of μ and ν in $\mathcal{P}(K|\pi)$.

2.6. The main theorem. We are now ready to formulate the main theorem of this paper.

Theorem 2.13. *Let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a fully dominated, regular, strongly ergodic HMM with limit measure π such that both $(S, \mathcal{F})$ and $(A, \mathcal{A})$ are complete, separable, metric spaces, and let $\mathbf{P}$ be the induced filter kernel. Suppose also that $\mathcal{H}$ fulfills Condition E. Then*

(A) *The filter kernel $\mathbf{P}$ is weakly contracting.*

If furthermore, either

(B) *there exists a measure $\mu \in \mathcal{P}(K, \mathcal{E})$ which is invariant with respect to $\mathbf{P}$, or*
(C) *there exists $x_0 \in K$ such that $\{\mathbf{P}^n(x_0, \cdot),\ n = 1, 2, \ldots\}$ is tight, or*
(D) *$\mathcal{H}$ is also uniformly ergodic,*

then the filter kernel $\mathbf{P}$ is weakly ergodic.

Remark 2.14. It was proved in [26] that if the state space of a strongly ergodic HMM with limit measure π is denumerable, then $\{\mathbf{P}^n(\pi, \cdot),\ n = 1, 2, \ldots\}$ is a tight sequence. We believe the same is true if the state space is a complete, separable, metric space. Therefore, we believe that the second part of the theorem could be omitted, and the conclusion in the first part of the theorem ought to be that the filter kernel $\mathbf{P}$ is weakly ergodic instead of just weakly contracting.

2.7. Some simple examples. To illustrate the consequences of Theorem 2.13 we present some simple examples.

EXAMPLE 2.15. We define a regular, fully dominated, ordinary HMM

$$\mathcal{H}_1 = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (r, \tau)\}$$

as follows. Let $S = [-1, 1]$, let δ_0 be the Euclidean metric on S, let $\mathcal{F}$ be the σ-algebra induced by δ_0, let λ be the Lebesgue–Borel measure on $(S, \mathcal{F})$ and let $p : S \times S \to [0, \infty)$ be a probability density kernel on $\{(S, \mathcal{F}), \lambda\}$ such that

$$\inf\{p(x, y) : (x, y) \in S \times S\} > 0.$$

We further let the observation space consist of just two points a and b, thus $A = \{a, b\}$, we let ϱ denote the discrete metric and let $\mathcal{A}$ be the power set of A. Finally, we let τ be the counting measure on $(A, \mathcal{A})$ and we define $r : S \times A$ simply by

$$r(t, a) = \begin{cases} 1 & \text{if } -1 \le t < 0, \\ 0 & \text{if } 0 \le t \le 1. \end{cases}$$

Since $A = \{a, b\}$ we have $r(t, a) + r(t, b) = 1$ for all $t \in S$, and hence $r(\cdot, b)$ is determined implicitly.

That the HMM $\mathcal{H}_1$ defined in this way is a fully dominated, regular and ordinary HMM is easily seen, and that the tr.pr.f. determined by (p, λ) is uniformly ergodic is easily proved and well-known. Furthermore, as we will prove in Chapter 9, the HMM $\mathcal{H}_1$ satisfies Condition E and therefore, by Theorem 2.13, the filter kernel induced by $\mathcal{H}_1$ is weakly ergodic.

A nice consequence of this result is the following. Let $g : [-1, 1] \to A$ be defined by

$$g(t) = \begin{cases} a & \text{if } t \in [-1, 0), \\ b & \text{if } t \in [0, 1]. \end{cases}$$

Let x_0 be a probability measure in $\mathcal{P}_\lambda(S, \mathcal{F})$, let $P : S \times \mathcal{F} \to [0, 1]$ be the tr.pr.f. determined by (p, λ) and let $\{X_n,\ n = 0, 1, 2, \ldots\}$ be the Markov chain generated by the initial distribution x_0 and the tr.pr.f. P. For $n = 0, 1, 2, \ldots$ define $Y_n = g(X_n)$. Let $H_R(Y_n)$ denote the entropy rate of the sequence $\{Y_n,\ n = 0, 1, 2, \ldots\}$ at time n. Then

$$\lim_{n\to\infty} H_R(Y_n) = -\int_K \|yM_a\| \log(\|yM_a\|)\, \mu(dy)/\log(2) \\ - \int_K \|yM_b\| \log(\|yM_b\|)\, \mu(dy)/\log(2), \tag{2.14}$$

where M_a and M_b denote the stepping functions determined by a and b respectively, and μ is the unique invariant measure of the filter kernel induced by the HMM $\mathcal{H}_1$ (see Corollary 9.4).

EXAMPLE 2.16. Again we shall determine a fully dominated, regular and ordinary HMM. Let $(S, \mathcal{F}, \delta_0)$ and (p, λ) be as in Example 2.15. Let $(A, \mathcal{A}, \varrho) = (S, \mathcal{F}, \delta_0)$ and $\tau = \lambda$. It remains to define $r : S \times A \to [0, \infty)$:

$$r(t,a) = \begin{cases} 5 & \text{if } -9/10 \le t \le 9/10 \text{ and } |t-a| < 1/10, \\ \dfrac{1}{1/10 + |t-a|} & \text{if } 9/10 < |t| \le 1 \text{ and } |t-a| < 1/10, \\ 0 & \text{elsewhere.} \end{cases}$$

We set $\mathcal{H}_2 = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (r, \tau)\}$.

Clearly, $\mathcal{H}_2$ is fully dominated and $\mathcal{H}_2$ is regular. In Chapter 9 we verify that Condition E is satisfied, and hence by Theorem 2.13 the filter kernel induced by $\mathcal{H}_2$ is weakly ergodic.

Our next two examples show that Condition E need not always hold.

Example 2.17. The fully dominated, regular and ordinary HMM

$$\mathcal{H}_3 = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (r, \tau)\}$$

in this example has the state space $(S, \mathcal{F}, \delta_0)$, the observation space $(A, \mathcal{A}, \varrho)$, the function $r : S \times A \to \infty$, the measure τ and the measure λ the same as in Example 2.15. The only difference is that we define the probability density kernel $p : S \times S \to [0, \infty)$ in a more complicated way:

$$p(s,t) = 1 \text{ if } (s,t) \in S_0, \qquad p(s,t) = 0 \text{ elsewhere,}$$

where $(s,t) \in S_0$ if one of the conditions (a)–(d) below is satisfied:

(a) $-1 < s \le -1/2$, and $-1 \le t \le -1/2$ or $0 \le t \le 1/2$,
(b) $-1/2 \le s \le 0$, and $-1/2 \le t \le 0$ or $1/2 \le t \le 1$,
(c) $0 \le s \le 1/2$, and $-1 \le t \le -1/2$ or $1/2 \le t \le 1$,
(d) $1/2 \le s \le 1$, and $-1/2 \le t \le 1/2$.

That $\mathcal{H}_3$ defined in this way is a fully dominated HMM is obvious, and since the observation space is finite, it follows that $\mathcal{H}_3$ is regular. That $\mathcal{H}_3$ is an ordinary HMM is also obvious from the definition.

It is also easily proved that the tr.pr.f. $P : S \times \mathcal{F} \to [0, 1]$ determined by (p, λ) is uniformly ergodic and has the uniform distribution on $[-1, 1]$ as limit measure. It is also not difficult to prove that the filter kernel induced by $\mathcal{H}_3$ does not have a unique invariant measure. (Compare with the example in [21, Section 10].)

Example 2.18. In our last example, we define an ordinary HMM $\mathcal{H}_4 = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (r, \tau)\}$ in such a way that $(S, \mathcal{F}, \delta_0)$, $(A, \mathcal{A}, \varrho)$, λ and (r, τ) are as in Example 2.17. The probability density kernel $p : S \times S \to [0, \infty)$ is defined in a slightly more complicated way:

$$p(s,t) = 2 \text{ if } (s,t) \in S_1, \qquad p(s,t) = 0 \text{ elsewhere,}$$

where $(s,t) \in S_1$ if one of conditions (a)–(h) below is satisfied:

(a) $-1 < s \le -3/4$, and $-1 \le t \le -3/4$ or $0 \le t \le 1/4$,
(b) $-3/4 \le s \le -1/2$, and $-3/4 \le t \le -1/2$ or $1/4 \le t \le 1/2$,
(c) $-1/2 \le s \le -3/4$, and $-3/4 \le t \le 0$ or $3/4 \le t \le 1$,
(d) $-3/4 \le s \le 0$, and $-1/2 \le t \le -1/4$ or $1/2 \le t \le 3/4$,

(e) $0 < s \leq 1/4$, and $-1 \leq t \leq -3/4$ or $3/4 \leq t \leq 1$,
(f) $1/4 \leq s \leq 1/2$, and $-3/4 \leq t \leq -1/2$ or $1/2 \leq t \leq 3/4$,
(g) $1/2 \leq s \leq 3/4$ and $-1/4 \leq t \leq 1/4$,
(h) $3/4 \leq s \leq 1$, and $-1/2 \leq t \leq -1/4$ or $1/4 \leq t \leq 1/2$.

It is again easily proved that $\mathcal{H}_4$ is a fully dominated, regular HMM such that the tr.pr.f. $P : S \times \mathcal{F} \to [0, 1]$ is uniformly ergodic with the limit measure equal to the uniform distribution on $[-1, 1]$. This time the filter kernel induced by $\mathcal{H}_4$ turns out to be periodic, which is easily verified. (Compare with [26, Example 11].)

2.8. Related results. Theorem 2.13 can be regarded as a result of filtering theory for filtering processes taking values in a nondenumerable state space.

In the classical paper [31] from 1971, H. Kunita considered filtering processes on a compact Hausdorff space, showed that the filtering process itself is a Markov process, and stated a condition under which the filter kernel of the filtering process is weakly ergodic. Kunita considered a continuous time process and assumed that the observation process was generated by a Wiener process.

The topology which determines the Borel field on the set of probability measures on the state space, which Kunita chose, was the weak topology.

In the paper [39] from 1989, Ł. Stettner generalised the work by Kunita to complete, separable, metric spaces. Stettner considered filtering processes in both continuous and discrete time.

Unfortunately there is a gap in one of the proofs in [31] and this gap also affects the results in [39] (see [4] for a discussion regarding this gap).

The fact that in this paper, as topology on the set of probabilities on the state space, we use the topology which is determined by the total variation metric implies that our notion of weak ergodicity is somewhat stronger than when weak ergodicity is defined by using the set of continuous functions determined by the weak topology.

Other papers considering the problem of finding conditions which guarantee weak ergodicity for the Markov kernel of the filtering process are e.g. by G. B. Di Masi and Ł. Stettner [11] and by R. van Handel [41]. In these papers and, as far as the author knows, in all other papers dealing with the existence of the limit distribution of the filtering process of a HMM—in contrast to our work—the σ-algebra used when defining the measurable space associated to the set of probabilities on the state space has been the Borel field induced by the weak topology.

Another assumption which is often made in papers on general HMM is that the probability density kernel $r : S \times A \to [0, \infty)$ determining the observations shall have $r(s, a) > 0$ for all $s \in S$ and all $a \in A$, a property which is not necessary in our set-up. Yet another assumption often made is that the tr.pr.f. that governs the hidden Markov chain shall be Feller continuous. We do not need this either.

2.9. The plan of the rest of the paper. In the next chapter we define compositions of HMMs, and we prove a universal inequality for fully dominated, regular HMMs. In Chapter 4 we show that for every fully dominated, regular HMM, there exists a random system with complete connections (RSCC) such that the filter kernel induced by the HMM is equal to the Markov kernel associated to this RSCC. The connection between RSCC and the conditional state distribution was already observed by D. Blackwell [5].

In Chapter 5 we prove some auxiliary theorems for Markov chains on bounded, complete, separable, metric spaces, not necessarily locally compact. In Chapter 6 we present some simple facts regarding the barycenters of the Markov chain generated by the filter kernel. In Chapter 7 we first prove an inequality for the Kantorovich distance between measures based on their barycenters, and then we finish the proof of the main theorem by verifying the hypotheses of the auxiliary theorems presented in Chapter 5.

The purpose of Chapter 8 is to find more explicit conditions that imply Condition E. In order to do this we shall need an estimate for iterations of integral kernels, which we prove using a classical result due to E. Hopf (see [17]). We then introduce Condition P, which can be regarded as a generalisation of Condition A mentioned in the introduction (see Section 1.2); then, by using (1) an estimate for iterations of integral kernels, (2) the connection to RSCCs and (3) the Vasershtein coupling of RSCCs, we prove that Condition P implies Condition E.

In Chapter 9 we present two examples satisfying Condition P. We also give an entropy formula for the observation sequence when the observation space is finite. Finally in Chapter 10 we raise a few questions and make a few comments.

3. On compositions of HMMs and a universal inequality

3.1. Compositions of HMMs. Let $\mathcal{H}_1 = \{(S,\mathcal{F}), P_1, (A_1,\mathcal{A}_1), M^1\}$ and $\mathcal{H}_2 = \{(S,\mathcal{F}), P_2, (A_2,\mathcal{A}_2), M^2\}$ be HMMs with the same state space $(S,\mathcal{F})$. Define $A^{1,2} = A_1 \times A_2$ and $\mathcal{A}^{1,2} = \mathcal{A}_1 \otimes \mathcal{A}_2$, define $M^{(1,2)} : S \times \mathcal{F} \times \mathcal{A}^{1,2} \to [0,1]$ by

$$M^{(1,2)}(s, F \times B_1 \times B_2) = \int_S M^1(s, dt, B_1) M^2(t, F, B_2),$$

define $P^{(1,2)} : S \times \mathcal{F} \to [0,1]$ by

$$P^{(1,2)}(s,F) = M^{(1,2)}(s, F \times A_1 \times A_2),$$

and set

$$\mathcal{H}^{1,2} = \{(S,\mathcal{F}), P^{(1,2)}, (A^{1,2},\mathcal{A}^{1,2}), M^{(1,2)}\}.$$

Obviously, $\mathcal{H}^{1,2}$ is also a HMM; we call $\mathcal{H}^{1,2}$ the *composition* of $\mathcal{H}_1$ and $\mathcal{H}_2$. For simplicity we write

$$\mathcal{H}^{1,2} = \mathcal{H}_1 * \mathcal{H}_2\,.$$

By Fubini's theorem it follows that if $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$ are HMMs with the same state space, then

$$(\mathcal{H}_1 * \mathcal{H}_2) * \mathcal{H}_3 = \mathcal{H}_1 * (\mathcal{H}_2 * \mathcal{H}_3).$$

If $\mathcal{H}$ is a HMM and $\mathcal{H}_n = \mathcal{H}$, $n = 1,\dots,N$, where $N \geq 2$, we set

$$\mathcal{H}^N = \mathcal{H}_1 * \cdots * \mathcal{H}_N.$$

We call $\mathcal{H}^N$ the *Nth iterate* of $\mathcal{H}$. Loosely speaking, the Nth iterate $\mathcal{H}^N$ of a HMM $\mathcal{H}$ is the HMM obtained from $\mathcal{H}$ when one collects the observations in groups of N instead of collecting them one by one.

3.2. On compositions of fully dominated HMMs. Let $\mathcal{H}_1 = \{(S,\mathcal{F},\delta_0),(p_1,\lambda),(A_1,\mathcal{A}_1,\varrho_1),(m_1,\tau_1)\}$ and $\mathcal{H}_2 = \{(S,\mathcal{F},\delta_0),(p_2,\lambda),(A_2,\mathcal{A}_2,\varrho_2),(m_2,\tau_2)\}$ be fully dominated HMMs with the same state space and with the same base measure $\lambda \in \mathcal{Q}^\infty(S,\mathcal{F})$. We define $m^{(1,2)} : S \times S \times A_1 \times A_2 \to [0,\infty)$ by

$$m^{(1,2)}(s,t,a_1,a_2) = \int_S m_1(s,s',a_1)m_2(s',t,a_2)\,\lambda(ds').$$

Again by Fubini's theorem, $\mathcal{H}_1 * \mathcal{H}_2$ is also a fully dominated HMM such that the HMM-kernel $M^{(1,2)} : S \times \mathcal{F} \otimes \mathcal{A}_1 \otimes \mathcal{A}_2 \to [0,1]$ satisfies

$$M^{(1,2)}(s, F \times B_1 \times B_2) = \int_F \int_{B_1} \int_{B_2} m^{(1,2)}(s,t,a_1,a_2)\,\lambda(dt)\,\tau_1(da_1)\,\tau_2(da_2).$$

Furthermore, if both $\mathcal{H}_1$ and $\mathcal{H}_2$ are regular, then it is elementary to prove that also $\mathcal{H}_1 * \mathcal{H}_2$ is regular.

If $a_1 \in A_1$, $a_2 \in A_2$ and $M^1_{a_1}$, $M^2_{a_2}$, $M^{(1,2)}_{(a_1,a_2)}$ are stepping functions determined by $a_1 \in A_1$, $a_2 \in A_2$ and $(a_1,a_2) \in A^{(1,2)}$ (see (2.7) for the definition of a stepping function), then clearly

$$M^1_{a_1} M^2_{a_2} = M^{(1,2)}_{(a_1,a_2)}. \tag{3.1}$$

Next, let $a_1 \in A_1$ and $a_2 \in A_2$, and consider the stepping functions $M^1_{a_1}, M^2_{a_2}, M^{(1,2)}_{(a_1,a_2)}$. The following scaling property holds by (3.1):

$$\frac{xM^{(1,2)}_{(a_1,a_2)}}{\|xM^{(1,2)}_{(a_1,a_2)}\|} = \frac{(xM^1_{a_1}/\|xM^1_{a_1}\|)M^2_{a_2}}{\|(xM^1_{a_1}/\|xM^1_{a_1}\|)M^2_{a_2}\|} \quad \text{if } \|xM^{(1,2)}_{(a_1,a_2)}\| > 0. \tag{3.2}$$

Furthermore, if (1) $\mathcal{H}_1$ and $\mathcal{H}_2$ are fully dominated, regular HMMs, (2) $\mathbf{P}_1$ and $\mathbf{P}_2$ denote the filter kernels induced by $\mathcal{H}_1$ and $\mathcal{H}_2$, (3) $\mathbf{T}_1$ and $\mathbf{T}_2$ denote the associated transition operators, (4) $\mathbf{P}^{(1,2)}$ denotes the filter kernel induced by $\mathcal{H}_1 * \mathcal{H}_2$, and (5) $\mathbf{T}^{(1,2)}$ denotes the transition operator associated to the kernel $\mathbf{P}^{(1,2)}$, then, by using the scaling property (3.2), it is not difficult to prove that

$$\mathbf{T}_1\mathbf{T}_2 = \mathbf{T}^{(1,2)}, \tag{3.3}$$

$$\mathbf{P}_1\mathbf{P}_2 = \mathbf{P}^{(1,2)}. \tag{3.4}$$

Since these relations are of importance for our proof of the main theorem (Theorem 2.13), we now prove them.

First, (3.4) follows from (3.3) by using (2.2). To prove (3.3), let $u \in B[K]$, set $u_2 = \mathbf{T}_2 u$, for $x \in K$ set $A_1(x) = \{a : \|xM^1_{a_1}\| > 0\}$, and for $x \in K$ and $a_1 \in A_1(x)$, set $x(a_1) = xM^1_{a_1}/\|xM^1_{a_1}\|$ and $A_2(x,a_1) = \{a_2 \in A_2 : \|x(a_1)M^2_{a_2}\| > 0\}$. From (2.12) we find that

$$u_2(x) = \int_{A_2(x,a_1)} u\left(\frac{xM^2_{a_2}}{\|xM^2_{a_2}\|}\right)\|xM^2_{a_2}\|\,\tau_2(da_2).$$

Hence

$$\begin{aligned}
&\mathbf{T}_1\mathbf{T}_2u(x)\\
&\quad = \int_{A_1(x)} u_2\Big(\frac{xM^1_{a_1}}{\|xM^1_{a_1}\|}\Big)\|xM^1_{a_1}\|\,\tau_1(da_1)\\
&\quad = \int_{A_1(x)}\int_{A_2(x,a_1)} u\Big(\frac{(xM^1_{a_1}/\|xM^1_{a_1}\|)M^2_{a_2}}{\|(xM_{a_1}/\|xM^1_{a_1}\|)M^2_{a_2}\|}\Big)\Big\|\frac{xM^1_{a_1}}{\|xM^1_{a_1}\|}M^2_{a_2}\Big\|\,\tau_2(da_2)\,\|xM^1_{a_1}\|\,\tau_1(da_1)\\
&\quad = \int_{A_1(x)}\int_{A_2(x,a_1)} u\Big(\frac{xM^1_{a_1}M^2_{a_2}}{\|xM^1_{a_1}M^2_{a_2}\|}\Big)\|xM^1_{a_1}M^2_{a_2}\|\,\tau_2(da_2)\,\tau_1(da_1).
\end{aligned}$$

It is easily checked that the set

$$B(x) = \{(a_1,a_2) \in A_1 \times A_2 : \|xM^1_{a_1}M^2_{a_2}\| > 0\}$$

satisfies

$$B(x) = \{(a_1,a_2)\} \in A_1 \times A_2 : a_1 \in A_1(x) \text{ and } a_2 \in A_2(x,a_1)\}.$$

Hence

$$\mathbf{T}_1\mathbf{T}_2u(x) = \int_{B(x)} u\Big(\frac{xM^1_{a_1}M^2_{a_2}}{\|M^1_{a_1}M^2_{a_2}\|}\Big)\|xM^1_{a_1}M^2_{a_2}\|\,\tau^2(da_1,da_2) = \mathbf{T}^{(1,2)}u(x)$$

and thus (3.3) holds.

Next, let $\mathcal{H} = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$ be a fully dominated, regular HMM. For $n = 1,2,\dots$ define $p^n : S \times S \to [0,\infty)$ recursively by $p^1(s,t) = p(s,t)$ and

$$p^{n+1}(s,t) = \int_S p^n(s,\sigma)p(\sigma,t)\,\lambda(d\sigma),$$

define $m^n : S \times S \times A^n \to [0,\infty)$ recursively by $m^1(s,t,a) = m(s,t,a)$ and

$$m^{n+1}(s,t,a^{n+1}) = \int_S m^n(s,\sigma,a^n)m(\sigma,t,a_{n+1})\,\lambda(d\sigma),$$

define $\varrho^{(n)} : A^n \times A^n \to [0,\infty)$ by

$$\varrho^{(n)}(a^n,b^n) = \sum_{i=1}^n \varrho((a^n)_i,(b^n)_i),$$

and define $\tau^n \in \mathcal{Q}^\infty(A^n,\mathcal{A}^n)$ by

$$\tau^n(B_1 \times \cdots \times B_n) = \prod_{i=1}^n \tau(B_i), \qquad B_i \in \mathcal{A},\ i = 1,\dots,n.$$

It is easy to prove that for $n = 1,2,\dots$ we have

$$\mathcal{H}^n = \{(S,\mathcal{F},\delta_0),(p^n,\lambda),(A^n,\mathcal{A}^n,\varrho^{(n)}),(m^n,\tau^n)\},$$

and so by induction $\mathcal{H}^n$ is a fully dominated regular HMM.

Furthermore, if we let $\mathbf{P}^{(N)}$ denote the filter kernel induced by $\mathcal{H}^N$ and let $\mathbf{T}^{(N)}$ denote the transition operator associated to $\mathbf{P}^{(N)}$, then (3.4) and (3.3) imply that

$$\mathbf{P}^N = \mathbf{P}^{(N)} \quad \text{and} \quad \mathbf{T}^N = \mathbf{T}^{(N)}. \tag{3.5}$$

The second of these equalities is used in order to prove that the filter kernel of a regular HMM is Lipschitz equicontinuous and not only Lipschitz continuous, a fact which is crucial to us when proving the main theorem.

From (3.5) and (2.11) it also follows that if $\mathbf{P}$ denotes the filter kernel induced by the fully dominated, regular HMM $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$, then for $n = 2, 3, \dots,$

$$\mathbf{P}^n(x, E) = \int_{A^n(x,E)} \|xM_{a_1} \cdots M_{a_n}\| \, \tau^n(da^n), \tag{3.6}$$

where

$$A^n(x, E) = \left\{(a_1, \dots, a_n) \in A^n : \|xM_{a_1} \cdots M_{a_n}\| > 0, \ \frac{xM_{a_1} \cdots M_{a_n}}{\|xM_{a_1} \cdots M_{a_n}\|} \in E\right\}.$$

3.3. A universal inequality

THEOREM 3.1. *Let* $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ *be a fully dominated, regular HMM. Let* $\mathbf{P} : K \times \mathcal{E} \to [0, 1]$ *be the filter kernel induced by* $\mathcal{H}$*. Then* $\mathbf{P}$ *is Lipschitz equicontinuous with bounding constant* 3.

We first prove the following lemma, which was first proved in [21] under the assumption that the HMM under consideration is a classical HMM determined by a lumping function.

LEMMA 3.2. *Let* $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ *be a fully dominated, regular HMM. Let* $\mathbf{P}$ *denote the filter kernel. Then* $\mathbf{P}$ *is Lipschitz continuous with bounding constant* 3.

Proof. Recall that the transition operator $\mathbf{T} : B[K] \to [B[K]$ is defined by

$$\mathbf{T}u(x) = \int_{A_x^+} u\left(\frac{xM_a}{\|xM_a\|}\right) \|xM_a\| \, \tau(da),$$

where M_a denotes the stepping function determined by $a \in A$ (see (2.7)).

We shall first show that for all $x, y \in K$ and all $u \in \mathrm{Lip}[K]$,

$$|\mathbf{T}u(x) - \mathbf{T}u(y)| \le (\|u\| + 2\gamma(u))\|x - y\|. \tag{3.7}$$

Recall that for $x \in K$, the set A_x^+ is defined as $A_x^+ = \{a : \|xM_a\| > 0\}$. Next note that if $x, y \in K$ and $a \in A$, then

$$|(\|xM_a\| - \|yM_a\|)| \le \|xM_a - yM_a\| = \|(x - y)M_a\|. \tag{3.8}$$

Furthermore, if x and y in K, and f and g in $B_u[S]$ are representatives of x and y respectively, we find that

$$\begin{aligned}\int_A \|(x - y)M_a\| \, \tau(da) &\le \int_A \int_S |f(s) - g(s)| \int_S m(s, t, a) \, \lambda(dt) \, \lambda(ds) \, \tau(da) \\ &= \int_S |f(s) - g(s)| \, \lambda(ds) = \|x - y\|. \end{aligned} \tag{3.9}$$

Next define $B \subset A$ by $B = \{a \in A : \|xM_a\|, \|yM_a\| > 0\}$. Clearly B is an open set, since we have assumed that $\mathcal{H}$ is regular. Define $B_1 = A_x^+ \setminus B$ and $B_2 = A_y^+ \setminus B$.

Obviously B, B_1, B_2 are disjoint, measurable sets. For $u \in \mathrm{Lip}[K]$ we now find that

$$
\begin{aligned}
&|\mathbf{T}u(x) - \mathbf{T}u(y)| \\
&\quad = \left| \int_{B\cup B_1} u\Big(\frac{xM_a}{\|xM_a\|}\Big)\|xM_a\|\,\tau(da) - \int_{B\cup B_2} u\Big(\frac{yM_a}{\|yM_a\|}\Big)\|yM_a\|\,\tau(da)\right| \\
&\quad \le \int_B \Big(\Big|u(\frac{xM_a}{\|xM_a\|}\Big) - u\Big(\frac{yM_a}{\|yM_a\|}\Big|\Big)\|xM_a\|\,\tau(da)| + \|u\|\int_B \big|\|xM_a\| - \|yM_a\|\big|\,\tau(da) \\
&\qquad + \|u\|\int_{B_1} \|xM_a\|\,\tau(da) + \|u\|\int_{B_2} \|yM_a\|\,\tau(da) \\
&\quad \le \gamma(u)\int_B \left\|\frac{xM_a}{\|xM_a\|} - \frac{yM_a}{\|yM_a\|}\right\|\|xM_a\|\,\tau(da) + \|u\|\int_A \big|\|xM_a\| - \|yM_a\|\big|\,\tau(da),
\end{aligned}
$$

and by Lemma 2.1, (3.8) and (3.9),

$$
\begin{aligned}
|\mathbf{T}u(x) - \mathbf{T}u(y)| &\le 2\gamma(u)\int_B \|xM_a - yM_a\|\,\tau(da) + \|u\|\int_A \|xM_a - yM_a\|\,\tau(da) \\
&\le (2\gamma(u) + \|u\|)\|x - y\|,
\end{aligned}
$$

proving (3.7).

From (3.7) it immediately follows that $\gamma(\mathbf{T}u) \le 2\gamma(u) + \|u\|$ for all u in $\mathrm{Lip}[K]$, therefore

$$\gamma(\mathbf{T}u) \le 3, \qquad \forall u \in \mathrm{Lip}_1[K], \tag{3.10}$$

since $\sup\{\|x - y\| : x, y \in K\} = 2$, and then from (3.10) we get

$$\gamma(\mathbf{T}u) \le 3\gamma(u), \qquad \forall u \in \mathrm{Lip}[K]. \tag{3.11}$$

That $\gamma(\mathbf{T}^n u) \le 3\gamma(u)$ also holds for $n \ge 2$ and all $u \in \mathrm{Lip}[K]$ is an immediate consequence of (3.5) and the fact that (3.10) holds for all fully dominated, regular HMMs. Hence the filter kernel $\mathbf{P}$ is Lipschitz equicontinuous with bounding constant 3. ∎

Remark 3.3. It is easy to construct an example which shows that a number strictly less than 2 cannot be a bounding constant (see [25]). We believe though that 2 is a bounding constant for any filter kernel $\mathbf{P}$.

4. On the relationship between HMMs and random systems with complete connections

The purpose of this section is to show that the filter kernel induced by a fully dominated regular HMM can be considered as the Markov kernel of the state sequence associated to a random system with complete connections.

4.1. On random systems with complete connections. We begin with the formal definition of the concept of random systems with complete connections as defined e.g. in [18, Section 1.1].

Definition 4.1. Let $(K, \mathcal{E})$ and $(A, \mathcal{A})$ be measurable sets. Let $h : K \times A \to K$ be a measurable function, and let $Q : K \times \mathcal{A} \to [0, 1])$ be a tr.pr.f. from $(K, \mathcal{E})$ to $(A, \mathcal{A})$. We call the 4-tuple

$$\mathcal{R} = \{(K, \mathcal{E}), (A, \mathcal{A}), h, Q\}$$

a *random system with complete connections* (abbreviated *RSCC*). We call h the *response function*, Q the *index probability function*, $(K, \mathcal{E})$ the *state space*, and $(A, \mathcal{A})$ the *index space*.

Associated to a RSCC $\mathcal{R} = \{(K, \mathcal{E}), (A, \mathcal{A}), h, Q\}$ and an initial distribution $\mu \in \mathcal{P}(K, \mathcal{E})$ there are two stochastic sequences, which we call the state sequence and the index sequence. A simple way to define these stochastic sequences is to first define the HMM associated to a RSCC.

Definition 4.2. Let $\mathcal{R} = \{(K, \mathcal{E}), (A, \mathcal{A}), h, Q\}$ be a RSCC. For $x \in K$ and $E \in \mathcal{E}$, we define $h^{-1}(x, E) \in \mathcal{A}$ by $\{a \in A : h(x, a) \in E\}$. We call the tr.pr.f. $M : K \times \mathcal{E} \otimes \mathcal{A} \to [0, 1]$ defined by

$$M(x, E \times B) = Q(x, h^{-1}(x, E) \cap B) \tag{4.1}$$

the *HMM-kernel associated to the RSCC* $\mathcal{R}$, we call the tr.pr.f. $P : K \times \mathcal{E} \to [0, 1]$ defined by

$$P(x, E) = Q(x, h^{-1}(x, E)) \, (= M(x, E \times A)) \tag{4.2}$$

the *Markov kernel associated to the RSCC* $\mathcal{R}$, and we call the Hidden Markov Model

$$\mathcal{H}_{\mathcal{R}} = \{(K, \mathcal{E}), P, (A, \mathcal{A}), M\},$$

where P and M are defined by (4.2) and (4.1) respectively, the *HMM associated to the RSCC* $\mathcal{R}$.

Furthermore, if $\{X_{n,\mu}, n = 0, 1, 2, \ldots\}$ and $\{Y_{n,\mu}, n = 1, 2, \ldots\}$ denote the hidden Markov chain and the observation sequence generated by the HMM $\mathcal{H}_{\mathcal{R}}$ and the initial distribution $\mu \in \mathcal{P}(K, \mathcal{E})$, we call $\{X_{n,\mu}, n = 0, 1, 2, \ldots\}$ the *state sequence* and $\{Y_{n,\mu}, n = 1, 2, \ldots\}$ the *index sequence* generated by the RSCC $\mathcal{R}$ and the initial distribution $\mu \in \mathcal{P}(K, \mathcal{E})$. If μ is the Dirac measure δ_x at $x \in K$, we usually write $X_n(x)$ instead of X_{n,δ_x}, and $Y_n(x)$ instead of Y_{n,δ_x}.

Remark 4.3. That both $M : K \times \mathcal{E} \otimes \mathcal{A} \to [0, 1]$ defined by (4.1) and $P : K \times \mathcal{E} \to [0, 1]$ defined by (4.2) are tr.pr.fs is well-known and follows for example from [27, Lemma 1.41].

If a RSCC $\mathcal{R} = \{(K, \mathcal{E}), (A, \mathcal{A}), h, Q\}$ is such that Q has a density, that is, if there exist a σ-finite measure τ on $(A, \mathcal{A})$ and a measurable function $q : K \times A \to [0, \infty)$ such that

$$Q(x, B) = \int_B q(x, a) \, \tau(da),$$

we usually denote the RSCC by $\mathcal{R} = \{(K, \mathcal{E}), (A, \mathcal{A}), h, (q, \tau)\}$, and call $q : K \times A \to [0, \infty)$ the *index probability density function.*

Remark 4.4. The classical name for the 4-tuple $\{(K, \mathcal{E}), (A, \mathcal{A}), h, Q\}$ is *random system with complete connections* (see e.g. the book [19] by M. Iosifescu and R. Theodorescu, or the book [18] by M. Iosefescu and S. Grigorescu). Another classical name is *learning model* (see e.g. the book [35] by F. Norman). A later terminology, introduced by M. Barnsley and coworkers, is *iterated function system with place-dependent probabilities* (see e.g. [2]). In the much cited paper by P. Diaconis and D. Freedman [10], the authors consider RSCCs for which the index probability function is independent of the state, and they call such a RSCC simply a *random function*. In learning model theory the index space is called the *event space* and the index sequence $\{Y_{n,\mu},\ n = 1, 2, \ldots\}$ is called the *event sequence* (see e.g. [35]).

The motivation for introducing the concept of RSCC in this paper is that there is a strong connection between the theory of filtering processes and the theory of RSCCs, which we shall describe in the next section.

The study of RSCCs has a long history (see e.g. [19], [35], [23], [18]); here we shall just present a few basic facts. Thus let $\mathcal{R} = \{(K, \mathcal{E}), (A, \mathcal{A}), h, Q\}$ be a given RSCC, let $P : K \times \mathcal{E} \to [0, 1]$ be the Markov kernel associated to $\mathcal{R}$, and let $T : B[K, \mathcal{E}] \to B[K, \mathcal{E}]$ be the transition operator associated to the Markov kernel P. Let us first note that

$$Tu(x) = \int_S u(y)\, P(x, dy) = \int_A u(h(x, a))\, Q(x, da).$$

We shall next introduce a notion which we call the *nth iterate* of a RSCC. Thus, let again $\{(K, \mathcal{E}), (A, \mathcal{A}), h, Q\}$ be a given RSCC. For $n = 1, 2, \ldots$, we let $A_n = A$ and $\mathcal{A}_n = \mathcal{A}$. We define $h^n : K \times A^n \to K$, $n = 1, 2, \ldots$, iteratively by first defining $h^1 = h$, and then setting

$$h^{n+1}(x, a^{n+1}) = h(h^n(x, a^n), a_{n+1}), \qquad n = 1, 2, \ldots,$$

and we define $Q^n : K \times \mathcal{A}^n \to [0, 1]$ iteratively by $Q^1 = Q$ and

$$Q^{n+1}(x, B' \times B_{n+1}) = \int_{B'} \int_{B_{n+1}} Q^n(x, da^n)\, Q(h^n(x, a^n), da_{n+1}), \qquad n = 1, 2, \ldots,$$

where $B' \in \mathcal{A}^n$ and $B_{n+1} \in \mathcal{A}_{n+1}$. It is well-known (see [18]) that h^n is measurable for each positive integer n and that $Q^n : K \times \mathcal{A}^n \to [0, 1]$ is a tr.pr.f. for each n. This implies that the set $\mathcal{R}^n = \{(K, \mathcal{E}), (A^n, \mathcal{A}^n), h^n, Q^n\}$ is also a RSCC for each n. For $n = 2, 3, \ldots$ we call $\mathcal{R}^n$ the *nth iterate* of $\mathcal{R}$, we call $h^n : K \times A^n \to K$ the *nth iterate* of $h : K \times A \to K$ and we call $Q^n : K \times \mathcal{A}^n \to [0, 1]$ the *nth iterate* of $Q : K \times \mathcal{A} \to [0, 1]$.

Now, if $P : K \times \mathcal{E} \to [0, 1]$ is the Markov kernel associated to the RSCC $\mathcal{R} = \{(K, \mathcal{E}), (A, \mathcal{A}), h, Q\}$ and $P^{(n)}$ denotes the Markov kernel associated to the nth iterate $\mathcal{R}^n = \{(K, \mathcal{F}), (A^n, \mathcal{A}^n), h^n, Q^n\}$ of $\mathcal{R}$, then it is easily proved that

$$P^n = P^{(n)}, \qquad n = 2, 3, \ldots.$$

Furthermore, if $u \in B[K, \mathcal{E}]$, $x \in K$, $\{X_n(x),\ n = 0, 1, 2, \ldots\}$ is the state sequence generated by $\mathcal{R}$ and x, and $\{Y_n(x),\ n = 1, 2, \ldots\}$ is the index sequence generated by $\mathcal{R}$ and x, then for $n = 1, 2, \ldots$,

$$\begin{aligned} T^n u(x) &= \int_K u(z)\, P^n(x, dz) = E[u(X_n(x))] = E[u(h^n(x, Y^n(x)))] \\ &= \int_{A^n} u(h^n(x, a^n))\, Q^n(x, da^n) = \int_K u(z)\, P^{(n)}(x, dz) \end{aligned}$$

where of course $Y^n(x) = (Y_1(x), \ldots, Y_n(x))$. We also have, for $n = 1, 2, \ldots$,

$$X_n(x) = h^n(x, Y^n(x)), \qquad x \in K,$$

a fact which we have already used in the previous string of equalities.

Next suppose that the index probability function $Q : K \times \mathcal{A} \to [0,1]$ has a density q with respect to a σ-finite measure τ on $(A, \mathcal{A})$. We then define $q^n : K \times A^n \to [0, \infty)$ iteratively by $q^1 = q$ and

$$q^{n+1}(x, a^{n+1}) = q^n(x, a^n) q(h^n(x, a^n), a_{n+1}), \qquad n = 1, 2, \dots,$$

where $a^{n+1} = (a^n, a_{n+1}) = (a_1, \dots, a_{n+1})$, and then we can express $Q^n : K \times \mathcal{A}^n \to [0,1]$, $n = 1, 2, \dots,$ by

$$Q^n(x, B) = \int_B q^n(x, a^n)\, \tau^n(da^n).$$

We call $q^n : K \times A^n \to [0, \infty)$ the *nth iterate* of $q : K \times A \to [0, \infty)$ and denote the nth iterate of $\mathcal{R}$ by $\{(K, \mathcal{E}), (A^n, \mathcal{A}^n), h^n, (q^n, \tau^n)\}$.

4.2. The RSCC induced by a fully dominated, regular HMM. Let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a fully dominated, regular HMM. As above, let $K = \mathcal{P}_\lambda(S, \mathcal{F})$ and let $\mathcal{E}$ denote the Borel field on K induced by the total variation distance. Furthermore, let M_a denote the stepping function determined by $a \in A$ (see (2.7)).

Now we define $g : K \times A \to [0, \infty)$ by

$$g(x, a) = \|xM_a\|, \tag{4.3}$$

$G : K \times \mathcal{A} \to [0,1]$ by

$$G(x, B) = \int_B g(x, a)\, \tau(da), \tag{4.4}$$

and $h : K \times A \to K$ by

$$h(x, a) = xM_a / \|xM_a\| \quad \text{if } \|xM_a\| > 0, \tag{4.5}$$

$$h(x, a) = x \quad \text{if } \|xM_a\| = 0. \tag{4.6}$$

Since $\mathcal{H}$ is assumed to be fully dominated and regular, it follows immediately that g is continuous. That G is a tr.pr.f. follows from the integral definition of G and the fact that

$$\int_A \|xM_a\|\, \tau(da) = 1, \qquad \forall x \in K.$$

That h is continuous on $\{(x, a) : \|xM_a\| > 0\}$ follows as a simple consequence of Lemma 2.1. Furthermore, since $\overline{M}(x, a)$ is a continuous function, it follows that the set $\{(x, a) : \|xM_a\| = 0\}$ is closed, and it is then easily proved, by using Lemma 2.1 again, that $\{(x, a) : h(x, a) \in B\} \in \mathcal{K} \otimes \mathcal{A}$ is a measurable set if B is an open set in $\mathcal{E}$, from which it follows that $h : K \times A \to K$ is a measurable function. Therefore the 4-tuple $\{(K, \mathcal{E}), (A, \mathcal{A}), h, (g, \tau)\}$ constitutes a RSCC.

Definition 4.5. Let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a regular HMM, and let $g : K \times A \to [0, \infty)$, $G : K \times \mathcal{A} \to [0,1]$ and $h : K \times A \to K$ be defined by (4.3), (4.4) and (4.5)–(4.6) respectively. We call the 4-tuple

$$\mathcal{R}_\mathcal{H} = \{(K, \mathcal{E}), (A, \mathcal{A}), h, (g, \tau)\}$$

the *RSCC induced by* $\mathcal{H}$ and call G the *tr.pr.f. determined by* (g, τ).

Next let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a fully dominated, regular HMM, let $\mathbf{P} : K \times \mathcal{E} \to [0, 1]$ denote the filter kernel induced by $\mathcal{H}$, let $\mathcal{R}_{\mathcal{H}} = \{(K, \mathcal{E}), (A, \mathcal{A}), h, (g, \tau)\}$ be the RSCC induced by $\mathcal{H}$, and let $\mathbf{Q} : K \times \mathcal{E} \to [0, 1]$ denote the Markov kernel associated to the RSCC $\mathcal{R}_{\mathcal{H}}$.

Observation 4.6.

$$\mathbf{Q}(x, E) = \mathbf{P}(x, E), \qquad \forall x \in K, \forall E \in \mathcal{E}.$$

Proof. Let $G : K \times \mathcal{A} \to [0, 1]$ be the tr.pr.f. determined by (g, τ). We have

$$\mathbf{Q}(x, E) = G(x, h^{-1}(x, E)) = \int_{h^{-1}(x,E)} \|xM_a\| \, \tau(da)$$
$$= \int_{\{a:h(x,a)\in E\}} \|xM_a\| \, \tau(da) = \int_{A(x,E)} \|xM_a\| \, \tau(da) = \mathbf{P}(x, E),$$

where $A(x, E)$ is defined by (2.10). ∎

As a trivial consequence of the preceding observation we have the following corollary.

Corollary 4.7. *Let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a fully dominated, regular, strongly ergodic HMM and let $\mathcal{R}_{\mathcal{H}} = \{(K, \mathcal{E}), (A, \mathcal{A}), h, (g, \tau)\}$ be the induced RSCC. Let $\mathbf{Q} : K \times \mathcal{E} \to [0, 1]$ be the Markov kernel associated to $\mathcal{R}_{\mathcal{H}}$.*

In order to prove part (A) *of Theorem* 2.13 *it suffices to prove that $\mathbf{Q}$ is weakly contracting, and to prove* (B)–(D) *it suffices to prove that $\mathbf{Q}$ is weakly ergodic.*

4.3. Some previous results connecting RSCCs and HMMs. Already in 1957, Blackwell proved the following theorem which he applied to the filtering process he was considering.

Theorem 4.8 (see [5, Theorem 2]). *Let $\mathcal{R} = \{(K, \mathcal{E}), (A, \mathcal{A}), h, (q, \tau)\}$ be a RSCC such that $(K, \mathcal{E})$ is a bounded, metric space with metric ϕ, A is a finite set, and τ is the counting measure on A. Further, let the index probability density function $q : K\times A \to [0, \infty)$ satisfy*

$$\inf\{q(x, a) : x \in K, \, a \in A\} > 0 \tag{4.7}$$

and

$$q(\cdot, a) \in \mathrm{Lip}[K], \qquad \forall a \in A, \tag{4.8}$$

and finally, suppose that there exists a number $\rho < 1$ such that

$$\phi(h(x, a), h(y, a)) \le \rho\phi(x, y)), \qquad \forall x, y \in K, \forall a \in A. \tag{4.9}$$

Then there exists at most one invariant measure for the Markov kernel associated to $\mathcal{R}$.

In Section 2.3.3.1 of the book [19] from 1969 the connection between partially observed Markov chains (HMMs) and random systems with complete connections is described, and also in [18] this connection is mentioned in several places.

In the paper [20] from 1973, a HMM with finite state space is considered, and it is proved that if the tr.pr.m. of the hidden Markov chain has strictly positive elements, then the induced RSCC is a so-called *distance diminishing model* as defined by Norman [35, Chapter 2], and from this fact it follows that the filtering process converges in distribution with geometric convergence rate. In the paper [1] from 2012 by C. Anton Popescu, the author considers a HMM with finite state space and general observation space, and gives conditions which imply that the induced RSCC is a distance diminishing model, and again it follows that the convergence to the limit measure has geometric rate.

The connection between filtering processes and random systems with complete connections is also utilized in [21].

5. Two auxiliary theorems

In this rather long chapter we shall state and prove two auxiliary theorems for Markov chains taking values in a bounded, complete, separable, metric space. We shall use the first of these to prove part (D) of Theorem 2.13 and the second to prove parts (A)–(C).

Let $Q : K \times \mathcal{E} \to [0,1]$ be a tr.pr.f. on a measurable space $(K, \mathcal{E})$, and $T : B[K] \to B[K]$ denote the transition operator associated to Q. We define $T^0 u(x) = u(x)$. Recall from the general theory of Markov chains that

$$\operatorname{osc}(T^{n+1}u) \le \operatorname{osc}(T^n u), \qquad n = 0, 1, 2, \dots, \; u \in B[K], \tag{5.1}$$

since T is an "averaging" operator.

Before stating our theorems we shall first introduce two properties which we call the shrinking property and the strong shrinking property.

Definition 5.1. Let Q be a tr.pr.f. on a metric space $(K, \mathcal{E}, \delta)$, and let T be the associated transition operator. If for every $\rho > 0$ there exists a number $0 < \alpha < 1$ and an integer N such that if $n \ge N$ then for all $u \in \operatorname{Lip}[K]$,

$$\operatorname{osc}(T^n u) \le \alpha\rho\gamma(u) + (1-\alpha)\operatorname{osc}(T^{n-N}u), \tag{5.2}$$

then we say that Q has the *strong shrinking property.*

Definition 5.2. Let Q be a tr.pr.f. on a metric space $(K, \mathcal{E}, \delta)$, and let T be the associated transition operator. If for every $\rho > 0$ there exists a number $0 < \alpha < 1$, such that for every nonempty, compact set $E \subset K$ and any $\eta, \kappa > 0$, there exist an integer N and another nonempty, compact set $F \subset K$ such that if $n \ge N$ then for all $u \in \operatorname{Lip}[K]$ we have

$$\operatorname{osc}_E(T^n u) \le \eta\gamma(u) + \kappa \operatorname{osc}(u) + \alpha\rho\gamma(u) + (1-\alpha)\operatorname{osc}_F(T^{n-N}u), \tag{5.3}$$

then we say that Q has the *shrinking property.*

Remark 5.3. We call the constant α occurring in (5.2) and (5.3) a *shrinking number* associated to ρ.

5.1. A simple auxiliary theorem. Our first auxiliary theorem is based on the strong shrinking property.

Theorem 5.4. *Let $(K, \mathcal{E})$ be a bounded, complete, separable, metric space with metric δ, and let Q be a tr.pr.f. on $(K, \mathcal{E})$. If Q has the strong shrinking property then Q is weakly ergodic.*

Proof. We shall first prove that

$$\lim_{n\to\infty} \sup\{\operatorname{osc}(T^n u) : u \in \mathrm{Lip}_1[K]\} = 0. \tag{5.4}$$

We define $\delta(K) = \sup\{\delta(x,y) : x, y \in K\}$. Let $\epsilon > 0$. From the strong shrinking property, we can find $\alpha > 0$ and an integer N such that if $u \in \mathrm{Lip}_1[K]$ and $n > N$, then

$$\operatorname{osc}(T^n u) \le \epsilon\alpha + (1-\alpha)\operatorname{osc}(T^{n-N} u). \tag{5.5}$$

Now define $M = \min\{m : (1-\alpha)^m < \epsilon/\delta(K)\}$. Then, if $n > NM$, it follows from (5.5), (5.1) and $\operatorname{osc}(u) \le \delta(K)$ that if $u \in \mathrm{Lip}_1[K]$, then

$$\begin{aligned}\operatorname{osc}(T^n u) &\le \epsilon\alpha + (1-\alpha)\operatorname{osc}(T^{n-N} u)\\ &\le \epsilon\alpha + (1-\alpha)\big(\epsilon\alpha + (1-\alpha)\operatorname{osc}(T^{n-2N} u)\big)\\ &\le \cdots < \epsilon\alpha\frac{1}{1-(1-\alpha)} + \delta(K)(1-\alpha)^M < 2\epsilon,\end{aligned}$$

and since ϵ is arbitrarily chosen, (5.4) follows.

To complete the proof we shall use the Kantorovich distance on $\mathcal{P}(K,\mathcal{E})$ defined by

$$d_K(\mu,\nu) = \inf\left\{\int_{K\times K} \delta(x,y)\,\tilde{\mu}(dxdy) : \tilde{\mu} \in \tilde{\mathcal{P}}(\mu,\nu,K\times K)\right\}, \tag{5.6}$$

which is well-defined since K is bounded (see Section 2.1 for the definition of $\tilde{\mathcal{P}}(\mu,\nu, K\times K)\})$.

It is well-known that the Kantorovich distance is a metric and that it is equal to, and sometimes defined by, the following supremum:

$$d_K(\mu,\nu) = \sup\left\{\int_K u(x)\,\mu(dx) - \int_K u(y)\,\nu(dy) : u \in \mathrm{Lip}_1[K]\right\}. \tag{5.7}$$

Another well-known fact is that the topology induced by the Kantorovich distance is equivalent to the weak topology (see e.g. [12, Chapter 11]).

Next, let $\mathcal{G}$ denote the Borel field on $\mathcal{P}(K,\mathcal{E})$ generated by the Kantorovich distance. To complete the proof, we shall also use the fact that the measurable space $(\mathcal{P}(K,\mathcal{E}),\mathcal{G})$ is a complete, separable, metric space, since $(K,\mathcal{E})$ is assumed to be a complete, separable space (see e.g. [12, Corollary 11.5.5 and Theorem 11.8.2]).

Next, let $x_0 \in K$. We shall now prove that for every $\epsilon > 0$ we can find an integer N such that, for every integer $m \ge 1$ and every integer $n \ge N$,

$$\sup\left\{\left|\int_K u(y)\,Q^n(x_0,dy) - \int_K u(y)\,Q^{n+m}(x_0,dy)\right| : u \in \mathrm{Lip}_1[K]\right\} < \epsilon. \tag{5.8}$$

Thus, let $\epsilon > 0$ and the integer $m \ge 1$ be given. Set $\nu_{x_0} = \delta_{x_0}Q^m$. Then, if $u \in \mathrm{Lip}_1[K]$, we find, for $n = 1, 2, \ldots,$ that

$$\left|\int_K u(y)\,Q^n(x_0,dy) - \int_K u(y)\,Q^{n+m}(x_0,dy)\right| \le \int_K |T^n u(x_0) - T^n u(y)|\,\nu_{x_0}(dy).$$

From the limit relation (5.4) it follows that we can find an integer N, independent of m, such that for any $u \in \mathrm{Lip}_1[K]$ and all $y \in K$ we have $|T^n u(x_0) - T^n u(y)| < \epsilon$ if $n \ge N$, which implies that (5.8) holds for all $n \geq N$. Now (5.8) and (5.7) yield

$$d_K(Q^n(x_0,\cdot), Q^m(x_0,\cdot)) < \epsilon$$

if $n, m \geq N$. This shows that $\{Q^n(x_0,\cdot),\ n = 1, 2, \ldots\}$ is a Cauchy sequence.

Since $(\mathcal{P}(K,\mathcal{E}), \mathcal{G}, d_K)$ is a complete, separable, metric space, there exists a probability measure μ, say, in $\mathcal{P}(K,\mathcal{E})$, such that

$$\lim_{n\to\infty} d_K(Q^n(x_0,\cdot),\mu) = 0.$$

But since $\lim_{n\to\infty} \sup\{\operatorname{osc}(T^n u) : u \in \mathrm{Lip}_1[K]\} = 0$ by (5.4), we also have

$$\lim_{n\to\infty} d_K(Q^n(x,\cdot),\mu) = 0, \qquad \forall x \in K,$$

which implies that for all $u \in \mathrm{Lip}_1[K]$,

$$\lim_{n\to\infty} \int_K u(y)\, Q^n(x,dy) - \int_K u(y)\, \mu(dy) = 0, \qquad \forall x \in K. \tag{5.9}$$

But if (5.9) holds for all $u \in \mathrm{Lip}_1[K]$, it also holds for all $u \in \mathrm{Lip}[K]$. Therefore by [6, proof of Theorem 2.1], we conclude that $\limsup_{n\to\infty} Q^n(x,F) \leq \mu(F)$ for all closed sets $F \in \mathcal{E}$, and then [6, Theorem 2.1] shows that (5.9) holds for all $u \in C[K]$ and all $x \in K$. Hence Q is weakly ergodic with limit measure μ. Thus Theorem 5.4 is proved. ∎

Next, we shall introduce a condition which together with Lipschitz equicontinuity implies the strong shrinking property.

Definition 5.5. Let $(K,\mathcal{E})$ be a complete, separable, metric space with metric δ, and let Q be a tr.pr.f. on $(K,\mathcal{E})$. If for every $\epsilon > 0$ there exists an integer N and a number $\alpha > 0$ such that for any two elements x and y in K there exists a coupling $\tilde{\mu}_{N,x,y}$ of $Q^N(x,\cdot)$ and $Q^N(y,\cdot)$ such that

$$\tilde{\mu}_{N,x,y}(D_\epsilon) \geq \alpha \tag{5.10}$$

where

$$D_\epsilon = \{z_1, z_2 \in K : \delta(z_1,z_2) < \epsilon\},$$

then we say that *Condition* C1 is satisfied.

Remark 5.6. For an early version of Condition C1 see e.g. [22].

Proposition 5.7. *Let $(K,\mathcal{E})$ be a complete, separable, metric space with metric δ, and Q be a tr.pr.f. on $(K,\mathcal{E})$. Suppose Q is Lipschitz equicontinuous and Condition* C1 *is satisfied. Then Q has the strong shrinking property.*

Proof. Let $\rho > 0$. We want to determine an integer N and an number α such that (5.2) holds. Let T denote the transition operator associated to Q. Since Q is Lipschitz equicontinuous, there exists a constant $C \geq 1$ such that for all $u \in \mathrm{Lip}[K]$,

$$\gamma(T^n u) \leq C\gamma(u), \qquad n = 1, 2, \ldots. \tag{5.11}$$

Define $\rho_1 = \rho/C$. Since Condition C1 holds, we can find an integer N and a number α such that for any pair $(x,y) \in K \times K$ we can find a coupling $\tilde{\mu}_{N,x,y}$ of $Q^N(x,\cdot)$ and $Q^N(y,\cdot)$ such that

$$\tilde{\mu}_{N,x,y}(D_{\rho_1}) \geq \alpha. \tag{5.12}$$

From (5.12) it follows that if $u \in \mathrm{Lip}[K]$ then

$$|T^N u(x) - T^N u(y)| = \left| \int_{K\times K} (u(z_1) - u(z_2))\, \tilde{\mu}_{N,x,y}(dz_1, dz_2) \right|$$
$$\leq \mathrm{osc}(u)(1-\alpha) + \alpha\gamma(u)\rho_1. \tag{5.13}$$

Now let $n \geq N$ and set $m = n - N$. From (5.1) and (5.11) it follows that

$$|T^n u(x) - T^n u(y)| \leq \mathrm{osc}(T^m u)(1-\alpha) + \alpha\gamma(T^m u)\rho_1$$
$$\leq \mathrm{osc}(T^m u)(1-\alpha) + \alpha\gamma(u)\rho.$$

Hence

$$\mathrm{osc}(T^n u) \leq \mathrm{osc}(T^{n-N} u)(1-\alpha) + \alpha\gamma(u)\rho$$

which is what we wanted to prove. ■

Remark 5.8. A tr.pr.f. Q on a metric space which is Feller continuous and satisfies Condition C1 need not be weakly ergodic. For two simple counterexamples see [23].

For the sake of completeness let us prove that if in Theorem 4.8 we assume that the state space of the RSCC is a bounded, complete, separable, metric space, then the associated filter kernel is weakly ergodic.

Theorem 5.9. *Let $\mathcal{R} = \{(K, \mathcal{E}), (A, \mathcal{A}), h, (q, \tau)\}$ be a RSCC having the same properties as the RSCC considered in Theorem* 4.8. *Assume also that $(K, \mathcal{E})$ is complete and separable. Let $P : K \times \mathcal{E} \to [0, 1]$ be the associated tr.pr.f. Then P is weakly ergodic.*

Proof. That Condition C1 holds follows from (4.7) and (4.9) together with the fact that the state space is bounded. That the tr.pr.f. P is Lipschitz equicontinuous with bounding constant $D + L/(1-\rho)$, where $L = \sup\{\sum_a |q(x, a) - q(y, a)|/\delta(x, y) : x \neq y,\ x, y \in K\}$ and $D = \sup\{\delta(x, y) : x, y \in K\}$, follows from (4.8) and (4.9). The conclusion then follows from Proposition 5.7 and Theorem 5.4. ■

Remark 5.10. If the state space of the RSCC considered in Theorems 4.8 and 5.9 is compact, then it follows from the general theory on RSCCs that the convergence rate is in fact geometric (see [35, Chapter 3]).

5.2. A second auxiliary theorem. In this section we prove a slightly more complicated auxiliary theorem.

Theorem 5.11. *Let $(K, \mathcal{E})$ be a complete, separable, bounded, metric space with metric δ, and let Q be a tr.pr.f. on $(K, \mathcal{E})$. Suppose that Q has the shrinking property. Then Q is weakly contracting.*

Proof. Set $D = \sup\{\delta(x, y) : x, y \in K\}$. Since K is assumed to be bounded we have $D < \infty$. Since furthermore it is not difficult to prove that the shrinking property also holds if we replace δ by $2\delta/D$, it is clearly no loss of generality to assume that $D = 2$.

In order to prove that Q is weakly contracting, we need to show that for all $x, y \in K$,

$$\lim_{n\to\infty} \sup\left\{ \left| \int_K u(z)\, Q^n(x, dz) - \int_K u(z)\, Q^n(y, dz) \right| : u \in \mathrm{Lip}_1[K] \right\} = 0. \tag{5.14}$$

Let $\epsilon > 0$, $x, y \in K$ and $u \in \mathrm{Lip}_1[K]$. In order to prove (5.14), we shall show that we can find an integer N, which may depend on x and y, but which does not depend on u, such that

$$\left| \int_K u(z)\, Q^n(x, dz) - \int_K u(z)\, Q^n(y, dz) \right| < 6\epsilon, \qquad \forall n \geq N. \tag{5.15}$$

This is not difficult to do if one uses the shrinking property. We first choose the number ρ sufficiently small, more precisely we set $\rho = \epsilon$. Next, let α be a shrinking number associated to ρ. Since $\{x, y\}$ is a compact set, it follows from the shrinking property that if we define $\eta = \eta_1 = \epsilon/2$ and $\kappa = \kappa_1 = \epsilon/2$, then we can find an integer N_1 and a compact set E_1 such that, if $n \geq N_1$, then

$$\begin{aligned} |\langle u, Q^n(x, \cdot)\rangle - \langle u, Q^n(y, \cdot)\rangle| &= |T^n u(x) - T^n u(y)| \\ &\leq \eta_1 + 2\kappa_1 + \alpha\epsilon + (1-\alpha) \sup_{z_1, z_2 \in E_1} |T^{n-N_1} u(z_1) - T^{n-N_1} u(z_2)|, \end{aligned}$$

where we have used the fact that $\gamma(u) \leq 1$, $\mathrm{osc}(u) \leq 2$ and $\rho = \epsilon$.

We now choose

$$M = \min\{m : (1-\alpha)^m < \epsilon\}.$$

For $i = 2, \dots, M$, we define $\eta_i = \epsilon/2^i$ and $\kappa_i = \epsilon/2^i$, and having defined the compact sets E_i, $i = 1, \dots, j-1$, and the integers N_i, $i = 1, \dots, j-1$, it follows from the shrinking property that we can find a compact set E_j and an integer N_j, such that

$$\begin{aligned} &\sup_{z_1, z_2 \in E_{j-1}} |T^n u(z_1) - T^n u(z_2)| \\ &\qquad \leq \eta_j + 2\kappa_j + \alpha\rho + (1-\alpha) \sup_{z_1, z_2 \in E_j} |T^{n-N_j} u(z_1) - T^{n-N_j} u(z_2)| \end{aligned} \tag{5.16}$$

if $n \geq N_j$. By using (5.16) repeatedly it follows that if $n \geq N_1 + \cdots + N_j$, then

$$\begin{aligned} &|T^n u(x) - T^n u(y)| \\ &\quad \leq \epsilon/2 + 2\epsilon/2 + \alpha\epsilon + (1-\alpha) \sup_{z_1, z_2 \in E_1} |T^{n-N_1} u(z_1) - T^{n-N_1} u(z_2)| \\ &\quad \leq \sum_{i=1}^{j} \epsilon/2^i + 2\sum_{i=1}^{j} \epsilon/2^i + \epsilon\alpha\big(1 + (1-\alpha) + (1-\alpha)^2 + \cdots + (1-\alpha)^{j-1}\big) \\ &\qquad + (1-\alpha)^j \sup_{z_1, z_2 \in E_j} |T^{n-(N_1+\cdots+N_j)} u(z_1) - T^{n-(N_1+\cdots+N_j)} u(z_2)|. \end{aligned}$$

In particular, if $j = M$ and $n \geq N_1 + \cdots + N_M$, then

$$\begin{aligned} &|T^n u(x) - T^n u(y)| \\ &\quad \leq \sum_{i=1}^{M} \epsilon/2^i + 2\sum_{i=1}^{M} \epsilon/2^i + \epsilon\alpha\big(1 + (1-\alpha) + (1-\alpha)^2 + \cdots + (1-\alpha)^{M-1}\big) \\ &\qquad + (1-\alpha)^M \sup_{z_1, z_2 \in E_M} |T^{n-N} u(z_1) - T^{n-N} u(z_2)|, \end{aligned}$$

where $N = N_1 + \cdots + N_M$, and by using $\mathrm{osc}(Tu) \leq \mathrm{osc}(u)$, $\mathrm{osc}(u) \leq 2$ and the fact that

$$\epsilon\alpha\big(1 + (1-\alpha) + (1-\alpha)^2 + \cdots + (1-\alpha)^M\big) < \epsilon,$$

we find that if $n \geq N$, then

$$|T^n u(x) - T^n u(y)| < \epsilon + 2\epsilon + \epsilon + 2(1-\alpha)^M \leq 4\epsilon + 2(1-\alpha)^M,$$

and since $(1-\alpha)^M < \epsilon$, it follows that

$$|T^n u(x) - T^n u(y)| = \left| \int_K u(z)\, Q^n(x, dz) - \int_K u(z)\, Q^n(y, dz) \right| < 6\epsilon$$

if $n \geq N$. Thus (5.15) holds, so (5.14) is satisfied. Hence Q is weakly contracting. ∎

Remark 5.12. The arguments used above are similar to arguments found in the paper by A. Lasota and J. Yorke [32] (see in particular [32, proof of Theorem 4.1]).

A natural conjecture is that if a tr.pr.f. is weakly contracting, then it can have at most one invariant measure. We shall prove this if also the tr.pr.f. is Lipschitz equicontinuous. We first prove the following lemma.

Lemma 5.13. *Let $(K, \mathcal{E})$ be a complete, separable, bounded, metric space with metric δ, let Q be a tr.pr.f. on $(K, \mathcal{E})$ and suppose that Q is weakly contracting. Suppose also that Q is Lipschitz equicontinuous. Then for every nonempty, compact set $E \in \mathcal{E}$ and every $\epsilon > 0$, we can find an integer N such that for any $u \in \mathrm{Lip}_1[K]$,*

$$\sup_{x,y \in E} \left| \int_K u(z)\, Q^n(x, dz) - \int_K u(z)\, Q^n(y, dz) \right| \leq \epsilon \tag{5.17}$$

for all $n \geq N$.

Proof. Let $E \in \mathcal{E}$ and $\epsilon > 0$, where E is a nonempty, compact set. Since we have assumed that Q is Lipschitz equicontinuous, there exists a constant $C > 1$ such that for all $n \geq 1$,

$$\left| \int_K u(z)\, Q^n(x, dz) - \int_K u(z)\, Q^n(y, dz) \right| \leq C\delta(x, y)\gamma(u) \tag{5.18}$$

for all $x, y \in K$ and all $u \in \mathrm{Lip}[K]$.

Next, set $\epsilon_1 = \epsilon/(3C)$. Since E is compact, we can find a finite set $\mathcal{M} = \{x_i,\ i = 1, \ldots, M\}$ such that for every $x \in E$, we have $\inf\{\delta(x, x_i) : x_i \in \mathcal{M}\} < \epsilon_1$. From (5.14) it follows that we can find an integer N such that if $n \geq N$, then

$$\left| \int_K u(z)\, Q^n(x_i, dz) - \int_K u(z)\, Q^n(x_j, dz) \right| < \frac{\epsilon}{3}$$

for all $u \in \mathrm{Lip}_1[K]$ and all pairs $(x_i, x_j) \in \mathcal{M} \times \mathcal{M}$.

Now let $x, y \in E$, choose $x_i \in \mathcal{M}$ such that $\delta(x, x_i) < \epsilon_1$, and $x_j \in \mathcal{M}$ such that $\delta(y, x_j) < \epsilon_1$. Let $u \in \mathrm{Lip}_1[K]$. Using the triangle inequality, (5.18) and $\epsilon_1 = \epsilon/(3C)$, we now find that if $n \geq N$, then

$$\begin{aligned}
&\left|\int_K u(z)\,Q^n(x,dz) - \int_K u(z)\,Q^n(y,dz)\right| \\
&\quad \le \left|\int_K u(z)\,Q^n(x,dz) - \int_K u(z)\,Q^n(x_i,dz)\right| + \left|\int_K u(z)\,Q^n(x_i,dz) - \int_K u(z)\,Q^n(x_j,dz)\right| \\
&\qquad + \left|\int_K u(z)\,Q^n(x_j,dz) - \int_K u(z)\,Q^n(y,dz)\right| \\
&\quad < C\delta(x,x_i) + \epsilon/3 + C\delta(x_j,y) = C\epsilon_1 + \epsilon/3 + C\epsilon_1 = \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.
\end{aligned}$$

Hence,

$$\sup\left\{\left|\int_K u(z)\,Q^n(x,dz) - \int_K u(z)\,Q^n(y,dz)\right| : x,y \in E,\ u \in \mathrm{Lip}_1[K]\right\} \le \epsilon$$

if $n \ge N$, and the lemma is proved. ■

The following proposition will be useful when proving part (B) of Theorem 2.13.

PROPOSITION 5.14. *Let $(K,\mathcal{E})$ be a complete, separable, bounded, metric space with metric δ, let Q be a tr.pr.f. on $(K,\mathcal{E})$ and suppose that Q is weakly contracting. Suppose also that Q is Lipschitz equicontinuous and μ is an invariant measure associated to Q. Then Q is weakly ergodic with limit measure μ.*

Proof. Let T denote the transition operator associated to Q. We want to show that for every $x \in K$ and every $u \in C[K]$,

$$\lim_{n\to\infty} T^n u(x) = \langle u,\mu\rangle. \tag{5.19}$$

We first prove (5.19) when $u \in \mathrm{Lip}_1[K]$. Thus, let $x \in K$, $u \in \mathrm{Lip}_1[K]$ and $\epsilon > 0$. Evidently, (5.19) holds trivially if $u \equiv 0$. Hence we can assume that $u \not\equiv 0$.

Next, let E be a compact set so large that $x \in E$ and

$$\mu(K \setminus E) < \epsilon/\mathrm{osc}(u).$$

From Lemma 5.13, we can find an integer N such that if $y \in K$ then

$$|T^n u(x) - T^n(y)| < \epsilon/2, \qquad n \ge N.$$

Hence, if $n \ge N$ we find that

$$\begin{aligned}
|T^n u(x) - \langle u,\mu\rangle| &\le \int_K |T^n u(x) - T^n u(y)|\,\mu(dy) \\
&\le \mathrm{osc}(u)\int_{K\setminus E}\mu(dy) + \int_E |T^n u(x) - T^n u(y)|\,\mu(dy) \le \epsilon/2 + \epsilon/2 = \epsilon,
\end{aligned}$$

so (5.19) holds for all $u \in \mathrm{Lip}_1[K]$ and all $x \in K$.

That (5.19) also holds for all $u \in \mathrm{Lip}[K]$ and all $x \in K$ follows from the fact that if $u \in \mathrm{Lip}[K]$ and $\gamma(u) > 0$ then $v = u/\gamma(u) \in \mathrm{Lip}_1[K]$. Then, by using the same argument as used in [6] when proving that (ii)⇒(iii) in [6, Theorem 2.1], it follows that $\limsup_{n\to\infty} Q^n(x,F) \le \nu(F)$ for all closed sets $F \in \mathcal{E}$. Now by [6, Theorem 2.1], we find that (5.19) holds for all $u \in C[K]$ and all $x \in K$. Hence Q is weakly ergodic with limit measure μ. ■

Corollary 5.15. *Let $(K,\mathcal{E})$ be a complete, separable, bounded, metric space with metric δ, let Q be a tr.pr.f. on $(K,\mathcal{E})$ and suppose that Q is weakly contracting. Suppose also that Q is Lipschitz equicontinuous. Then Q has at most one invariant measure.*

Proof. Follows immediately from Proposition 5.14. ■

Before concluding this section we prove the fairly obvious fact that tightness and Lipschitz continuity imply the existence of an invariant measure.

Proposition 5.16. *Let $(K,\mathcal{E})$ be a complete, separable, bounded, metric space with metric δ, and let Q be a tr.pr.f. on $(K,\mathcal{E})$. Suppose also that Q is Lipschitz continuous and there exists $x^* \in K$ such that $\{Q^n(x^*,\cdot),\ n=1,2,\ldots\}$ is a tight sequence. Then Q has an invariant probability measure.*

Proof. We shall use a classical argument due to Krylov and Bogolyubov (see e.g. [34, Section 32.2]) together with the fact that Q is Lipschitz continuous.

Let T denote the transition operator associated to Q. For $n=1,2,\ldots$ we define $T^{(n)} = (1/n)\sum_{k=1}^n T^k$ and $Q^{(n)} = (1/n)\sum_{k=1}^n Q^k$. Now, since $\{Q^n(x^*,\cdot),\ n=1,2,\ldots\}$ is a tight sequence, so is $\{Q^{(n)}(x^*,\cdot),\ n=1,2,\ldots\}$. Therefore we can extract a subsequence n_j, $j=1,2,\ldots$, such that $\{Q^{(n_j)}(x^*,\cdot),\ j=1,2,\ldots\}$ converges weakly towards a probability measure ν, say. Hence

$$\lim_{j\to\infty} T^{(n_j)}u(x^*) = \langle u,\nu\rangle \tag{5.20}$$

for all $u\in C[K]$.

Now assume that $u\in \mathrm{Lip}[K]$. By considering the sequence $\{T^{(n_j+1)}u(x^*),\ j=1,2,\ldots\}$ it is easily proved that on the one hand,

$$\lim_{j\to\infty} T^{(n_j+1)}u(x^*) = \lim_{j\to\infty} T^{(n_j)}u(x^*) = \langle u,\nu\rangle,$$

and on the other hand,

$$\lim_{j\to\infty} T^{(n_j+1)}u(x^*) = \lim_{j\to\infty} T^{(n_j)}Tu(x^*) = \langle Tu,\nu\rangle = \langle u,\nu Q\rangle,$$

where we have used the fact that $Tu\in \mathrm{Lip}[K]$ if $u\in\mathrm{Lip}[K]$.

Hence, if $u\in\mathrm{Lip}[K]$, then

$$\langle u,\nu Q\rangle = \langle u,\nu\rangle, \tag{5.21}$$

and since the set of Lipschitz continuous functions is measure determining, it follows that (5.21) also holds for $u\in C[K]$, which we wanted to prove. ■

Corollary 5.17. *Let $(K,\mathcal{E})$ be a complete, separable, bounded, metric space with metric δ, let Q be a tr.pr.f. on $(K,\mathcal{E})$ and suppose that Q is weakly contracting. Suppose also that Q is Lipschitz equicontinuous and there exists $x^*\in K$ such that $\{Q^n(x^*,\cdot),\ n=1,2,\ldots\}$ is a tight sequence. Then Q is weakly ergodic.*

Proof. Follows immediately from Propositions 5.14 and 5.16. ■

Remark 5.18. We end this section by raising the following general question: *Does Lipschitz continuity imply Feller continuity?*

5.3. Some further auxiliary results. The main purpose of this section is to introduce another contact condition (Condition C2 below) and to show how this condition together with Lipschitz equicontinuity implies the shrinking property.

We first define Condition E in a slightly more general setting.

DEFINITION 5.19. Let $(K, \mathcal{E}, \delta)$ be a complete, separable, metric space and let Q be a tr.pr.f. on $(K, \mathcal{E})$. Suppose that the nonvoid subset $\mathcal{P}_0$ of $\mathcal{P}(K, \mathcal{E})$ is such that for every $\rho > 0$, there exist an integer N and a number α such that for any measures μ and ν in $\mathcal{P}_0$, there exists a coupling $\tilde{\mu}_N$ of μQ^N and νQ^N such that

$$\tilde{\mu}_N(\{(x, y) \in K \times K : \delta(x, y) < \rho\}) \geq \alpha.$$

We then say that the pair $(\mathcal{P}_0, Q)$ satisfies *Condition E*, or simply that $\mathcal{P}_0$ satisfies Condition E.

We first prove the following lemma.

LEMMA 5.20. *Let $(K, \mathcal{E}, \delta)$ be a complete, separable, bounded, metric space, let Q be a tr.pr.f. on $(K, \mathcal{E})$ which is Lipschitz equicontinuous, let T be the associated transition operator, and suppose that the nonempty subset $\mathcal{P}_0 \subset \mathcal{P}(K, \mathcal{E})$ is such that $(\mathcal{P}_0, Q)$ satisfies Condition E. Then:*

A. *for every $\rho > 0$, there exists $\alpha > 0$ and an integer N such that for any two probability measures μ and ν in $\mathcal{P}_0$,*

$$|\langle u, \mu Q^n \rangle - \langle u, \nu Q^n \rangle| \leq \alpha\gamma(u)\rho + (1 - \alpha) \operatorname{osc}(T^{n-N} u)$$

if $u \in \mathrm{Lip}[K]$ and $n \geq N$;

B. *for every $\rho > 0$, there exists $\alpha > 0$ and an integer N such that for any probability measures μ and ν in $\mathcal{P}_0$ and any $\kappa > 0$, there exists a compact set F such that*

$$|\langle u, \mu Q^n \rangle - \langle u, \nu Q^n \rangle| \leq \alpha\gamma(u)\rho + \kappa \operatorname{osc}(u) + (1 - \alpha) \operatorname{osc}_F(T^{n-N} u)$$

if $u \in \mathrm{Lip}[K]$ and $n \geq N$.

Proof. Since Q is Lipschitz equicontinuous there exists $C \geq 1$ such that for $u \in \mathrm{Lip}[K]$,

$$\gamma(T^n u) \leq C\gamma(u), \qquad n = 1, 2, \ldots. \tag{5.22}$$

Next, let $\rho > 0$ and set $\rho_1 = \rho/C$. Since $\mathcal{P}_0$ satisfies Condition E, there exist $\alpha > 0$ and an integer N such that for any two probability measures μ and ν in $\mathcal{P}_0$, there exists a coupling $\tilde{\mu}_N$ of μQ^N and νQ^N such that if $D_{\rho_1} = \{(x, y) \in K^2 : \delta(x, y) < \rho_1\}$, then

$$\tilde{\mu}_N(D_{\rho_1}) \geq \alpha.$$

Hence, if $u \in \mathrm{Lip}[K]$, $n \geq N$ and we set $v = T^{n-N} u$, we find that

$$\begin{aligned}
|\langle u, \mu Q^n \rangle - \langle u, \nu Q^n \rangle| &= \left| \int_K T^n u(z)\, \mu(dz) - \int_K T^n u(z)\, \nu(dz) \right| \\
&= \left| \int_K T^{n-N} u(z) \mu\, Q^N(dz) - \int_K T^{n-N} u(z) \nu\, Q^N(dz) \right| \\
&= \left| \int_{K \times K} (v(z) - v(z'))\, \tilde{\mu}_N(dz, dz') \right|.
\end{aligned} \tag{5.23}$$

Next set

$$B_1 = \{(z, z') \in K^2 : \delta(z, z') < \rho_1\} \quad \text{and} \quad B_2 = \{(z, z') \in K^2 : \delta(z, z_1) \geq \rho_1\}.$$

Using the fact that $\gamma(T^m u) \leq C\gamma(u)$ for all $m \geq 1$ and that

$$b \min\{\epsilon, \Theta\} + (1-b)\Theta \leq a\epsilon + (1-a)\Theta \tag{5.24}$$

if

$$0 < a \leq b \leq 1, \qquad \epsilon > 0 \quad \text{and} \quad \Theta > 0,$$

we obtain

$$\begin{aligned}
&\left| \int_{K \times K} (v(z) - v(z'))\, \tilde{\mu}_N(dz, dz') \right| \\
&\qquad \leq \left| \int_{B_1} (v(z) - v(z'))\, \tilde{\mu}_N(dz, dz') \right| + \left| \int_{B_2} (v(z) - v(z'))\, \tilde{\mu}_N(dz, dz') \right| \\
&\qquad \leq \min\{\operatorname{osc}(v), \gamma(v)(\rho/C)\}\tilde{\mu}_N(B_1) + \operatorname{osc}(v)(1 - \tilde{\mu}_N(B_1)) \\
&\qquad \leq \gamma(v)(\rho/C)\alpha + (1-\alpha)\operatorname{osc}(v) \leq \gamma(u)\rho\alpha + (1-\alpha)\operatorname{osc}(T^{n-N}u),
\end{aligned}$$

which combined with (5.23) implies that

$$|\langle u, \mu Q^n \rangle - \langle u, \nu Q^n \rangle| \leq \alpha\gamma(u)\rho + (1-\alpha)\operatorname{osc}(T^{n-N}u),$$

and hence part A is proved.

Next, let $\kappa > 0$. Since $(K, \mathcal{E})$ is a complete, separable, metric space, there exists a compact set $F \in \mathcal{E}$ such that

$$\tilde{\mu}((K \setminus F) \times (K \setminus F)) \leq \kappa. \tag{5.25}$$

Further, define

$$\begin{aligned}
B_3 &= \{(z, z') \in K \times K : \delta(z, z') < \rho_1,\ z, z' \in F\}, \\
B_4 &= \{(z, z') \in K \times K : \delta(z, z') \geq \rho_1,\ z, z' \in F\}, \\
B_5 &= K \times K \setminus (B_3 \cup B_4).
\end{aligned}$$

Then

$$\begin{aligned}
&\left| \int_{K \times K} (v(z) - v(z'))\, \tilde{\mu}_N(dz, dz') \right| \\
&\qquad \leq \left| \int_{B_3} (v(z) - v(z'))\, \tilde{\mu}_N(dz, dz') \right| \\
&\qquad\quad + \left| \int_{B_4} (v(z) - v(z'))\, \tilde{\mu}_N(dz, dz') \right| + \left| \int_{B_5} (v(z) - v(z'))\, \tilde{\mu}_N(dz, dz') \right| \\
&\qquad \leq \min\{\operatorname{osc}_F(v), \rho_1\gamma(v)\}\tilde{\mu}(B_3) + \operatorname{osc}_F(v)(1 - \tilde{\mu}(B_3)) + \operatorname{osc}(v)\tilde{\mu}(B_5),
\end{aligned}$$

and by using (5.23)–(5.25), $\gamma(v) \leq C\gamma(u)$ and the fact that $\operatorname{osc}(T^n u) \leq \operatorname{osc}(u)$ for all integers $n \geq 1$, we find that

$$\left| \int_{K \times K} (v(z) - v(z'))\, \tilde{\mu}(dz, dz') \right| \leq \alpha\gamma(u)\rho + (1-\alpha)\operatorname{osc}_F(v) + \kappa \operatorname{osc}(u)$$

which together with (5.23) and $v = T^{n-N}u$ implies that

$$|\langle u, \mu Q^n \rangle - \langle u, \nu Q^n \rangle| \leq \alpha\gamma(u)\rho + (1-\alpha)\operatorname{osc}_F(T^{n-N}u) + \kappa \operatorname{osc}(u),$$

hence part B is proved and the proof of Lemma 5.20 is complete. ∎

Before we introduce two further "contact" conditions, recall that $d_K : \mathcal{P}(K,\mathcal{E}) \times \mathcal{P}(K,\mathcal{E}) \to [0,2]$ denotes the Kantorovich distance on $\mathcal{P}(K,\mathcal{E})$ (see Section 5.1).

DEFINITION 5.21. Let $(K,\mathcal{E},\delta)$ be a complete, separable, metric space and let Q be a tr.pr.f. on $(K,\mathcal{E})$. We say that *Condition* C2 is satisfied if there exists a nonvoid subset $\mathcal{P}_0$ of $\mathcal{P}(K,\mathcal{E})$ satisfying Condition E and such that for every $\epsilon > 0$ and every $x \in K$ there exists an integer N such that if $n \geq N$ then

$$\inf\{d_K(Q^n(x,\cdot),\nu) : \nu \in \mathcal{P}_0\} < \epsilon. \tag{5.26}$$

If also N is independent of $x \in K$, we say that *Condition* C3 is satisfied.

PROPOSITION 5.22. *Let $(K,\mathcal{E},\delta)$ be a complete, separable, metric space, and let Q be a tr.pr.f. on $(K,\mathcal{E})$. If Q is Lipschitz equicontinuous and Condition* C2 *is satisfied, then Q has the shrinking property. If also Condition* C3 *is satisfied, then Q has the strong shrinking property.*

Proof. We first prove that Q has the strong shrinking property if Condition C3 is satisfied.

Let $\rho > 0$ and T denote the transition operator associated to Q. We want to prove that we can find an integer N and $\alpha > 0$ such that if $n \geq N$, then for all $u \in \mathrm{Lip}[K]$,

$$\mathrm{osc}(T^n u) \leq \alpha\rho\gamma(u) + (1-\alpha)\,\mathrm{osc}(T^{n-N}u).$$

Since Condition C3 is satisfied, there exists a set $\mathcal{P}_0 \subset \mathcal{P}(K,\mathcal{E})$ satisfying Condition E and such that for every $\epsilon > 0$ we can find an integer N such that for all $x \in K$, (5.26) holds.

Since Q is Lipschitz equicontinuous there exists $C \geq 1$ such that for $u \in \mathrm{Lip}[K]$,

$$\gamma(T^n u) \leq C\gamma(u), \qquad n = 1,2,\ldots. \tag{5.27}$$

Now set $\rho_1 = \rho/2C$. Since $\mathcal{P}_0$ satisfies Condition E, Lemma 5.20 yields $\alpha > 0$ and an integer N_2 such that for any probability measures μ and ν in $\mathcal{P}_0$, we have

$$|\langle u,\mu Q^n\rangle - \langle u,\nu Q^n\rangle| \leq \alpha\gamma(u)\rho_1 + (1-\alpha)\,\mathrm{osc}(T^{n-N_2}u) \tag{5.28}$$

if $u \in \mathrm{Lip}[K]$ and $n \geq N_2$.

Next, let $x,y \in K$. From Condition C3 it follows that we can find an integer N_1 and probability measures ν_x and ν_y in $\mathcal{P}_0$ such that if $n \geq N_1$ then

$$d_K(Q^n(x,\cdot),\nu_x) < \alpha\rho_1/2 \quad \text{and} \quad d_K(Q^n(y,\cdot),\nu_y) < \alpha\rho_1/2. \tag{5.29}$$

Now set $N = N_1 + N_2$, let $n \geq N$ and set $m = n - N_1$. From (5.27)–(5.29) and the definition of the Kantorovich distance, it follows that

$$\begin{aligned}
|T^n u(x) - T^n u(y)| &= |\langle T^m u, \delta_x Q^{N_1}\rangle - \langle T^m u, \delta_y Q^{N_1}\rangle| \\
&\leq |\langle T^m u, \nu_x\rangle - \langle T^m u, \nu_y\rangle| + \gamma(T^m u)\alpha\rho_1 \\
&\leq |\langle u, \nu_x Q^m\rangle - \langle u, \nu_y Q^m\rangle| + C\gamma(u)\alpha\rho_1 \\
&\leq \alpha\gamma(T^m u)\rho_1 + (1-\alpha)\,\mathrm{osc}(T^{m-N_2}u) + \gamma(u)\alpha\rho/2 \\
&\leq \alpha C\gamma(u)\rho/2C + (1-\alpha)\,\mathrm{osc}(T^{n-N}u) + \gamma(u)\alpha\rho/2 \\
&= \alpha\gamma(u)\rho + (1-\alpha)\,\mathrm{osc}(T^{n-N}u).
\end{aligned}$$

Hence

$$\operatorname{osc}(T^n u) \le \alpha\gamma(u)\rho + (1-\alpha)\operatorname{osc}(T^{n-N}u),$$

and thus the strong shrinking property holds.

We shall next prove that if only Condition C2 is satisfied, then Q has the shrinking property.

Let $\rho > 0$. We want to prove that if Condition C2 is satisfied, then we can find $\alpha > 0$ such that for every nonempty, compact set $E \in \mathcal{E}$ and any $\eta, \kappa > 0$, we can find a nonempty compact set F and an integer N such that

$$\operatorname{osc}_E(T^n u) \le \eta\gamma(u) + \kappa\operatorname{osc}(u) + \alpha\rho\gamma(u) + (1-\alpha)\operatorname{osc}_F(T^{n-N}u) \tag{5.30}$$

for all $u \in \mathrm{Lip}[K]$.

So, let $E \in \mathcal{E}$ be a nonempty, compact set, and let $\eta, \kappa > 0$. Set $\eta_1 = \eta/(4C)$ where C is the constant in (5.27).

Since E is a nonempty, compact set in a metric space, we can find a finite set $\mathcal{M} = \{x_i,\ i = 1, \dots, M\} \subset K$ such that

$$\sup_{x\in E} \min\{\delta(x, x_i) : x_i \in \mathcal{M}\} < \eta_1.$$

Since $\mathcal{M}$ is finite, it follows from Condition C2 that there exists an integer N_1 such that for every x_i in $\mathcal{M}$ there exists a measure $\nu_i \in \mathcal{P}_0$ such that for any $u \in \mathrm{Lip}[K]$,

$$|\langle u, \delta_{x_i} Q^{N_1}\rangle - \langle u, \nu_i\rangle| < \eta_1\gamma(u). \tag{5.31}$$

Next set $\mathcal{V} = \{\nu_1, \dots, \nu_M\}$. From the fact that $\mathcal{P}_0$ satisfies Condition E, it follows from Lemma 5.20 that we can choose $\alpha > 0$ and an integer N_2 in such a way that if ν_i and ν_j belong to $\mathcal{V}$, then there exists a compact set $F_{i,j} \in \mathcal{E}$ such that

$$|\langle u, \nu_i Q^m\rangle - \langle u, \nu_j Q^m\rangle| < \alpha\gamma(u)\rho + \kappa\operatorname{osc}(u) + (1-\alpha)\operatorname{osc}_{F_{i,j}}(T^{m-N_2}u)$$

if $u \in \mathrm{Lip}[K]$ and $m \ge N_2$.

By defining $F = \bigcup_{1\le i<j\le M} F_{i,j}$, it clearly follows that also

$$\begin{aligned}&|\langle u, \nu_i Q^m\rangle - \langle u, \nu_j Q^m\rangle| \\ &\quad = |\langle T^m u, \nu_i\rangle - \langle T^m u, \nu_j\rangle| < \alpha\gamma(u)\rho + \kappa\operatorname{osc}(u) + (1-\alpha)\operatorname{osc}_F(T^{m-N_2}u)\end{aligned} \tag{5.32}$$

if $u \in \mathrm{Lip}[K]$, $m \ge N_2$ and $\nu_i, \nu_j \in \mathcal{V}$.

Now set $N = N_1 + N_2$, let $n \ge N$, set $m = n - N_1$, and let x and y be probability measures in E. Let $x_i \in \mathcal{M}$ satisfy $\delta(x, x_i) < \eta_1$, and $x_j \in \mathcal{M}$ satisfy $\delta(y, x_j) < \eta_1$. By the triangle inequality,

$$|T^n u(x) - T^n u(y)| \le |T^n u(x_i) - T^n u(x_j)| + 2\eta_1\gamma(T^n u). \tag{5.33}$$

From (5.31) and the triangle inequality it follows also that

$$|T^n u(x_i) - T^n u(x_j)| \le |\langle T^m u, \nu_i\rangle - \langle T^m u, \nu_j\rangle| + 2\eta_1\gamma(T^m u). \tag{5.34}$$

By combining (5.32)–(5.34) we find that

$$\begin{aligned}&|T^n u(x) - T^n u(y)| \\ &\quad \le 2\eta_1\gamma(T^n u) + 2\eta_1\gamma(T^m u) + \alpha\gamma(u)\rho + \kappa\operatorname{osc}(u) + (1-\alpha)\operatorname{osc}_F(T^{m-N_2}u).\end{aligned}$$

Since x and y are arbitrarily chosen in E, and $\gamma(T^n u) \le C\gamma(u)$ for all $n \ge 1$, it follows that

$$\operatorname{osc}_E(T^n u) \le 4C\eta_1\gamma(u) + \alpha\gamma(u)\rho + \kappa \operatorname{osc}(u) + (1-\alpha)\operatorname{osc}_F(T^{m-N_2}u),$$

and since $\eta_1 = \eta/(4C)$ and $m - N_2 = n - N$, we find that

$$\operatorname{osc}_E(T^n u) \le \eta\gamma(u) + \alpha\gamma(u)\rho + \kappa \operatorname{osc}(u) + (1-\alpha)\operatorname{osc}_F(T^{n-N}u)$$

if $u \in \mathrm{Lip}[K]$, which we wanted to prove. Thus Proposition 5.22 is proved. ∎

6. On the barycenter and the filter kernel

From the results proved in Chapter 5 it follows that in order to prove parts (A)–(C) of Theorem 2.13 it now suffices to prove that Condition C2 is satisfied, and to prove (D) it remains to prove that Condition C3 is satisfied.

In order to accomplish this we shall need two results on barycenters. The first is

THEOREM 6.1. *Let $\mathcal{H} = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$ be a fully dominated, regular HMM and let $\mathbf{P}$ be the filter kernel. Let P be the Markov kernel of $\mathcal{H}$. Then for all $x \in K$,*

$$\bar{b}(\mathbf{P}^n(x,\cdot)) = xP^n, \qquad n = 1,2,\ldots.$$

REMARK 6.2. The theorem is essentially due to Kunita [31].

Proof. Let $F \in \mathcal{F}$, let $I_F : S \to \{0,1\}$ denote the indicator function of F, define $U_F : K \to \mathbb{R}$ by $U_F(x) = \langle I_F, x\rangle$, and let $\mathbf{T}$ denote the transition operator associated to $\mathbf{P}$ as defined by (2.12). From the definition of the barycenter we find that

$$\begin{aligned}
\bar{b}(\delta_x\mathbf{P})(F) &= \int_K \int_F y(ds)\,\delta_x\mathbf{P}(dy) = \int_K \langle I_F, y\rangle\,\delta_x\mathbf{P}(dy) = \langle U_F, \delta_x\mathbf{P}\rangle \\
&= \langle \mathbf{T}U_F, \delta_x\rangle = \mathbf{T}U_F(x) = \int_{A_x^+} U_F\Big(\frac{xM_a}{\|xM_a\|}\Big)\|xM_a\|\,\tau(da) \\
&= \int_{A_x^+} \Big\langle I_F, \frac{xM_a}{\|xM_a\|}\Big\rangle \|xM_a\|\,\tau(da) = \int_{A_x^+} \langle I_F, xM_a\rangle\,\tau(da) \\
&= \int_{A_x^+}\int_F x\,M_a(dt)\,\tau(da) = \int_{A_x^+}\int_F\int_S m(s,t,a)\,x(ds)\,\lambda(dt)\,\tau(da) \\
&= \int_F\int_S p(s,t)\,x(ds)\,\lambda(dt) = \int_F (xP)(dt) = xP(F),
\end{aligned}$$

from which it follows that $\bar{b}(\delta_x\mathbf{P}) = xP$. That $\bar{b}(\delta_x\mathbf{P}^n) = xP^n$ for $n \ge 2$ follows from the relation (3.4). ∎

The following lemma is not needed in the proof of the main theorem, but will be needed later, when we want to verify that Condition E holds. We present it here, since it gives some insight into sets of probability measures with equal barycenter.

LEMMA 6.3. *Let $(S,\mathcal{F},\delta_0)$ be a complete, separable metric space, let λ be a positive σ-finite measure on $(S,\mathcal{F})$, let $K = \mathcal{P}_\lambda(S,\mathcal{F})$, let $\mathcal{E}$ denote the σ-algebra generated by the total variation metric, let $\mathcal{P}(K,\mathcal{E})$ be the set of probability measures on $(K,\mathcal{E})$, let $\pi \in K$ and let $\mathcal{P}(K|\pi)$ denote the subset of $\mathcal{P}(K,\mathcal{E})$ consisting of those probability measures that have π as barycenter. For $F \in \mathcal{F}$ define*

$$E(F) = \{x \in K : x(F) \geq \pi(F)/2\}. \tag{6.1}$$

Then for all $\mu \in \mathcal{P}(K|\pi)$ and all $F \in \mathcal{F}$,

$$\mu(E(F)) \geq \pi(F)/2. \tag{6.2}$$

Proof. The inequality (6.2) holds trivially if $\pi(F) = 0$. Thus assume $F \in \mathcal{F}$ is such that $\pi(F) > 0$. Clearly $E(F) \in \mathcal{E}$. Set $E(F) = E$. Since $\mu \in \mathcal{P}(K|\pi)$ we have $\int_K \langle I_F, x\rangle\, \mu(dx) = \pi(F)$. Hence

$$\begin{aligned}\pi(F) &= \int_E \langle I_F, x\rangle\, \mu(dx) + \int_{K\setminus E} \langle I_F, x\rangle\, \mu(dx) \\ &= \int_E \int_F x(ds)\, \mu(dx) + \int_{K\setminus E} \int_F x(ds)\, \mu(dx) \leq \mu(E) + (1-\mu(E))\pi(F)/2.\end{aligned}$$

Thus $\mu(E)(1-\pi(F)/2) \geq \pi(F)/2$ and hence $\mu(E(F)) > \pi(F)/2$. Therefore (6.2) holds for all $\mu \in \mathcal{P}(K|\pi)$ and all $F \in \mathcal{F}$. ∎

We end this section with yet another simple observation, which we believe can be utilized to prove the existence of an invariant measure associated to a filter kernel.

PROPOSITION 6.4. *Let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a fully dominated, regular HMM and let $\mathbf{P}$ be the filter kernel. Let P be the Markov kernel of $\mathcal{H}$, and suppose that π is an invariant probability measure for P. Then for all $\mu \in \mathcal{P}(K|\pi)$,*

$$\bar{b}(\mu\mathbf{P}) = \pi.$$

Proof. Let $F \in \mathcal{F}$, let $I_F : S \to \{0,1\}$ denote the indicator function of F, define $U_F : K \to \mathbb{R}$ by $U_F(x) = \langle I_F, x\rangle$, let $\mathbf{T}$ denote the transition operator associated to the filter kernel $\mathbf{P}$ as defined by (2.12) and T denote the transition operator associated to the Markov kernel P. Let $\mu \in \mathcal{P}(K|\pi)$. From the definition of the barycenter and (2.13) we find that

$$\begin{aligned}\bar{b}(\mu\mathbf{P})(F) &= \int_K \int_F y(ds)\, \mu\mathbf{P}(dy) = \int_K \langle I_F, y\rangle\, \mu\mathbf{P}(dy) = \langle U_F, \mu\mathbf{P}\rangle \\ &= \langle \mathbf{T}U_F, \mu\rangle = \int_K \mathbf{T}U_F(x)\, \mu(dx) = \int_K \int_{A_x^+} U_F\left(\frac{xM_a}{\|xM_a\|}\right)\|xM_a\|\, \tau(da)\, \mu(dx) \\ &= \int_K \int_{A_x^+} \left\langle I_F, \frac{xM_a}{\|xM_a\|}\right\rangle \|xM_a\|\, \tau(da)\, \mu(dx) = \int_K \int_{A_x^+} \langle I_F, xM_a\rangle\, \tau(da)\, \mu(dx) \\ &= \int_K \int_{A_x^+} \int_F x\, M_a(dt)\, \tau(da)\, \mu(dx) \\ &= \int_K \int_{A_x^+} \int_F \int_S m(s,t,a)\, x(ds)\, \lambda(dt)\, \tau(da)\, \mu(dx) \\ &= \int_K \int_F \int_S p(s,t)\, x(ds)\, \lambda(dt)\, \mu(dx) = \int_K \int_F (xP)(dt)\, \mu(dx) \\ &= \int_K \langle I_F, xP\rangle\, \mu(dx) = \int_K \langle TI_F, x\rangle\, \mu(dx) \\ &= \langle TI_F, \pi\rangle = \langle I_F, \pi P\rangle = \langle I_F, \pi\rangle = \pi(F).\end{aligned}$$

Hence $\bar{b}(\mu\mathbf{P})(F) = \pi(F)$ for all $F \in \mathcal{F}$, and thus $\bar{b}(\mu\mathbf{P}) = \pi$. ∎

7. Completing the proof of the main theorem

Let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a fully dominated, regular, strongly ergodic HMM with limit measure π and let $\mathbf{P}$ be the filter kernel. Let $\mathcal{P}(K|\pi)$ be the set of probability measures having π as barycenter and suppose that $\mathcal{P}(K|\pi)$ satisfies Condition E. From Proposition 5.22 and the definitions of Conditions C2 and C3 we can conclude that to prove parts (A)–(C) of Theorem 2.13, it remains to show that for every $\epsilon > 0$ and every $x \in K$ there exists an integer N such that

$$\inf\{d_K(Q^n(x,\cdot),\nu) : \nu \in \mathcal{P}(K|\pi)\} < \epsilon, \qquad n \geq N, \tag{7.1}$$

and to prove part (D) it remains to show that if also the HMM $\mathcal{H}$ is uniformly ergodic then for every $\epsilon > 0$ there exists an integer N such that for all $x \in K$ the inequality (7.1) holds.

To do so, we shall show in this chapter that if $\mu \in \mathcal{P}(K|q)$ and $\|q - \pi\|$ is small, then we can find a measure $\nu \in \mathcal{P}(K|\pi)$ such that $d_K(\mu, \nu)$ is also small. Then Theorem 6.1 implies immediately that Condition C2 holds if $\mathcal{H}$ is strongly ergodic, and that Condition C3 holds if $\mathcal{H}$ is uniformly ergodic, and then the conclusions of Theorem 2.13 follow from the results of Chapter 5.

7.1. On the Kantorovich distance between sets with different barycenters. Let $(S, \mathcal{F})$ be a complete, separable, measurable space with metric δ_0, let λ denote a σ-finite, positive measure on $(S, \mathcal{F})$ and set $K = \mathcal{P}_\lambda(S, \mathcal{F})$. As before, let δ_{TV} denote the metric on K induced by the total variation and let $\mathcal{E}$ denote the σ-algebra generated by δ_{TV}. Instead of writing $\delta_{TV}(x, y)$, in this section we shall usually write $\|x - y\|$. Let $\mathcal{P}(K, \mathcal{E})$ denote the set of probability measures on $(K, \mathcal{E})$, let $\mathcal{Q}(K, \mathcal{E})$ denote the set of positive and finite measures on $(K, \mathcal{E})$, and for $r > 0$ let $\mathcal{Q}^r(K, \mathcal{E})$ denote the set of positive, finite measures on $(K, \mathcal{E})$ with total mass equal to r.

Let $d_K : \mathcal{P}(K, \mathcal{E}) \times \mathcal{P}(K, \mathcal{E}) \to [0, 2]$ denote the Kantorovich distance on $\mathcal{P}(K, \mathcal{E})$. Recall that the Kantorovich distance on $\mathcal{P}(K, \mathcal{E})$ has two equivalent definitions, (5.6) and (5.7).

For the set $\mathcal{Q}^r(K, \mathcal{E})$ we also define a metric, which we also denote by d_K, simply by

$$d_K(\mu, \nu) = r d_K(\mu/r, \nu/r), \qquad \mu, \nu \in \mathcal{Q}^r(K, \mathcal{E}).$$

Also in this case we call d_K the *Kantorovich distance*.

As above, we let $\mathcal{P}(K|x)$ denote the set of probability measures on $(K, \mathcal{E})$ with barycenter x. For $\mu \in \mathcal{Q}^r(K, \mathcal{E})$ we also define the barycenter $\bar{b}(\mu)$ simply by

$$\bar{b}(\mu) = r\bar{b}(\mu/r).$$

Thus, if $\mu \in \mathcal{Q}^r(K, \mathcal{E})$ then $\bar{b}(\mu) \in \mathcal{Q}_\lambda(S, \mathcal{F})$ and $\|\bar{b}(\mu)\| = r$. For $x \in K$ and $r > 0$, we let $\mathcal{Q}^r(K|x)$ denote the set of measures in $\mathcal{Q}^r(K, \mathcal{E})$ with barycenter rx.

The purpose of this section is to prove the following result:

THEOREM 7.1. *Let $r > 0$, let $x, y \in K$ and let $\mu \in \mathcal{Q}^r(K|x)$. Then*

$$\inf\{d_K(\mu, \nu) : \nu \in \mathcal{Q}^r(K|y)\} = r\|x - y\|.$$

Proof. First note that if $x, y \in K$, then $d_K(\delta_x, \delta_y) = \|x - y\|$, where δ_x and δ_y denote the Dirac measures at x and y respectively. This follows from (5.6) and the fact that $\tilde{\mathcal{P}}(\delta_x, \delta_y, K \times K) = \{\tilde{\delta}_{(x,y)}\}$, where $\tilde{\delta}_{(x,y)}$ denotes the Dirac measure at $(x, y) \in K \times K$.

The following lemma gives a lower bound for the Kantorovich distance between two measures in $\mathcal{Q}^r(K, \mathcal{E})$ in terms of their barycenters.

LEMMA 7.2. *Let $r > 0$ and $\mu, \nu \in \mathcal{Q}^r(K, \mathcal{E})$. Then*

$$d_K(\mu, \nu) \geq \|\overline{b}(\mu) - \overline{b}(\nu)\|.$$

Proof. The conclusion is trivially true if $\overline{b}(\mu) = \overline{b}(\nu)$. So assume that $\overline{b}(\mu) \neq \overline{b}(\nu)$. From the definition of the Kantorovich distance in $\mathcal{Q}^r(K, \mathcal{E})$ and the definition of the barycenter of a measure in $\mathcal{Q}^r(K, \mathcal{E})$, it follows that it suffices to prove the inequality if $r = 1$, that is, when $\mu, \nu \in \mathcal{P}(K, \mathcal{E})$.

Thus, let $\mu, \nu \in \mathcal{P}(K, \mathcal{E})$ and set $x = \overline{b}(\mu)$ and $y = \overline{b}(\nu)$. Let $F_1, F_2 \in \mathcal{F}$ be such that $F_2 = S \setminus F_1$ and $x(F \cap F_1) \geq y(F \cap F_1)$ for all $F \in \mathcal{F}$ with $F \subset F_1$, and $x(F \cap F_2) \leq y(F \cap F_2)$ for all $F \in \mathcal{F}$ with $F \subset F_2$ (F_1 and F_2 constitute a Hahn decomposition). Define $J : S \to [-1, 1]$ by

$$J(s) = I_{F_1}(s) - I_{F_2}(s), \tag{7.2}$$

where I_{F_1} and I_{F_2} denote the respective indicator functions.

Next, define $v \in B[K]$ by $v(z) = \langle J, z\rangle$. Since $\operatorname{osc}(J) \leq 2$, it follows from (2.1) that

$$|v(z_1) - v(z_2)| = |\langle J, z_1\rangle - \langle J, z_2\rangle| \leq \operatorname{osc}(J)\|z_1 - z_2\|/2 \leq \|z_1 - z_2\|,$$

and hence $v \in \mathrm{Lip}_1[K]$. The definition of the Kantorovich distance then yields

$$d_K(\mu, \nu) \geq \left|\int_K v(z)\,\mu(dz) - \int_K v(z)\,\nu(dz)\right|, \tag{7.3}$$

and from the definition of the barycenter and (7.2),

$$\begin{aligned}\left|\int_K v(z)\,\mu(dz) - \int_K v(z)\,\nu(dz)\right| &= \left|\int_K \langle J, z\rangle\,\mu(dz) - \int_K \langle J, z\rangle\,\nu(dz)\right| \\ &= |\langle I_{F_1}, \overline{b}(\mu)\rangle - \langle I_{S\setminus F_1}, \overline{b}(\mu)\rangle - \langle I_{F_1}, \overline{b}(\nu)\rangle + \langle I_{S\setminus F_1}, \overline{b}(\nu)\rangle| \\ &= |x(F_1) - y(F_1) + y(S \setminus F_1) - x(S \setminus F_1)| = \|x - y\| = \|\overline{b}(\mu) - \overline{b}(\nu)\|,\end{aligned}$$

which together with (7.3) implies that $d_K(\mu, \nu) \geq \|\overline{b}(\mu) - \overline{b}(\nu)\|$. ∎

We now continue our proof of Theorem 7.1 by proving that if $\mu \in \mathcal{Q}(K, \mathcal{E})$ is a weighted finite sum of Dirac measures, then for every $y \in K$ we can find $\nu \in \mathcal{Q}(K, \mathcal{E})$ such that $\mu(K) = \nu(K)$, $\overline{b}(\nu) = y\mu(K)$ and $d_K(\mu, \nu) = \|\overline{b}(\mu) - \overline{b}(\nu)\|$. As usual, if $\xi \in K$, we let δ_ξ denote the Dirac measure at ξ.

LEMMA 7.3. *Let N be a positive integer and let ξ_k, $k = 1, \ldots, N$, be elements in K. Let $\beta_k > 0$, $k = 1, \ldots, N$, let $\varphi \in \mathcal{Q}(K, \mathcal{E})$ be defined by $\varphi = \sum_{k=1}^N \beta_k \delta_{\xi_k}$, and define $a \in \mathcal{Q}_\lambda(S, \mathcal{F})$ by $a = \sum_{k=1}^N \beta_k \xi_k$. Let $b \in \mathcal{Q}_\lambda(S, \mathcal{F})$ satisfy $\|b\| = \|a\|$. Then there exist elements ζ_k, $k = 1, \ldots, N$, in K such that $b = \sum_{k=1}^N \beta_k \zeta_k$ and if we define $\Psi = \sum_{k=1}^N \beta_k \delta_{\zeta_k}$, then*

$$d_K(\varphi, \Psi) = \|a - b\|.$$

Proof. First observe that if $\psi \in \mathcal{Q}(K,\mathcal{E})$ is defined by $\psi = \sum_{k=1}^{N} \beta_k \delta_{\zeta_k}$, where β_k, $k = 1,\dots,N$, is a positive number, and ζ_k, $k = 1,\dots,N$, belongs to K, then the barycenter of ψ satisfies

$$\bar{b}(\psi) = \sum_{k=1}^{N} \beta_k \zeta_k. \tag{7.4}$$

This follows from the fact that if $\mu \in \mathcal{Q}(K,\mathcal{E})$ is defined by $\mu = \delta_{z_0}$ and $F \in \mathcal{F}$, then $\int_K \langle I_F, z\rangle\, \mu(dz) = \langle I_F, z_0\rangle = z_0(F)$.

Next, let $\zeta_1,\dots,\zeta_N \in K$ and define $\theta \in \mathcal{Q}(K,\mathcal{E})$ by $\theta = \sum_{k=1}^{N} \beta_k \delta_{\zeta_k}$. Clearly $\theta(K) = \sum_{k=1}^{N} \beta_k$ and hence $\theta(K) = \varphi(K) = \|a\|$. We now define the measure $\tilde{\varphi}$ on $(K^2,\mathcal{E}^2)$ by $\tilde{\varphi}(\{(\xi_k,\zeta_k)\}) = \beta_k$, $k = 1,\dots,N$. Then clearly $\tilde{\varphi}(A \times K) = \varphi(A)$ and $\tilde{\varphi}(K \times A) = \theta(A)$, for all $A \in \mathcal{E}$, from which it follows that

$$d_K(\varphi,\theta) \le \sum_{k=1}^{N} \beta_k \|\xi_k - \zeta_k\|, \tag{7.5}$$

since

$$d_K(\varphi,\theta) \le \int_{K\times K} \|x-y\|\, \tilde{\varphi}(dx,dy) = \sum_{k=1}^{N} \beta_k \|\xi_k - \zeta_k\|.$$

By combining (7.5) and (7.4) with Lemma 7.2, it follows that to prove Lemma 7.3, it suffices to find probability measures ζ_k, $k = 1,\dots,N$, belonging to K such that

$$b = \sum_{k=1}^{N} \beta_k \zeta_k \tag{7.6}$$

and also

$$\sum_{k=1}^{N} \beta_k \|\xi_k - \zeta_k\| = \|a - b\|. \tag{7.7}$$

That we can do this when $N = 1$, that is, when $\varphi = \beta_1 \delta_{\xi_1}$, is trivial: simply define $\zeta_1 = b/\beta_1$; then $\beta_1\|\xi_1 - \zeta_1\| = \|a - b\|$, as desired. The case when $b = a$ is also trivial: just take $\zeta_k = \xi_k$, $k = 1,\dots,N$. In the remaining part of the proof we therefore assume that $a \neq b$.

We now prove by induction that we can find probability measures $\zeta_k \in K$, $k = 1,\dots,N$, such that (7.6) and (7.7) hold. Thus, let $M \ge 2$, and assume that if $N = M-1$, then if $a = \sum_{k=1}^{N} \beta_k \xi_k$ where $\beta_k > 0$ and $\xi_k \in K$, $k = 1,\dots,N$, and if the measure b belongs to $\mathcal{Q}_\lambda(S,\mathcal{F})$ and satisfies $\|b\| = \|a\|$, then we can find ζ_k, $k = 1,\dots,N$, in K such that (7.6) and (7.7) hold.

Now, let $N = M$, let $\beta_k > 0$, $k = 1,\dots,M$, let $\xi_k \in K$, $k = 1,\dots,M$, set $a = \sum_{k=1}^{M} \beta_k \xi_k$ and suppose that $b \in \mathcal{Q}_\lambda(S,\mathcal{F})$ satisfies $\|b\| = \|a\|$. Our aim is to find ζ_k, $k = 1,\dots,M$, in K, such that (7.6) and (7.7) hold.

Recall that we have assumed that $a \neq b$ and hence $\|a - b\| \neq 0$. We define

$$\Delta = \|a - b\|/2.$$

Define $a_1 \in \mathcal{Q}_\lambda(S,\mathcal{F})$ by $a_1 = \sum_{k=1}^{M-1} \beta_k \xi_k$. Clearly $\|a_1\| = \|a\| - \beta_M$.

Suppose that we can find a probability measure $\zeta_M \in K$ such that if we define

$$b_1 = b - \beta_M \zeta_M, \tag{7.8}$$

then

$$b_1 \in \mathcal{Q}_\lambda(S, \mathcal{F}) \tag{7.9}$$

and

$$\|a - b\| = \|a_1 - b_1\| + \beta_M \|\xi_M - \zeta_M\|. \tag{7.10}$$

From (7.9) and the definition of b_1 it follows that $\|b_1\| = \|b\| - \beta_M = \|a\| - \beta_M = \|a_1\|$ and so, using the induction hypothesis, we can find probability measures ζ_k, $k = 1, \dots, M-1$, such that $b_1 = \sum_{k=1}^{M-1} \beta_k \zeta_k$ and

$$\sum_{k=1}^{M-1} \beta_k \|\xi_k - \zeta_k\| = \|a_1 - b_1\|; \tag{7.11}$$

consequently, by (7.10) and (7.11),

$$\|a - b\| = \sum_{k=1}^{M-1} \beta_k \|\xi_k - \zeta_k\| + \beta_M \|\xi_M - \zeta_M\| = \sum_{k=1}^{M} \beta_k \|\xi_k - \zeta_k\|,$$

and hence (7.6) and (7.7) hold with $N = M$.

To determine $\zeta_M \in K$ such that if we define b_1 by (7.8), then (7.9) and (7.10) hold, we proceed as follows.

First, let $F_1, F_2 \in \mathcal{F}$ be such that $F_2 = S \setminus F_1$ and $a(F \cap F_1) \geq b(F \cap F_1)$ for all $F \in \mathcal{F}$ satisfying $F \subset F_1$, and $a(F \cap F_2) \leq b(F \cap F_2)$ for all $F \in \mathcal{F}$ with $F \subset F_2$. We write $\mathcal{F}_1 = \{F \in \mathcal{F} : F \subset F_1\}$ and $\mathcal{F}_2 = \{F \in \mathcal{F} : F \subset F_2\}$. Note that $2\Delta = \|a - b\| = \sup\{a(F) - b(F) : F \in \mathcal{F}\} + \sup\{b(F) - a(F) : F \in \mathcal{F}\} = a(F_1) - b(F_1) + b(F_2) - a(F_2)$, and since $F_1 \cup F_2 = S$, $F_1 \cap F_2 = \emptyset$ and $\|a\| = \|b\|$, it is clear that

$$\Delta = a(F_1) - b(F_1). \tag{7.12}$$

Next, define a measure $c \in \mathcal{Q}_\lambda(S, \mathcal{F})$ by

$$c(F) = ((a - a_1) \wedge (a - b))(F \cap F_1), \qquad F \in \mathcal{F}, \tag{7.13}$$

and set

$$\Delta_0 = c(F_1).$$

Since obviously $c(F) \leq a(F) - b(F)$ if $F \in \mathcal{F}_1$, it follows that $c(F_1) \leq a(F_1) - b(F_1) = \Delta$ because of (7.12), and hence $\Delta_0 \leq \Delta$. We now define ζ_M as follows:

$$\begin{aligned} \zeta_M(F) &= \xi_M(F) - c(F)/\beta_M && \text{if } F \in \mathcal{F}_1, \\ \zeta_M(F) &= \xi_M(F) + (\Delta_0/\Delta)(b(F) - a(F))/\beta_M && \text{if } F \in \mathcal{F}_2. \end{aligned}$$

We have to verify that $\zeta_M \in K$. We first show that $\zeta_M \in \mathcal{Q}_\lambda(S, \mathcal{F})$. For $F \in \mathcal{F}_1$ we find from the definition of c (see (7.13)) that

$$\zeta_M(F) = \xi_M(F) - c(F)/\beta_M = (a(F) - a_1(F) - c(F))/\beta_M \geq 0,$$

and if $F \in \mathcal{F}_2$, then obviously $\zeta_M(F) \geq 0$. Hence $\zeta_M \in \mathcal{Q}(S, \mathcal{F})$. Since a, b, c and ξ_M belong to $\mathcal{Q}_\lambda(S, \mathcal{F})$, it follows that also $\zeta_M \in \mathcal{Q}_\lambda(S, \mathcal{F})$.

To prove $\zeta_M \in K$, we need to show that $\zeta_M(S) = 1$. Since

$$\zeta_M(F_1) = \xi_M(F_1) - \Delta_0/\beta_M$$

and

$$\zeta_M(F_2) = \xi_M(F_2) + (\Delta_0/\Delta)(b(F_2) - a(F_2))/\beta_M = \xi_M(F_2) + \Delta_0/\beta_M,$$

we have $\zeta_M(S) = \xi_M(F_1) + \xi_M(F_2) = 1$, and hence $\zeta_M \in K$. We also find that

$$\begin{aligned}\|\xi_M - \zeta_M\| &= \xi_M(F_1) - \zeta_M(F_1) + \zeta_M(F_2) - \xi_M(F_2) \\ &= c(F_1)/\beta_M + c(F_1)/\beta_M = 2\Delta_0/\beta_M.\end{aligned} \tag{7.14}$$

Furthermore, if b_1 is defined by (7.8), we find that if $F \in \mathcal{F}_1$, then

$$\begin{aligned}b_1(F) &= b(F) - \beta_M\zeta_M(F) \\ &= b(F) - \beta_M\xi_M(F) + c(F) = b(F) - a(F) + a_1(F) + c(F) \\ &= b(F) + a_1(F) + ((a - a_1) \wedge (a - b))(F) - a(F) \\ &= b(F) + a_1(F) - (a_1 \vee b)(F) \geq 0,\end{aligned}$$

and if $F \in \mathcal{F}_2$, then since $\Delta_0 \leq \Delta$ we obtain

$$\begin{aligned}b_1(F) &= b(F) - \beta_M\xi_M(F) - (b(F) - a(F))\Delta_0/\Delta \\ &\geq b(F) - a(F) + a_1(F) - (b(F) - a(F)) \geq a_1(F).\end{aligned}$$

Hence (7.9) is satisfied.

It thus remains to show that (7.10) is satisfied. Since

$$b_1(F) = b(F) + a_1(F) - (a_1 \vee b)(F) \leq a_1(F)$$

if $F \in \mathcal{F}_1$, and as we just showed $b_1(F) \geq a_1(F)$ if $F \in \mathcal{F}_2$, we find that

$$\begin{aligned}\|a_1 - b_1\| &= a_1(F_1) - b_1(F_1) + b_1(F_2) - a_1(F_2) \\ &= a(F_1) - \beta_M\xi_M(F_1) - b(F_1) + \beta_M\xi_M(F_1) - c(F_1) \\ &\quad + b(F_2) - \beta_M\xi_M(F_2) - (\Delta_0/\Delta)(b(F_2) - a(F_2)) - a(F_2) + \beta_M\xi_M(F_2) \\ &= a(F_1) - b(F_1) - \Delta_0 + b(F_2) - a(F_2) - \Delta_0 = 2\Delta - 2\Delta_0,\end{aligned}$$

and since $\|a - b\| = 2\Delta$ and $\beta_M\|\xi_M - \zeta_M\| = 2\Delta_0$ because of (7.14), the equality (7.10) holds and thus the proof of Lemma 7.3 is complete. ∎

Using Lemmas 7.3 and 7.2 it is now easy to conclude the proof of Theorem 7.1. Thus let $x, y \in K$ and suppose $\mu \in \mathcal{Q}^r(K|x)$. We want to prove that for every $\epsilon > 0$ we can find a measure $\nu \in \mathcal{Q}^r(K|y)$ such that

$$d_K(\mu, \nu) < r\|x - y\| + \epsilon.$$

Thus, let $\epsilon > 0$. From the general theory of measures we know, since $(K, \mathcal{E})$ is a complete, separable, metric space, that we can find a measure $\mu_1 \in \mathcal{Q}^r(K, \mathcal{E})$ of the form $\mu_1 = \sum_{k=1}^{N} \beta_k\delta_{\xi_k}$ such that $d_K(\mu, \mu_1) < \epsilon/2$, where ξ_k, $k = 1, \ldots, N$, belong to K, and $\beta_k > 0$ for $k = 1, \ldots, N$ (see e.g. [12, Chapter 9]). From Lemma 7.2 it now follows that

$$\epsilon/2 > d_K(\mu, \mu_1) \geq \|rx - \bar{b}(\mu_1)\|,$$

and from Lemma 7.3 it follows that we can find a measure $\nu \in \mathcal{Q}^r(K|y)$ such that

$$d_K(\mu_1, \nu) = \|\bar{b}(\mu_1) - ry\|.$$

Then from the triangle inequality we have

$$\begin{aligned} d_K(\mu,\nu) &\le d_K(\mu,\mu_1) + d_K(\mu_1,\nu) < \epsilon/2 + \|\bar{b}(\mu_1) - ry\| \\ &\le \epsilon/2 + \|\bar{b}(\mu_1) - rx\| + r\|x-y\| \le \epsilon/2 + \epsilon/2 + r\|x-y\|. \end{aligned}$$

Hence, $d_K(\mu,\nu) < r\|x-y\| + \epsilon$ and thus Theorem 7.1 is proved. ■

7.2. Completing the proof. By using Theorems 6.1 and 7.1 it is now easy to finish the proof of the main theorem. We first prove the following corollary to Theorem 7.1.

Corollary 7.4. *Let* $\mathcal{H} = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$ *be a fully dominated, regular HMM and let* $\mathbf{P}$ *be the filter kernel.*

(A) *Suppose* $\mathcal{H}$ *is strongly ergodic with limit measure* π. *Then for every* $\eta > 0$ *we can find an integer* N *such that for every* $x \in \mathcal{M}$ *there exists a probability* $\nu_x \in \mathcal{P}(K|\pi)$ *such that for every* $u \in \mathrm{Lip}[K]$ *and every integer* $n \ge N$,

$$|\langle u, \delta_x \mathbf{P}^n\rangle - \langle u, \nu_x\rangle| < \eta\gamma(u). \tag{7.15}$$

(B) *If furthermore* $\mathcal{H}$ *is uniformly ergodic, then for every* $\eta > 0$ *we can find an integer* N *such that for every* $x \in K$ *there exists a measure* $\nu_x \in \mathcal{P}(K|\pi)$ *such that* (7.15) *holds for every* $u \in \mathrm{Lip}[K]$ *and every* $n \ge N$.

Proof. Let $x \in K$ and $\eta > 0$. Let P denote the Markov kernel determined by (p,λ). Since $\mathcal{H}$ is strongly ergodic with limit measure π, we can find an integer N such that if $n \ge N$, then

$$\delta_{TV}(xP^n, \pi) < \eta.$$

From Theorem 6.1 we deduce that

$$\delta_{TV}(\bar{b}(\delta_x \mathbf{P}^n), \pi) < \eta \quad \text{if } n \ge N, \tag{7.16}$$

and Theorem 7.1 then yields $\nu \in \mathcal{P}(K|\pi)$ such that $d_K(\delta_x\mathbf{P}^n, \nu) < \eta$ if $n \ge N$, from which we conclude that (7.15) holds if $u \in \mathrm{Lip}[K]$. Thus part (A) is proved.

Next suppose that $\mathcal{H}$ is also uniformly ergodic with limit measure π. Let $\eta > 0$. We can then find an integer N such that if $n \ge N$, then

$$\delta_{TV}(xP^n, \pi) < \eta$$

for all $x \in K$. Theorem 6.1 shows that

$$\delta_{TV}(\bar{b}(\delta_x \mathbf{P}^n), \pi) < \eta \quad \text{if } n \ge N \tag{7.17}$$

for all $x \in K$. From Theorem 7.1 it then follows that for every $x \in K$, we can find $\nu_x \in \mathcal{P}(K|\pi)$ such that $d_K(\delta_x\mathbf{P}^n, \nu_x) < \eta$ if $n \ge N$, from which it follows that (7.15) holds for all $x \in K$ if $u \in \mathrm{Lip}[K]$. Thus part (B) is also proved. ■

Corollary 7.5. *Let* $\mathcal{H} = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$ *be a fully dominated, regular, strongly ergodic HMM with limit measure* π, *let* $\mathbf{P}$ *be the filter kernel and suppose that Condition* E *is satisfied. Then:*

(A) *The filter kernel* $\mathbf{P}$ *satisfies Condition* C2.
(B) *If furthermore* $\mathcal{H}$ *is uniformly ergodic, then* $\mathbf{P}$ *satisfies Condition* C3.

Proof. Let $\mathcal{P}_0 = \mathcal{P}(K|\pi)$. By assumption, $\mathcal{P}_0$ satisfies Condition E. Condition C2 then follows from Corollary 7.4(A). If furthermore $\mathcal{H}$ is uniformly ergodic, it follows from Corollary 7.4(B) that also Condition C3 holds. ∎

Finally, the conclusions of Theorem 2.13 now follow by combining Corollary 7.5, Proposition 5.22, Theorem 5.11, Proposition 5.14, Corollary 5.17, Theorem 5.4 and Theorem 3.1. Thus the proof of the main theorem is complete. ∎

REMARK 7.6. Consider the following condition introduced by T. Szarek [40].

DEFINITION 7.7. Let $(K, \mathcal{E}, \delta)$ be a complete, separable, metric space and let P be a tr.pr.f. on $(K, \mathcal{E}, \delta)$. If there exists $x_0 \in K$ such that for every open set O containing x_0 there exists $x \in K$ such that

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P^n(x, O) > 0,$$

then we say that *Condition $\mathcal{E}$ holds.*

Now from [40, proof of Proposition 2.1] and Theorem 3.1, it follows that if a HMM is a fully dominated, regular, strongly ergodic HMM and Condition $\mathcal{E}$ holds, then there exists $x_0 \in K$ such that $\{\mathbf{P}^n(x_0, \cdot), n = 1, 2, \ldots\}$ is a tight sequence, and hence hypothesis (B) of Theorem 2.13 holds. Therefore, if we could show that the filter kernel $\mathbf{P}$ induced by a fully dominated, regular HMM satisfying Condition E also satisfies Condition $\mathcal{E}$, we would be able to replace "weakly contracting" by "weakly ergodic" in part (A) of Theorem 2.13, and we could omit the rest of the theorem.

8. On Condition E

Theorem 2.13 has two evident weaknesses. The first is that the conclusion in part (A) is only weak contraction and not weak ergodicity. This we have not been able to surmount.

The other weakness is that it is not easy to tell, by looking at the HMM under consideration, whether Condition E is satisfied or not. We do believe that for most fully dominated, regular HMMs Condition E does indeed hold, but as Examples 2.17 and 2.18 show, there are exceptions.

The main purpose of this chapter is to introduce more easily verifiable conditions that imply Condition E. To this end we shall use some estimates for iterations of integral kernels.

8.1. Estimates of iterations of integral kernels. Let $(S, \mathcal{F}, \delta)$ be a complete, separable, metric space and let λ be a positive, σ-finite measure on $(S, \mathcal{F})$. If $k : S \times S \to [0, \infty)$ is a nonnegative, measurable function defined on $S \times S$ such that

$$\sup\left\{ \int_S k(s,t)\,\lambda(dt) : s \in S \right\} < \infty,$$

then we call k a *density kernel* with respect to λ. We denote by $D_\lambda[S \times S]$ the set of all density kernels with respect λ.

Definition 8.1. Let $k \in D_\lambda[S \times S]$. We say that k has *rectangular support* if there exist $F, G \in \mathcal{F}$ such that $\lambda(F), \lambda(G) > 0$, and if $(s,t) \in F \times G$ then $k(s,t) > 0$, while if $(s,t) \notin F \times G$ then $k(s,t) = 0$. We call $F \times G$ the *rectangular support* of k.

Remark 8.2. In case S is a finite set, the notion of rectangular support is equivalent to the notion of subrectangular matrix introduced in [21] and used implicitly in the definition of Condition A presented in the introduction (Section 1.2).

The following theorem is a generalisation of [21, Lemma 6.2].

Theorem 8.3. *Let $(S, \mathcal{F}, \delta)$ be a complete, separable, metric space and let λ be a positive, σ-finite measure on $(S, \mathcal{F})$. Let k_m, $m = 1, \dots, n$, $n \geq 1$, be density kernels belonging to $D_\lambda[S \times S]$ having rectangular supports $F_m \times G_m$, $m = 1, \dots, n$, where $\lambda(F_m)\lambda(G_m) > 0$, $m = 1, \dots, n$. Let $K_m : S \times \mathcal{F} \to [0, \infty)$ be defined by*

$$K_m(s, E) = \int_E k_m(s,t)\, \lambda(dt),$$

and for $m = 1, \dots, n$, define $K^{m,n} : S \times \mathcal{F} \to [0, \infty)$ recursively by $K^{n,n} = K_n$ and

$$K^{m-1,n}(s, E) = \int_S k_{m-1}(s,t) K^{m,n}(t, E)\, \lambda(dt), \qquad m = n, n-1, \dots, 2. \tag{8.1}$$

Set $K^n = K^{1,n}$, and for $x \in \mathcal{P}(S, \mathcal{F})$, let $xK^n \in \mathcal{Q}(S, \mathcal{F})$ be defined by $xK^n(E) = \int_S K^n(s, E)\, x(ds)$.

Now suppose that there exist $\kappa_m \geq 1$ such that for $1 \leq m \leq n$,

$$\sup\left\{ \frac{k_m(s_1,t_1) k_m(s_2,t_2)}{k_m(s_2,t_1) k_m(s_1,t_2)} : s_1, s_2 \in F_m,\ t_1, t_2 \in G_m \right\} \leq \kappa_m^2. \tag{8.2}$$

Suppose also that

$$K^n(s, S) > 0 \tag{8.3}$$

for all $s \in F_1$.

If $x, y \in \mathcal{Q}(S, \mathcal{F})$ are such that $x(F_1), y(F_1) > 0$ and $n \geq 1$, then

$$\left\| \frac{xK^n}{\|xK^n\|} - \frac{yK^n}{\|yK^n\|} \right\| \leq 2 \prod_{m=1}^{n} \frac{\kappa_m - 1}{\kappa_m + 1}. \tag{8.4}$$

Proof. We first state the following lemma.

Lemma 8.4. *Let $n \geq 1$, and let k_m, K_m, $K^{m,n}$, $m = 1, \dots, n$, and K^n be as in Theorem 8.3. Then*

$$\sup\left\{ \left| \frac{K^n(s_1, E)}{K^n(s_1, G_n)} - \frac{K^n(s_2, E)}{K^n(s_2, G_n)} \right| : s_1, s_2 \in F_1,\ E \in \mathcal{F} \right\} \leq \prod_{m=1}^{n} \frac{\kappa_m - 1}{\kappa_m + 1}. \tag{8.5}$$

Proof. The lemma is a simple consequence of the following proposition, which is a special version of a result due to E. Hopf from 1963 (see [17, Theorem 1]).

Proposition 8.5. *Let $(S, \mathcal{F}, \delta)$ be a complete, separable, metric space, let λ be a positive, σ-finite measure on $(S, \mathcal{F})$ and let $k \in D_\lambda[S \times S]$ be a density kernel with rectangular support $F \times G$. Suppose that there exists $\kappa \geq 1$ such that*

$$\sup\left\{ \frac{k(s_1,t_1) k(s_2,t_2)}{k(s_2,t_1) k(s_1,t_2)} : s_1, s_2 \in F,\ t_1, t_2 \in G \right\} \leq \kappa^2.$$

Let $v \in B[S]$ be nonnegative, let $u \in B[S]$ be nonnegative such that $u(t) > 0$ if $t \in G$, and suppose that $\sup\{v(t)/u(t) : t \in G\} < \infty$. *Define $u_1, v_1 : S \to [0, \infty)$ by*

$$u_1(s) = \int_S k(s,t)u(t)\,\lambda(dt) \quad \text{and} \quad v_1(s) = \int_S k(s,t)v(t)\,\lambda(dt).$$

Then

$$\operatorname{osc}_F\left(\frac{v_1}{u_1}\right) \le \frac{\kappa-1}{\kappa+1}\operatorname{osc}_G\left(\frac{v}{u}\right).$$

By Proposition 8.5, for every $E \in \mathcal{F}$,

$$\operatorname{osc}_{F_{n-1}}\left(\frac{K_n(\cdot,E)}{K_n(\cdot,G_n)}\right) \le \frac{\kappa_n-1}{\kappa_n+1};$$

using the integral representation (8.1) and Proposition 8.5, the inequality (8.5) now follows easily by induction, proving Lemma 8.4. ∎

To conclude the proof of Theorem 8.3 we argue as follows. (The argument is inspired by an argument in [13].)

Let $x, y \in \mathcal{Q}(S,\mathcal{F})$ with $x(F_1), y(F_1) > 0$. We write $K^n = U$. We want to prove that if $n \ge 1$, then

$$\left\|\frac{xU}{\|xU\|} - \frac{yU}{\|yU\|}\right\| \le 2\prod_{m=1}^{n}\frac{\kappa_m-1}{\kappa_m+1}.$$

Let $E \in \mathcal{F}$. Then

$$\frac{xU(E)}{\|xU\|} = \int_{F_1}\frac{U(s,E)}{xU(G_n)}\,x(ds) = \int_{F_1}\frac{U(s,E)}{U(s,G_n)}\,\alpha(ds),$$

where

$$\alpha(ds) = \frac{U(s,G_n)}{xU(G_n)}\,x(ds).$$

Evidently $\alpha \in \mathcal{P}(S,\mathcal{F})$.

In a similar manner we can write

$$\frac{yU(E)}{\|yU\|} = \int_{F_1}\frac{U(s,E)}{U(s,G_n)}\,\beta(ds),$$

where $\beta \in \mathcal{P}(S,\mathcal{F})$ is defined by

$$\beta(ds) = \frac{U(s,G_n)}{yU(G_n)}\,y(ds).$$

Hence, by (2.1),

$$\begin{aligned}\left|\frac{xU(E)}{\|xU\|} - \frac{yU(E)}{\|yU\|}\right| &= \left|\int_{F_1}\frac{U(s,E)}{U(s,G_n)}\,\alpha(ds) - \int_{F_1}\frac{U(s,E)}{U(s,G_n)}\,\beta(ds)\right| \\ &\le \sup\left\{\frac{U(s_1,E)}{U(s_1,G_n)} - \frac{U(s_2,E)}{U(s_2,G_n)} : s_1, s_2 \in F_1\right\}\frac{1}{2}\|\alpha-\beta\|, \qquad (8.6)\end{aligned}$$

and since $\|\alpha - \beta\| \le 2$ and (8.6) holds for all $E \in \mathcal{F}$, Lemma 8.4 implies (8.4). ∎

8.2. Couplings of RSCCs. As described in Chapter 4, the filter kernel induced by a fully dominated regular HMM is equal to the Markov kernel associated to the RSCC induced by the HMM (see Observation 4.6).

In this section we shall define couplings of RSCCs, and in the next section we shall define the Vasershtein coupling of a RSCC.

Let $\mathcal{R} = \{(K,\mathcal{E}),(A,\mathcal{A}),h,Q\}$ be a RSCC. If a RSCC

$$\tilde{\mathcal{R}} = \{(K^2,\mathcal{E}^2),(A^2,\mathcal{A}^2),\tilde{h},\tilde{Q}\}$$

has $(K^2,\mathcal{E}^2)$ as state space, has $(A^2,\mathcal{A}^2)$ as index space, the response function $\tilde{h} : K^2 \times A^2 \to K^2$ satisfies

$$\tilde{h}((x,y),(a,b)) = (h(x,a),h(y,b)), \tag{8.7}$$

and the index probability function $\tilde{Q} : K^2 \times \mathcal{A}^2 \to [0,1]$ satisfies

$$\tilde{Q}((x,y),B\times A) = Q(x,B), \quad \forall (x,y)\in K^2,\ \forall B\in\mathcal{A}, \tag{8.8}$$

$$\tilde{Q}((x,y),A\times B) = Q(y,B), \quad \forall (x,y)\in K^2,\ \forall B\in\mathcal{A}, \tag{8.9}$$

then we call $\tilde{\mathcal{R}}$ a *coupling* of $\mathcal{R}$.

If the index probability function $\tilde{Q} : K^2 \times \mathcal{A}^2 \to [0,1]$ satisfies

$$\tilde{Q}((x,y),B_1\times B_2) = Q(x,B_1)Q(y,B_2)), \quad \forall x,y\in K,\ \forall B_1,B_2\in\mathcal{A},$$

then we call $\tilde{\mathcal{R}}$ the *trivial coupling* of $\mathcal{R}$.

Now let $P : K\times\mathcal{E}\to[0,1]$ be the Markov kernel (see (4.2)) associated to a RSCC $\mathcal{R} = \{(K,\mathcal{E}),(A,\mathcal{A}),h,Q\}$ and let $\tilde{P} : K^2\times\mathcal{E}^2\to[0,1]$ be the Markov kernel associated to any coupling $\tilde{\mathcal{R}}$ of $\mathcal{R}$. Then clearly for any pair (μ,ν) in $\mathcal{P}(K,\mathcal{E})$ and any $n\ge 1$, if we define $\tilde{\mu} = \mu\otimes\nu$, then $\tilde{\mu}\tilde{P}^n$ is a coupling of μP^n and νP^n.

Since the filter kernel $\mathbf{P}$ induced by a fully dominated, regular HMM $\mathcal{H}$ is equal to the Markov kernel associated to the RSCC induced by $\mathcal{H}$, a natural approach to verify Condition E is to find a useful coupling of the RSCC associated to the HMM under consideration.

It turns out that the trivial coupling is a good candidate in many cases. However, there is another coupling which is useful; we call it the Vasershtein coupling of a RSCC, and define it in the next section. For simplicity we restrict ourselves to RSCCs for which the index tr.pr.f. Q is determined by a continuous density function $q : K\times A\to[0,\infty)$ and a base measure $\tau\in\mathcal{Q}^\infty(A,\mathcal{A})$.

8.3. The Vasershtein coupling of a RSCC. Let $(K,\mathcal{E},\delta)$ and $(A,\mathcal{A},\varrho)$ be complete, separable, metric spaces, let $\mathcal{R} = \{(K,\mathcal{E}),(A,\mathcal{A}),h,(q,\tau)\}$ be a RSCC with a continuous index probability density function $q : K\times A\to[0,\infty)$ and base measure τ, let Q denote the index tr.pr.f. determined by (q,τ), and let P denote the associated Markov kernel defined by (4.2). Let $D = \{(a,b)\in A^2 : a = b\}$. The set D is measurable, since $(A,\mathcal{A},\varrho)$ is a complete, separable, metric space. For $x,y\in K$, define $C_1(x,y) = \{a : q(x,a)\ge q(y,a)\}$, $C_2(x,y) = A\setminus C_1(x,y)$ and $C^2(x,y) = \{(a,b)\in A^2 : a\in C_1(x,y),\ b\in C_2(x,y)\}$. For $B\in\mathcal{A}^2$, we set $\Pi(B) = \{a\in A : (a,a)\in B\}$. Then $A_1(x,y)$ and $A_2(x,y)$ are measurable since q is continuous, and $\Pi(B)\in\mathcal{A}$ since D is measurable and so is the mapping $\vartheta : A\to A^2$ defined by $\vartheta(a) = (a,a)$.

Next, define $\check{q} : K\times K\times A\to[0,\infty)$ by $\check{q}(x,y,a) = \min\{q(x,a),q(y,a)\}$, and for $x,y\in K$ set $\Delta(x,y) = \int_A (q(x,a)-\check{q}(x,y,a))\,\tau(da)$. We define $\tilde{Q}_V : K^2\times\mathcal{A}^2\to[0,1]$ by

$$\tilde{Q}_V((x,y),B) = \int_{\Pi(B)} \check{q}((x,y),a)\,\tau(da)$$
$$+ \iint_{B\cap C^2(x,y)} (q(x,a)-\check{q}(x,y,a))\big(q(x,b)-\check{q}(x,y,b)\big)\,\tau(da)\,\tau(db)/\Delta(x,y), \qquad (8.10)$$

where the last term is omitted if $\Delta(x,y)=0$.

It is easy to verify that $\tilde{Q}_V$ is a tr.pr.f. from $(K^2,\mathcal{E}^2)$ to $(A^2,\mathcal{A}^2)$, and $\tilde{Q}_V((x,y),\cdot)$ is well-known to be a coupling of $Q(x,\cdot)$ and $Q(y,\cdot)$ for all $x,y\in K$ (see [33, Section I.5]). We call $\tilde{Q}_V : K^2\times\mathcal{A}^2\to[0,1]$ the *Vasershtein coupling* of (q,τ) (or of Q), and call the RSCC

$$\tilde{\mathcal{R}}_V = \{(K^2,\mathcal{E}^2),(A^2,\mathcal{A}^2),\tilde{h},\tilde{Q}_V\},$$

where $\tilde{h} : K^2\times A^2\to K^2$ is defined by (8.7) and $\tilde{Q}_V$ is defined by (8.10), the *Vasershtein coupling* of the RSCC $\mathcal{R}$. We denote by $\tilde{P}_V$ the Markov kernel associated to $\tilde{\mathcal{R}}_V$.

REMARK 8.6. The original paper using the Vasershtein coupling is [42]. For an early application of the Vasershtein coupling to RSCCs see [23], where it is used in proving convergence in distribution, the law of large numbers and the central limit theorem for the state sequence of a RSCC. In [24, Section 7], the Vasershtein coupling is used to prove a classical result by Karlin (see [29]), and in [24] it is also applied to convergence rate problems for continued fraction expansions. M. Ślęczka [38] used the Vasershtein coupling to prove that the rate of convergence of the distributions of the state sequence of a RSCC to a unique limit distribution is geometric for RSCCs with a complete, separable, metric state space and a finite index space, if moreover the index probability function is strictly positive and an arithmetic mean contraction property holds.

An important property of the Vasershtein coupling $\tilde{Q}_V$ is described in the next proposition, which follows immediately from the definition.

PROPOSITION 8.7. *Let $(K,\mathcal{E},\delta)$ and $(A,\mathcal{A},\varrho)$ be complete, separable, metric spaces, let $\{(K,\mathcal{E}),(A,\mathcal{A}),h,(q,\tau)\}$ be a RSCC such that $q : K\times A\to[0,\infty)$ is continuous, and let*

$$\tilde{\mathcal{R}}_V = \{(K^2,\mathcal{E}^2),(A^2,\mathcal{A}^2),\tilde{h},\tilde{Q}_V\}$$

be the Vasershtein coupling of $\mathcal{R}$.

Suppose that there exist a measurable set $K_0\subset K$, a measurable set $B\subset A$ and positive numbers η and β such that

(1) $\tau(B)=\beta$,
(2) $\inf\{q(x,a) : (x,a)\in K_0\times B\}=\eta$.

Then

$$\tilde{Q}_V((x,y),\{(a,a) : a\in B\}) \ge \eta\beta, \qquad \forall x,y\in K_0.$$

Proof. Let $x,y\in K_0$. From the definition (8.10) of $\tilde{Q}_V$ it follows that

$$\tilde{Q}_V((x,y),\{(a,a) : a\in B\}) = \int_B \min\{q(x,a),q(y,a)\}\,\tau(da) \ge \eta\beta. \ \blacksquare$$

In order to find conditions implying Condition E of Theorem 2.13, we will use Proposition 8.9 below.

We first prove the following lemma. Recall that in Section 5.3 we gave a slightly broader definition of Condition E (see Definition 5.19) than in Section 2.5. The definition of the nth iterate of a RSCC was given in Section 4.1.

LEMMA 8.8. *Let $(K,\mathcal{E},\delta)$ and $(A,\mathcal{A},\varrho)$ be complete, separable, metric spaces, let $\mathcal{R} = \{(K,\mathcal{E}),(A,\mathcal{A}),h,(q,\tau)\}$ be a RSCC such that $q : K \times A \to [0,\infty)$ is continuous, and let $P : K \times \mathcal{E} \to [0,1]$ be the associated Markov kernel. For $n = 2,3,\dots,$ let $\mathcal{R}^n = \{(K,\mathcal{E}),(A^n,\mathcal{A}^n),h^n,(q^n,\tau^n)\}$ be the nth iterate of $\mathcal{R}$ and assume that $q^n : K \times A^n \to [0,\infty)$ is continuous for all $n \geq 2$.*

Suppose that $\mathcal{P}_0 \subset \mathcal{P}(K,\mathcal{E})$ is such that for every $\rho > 0$, there exist an integer N, $K_0 \in \mathcal{E}$, $B \in \mathcal{A}^N$, and positive constants ξ, β and η such that

(1) *$\mu(K_0) \geq \xi$ for all $\mu \in \mathcal{P}_0$,*
(2) *$\tau^N(B) \geq \beta$,*
(3) *if $x \in K_0$ and $a^N \in B$, then*
$$q^N(x,a^N) \geq \eta, \tag{8.11}$$
(4) *if $x,y \in K_0$ and $a^N \in B$, then*
$$\delta(h^N(x,a^N),h^N(y,a^N)) < \rho. \tag{8.12}$$

Then $(\mathcal{P}_0,P)$ satisfies Condition E.

Proof. Let $\rho > 0$. Choose the integer N, $K_0 \in \mathcal{E}$, $B \in A^N$ and $\xi,\beta,\eta > 0$ such that hypotheses (1)–(4) hold.

Let $\mathcal{R}^N = \{(K,\mathcal{E}),(A^N,\mathcal{A}^N),h^N,(q^N,\tau^N)\}$ be the Nth iterate of $\mathcal{R}$. From the hypotheses of the lemma we know that $q^N : K \times A^N \to [0,\infty)$ is continuous. Therefore we can define the Vasershtein coupling $\tilde{\mathcal{R}}^N_V$ of $\mathcal{R}^N$.

Now let $\mu,\nu \in \mathcal{P}_0$, and let $\tilde{\mathcal{R}}^N_V = \{(K^2,\mathcal{E}^2),(A^{2N},\mathcal{A}^{2N}),\tilde{h}^N,\tilde{Q}^N_V\}$ be the Vasershtein coupling of $\mathcal{R}^N$. Set
$$\tilde{B} = \{(a^N,b^N) \in A^N \times A^N : a^N = b^N,\ a^N \in B\}.$$
Since $\tau^N(B) \geq \beta$ and $q^N(x,a^N) \geq \eta$ if $x \in K_0$ and $a^N \in B$, Proposition 8.7 shows that
$$\tilde{Q}^N_V((x,y),\tilde{B}) \geq \eta\beta$$
if $x,y \in K_0$. Now let
$$D_\rho = \{(z_1,z_2) \in K \times K : \delta(z_1,z_2) < \rho\},$$
$$\tilde{A}^N(D_\rho) = \{(a^N,b^N) \in A^N \times A^N : (h^N(x,a^N),h^N(y,b^N)) \in D_\rho\},$$
and let $\tilde{P}_{V,N}$ be the Markov kernel associated to the RSCC $\tilde{\mathcal{R}}^N_V$. From the definition (4.2) and the fact that $\tilde{B} \subset \tilde{A}^N(D_\rho)$, it follows that
$$\tilde{P}_{V,N}((x,y),D_\rho) = \tilde{Q}^N_V((x,y),\tilde{A}^N(D_\rho)) \geq \tilde{Q}^N_V((x,y),\tilde{B}) \geq \beta\eta.$$
Hence, if we define $\tilde{\mu} = \mu \otimes \nu$ and set $\alpha = \xi^2\beta\eta$, then
$$\tilde{\mu}\tilde{P}^N_V(D_\rho) \geq \xi^2\beta\eta = \alpha,$$
since $\tilde{\mu}(K_0 \times K_0) \geq \xi^2$. Since $P^{(N)} = P^N$, where $P^{(N)}$ denotes the Markov kernel associated to $\mathcal{R}^N$, and $\tilde{\mu}\tilde{P}_{V,N}$ is a coupling of $\mu P^{(N)}$ and $\nu P^{(N)}$, it follows that Condition E holds. ∎

The following proposition now follows almost immediately from Lemma 8.8. Recall that the nth iterate of a fully dominated HMM was defined in Section 3.1.

PROPOSITION 8.9. *Suppose that* $\mathcal{H} = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$ *is a fully dominated, regular, strongly ergodic HMM with limit measure* π, *and for* $n = 1,2,\ldots$ *let* $\mathcal{H}^n = \{(S,\mathcal{F},\delta_0),(p^n,\lambda),(A^n,\mathcal{A}^n,\varrho^{(n)}),(m^n,\tau^n)\}$ *be the nth iterate of* $\mathcal{H}$. *Let* $M_{a^n} : \mathcal{Q}_\lambda(S,\mathcal{F}) \to \mathcal{Q}_\lambda(S,\mathcal{F})$ *denote the stepping function (for* $\mathcal{H}^n$*) determined by* $a^n \in A^n$.

Suppose that for every $\rho > 0$ *there exist an integer* N, $K_0 \in \mathcal{E}$, $B \in \mathcal{A}^N$ *and positive constants* ξ, β *and* η *such that*

(a) $\mu(K_0) \geq \xi$ *for all* $\mu \in \mathcal{P}(K|\pi)$,
(b) $\tau^N(B) \geq \beta$,
(c) *if* $x \in K_0$ *and* $a^N \in B$, *then*

$$\|xM_{a^N}\| \geq \eta, \tag{8.13}$$

(d) *if* $x, y \in K_0$ *and* $a^N \in B$, *then*

$$\left\| \frac{xM_{a^N}}{\|xM_{a^N}\|} - \frac{xM_{a^N}}{\|xM_{a^N}\|} \right\| < \rho. \tag{8.14}$$

Then Condition E *is satisfied.*

Proof. For $n = 1,2,\ldots$, let $\mathcal{R}_{\mathcal{H}^n} = \{(K,\mathcal{E}),(A^n,\mathcal{A}^n),h^{(n)},(g^{(n)},\tau^n)\}$ denote the RSCC induced by $\mathcal{H}^n$. Let $\mathbf{P} : K \times \mathcal{E} \to [0,1]$ be the filter kernel induced by $\mathcal{H}$, and let $\mathbf{Q}$ be the tr.pr.f. associated to $\mathcal{R}_{\mathcal{H}^1}$.

Since, for $n \geq 2$, $\mathcal{H}^n$ is a fully dominated, regular HMM if so is $\mathcal{H}$, it follows that $g^{(n)} : K \times A^n \to [0,\infty)$ is continuous for all $n \geq 2$.

Furthermore since

$$g^{(N)}(x,a^N) = \|xM_{a^N}\| \quad \text{and} \quad h^{(N)}(x,a^N) = \frac{xM_{a^N}}{\|xM_{a^N}\|}$$

if $\|xM_{a^N}\| > 0$, the hypotheses of Lemma 8.8 are satisfied with $\mathcal{P}_0$ replaced by $\mathcal{P}(K|\pi)$. Hence Lemma 8.8 implies that $(\mathcal{P}(K|\pi),\mathbf{Q})$ satisfies Condition E. Since the filter kernel $\mathbf{P}$ is equal to $\mathbf{Q}$, Condition E of Theorem 2.13 is satisfied. ∎

8.4. On HMMs with finite or denumerable state space. The purpose of this section is to verify that both Theorem 1.2 of Section 1.1 and Theorem 1.1 of [26] are special cases of Theorem 2.13.

PROPOSITION 8.10. *Let* $\mathcal{H} = \{S,P,A,R\}$ *be an ordinary HMM such that* S *and* A *are finite sets and the tr.pr.m.* P *is aperiodic and irreducible. Suppose also that Condition* KR *is satisfied. Then Condition* E *is satisfied.*

Proof. As in Section 1.1, for each $a \in A$ we define the stepping matrix $M(a)$ induced by $a \in A$ as

$$(M(a))_{i,j} = (P)_{i,j}(R)_{i,a}, \quad \forall i,j \in S.$$

For $a^n \in A^n$ we write $M(a^n) = M(a_1)\cdots M(a_n)$.

Since the hidden Markov chain is an aperiodic, irreducible Markov chain on a finite state space, it has a unique stationary probability vector, which we denote by π. Using moreover Condition KR, it is not difficult to prove that for every $\rho > 0$ we can find an integer N, elements $b_1, \ldots, b_N$ in A, an element $i \in S$ and a number $\eta_1 > 0$ such that

- the (i,i)th entry of the matrix $M(b^N)$ satisfies
$$(M(b^N))_{i,i} = \eta_1,$$
- if $x, y \in K$ are such that $(x)_i \geq (\pi)_i/2$ and $(y)_i \geq (\pi)_i/2$, then
$$\left\| \frac{xM(b^N)}{\|xM(b^N)\|} - \frac{yM(b^N)}{\|yM(b^N)\|} \right\| < \rho.$$

(Note that $(\pi)_j > 0$ for all $j \in S$ since the hidden Markov chain is irreducible.)

Therefore, if $\{(K, \mathcal{E}), (A^N, \mathcal{A}^N), h^{(N)}, (g^{(N)}, \tau^N)\}$ denotes the RSCC associated to the Nth iterate of $\mathcal{H}$, and we define $B \subset A^N$ by $B = \{(b_1, \ldots, b_N)\}$, then clearly $\tau^N(B) = 1$, since we assume that τ is the counting measure when the observation space is finite. If we define
$$K_0 = \{x \in K : (x)_i \geq (\pi)_i/2\}$$
and set $(\pi)_i/2 = \xi$, then $\mu(K_0) \geq \xi$ by Lemma 6.3. Furthermore, if we set $\eta = \eta_1 \xi$, we find that if $x \in K_0$ and $a^N \in B$, then
$$g^{(N)}(x, a^N) = \|xM(a^N)\| \geq \xi \eta_1 = \eta,$$
and if also $y \in K_0$, then
$$\|h^{(N)}(x, a^N) - h^{(N)}(y, a^N)\| < \rho.$$
Hence the hypotheses of Proposition 8.9 are fulfilled, and thus Condition E is satisfied. ■

In order to prove a similar result for the case when the state space is denumerable, we need to replace Condition KR by a condition more suitable for denumerable state spaces.

One such condition is the following one, introduced in [26].

DEFINITION 8.11. Let $\mathcal{H} = \{S, P, A, M\}$ be a HMM such that S and A are denumerable sets and the tr.pr.m. P is irreducible, strongly ergodic with limit distribution π. We say that $\mathcal{H} = \{S, P, A, M\}$ satisfies *Condition* B if the following holds:

For every $\rho > 0$, there exists $i_0 \in S$ such that for any compact set $C \subset K$ satisfying
$$\mu(C \cap \{x : (x)_{i_0} \geq (\pi)_{i_0}/2\}) \geq (\pi)_{i_0}/3, \quad \forall \mu \in \mathcal{P}(K|\pi), \tag{8.15}$$
there exist an integer N and a sequence $b_1, \ldots, b_N$ such that

- $\|\delta_{i_0} M(b^N)\| > 0$,
- if $x \in C \cap \{x : (x)_{i_0} \geq (\pi)_{i_0}/2\}$ then
$$\left\| \frac{xM(b^N)}{\|xM(b^N)\|} - \frac{\delta_{i_0} M(b^N)}{\|\delta_{i_0} M(b^N)\|} \right\| < \rho,$$
where as above $M(b^N) = M(b_1) \cdots M(b_N)$ and $M(b_n), n = 1, \ldots, N$, denotes the stepping matrix associated to b_n.

REMARK 8.12. Since $(K, \mathcal{E})$ is a complete, separable metric space when S is denumerable, and $\mathcal{P}(K|\pi)$ is a tight set if π is finite-dimensional, it follows easily from Theorem 7.1

that $\mathcal{P}(K|\pi)$ is also tight when π is infinite-dimensional. Lemma 6.3 then shows that we can always find a compact set C such that (8.15) holds.

PROPOSITION 8.13. *Let $\mathcal{H} = \{S, P, A, M\}$ be a HMM such that S and A are denumerable sets and the tr.pr.m. P is aperiodic, irreducible and strongly ergodic with limit distribution π. Suppose also that Condition* B *is satisfied. Then Condition* E *is satisfied.*

Proof. Let $\rho > 0$. Set $\rho_1 = \rho/2$. Choose $i_0 \in S$ and a compact set $C \subset K$ such that (8.15) holds. We can do this since Condition B is satisfied. Let $K_0 \in \mathcal{E}$ be defined by $K_0 = C \cap \{x : (x)_{i_0} \geq (\pi)_{i_0}/2\}$. Note that $(\pi)_{i_0} > 0$ since the hidden Markov chain is irreducible. From Condition B it follows that we can find an integer N and elements $b_1, \ldots, b_N$ in A such that if we define $M(b^N) = M(b_1) \cdots M(b_N)$ then $\|\delta_{i_0} M(b^N)\| > 0$, and if $x \in K_0$ then

$$\left\| \frac{xM(b^N)}{\|xM(b^N)\|} - \frac{\delta_{i_0} M(b^N)}{\|\delta_{i_0} M(b^N)\|} \right\| < \rho_1.$$

Now let $\{(K, \mathcal{E}), (A^N, \mathcal{A}^N), h^{(N)}, (g^{(N)}, \tau^N)\}$ denote the RSCC associated to the Nth iterate of $\mathcal{H}$, and define $B \subset A^N$ by $B = \{(b_1, \ldots, b_N)\}$. Then clearly $\tau^N(B) = 1$, since we assume that τ is the counting measure when the observation space is denumerable. Moreover, if we define $\xi = (\pi)_{i_0}/3$, then

$$\mu(K_0) \geq \xi, \quad \forall \mu \in \mathcal{P}(K|\pi),$$

because of Condition B.

Therefore, if we define

$$\eta = \|\delta_{i_0} M(b^N)\| (\pi)_{i_0}/2,$$

we find that if $a^N \in B$ and $x \in K_0$, then $g^{(N)}(x, a^N) = \|xM(b^N)\| \geq \eta$ and

$$\begin{aligned} \|h^{(N)}(x, a^N) - h^{(N)}(y, a^N)\| &= \left\| \frac{xM(b^N)}{\|xM(b^N)\|} - \frac{yM(b^N)}{\|yM(b^N)\|} \right\| \\ &\leq \left\| \frac{xM(b^N)}{\|xM(b^N)\|} - \frac{\delta_{i_0} M(b^N)}{\|\delta_{i_0} M(b^N)\|} \right\| + \left\| \frac{yM(b^N)}{\|yM(b^N)\|} - \frac{\delta_{i_0} M(b^N)}{\|\delta_{i_0} M(b^N)\|} \right\| \leq 2\rho_1 = \rho. \end{aligned}$$

Hence all the hypotheses of Proposition 8.9 are fulfilled, and thus Condition E is satisfied. ∎

8.5. Condition P. Although the hypotheses of Proposition 8.9 are more explicit than Condition E, in concrete situations it is not yet obvious how to verify them.

The purpose of this section is to introduce another set of conditions for fully dominated, regular HMMs which imply Condition E and which in concrete situations might be easier to check.

DEFINITION 8.14. Let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0\}, (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a fully dominated, regular HMM with stationary measure π. If there exist $F_0 \in \mathcal{F}$ and $B_0 \in \mathcal{A}$ such that

(1) $\pi(F_0) > 0$,
(2) $\tau(B_0) > 0$,
(3) there exist positive numbers d_0, D_0 and β_0 such that for every $a \in B_0$ there exists $F_1(a) \in \mathcal{F}$ such that

(a) $F_1(a) \subset F_0$,
(b) $\lambda(F_1(a)) \geq \beta_0$,
(c) $d_0 \leq m(s,t,a) \leq D_0$ for all $(s,t) \in F_0 \times F_1(a)$,
(d) $m(s,t,a) = 0$ for all $(s,t) \in F_0 \times (F_0 \setminus F_1(a))$,

then we say that $\mathcal{H}$ *satisfies Condition* P.

REMARK 8.15. The idea of Condition P comes from Kochman and Reeds [30] and their proof of the fact that Condition A of [21] (see Section 1.2) implies that their "rank one condition" holds.

REMARK 8.16. Condition P, as introduced above, is a rather straightforward generalisation of a similar condition introduced in [25, Section 9], for a HMM with denumerable state space.

THEOREM 8.17. *Let $\mathcal{H} = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$ be a fully dominated, regular, strongly ergodic HMM with stationary measure π. For $n = 1,2,\ldots$, let $\mathcal{H}^n = \{(S,\mathcal{F},\delta_0),(p^n,\lambda),(A^n,\mathcal{A}^n,\varrho^{(n)}),(m^n,\tau^n)\}$ be the nth iterate of $\mathcal{H}$. Suppose there exists an integer N_0 such that $\mathcal{H}^{N_0}$ satisfies Condition P. Then Condition E is satisfied.*

Proof. Obviously we may assume that $N_0 = 1$. Let F_0, B_0, $m : S \times S \times A \to [0,\infty)$, d_0, D_0, η_0 and $F_1(a)$, $a \in B_0$, be such that the hypotheses of Condition P are satisfied. For $n = 1,2,\ldots$, let $M_{a^n} : \mathcal{Q}_\lambda(S,\mathcal{F}) \to \mathcal{Q}_\lambda(S,\mathcal{F})$ denote the stepping function (for $\mathcal{H}^n$) determined by $a^n \in A^n$.

Now let $\rho > 0$. In view of Proposition 8.9 we want to prove that there exist an integer N, a set K_0, a number $\xi > 0$, a set $B \in \mathcal{A}^N$, and numbers $\beta,\eta > 0$ such that

(i) $\mu(K_0) \geq \xi$ for all $\mu \in \mathcal{P}(K|\pi)$,
(ii) $\tau^N(B) \geq \beta$,
(iii) for all $x \in K_0$ and all $a^N \in B$ we have $\|xM_{a^N}\| \geq \eta$,
(iv) for all $x,y \in K_0$ and $a^N \in B$,

$$\left\| \frac{xM_{a^N}}{\|xM_{a^N}\|} - \frac{yM_{a^N}}{\|yM_{a^N}\|} \right\| < \rho. \tag{8.16}$$

The choice of K_0 is simple: $K_0 = \{x \in K : x(F_0) \geq \pi(F_0)/2\}$, where F_0 is determined by Condition P. Since $\pi(F_0) > 0$, Lemma 6.3 shows that if we set $\xi = \pi(F_0)/2$, then $\mu(K_0) \geq \xi$ if $\mu \in \mathcal{P}(K|\pi)$, and hence hypothesis (a) of Proposition 8.9 is fulfilled.

Next, set $\kappa = D_0/d_0$ where d_0 and D_0 occur in hypothesis (c) of Condition P, and let $I_{F_0} : S \to \{0,1\}$ denote the indicator function of F_0. From the hypotheses of Condition P it follows that if $a \in B_0$ and we define $m_a : S \times S \to [0,\infty)$ by $m_a(s,t) = m(s,t,a)I_{F_0}(s)$, then m_a has rectangular support $F_0 \times F_1(a)$ and

$$\sup\left\{ \frac{m_a(s_1,t_1)m_a(s_2,t_2)}{m_a(s_2,t_1)m_a(s_1,t_2)} : s_1,s_2 \in F_0,\ t_1,t_2 \in F_1(a) \right\} \leq \kappa^2. \tag{8.17}$$

We now simply define

$$N = \min\left\{ n \geq 1 : 2\left(\frac{\kappa-1}{\kappa+1}\right)^n < \rho \right\}, \tag{8.18}$$

and define the set B in $\mathcal{A}^N$ by $B = B_1 \times \cdots \times B_N$, where $B_i = B_0$, $i = 1,\ldots,N$.

By setting $\beta = \tau(B_0)^N$ we find that $\tau^N(B) = \beta > 0$, and hence hypothesis (b) of Proposition 8.9 is fulfilled.

Next, let $x \in K_0$ and $a^N \in B$. Then $\|xM_{a^N}\| = \int_S \int_S m^N(s,t,a^N)\, x(ds)\, \lambda(dt)$. From hypothesis (c) of Condition P it follows that if $s \in F_0$, then

$$\int_S m^N(s,t,a^N)\, \lambda(dt) \geq d_0^N \prod_{i=1}^{N} \lambda(F(a_i)) \geq d_0^N \beta_0^N.$$

Therefore, if we define

$$\eta = (\pi(F)/2) d_0^N \beta_0^N$$

and use the fact that $x(F) \geq \pi(F)/2$ if $x \in K_0$, we find that

$$\|xM_{a^N}\| \geq \int_F \int_S m^N(s,t,a^N)\, x(ds)\, \lambda(dt) \geq \eta.$$

Hence hypothesis (c) of Proposition 8.9 is fulfilled.

It remains to show that if $x, y \in K_0$ and $(a_1,\ldots,a_N) = a^N \in B$, then (8.16) holds. But this follows immediately from Theorem 8.3 and the definition of N (see (8.18)). Hence also hypothesis (d) of Proposition 8.9 is satisfied, and hence Condition E holds by Proposition 8.9. ∎

9. Examples

Our first example is obtained by making a denumerable partition of the state space.

Example 9.1. Let $\mathcal{H}_1 = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$ be a fully dominated, regular, strongly ergodic HMM with limit measure π such that A is a denumerable set, ϱ is the discrete metric and τ is the counting measure. Suppose also that

- for each $a \in A$ there exists a set $S_a \in \mathcal{F}$ such that $\lambda(S_a) > 0$,
- $\bigcup_a S_a = S$ and $S_a \cap S_b = \emptyset$ if $a \neq b$,
- for each $a \in A$,

$$m(s,t,a) = p(s,t) I_{S_a}(s),$$

where $I_{S_a} : S \to \{0,1\}$ denotes the indicator function of the set S_a.

Theorem 9.2. *Let $\mathcal{H}_1 = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$ be the HMM defined in Example 9.1, let $\mathbf{P}$ denote the induced filter kernel, and for $n = 2,3,\ldots$, let $\mathcal{H}_1^n = \{(S,\mathcal{F},\delta_0),(p^n,\lambda),(A^n,\mathcal{A}^n,\varrho^{(n)}),(m^n,\tau^n)\}$ be the nth iterate of $\mathcal{H}_1$.*

Suppose that there exist an integer N, $a_0 \in A$, $b^N = (b_1,\ldots,b_N) \in A^N$ such that $b_N = a_0$, and positive numbers d_0, D_0 satisfying $d_0 \leq D_0$, such that $\pi(S_{a_0}) > 0$ and

$$d_0 \leq m^N(s,t,b^N) \leq D_0, \qquad \forall (s,t) \in S_{a_0} \times S_{a_0}.$$

Then the filter kernel $\mathbf{P}$ is weakly ergodic.

Proof. We shall first verify that $\mathcal{H}_1^N$ fulfills the hypotheses of Condition P.

First, let $F_0 = S_{a_0}$. By assumption $\pi(S_{a_0}) > 0$, and therefore obviously $\pi(F_0) > 0$. Hence hypothesis (1) of Condition P is satisfied with this choice of F_0.

Next, set $B = \{b^N\}$. Since τ is the counting measure, $\tau^N(B) = 1 > 0$; hence hypothesis (2) of Condition P holds.

Now define $F_1(a_0) = F_0$. Evidently $F_1(a_0) \subset F_0$. Since $\pi(F_0) > 0$ and

$$\pi(F_0) = \int_{F_0} p(s,t)\,\pi(ds)\,\lambda(dt) \le D_0 \lambda(F_0)\pi(F_0),$$

it follows that $\lambda(F_0) > 0$. Hence (3a)–(3b) of Condition P are satisfied.

Further, since $m(s,t,a_0) = p(s,t)$ if $(s,t) \in F_0 \times F_0$, and $m(s,t,a_0) = 0$ if $(s,t) \in F_0 \times (S \setminus F)$, it is clear that (3c)–(3d) hold. Hence Condition P is satisfied.

Theorem 8.17, Proposition 8.9 and Theorem 2.13 now imply that the filter kernel $\mathbf{P}$ is weakly contracting. If furthermore the Markov chain is uniformly ergodic, then hypothesis (D) of Theorem 2.13 is fulfilled and hence $\mathbf{P}$ is weakly ergodic.

In order to prove that the filter kernel is weakly ergodic without this extra assumption, we shall use a result of [40]. We shall show that the following condition is satisfied.

CONDITION $\mathcal{E}1$. *There exists $x_0 \in K$ such that for every $\epsilon > 0$,*

$$\liminf_{n\to\infty} \mathbf{P}^n(x, B(x_0,\epsilon)) > 0, \qquad \forall x \in K, \tag{9.1}$$

where $B(x_0,\epsilon) = \{y \in K : \delta_{TV}(x_0,y) < \epsilon\}$.

Once we have verified Condition $\mathcal{E}1$, it follows from [40, Proposition 2.1 and Theorem 3.1] that $\{\mathbf{P}^n(x_0,\cdot),\ n = 1,2,\dots\}$ is a tight sequence, since obviously Condition $\mathcal{E}1$ implies Condition $\mathcal{E}$ of [40]. (Condition $\mathcal{E}$ is also formulated at the end of Section 7.2.) Then the filter kernel is weakly ergodic by Theorem 2.13(C).

To verify Condition $\mathcal{E}1$ we argue as follows. Set $F_0 = S_{a_0}$, define $k : F_0 \times F_0 \to [0,\infty)$ by $k(s,t) = m^N(s,t,b^N)$, define $K : S \times \mathcal{F} \to [0,\infty)$ by $K(s,F) = \int_S k(s,t)\,\lambda(dt)$ and set $\kappa = D_0/d_0$. Since $d_0 \le m^N(s,t,b^N) \le D_0$ if $(s,t) \in F_0 \times F_0$, there exist a function $q : F_0 \to (0,\infty)$ satisfying $\int_{F_0} q(t)\,\lambda(dt) = 1$ and a number $\beta > 0$ such that $\int_{F_0} k(s,t)q(t)\,\lambda(dt) = \beta q(s)$ (see e.g. [17]). Moreover, if we define $x_0 \in K$ by

$$x_0(F) = \int_F q(t)\,\lambda(dt),$$

it follows from Theorem 8.3 that for any $x \in K$ such that $x(F_0) > 0$,

$$\left\| \frac{xK^n}{\|xK^n\|} - x_0 \right\| \le 2\left(\frac{\kappa-1}{\kappa+1}\right)^n.$$

Now let $\epsilon > 0$. Define

$$N_0 = \min\left\{ n \ge 2 : 2\left(\frac{\kappa-1}{\kappa+1}\right)^{n-1} < \epsilon \right\}, \qquad \alpha = d_0^{N_0}\lambda(F_0)^{N_0}.$$

It follows that if $\mu \in \mathcal{P}(K,\mathcal{E})$ satisfies

$$\mu(\{z \in K : z(F_0) > 0\}) \ge \pi(F_0)/3, \tag{9.2}$$

then

$$\mu\mathbf{P}^{NN_0}(B(x_0,\epsilon)) \ge \alpha\pi(F_0)/3. \tag{9.3}$$

Therefore, if for every $x \in K$ we could find an integer N_1, which may depend on x, such that

$$\mathbf{P}^n(x, \{z : z(F_0) > 0\}) \geq \pi(F_0)/3, \qquad \forall n \geq N_1, \tag{9.4}$$

then (9.1) would follow and hence Condition $\mathcal{E}1$ would be satisfied.

Thus let $x \in K$. To find N_1 such that (9.4) holds, we shall use Theorems 6.1 and 7.1.

From Theorem 6.1 it follows that we can find an integer N_1 such that

$$\|\bar{b}(\delta_x \mathbf{P}^n) - \pi\| < \pi(F_0)/6, \qquad \forall n \geq N_1,$$

and Theorem 7.1 shows that

$$\inf\{d_K(\delta_x \mathbf{P}^n, \nu) : \nu \in \mathcal{P}(K|\pi)\} < \pi(F_0)/6, \qquad n \geq N_1. \tag{9.5}$$

From Lemma 6.3 we also know that $\nu\{z : z(F_0) \geq \pi(F_0)/2\} \geq \pi(F_0)/2$ for all $\nu \in \mathcal{P}(K|\pi)$, which together with (9.5) implies that

$$\mathbf{P}^n(x, \{z : z(F_0) \geq \pi(F_0)/2\}) \geq \pi(F_0)/2 - \pi(F_0)/6 = \pi(F_0)/3 \tag{9.6}$$

for $n \geq N_1$. Thus (9.4) is proved and the proof of Theorem 9.2 is complete. ∎

Remark 9.3. It is clear that the HMMs of Examples 2.15, 2.17 and 2.18 all fulfill the hypotheses of Example 9.1, but only the HMM of Example 2.15 also fulfills the hypotheses of Theorem 9.2.

Corollary 9.4. *Suppose $\mathcal{H}_1 = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ satisfies the hypotheses of Theorem 9.2. Suppose also that A is finite. Set $\mathcal{H}_1^1 = \mathcal{H}_1$, and for $n = 2, 3, \ldots$, let $\mathcal{H}_1^n = \{(S, \mathcal{F}, \delta_0), (p^n, \lambda), (A^n, \mathcal{A}^n, \varrho^{(n)}), (m^n, \tau^n)\}$ denote the nth iterate of $\mathcal{H}_1$, and for $n = 1, 2, \ldots$ and $a^n \in A^n$, let $M_{a^n} : \mathcal{Q}_\lambda(S, \mathcal{F}) \to \mathcal{Q}_\lambda(S, \mathcal{F})$ be the stepping function determined by $a^n \in A^n$.*

Let $h : [0, 1] \to [0, 1/(e \cdot \ln(2))]$ be defined by

$$h(t) = -t \ln(t)/\ln(2) \text{ if } 0 < t \leq 1 \quad \text{and} \quad h(0) = 0. \tag{9.7}$$

For $x \in \mathcal{P}_\lambda(S, \mathcal{F})$, let $\{Y_{n,x}, n = 1, 2, \ldots\}$ denote the observation sequence generated by the HMM $\mathcal{H}_1$ and the initial distribution x, and for $n = 1, 2, \ldots$ and $x \in K$, define the entropy of $Y_{n,x}$ by

$$H^n(Y; x) = \sum_{a^n \in A^n} h(\|x M_{a^n}\|),$$

and for $n = 2, 3, \ldots$, define the entropy rate $H_R^n(Y; x)$ of $Y_{n,x}$ by

$$H_R^n(Y; x) = H^{n+1}(Y; x) - H^n(Y; x).$$

Let $K = \mathcal{P}_\lambda(S, \mathcal{F})$, let $\mathcal{E}$ be the Borel field induced by the total variation distance, and let $\mathbf{P} : K \times \mathcal{E} \to [0, 1]$ be the filter kernel induced by $\mathcal{H}_1$.

Then

(a)

$$H_R^n(Y; x) = \sum_{a \in A} \int_K h(\|z M_a\|)\, \mathbf{P}^n(x, dz), \tag{9.8}$$

(b) *there exists a unique measure* $\mu \in \mathcal{P}(K, \mathcal{E})$ *such that for* $x \in K$,

$$\lim_{n\to\infty} H_R^n(Y; x) = \sum_{a\in A} \int_K h(\|zM_a\|)\, \mu(dz).$$

Proof. (a) follows easily from the scaling property (3.2) and formula (3.6); and (b) follows by combining (9.8), Theorem 9.2, and the fact that for each $a \in A$ the function $h_a : K \to [0, 1]$ defined by $h_a(y) = h(\|yM_a\|)$ is continuous. ∎

REMARK 9.5. Since the HMM considered in Example 2.15 fulfills the hypotheses of Theorem 9.2, it is clear that the entropy formula (2.14) holds.

Before we present our next example we recall the following notation. If A and B are two sets we let $A \triangle B$ be the set consisting of those elements that belong either to A or to B, but not to both.

EXAMPLE 9.6. Let $\mathcal{H}_2 = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a fully dominated HMM such that the probability density kernel $m : S \times S \times A \to [0, \infty)$ can be written as

$$m(s, t, a) = p(s, t)r(t, a),$$

where $r : S \times A \to [0, \infty)$ is a measurable function satisfying

$$\int_A r(t, a)\, \tau(da) = 1, \qquad \forall t \in S.$$

($\mathcal{H}_2$ is thus an ordinary HMM (see Section 2.3).)

We assume that

$$\sup\{p(s, t) : s, t \in S\} < \infty, \qquad \sup\{r(t, a) : t \in S,\ a \in A\} < \infty.$$

For each $a \in A$, set $S_+(a) = \{t : r(t, a) > 0\}$. We assume that $\lambda(S_+(a)) > 0$ for all $a \in A$. We also assume that the probability density kernel r is such that for every $\epsilon > 0$ we can find an $\eta > 0$ such that if $\varrho(a, b) < \eta$, then

$$\lambda(S_+(a) \triangle S_+(b)) < \epsilon \tag{9.9}$$

and

$$|r(t, a) - r(t, b)| < \epsilon, \qquad \forall t \in S_+(a) \cap S_+(b). \tag{9.10}$$

PROPOSITION 9.7. *Let* $\mathcal{H}_2$ *be as in Example* 9.6. *Then* $\mathcal{H}_2$ *is regular.*

Proof. We need to prove that $\overline{M} : \mathcal{Q}_\lambda(S, \mathcal{F}) \times A \to \mathcal{Q}_\lambda(S, \mathcal{F})$ is a continuous function where

$$\overline{M}(x, a)(F) = \int_S \int_F p(s, t)r(t, a)\, \lambda(dt)\, x(ds).$$

That $\overline{M}$ is continuous in the first variable follows easily from the boundedness condition regarding the probability density kernel r.

That it is also continuous in the second variable follows easily from (9.9) and (9.10) together with the hypothesis that both $r : S \times A \to [0, \infty)$ and $p : S \times S \to [0, \infty)$ are uniformly bounded. Since the proof is elementary we omit the details. ∎

THEOREM 9.8. *Let* $\mathcal{H}_2 = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ *be the HMM defined in Example* 9.6*, set* $\mathcal{H}_2 = \mathcal{H}_2^1$*, for* $n = 2, 3, \ldots$ *let*

$$\mathcal{H}_2^n = \{(S, \mathcal{F}, \delta_0), (p^n, \lambda), (A^n, \mathcal{A}^n, \varrho^{(n)}), (m^n, \tau^n)\}$$

be the nth iterate of $\mathcal{H}_2$*, and let* $\mathbf{P}$ *denote the filter kernel induced by* $\mathcal{H}_2$*. Suppose that*

(a) *the HMM* $\mathcal{H}_2$ *is* strongly ergodic *with stationary measure* π*;*
(b) *there exists* $F_0 \in \mathcal{F}$*, an integer* N*,* $B_0 \in \mathcal{A}^N$*, and numbers* $\beta_0, c > 0$ *such that*

  (i) $\pi(F_0) > 0$,
  (ii) $\tau^N(B_0) > 0$,
  (iii) $S_+((a^N)_N) \subset F_0$ *for all* $a^N \in B_0$*,*
  (iv) $\lambda(S_+((a^N)_N)) \geq \beta_0$ *for all* $a^N \in B_0$*,*
  (v) $m^N(s, t, a^N) \geq d_0$ *for all* $(s, t) \in F_0 \times S_+((a^N)_N)$ *and all* $a^N \in B_0$*.*

Then the filter kernel $\mathbf{P}$ *is weakly contracting.*

If furthermore $\mathcal{H}_2$ *is uniformly ergodic, then* $\mathbf{P}$ *is weakly ergodic.*

Proof. Let N be the integer mentioned in the theorem. It suffices to verify that the HMM $\mathcal{H}_2^N$ satisfies hypotheses (1)–(3) of Condition P.

We shall verify these hypotheses when F_0, B_0, β_0, d_0 are as in the hypotheses of Theorem 9.8, $F_1(a^N) = S_+((a^N)_N)$, and

$$D_0 = \sup\{(p(s,t)r(t,a))^N : s, t \in S,\ a \in A\}. \tag{9.11}$$

Since $\pi(F_0), \tau^N(B_0) > 0$, hypotheses (1) and (2) of Condition P are satisfied. Since $\lambda(F_1(a^N)) \geq \beta_0$ for all $a^N \in B_0$ because of (iv), hypothesis (3b) of Condition P is satisfied. From (iii) we also know that $F_1(a^N) \subset F_0$ if $a^N \in B$, and hence (3a) of Condition P is satisfied. Furthermore, from the definition of $S^+(a)$ we can also conclude that if $a^N \in B$ and $t \notin F_1(a^N)$, then $m(s, t, a^N) = 0$ for all $s \in S$ and in particular for all $s \in F_0$. Hence (3d) of Condition P is satisfied. Finally it is obvious that (3c) of Condition P holds when d_0 is as in (v) and D_0 is defined by (9.11). Hence, also hypothesis (3) of Condition P is satisfied. The conclusion of the theorem now follows from Theorems 8.17 and 2.13. ∎

REMARK 9.9. It is easy to show that the HMM considered in Example 2.16 of Section 2.7 fulfills the hypotheses of Theorem 9.8.

10. Discussion

10.1. On entropy. Consider a fully dominated, regular, strongly ergodic HMM

$$\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$$

with finite state space. Let $x \in \mathcal{P}_\lambda(K, \mathcal{E})$ and let $\{Y_{n,x},\ n = 1, 2, \ldots\}$ denote the observation sequence generated by the HMM $\mathcal{H}$ and the initial distribution x. Let $\mathbf{P}$ denote the induced filter kernel, let $h : [0, 1] \to [0, 1/(e \cdot \ln(2))]$ be defined by (9.7), and for

$n = 1, 2, \ldots$ define

$$H_R^n(Y;x) = \sum_{a\in A} \int h(\|zM_a\|)\,\mathbf{P}^n(x, dz).$$

Is it always true that there exists a constant h_0 such that

$$\lim_{n\to\infty} H_R^n(Y;x) = h_0, \qquad \forall x \in \mathcal{P}_\lambda(K, \mathcal{E})? \tag{10.1}$$

Recall that Blackwell's conjecture was that if a HMM has a finite state space and is determined by a lumping function and the tr.pr.m. determining the hidden Markov chain is indecomposable, then the filter kernel has a unique invariant measure. As pointed out in the introduction this is not in general true, but it seems likely that the exceptional cases are so special that the relation (10.1) still might be true.

10.2. Convergence rates. Consider again a fully dominated, regular, uniformly ergodic HMM $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ with limit measure π, let $P : S\times\mathcal{F} \to [0, 1]$ be the tr.pr.f. determined by (p, λ), and let $\mathbf{P}$ denote the induced filter kernel. Let $\alpha > 0$ and suppose that

$$\lim_{n\to\infty} \sup\{n^\alpha \|P^n(s, \cdot) - \pi\| : s \in S\} = 0$$

and $\mathcal{H}$ satisfies Condition P. Let μ be the unique invariant of the filter kernel $\mathbf{P}$. Does it then follow that

$$\lim_{n\to\infty} n^\beta d_K(\mathbf{P}^n(x, \cdot), \mu) = 0$$

if $\beta < \alpha$? If also

$$\lim_{n\to\infty} \sup\{e^{n\alpha} \|P^n(s, \cdot) - \pi\| : s \in S\} = 0,$$

does it follow that there exists a $\beta > 0$ such that for all $x \in \mathcal{P}_\lambda(S, \mathcal{F})$,

$$\lim_{n\to\infty} e^{n\beta} d_K(\mathbf{P}^n(x, \cdot), \mu) = 0?$$

M. Ślęczka [38] and H. Mairer [16] consider RSCCs and prove geometric convergence of the distributions of the state sequence to the unique invariant limit distribution. The former paper uses the Vasershtein coupling. In the latter a very interesting class of couplings is introduced for RSCCs for which the index probability distribution is independent of the state space. This class of couplings, generalised to ordinary RSCCs, might be quite useful, both for verifying Condition E and for proving convergence rates for the distributions of the Markov chain generated by the filter kernel of a fully dominated, regular HMM.

10.3. On HMMs with finite state space and observation space. When the HMM $\{S, A, P, M\}$ under consideration is uniformly ergodic and both the state space and the observation space are finite, then Condition KR (see Section 1.1) is both a sufficient and necessary condition for weak ergodicity of the filter kernel. In practice though it seems that Condition A is somewhat easier to verify.

An important related problem is to classify those ergodic, aperiodic transition probability matrices P for which there exists a partition of P such that Condition KR is *not* satisfied. In [26, Section 11], some partial result on this problem was given, but a full classification is still lacking. Loosely speaking, it seems as if the class of uniformly ergodic HMMs with finite state and observation space for which the filter kernel is not weakly ergodic, is of the same size as the class of those HMMs which satisfy Condition KR but do not satisfy Condition A.

10.4. On HMMs with denumerable state space and denumerable observation space. To prove weak ergodicity for a given HMM, which has a denumerable and infinite state space but for which the observation space is finite, seems in general a rather complicated task, the reason being that the stepping matrices that occur have in general infinitely many nonzero rows and columns, and therefore it seems difficult to verify Condition E or B. What one would probably need is some kind of generalisation of Perron's theorem (see e.g. [14]) for positive, finite-dimensional matrices to nonnegative infinite-dimensional matrices. (Perhaps some of the papers by D. Vere-Jones from the 1960s can be useful for this problem—see e.g. [43]).

If instead the observation space is infinite, then it is more likely that a stepping matrix has only finitely many nonzero columns, and then in concrete examples it is more likely that for example Condition P is satisfied for some iteration of the HMM under consideration.

10.5. More on exceptional cases. In Section 2.7 we gave two examples such that the induced filter kernel of the given fully dominated and regular HMM is not weakly ergodic in spite of the fact that the tr.pr.f. of the hidden Markov chain is uniformly ergodic. Is it possible to classify all exceptional cases?

References

[1] C. Anton Popescu, *Applications of dependence with complete connections to hidden Markov models*, Bull. Transilv. Univ. Braşov Ser. III 5, (2012), 23–32.

[2] M. Barnsley, S. Demko, J. Elton and J. Gerinomo, *Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities*, Ann. Inst. H. Poincaré Probab. Statist. 24 (1988), 367–394.

[3] L. E. Baum and T. Petrie, *Statistical inference for probabilistic functions of finite state Markov chains*, Ann. Math. Statist. 37 (1966), 1554–1563.

[4] P. Baxendale, P. Chigansky and R. Liptzer, *Asymptotic stability of the Wonham filter: Ergodic and nonergodic signals*, SIAM J. Control Optim. 43 (2004), 643–669.

[5] D. Blackwell, *The entropy of functions of finite-state Markov chains*, in: Trans. First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Praha, 1957, 13–20.

[6] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

[7] O. Cappé, E. Moulines and T. Ryden, *Inference in Hidden Markov Models*, Springer Ser. Statist., Springer, 2005.

[8] P. Chigansky and R. van Handel, *A complete solution to Blackwell's unique ergodicity problem for hidden Markov chains*, Ann. Appl. Probab. 20 (2010), 2318–2345.

[9] G. Choquet, *Lectures on Analysis, Vol. II, Representation Theory*, Benjamin, New York, 1969.

[10] P. Diaconis and D. Freedman, *Iterated random functions*, SIAM Rev. 41 (1989), 45–76.

[11] G. B. Di Masi and Ł. Stettner, *Ergodicity of hidden Markov models*, Math. Control Signals Systems 17 (2005), 269–296.

[12] R. M. Dudley, *Real Analysis and Probability*, Cambridge Univ. Press, 2002.

[13] H. Furstenberg and H. Kesten, *Products of random matrices*, Ann. Math. Statist. 31 (1960), 457–469.

[14] F. R. Gantmacher, *The Theory of Matrices*, Chelsea, New York, 1965.

[15] A. J. Goldsmith and P. P. Varaiya, *Capacity, mutual information and coding for finite-state Markov channels*, IEEE Trans. Inform. Theory 42 (1996), 868–886.

[16] M. Hairer, *Exponential mixing properties of stochastic PDEs through asymptotic coupling*, Probab. Theory Related Fields 124 (2002), 345–380.

[17] E. Hopf, *An inequality for positive linear integral operators*, J. Math. Mech. 12 (1963), 683–692.

[18] M. Iosifescu and S. Grigorescu, *Dependence with Complete Connections and Its Applications*, Cambridge Univ. Press, Cambridge, 1990.

[19] M. Iosifescu and R. Theodorescu, *Random Processes and Learning*, Springer, Berlin, 1969.

[20] T. Kaijser, *A limit theorem for the conditional distributions of Markov chains with incomplete state information*, report LiH-Mat-R-73-7, Linköping Univ., Linköping, 1973.

[21] T. Kaijser, *A limit theorem for partially observed Markov chains*, Ann. Probab. 3 (1975), 677–696.

[22] T. Kaijser, *A limit theorem for Markov chains in compact metric spaces with applications to products of random matrices*, Duke Math. J. 45 (1978), 311–349.

[23] T. Kaijser, *On a new contraction condition for random systems with complete connections*, Rev. Roumaine Math. Pures Appl. 26 (1981), 1075–1117.

[24] T. Kaijser, *On a theorem of Karlin*, Acta Appl. Math. 34 (1994), 51–69.

[25] T. Kaijser, *On Markov chains induced by partitioned transition probability matrices*, arXiv: 0907.4502v1 (2009).

[26] T. Kaijser, *On Markov chains induced by partitioned transition probability matrices*, Acta Math. Sinica 20 (2011), 441–476.

[27] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed., Springer, New York, 2002.

[28] L. V. Kantorovich, *On the translocation of masses*, Dokl. Akad. Nauk USSR 37 (1942), 199–201 (in Russian); English transl.: Managements Science 5 (1958), 1–4.

[29] S. Karlin, *Some random walks arising in learning models. I*, Pacific J. Math. 3 (1953), 139–159.

[30] F. Kochman and J. Reeds, *A simple proof of Kaijser's unique ergodicity result for hidden Markov α-chains*, Ann. Appl. Probab. 16 (2006), 1805–1815.

[31] H. Kunita, *Asymptotic behaviour of the nonlinear filtering errors of Markov processes*, J. Multivariate Anal. 1 (1971), 365–393.

[32] A. Lasota and J. Yorke, *A lower bound for Markov operators and iterated function systems*, Random Comput. Dynam. 2 (1994), 41–77.

[33] T. Lindwall, *Lectures on the Coupling Method*, Wiley, New York, 1992.

[34] M. Loève, *Probability Theory*, 3rd ed., van Nostrand, 1963.

[35] F. Norman, *Markov Processes and Learning Models*, Academic Press, New York, 1972.

[36] L. R. Rabiner and B. H. Juang, *An introduction to hidden Markov models*, IEEE ASSP Mag. 3 (1986), 4–16.

[37] D. Revuz, *Markov Chains*, North-Holland, Amsterdam, 1975.

[38] M. Ślęczka, *The rate of convergence for iterated function systems*, Studia Math. 205 (2011), 201–214.

[39] Ł. Stettner, *On invariant measures of filtering processes*, in: Stochastic Differential Systems (Bad Honnef, 1988), Lecture Notes in Control and Inform. Sci. 126 (1989), 9–24.

[40] T. Szarek, *Feller proceses on nonlocally compact spaces*, Ann. Probab. 34 (2006), 1849–1863.

[41] R. van Handel, *The stability of conditional Markov processes and Markov chains in random environments*, Ann. Probab. 37 (2009), 1876–1925.

[42] L. N. Vasershtein, *Markov processes over denumerable products of spaces describing large systems of automata*, Problemy Peredachi Informatsii 5 (1969), 64–72 (in Russian).

[43] D. Vere-Jones, *Ergodic properties of nonnegative matrices*, Pacific J. Math. 22 (1967), 361–386.