Convergence in distribution for filtering processes associated to Hidden Markov Models with densities.

Thomas Kaijser

Department of Mathematics, Linköping University,

S-581 83 Linköping, Sweden ; *thkai@mai.liu.se*

Abstract

A Hidden Markov Model generates two basic stochastic processes, a Markov chain, which is hidden, and an observation sequence. The filtering process of a Hidden Markov Model is, roughly speaking, the sequence of conditional distributions of the hidden Markov chain that is obtained as new observations are received.

It is well-known, that the filtering process itself, is also a Markov chain. A classical, theoretical problem is to find conditions which implies that the distributions of the filtering process converge towards a unique limit measure.

This problem goes back to a paper of D Blackwell for the case when the Markov chain takes its values in a finite set and it goes back to a paper of H Kunita for the case when the state space of the Markov chain is a compact Hausdorff space.

Recently, due to work by F Kochmann, J Reeds, P Chigansky and R van Handel, a necessary and sufficient condition for the convergence of the distributions of the filtering process has been found for the case when the state space is finite. This condition has since been generalised to the case when the state space is denumerable.

In this paper we generalise some of the previous results on convergence in distribution to the case when the Markov chain and the observation sequence of a Hidden Markov Model take their values in complete, separable, metric spaces; it has though been necessary to assume that both the transition probability function of the Markov chain and the transition probability function that generates the observation sequence have densities.

Keywords: Hidden Markov Models, filtering processes, Markov chains on nonlocally compact spaces, convergence in distribution, barycenter, ergodicity, Kantorovich metric.

Mathematics Subject Classification (2000): Primary 60J05; Secondary 60F05.

1 Introduction

Let $\{X_n, n = 0, 1, 2, ...\}$ be an aperiodic, irreducible Markov chain with finite state space S, transition probability matrix (tr.pr.m) P and initial distribution p_0. Let $g : S \to A$ be a mapping from S to another space A and, for $n = 1, 2, ...$, define $Y_n = g(X_n)$. (The function g is sometimes called a *lumping function.*) For each $s \in S$ and every integer $n \geq 1$ define

$$Z_{n,s} = Pr[X_n = s|Y_0, Y_1, Y_2, ..., Y_n]$$

and set $Z_n = (Z_{n,s}, s \in S)$. Clearly Z_n is a probability vector on the finite set S. It is also random. Hence $\{Z_n,\ n = 1, 2, ...\}$ is a sequence of random probability vectors. It is well-known that the sequence $\{Z_n, n = 1, 2, ...\}$ of conditional distributions is also a Markov chain. Let $\mathbf{P}$ denote the transition probability function (tr.pr.f) for this chain. A rather natural question to ask 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.

In the paper [17] from 2006 by F Kochman and J Reeds, the authors gave a sufficient condition for ergodicity and in the paper [6] from 2010 by P Chigansky and R van Handel the authors proved that this condition is also necessary.

In the classical paper [4] from 1957, D Blackwell conjectured that if the Markov chain $\{X_n, n = 0, 1, 2...\}$ is indecomposable, then the tr.pr.f $\mathbf{P}$ has a unique invariant measure. However, in the paper [13] from 1975 a counterexample to Blackwell's conjecture was given and in the same paper a rather weak sufficient condition for unicity was presented, a condition only slightly stronger than the condition introduced in [17] by Kochman and Reeds.

In the paper [16] from 2011 (see also [15]) the result obtained by Kochman and Reeds was generalised to the case when the state space S is denumerable. In the paper [6] by Chigansky and van Handel the main results of [15] and [16] are proved with different methods.

We shall now generalise the "set-up" described above, somewhat. This time we let S be an arbitrary set, let $\mathcal{F}$ be a $\sigma - algebra$ on S and let $\{X_n, n = 0, 1, 2, ...\}$ be an aperiodic and ergodic Markov chain taking values in S and which is generated by a tr.pr.f $P : S \times \mathcal{F} \to [0, 1]$ and an initial probability distribution p_0 on $(S, \mathcal{F})$. We shall denote the set of all probability measures on $(S, \mathcal{F})$ by $\mathcal{P}(S, \mathcal{F})$. As above, we let $g : S \to A$ be a mapping from S onto another set A, which we assume is a denumerable set. We also assume that g is measurable. This implies that for each $a \in A$ we can define a subset $S_a \in \mathcal{F}$ by $S_a = \{s : g(s) = a\}$. Clearly $\cup_{a \in A} S_a = S$.

Let us again - as above - for $n = 1, 2, ...$, define $Y_n = g(X_n)$ and let again Z_n denote the *conditional distribution* of X_n given $Y_1, Y_2, ..., Y_n$. But, what do we more precisely mean by saying that Z_n denotes the conditional distribution of X_n given $Y_1, Y_2, ...Y_n$?

Well, to illustrate, let us consider the distribution of Z_1. If we assume that $Y_1 = a$, it follows that Z_1 is the probability distribution $z_{1,a} \in \mathcal{P}(S, \mathcal{F})$, say, defined by

$$z_{1,a}(F) = \frac{\int_S P(s, F \cap S_a) p_0(ds)}{\int_S P(s, S_a) p_0(ds)}, \quad F \in \mathcal{F},$$

and the probability that Z_1 will take the value $z_{1,a}$ is of course equal to the probability that $Y_1 = a$ which is equal to

$$\int_S P(s, S_a)p_0(ds).$$

Thus, the distribution of Z_1 is a discrete distribution - a *mass distribution* - on the set $\mathcal{P}(S, \mathcal{F})$.

The distribution of Z_n for $n = 2, 3, ...$ can be described similarly, although the integral formulas describing the distribution become slightly more complicated.

Now, let us again raise the question under which conditions the sequence $\{Z_n, n = 1, 2...\}$ converges in distribution. In order to be able to give an answer to this question we need to define a topology for the set $\mathcal{P}(S, \mathcal{F})$. In this paper we shall make such assumptions, that it will be convenient to use the topology generated by the metric induced by the *total variation norm* as the topology for the set $\mathcal{P}(S, \mathcal{F})$. In order to make this choice suitable we shall assume 1) that $\{S, \mathcal{F}\}$ is a complete, separable, metric space with a metric δ_0 and that $\mathcal{F}$ is the Borel field generated by δ_0, and 2) that there exists a $\sigma - finite$ measure λ on $(S, \mathcal{F})$ such that the tr.pr.f $P : S \times \mathcal{F} \to [0, 1]$ has a density kernel $p : S \times S \to [0, \infty)$ with respect to the measure λ, that is we have

$$P(s, F) = \int_F p(s, t)\lambda(dt), \ \ \forall\, s \in S, \ \forall F \in \mathcal{F}.$$

Having made these assumptions, we can restrict our considerations to the subset of probability measures in $\mathcal{P}(S, \mathcal{F})$ which have a density with respect to the base measure λ; we denote this set $\mathcal{P}_\lambda(S, \mathcal{F})$.

Next, let us return to the process $\{Z_n, n = 1, 2, ...\}$ of conditional distributions, that we described above. If we make the assumptions that the tr.pr.f P has a density kernel p with respect to λ and that the initial distribution p_0 belongs to the set $\mathcal{P}_\lambda(S, \mathcal{F})$, then it is not difficult to prove that the process $\{Z_n, n = 1, 2, ...\}$ is a Markov chain with state space $\mathcal{P}_\lambda(S, \mathcal{F})$.

Now, let μ_n denote the distribution of $Z_n, n = 1, 2, ...$. In order to illustrate the kind of results that we shall prove in this paper, we shall now give sufficient conditions implying that the sequence $\{\mu_n, n = 1, 2, ...\}$ will converge in distribution towards a unique limit measure which is independent of p_0. For, if 1) S is a compact, metric space, 2) π is a unique invariant measure for the tr.pr.f P which satisfies

$$\lim_{n\to\infty} \sup_{x\in\mathcal{P}(S,\mathcal{F})} ||xP^n - \pi|| = 0,$$

(here $||\cdot||$ denotes the total variation norm), and 3) there exists an element $a \in A$ such that the density kernel p satisfies

$$0 < d_0 \le p(s, t) \le D_0 < \infty, \ \forall\, s, t, \in S_a$$

then $\{\mu_n, n = 1, 2, ...\}$ converges in distribution towards a unique limit measure on $\mathcal{P}_\lambda(S, \mathcal{F})$, which is independent of the initial distribution p_0.

Next a few words on hidden Markov models. A hidden Markov model (HMM), as described in the classical paper [22] by L R Rabiner and B H Juang, 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. Roughly speaking, this classical definition of a HMM is obtained by replacing the deterministic "lumping function" g introduced above by a "stochastic mapping" R. In the more modern literature, see e.g [5] by O Cappé, E Moulines and T Ryden, 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, of course, the transition probability matrices P and R must be replaced by transition probability functions.

In this paper our definition of an HMM will be slightly different from the one given in for example [5], and will be based on a transition probability kernel from the state space $(S, \mathcal{F})$ to the product space $(S \times A, \mathcal{F} \otimes \mathcal{A})$. (See Definition 4.2 below.)

As is well-known a HMM gives rise to two stochastic process $\{X_n, n = 0, 1, 2, ...\}$ and $\{Y_n, n = 1, 2, ...\}$ where the former process, - often called the *hidden Markov chain* -, is determined by the tr.pr.f P and the initial distribution p_0, whereas the latter process - usually called the *observation sequence* - is determined by using the tr.pr.f R from S to A. (We have chosen to start the observation sequence at $n = 1$ for sake of convenience.)

Next, for $n = 1, 2, ...,$ let Z_n denote the conditional law of X_n given the observations $Y_1, Y_2, ..., Y_n$. The process $\{Z_n, n = 1, 2, ...\}$ is usually called the *filtering process* (or the *filter process*).

One question regarding the filtering process $\{Z_n, n = 1, 2...\}$ often treated in the literature is whether Z_n's dependence on the initial distribution p_0 disappears as $n \to \infty$. This property is often called the *forgetting property.*

It is though not this question which is in focus in this paper. Our main focus will instead be on how the **distribution**, μ_n say, of Z_n depends on the initial distribution p_0, and more precisely whether μ_n's dependence of p_0 vanishes as $n \to \infty$ and whether the sequence $\{\mu_n, n = 1, 2, ...\}$ of distributions converges in distribution to a limit distribution independent of the initial distribution p_0, as $n \to \infty$.

Let us here also mention that if both the state space and the observation space are denumerable, then, as pointed out in the paper [3] by L Baum and T Petrie, a HMM can be "transformed" to the "set-up" with a lumping function, if one uses the two tr.pr.ms P and R to construct another tr.pr.m having the product $S \times A$ as state space and if one defines a lumping function $g : S \times A \to A$ simply by $g(s, a) = a$.

The plan of the paper is as follows. In Section 2 we introduce some basic notations, present some simple relations to be used later and define various types of ergodicity; *weak ergodicity*, *strong ergodicity* and *uniform ergodicity.*

In Section 3 we introduce the notion *"partition of transition probability functions"*, a generalisation of the notion "partition of transition probability matrices" introduced in [16].

In Section 4 we first introduce a notion we call *HMM-kernel* and then we make our definition of a *Hidden Markov Model* (HMM). Our definition covers the usual definition of a HMM with finite state space and finite observation space, and also covers the usual definition of a HMM with general state space and general observation space (sometimes also called "state space model"). We also introduce a smaller class of HMMs which we call *HMMs with densities* consisting of such HMMs for which both the state space and the observation space are complete, separable, metric spaces and the HMM-kernel is determined by a density.

In Section 4 we also introduce a notion which we have chosen to call *random mapping*. The classical name is *random system with complete connections*. Other names for this concept is e.g. *learning model* or *iterated function system with place-dependent probabilities*. One important property of a random mapping is that it induces a *Markov kernel*.

In Section 5 we define the **filter kernel**, induced by a HMM with densities. The filter kernel we introduce is a tr.pr.f on the subset $\mathcal{P}_\lambda(S, \mathcal{F})$ of $\mathcal{P}(S, \mathcal{F})$ consisting of those probability measures which have a density with respect to a fixed $\sigma - finite$ base measure λ on $(S, \mathcal{F})$. We denote the filter kernel by $\mathbf{P}$. To simplify notations we shall usually denote the set $\mathcal{P}_\lambda(S, \mathcal{F})$ by K. The *topology* which we shall use on K will be the topology induced by the metric determined by the total variation between probabilities.

In order to be able to prove that certain sets that occur in our investigation are measurable, we introduce a property which we call a **regularity property** and which means that the kernel determining the HMM satisfies a certain continuity condition. (See Definition 5.1.)

The problems we want to solve are thus 1) to find conditions on the HMM such that the dependence of the initial distribution for the distributions of the filtering process vanishes (*weak ergodicity*) and 2) to find conditions such that there exists a unique probability measure on the set K towards which the distributions of the filtering process converge for all choices of initial distributions. (*Weak ergodicity with stationary measure.*)

In Section 6 we define the *random mapping associated to a HMM with densities*, and show that the filter kernel can be considered as the Markov kernel induced by this random mapping. This is an old observation and goes at least back to the paper [4] from 1957 by Blackwell.

In Section 7 we recall the definition of the *Kantorovich distance* between probability measures.

In Section 8 we introduce the set of probability measures on K which have the same *barycenter*. It was probably H Kunita who first observed the usefulness of the concept barycenter when studying the limit behaviour of filtering processes. (See [18].)

In the paper [16] it was proved, that for a denumerable state space S, the set of probabilities on K with *equal* barycenter is a *tight* set, and by using this property, it is not difficult to prove that if the hidden Markov chain $\{X_n, n = 0, 1, 2...\}$ has a stationary probability measure $\pi \in K$ and the probability measure μ on K has barycenter π then the sequence $\{\mu\mathbf{P}^n, n = 1, 2, ...\}$ is a tight sequence, where thus $\mathbf{P}$ denotes the filter kernel. Unfortunately, when we tried to generalise this result from the case when the state space is denumerable to the case when the state space is a complete, separable, metric space, we failed.

In Section 9 we recall the well-known notion called *coupling*, and we also introduce the well-known *Vaserstein coupling*, and in Section 10 we describe in detail what we mean by the *Vaserstein coupling of a transition probability function with density.*

In Section 11 we present the **main theorem** of this paper. In the theorem we introduce a condition, which we call Condition E, and which reads as follows: *To every $\rho > 0$ there exist an integer N and a number $\alpha > 0$ such that for any two probability measures μ and ν on K having barycenter equal to π (the stationary distribution of the hidden Markov chain $\{X_n, n = 0, 1, 2...\}$) we can* *find a coupling* $\tilde{\mu}_N$ *of* $\mu\mathbf{P}^N$ *and* $\nu\mathbf{P}^N$ *such that*

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

where

$$D_\rho = \{(x,y) \in K \times K : ||x-y|| < \rho\}.$$

One result of the main theorem is that if the HMM under consideration has a density which fulfills the regularity condition introduced in Section 5, and there exists a unique probability measure π in $\mathcal{P}(S,\mathcal{F})$ which is invariant with respect to the tr.pr.f P of the hidden Markov chain and is such that

$$\lim_{n\to\infty} \sup_{x\in\mathcal{P}(S,\mathcal{F})} ||xP^n - \pi|| = 0, \tag{1}$$

then the distributions of the filtering process converge in distribution towards a unique limit distribution, if Condition E is satisfied. In case the tr.pr.f P of the hidden Markov chain $\{X_n, n = 0,1,2,...\}$ is such that

$$\lim_{n\to\infty} ||xP^n - \pi|| = 0, \ \forall x \in \mathcal{P}(S,\mathcal{F})$$

and also Condition E is satisfied, then we have not been able to prove the existence of a unique invariant measure for the filtering process; we have only been able to prove that the Kantorovich distance between the distributions of two filtering processes generated by two different initial distributions tends to zero as $n \to \infty$.

In Sections 12 to 15 the proof of the main theorem is given. In Section 12 we prove two auxiliary theorems; both theorems are for a Markov chain taking its values in a complete, separable, metric, bounded space. The first theorem is based on two properties which were introduced in the paper [16] namely a property we call the *shrinking property* and a property we call *Lipschitz equicontinuity*. The second theorem is similar and based on a stronger property which we call the *strong shrinking property*.

In Section 13 we prove that the filtering process induced by a HMM with densities has the Lipschitz equicontinuity property, and in Section 14 we prove that the Kantorovich distance between a probability measure on the set K with barycenter x and the set of probability measures with barycenter y is equal to the total variation between x and y.

In Section 15 we conclude the proof of the main theorem by verifying that the hypotheses of the auxiliary theorems are satisfied under the various hypotheses of the main theorem.

In Section 16 we prove some inequalities for the *nth composition* of positive integral kernels. The proofs of these inequalities are based on some tricks used in the classical paper [9] on products of random matrices by H Furstenberg and H Kesten.

In Section 17 we introduce some further conditions and show, by using the inequalities proved in Section 16, that these conditions imply Condition E. In concrete situations it may be easier to verify the conditions introduced in Section 17 then to verify Condition E directly. In this section we also verify that the results in [17] and [16] are covered by the results of the present paper.

In Section 18 finally, we first apply our theorems to two examples. We end our paper with an example which shows that even if the hidden Markov chain of a HMM is uniformly ergodic (see (1)), the filtering process need not be weakly ergodic. In fact, the filtering process may even become a *periodic* process.

2 Basic notations

Let $(S, \mathcal{F})$ be a measurable space. We let $B[S]$ - or $B[S, \mathcal{F}]$ if we want to emphasize the $\sigma - algebra$ $\mathcal{F}$, denote the set of real, bounded, $\mathcal{F}$- measurable functions on S, and we let $B_u[S]$ or $B_u[S, \mathcal{F}]$ denote the set of real $\mathcal{F}$- measurable functions on S, thus not necessarily bounded. For $u \in B[S]$ we define

$$||u|| = \sup\{|u(s)| : s \in S\},$$

we define

$$osc(u) = \sup\{u(s) - u(t) : s, t \in S\}$$

and, if $F \subset S$, $F \neq \emptyset$, we define

$$osc_F(u) = \sup\{u(s) - u(t) : s, t \in F\}.$$

We let $B[(S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2)]$ denote the set of measurable mappings from $(S_1, \mathcal{F}_1)$ to $(S_2, \mathcal{F}_2)$; we let $\mathcal{P}(S, \mathcal{F})$ denote the set of probability measures on $(S, \mathcal{F})$, we let $\mathcal{M}(S, \mathcal{F})$ denote the set of finite signed measures on $(S, \mathcal{F})$, we let $\mathcal{Q}(S, \mathcal{F})$ denote the set of finite non-negative measures on $(S, \mathcal{F})$ and we let $\mathcal{Q}^\infty(S, \mathcal{F})$ denote the set of positive $\sigma - finite$ measures on $(S, \mathcal{F})$.

As is well-known, $\mathcal{M}(S, \mathcal{F})$ is a vector space. We use the notation δ_s to denote the Dirac measure at $s \in S$. Furthermore, if $F \in \mathcal{F}$ we let $(F, \mathcal{F}_F)$ denote the measurable space $\{F' \in \mathcal{F} : F' \subset F\}$.

If $u \in B_u[S]$ and $x \in \mathcal{M}(S, \mathcal{F})$ we usually write

$$\langle u, x \rangle = \int_S u(s)x(ds)$$

whenever the integral exists. Furthermore, if $F \subset S$, we let $I_F : S \to \{0, 1\}$ denote the indicator function of the set F.

Next, if $x, y \in \mathcal{M}(S, \mathcal{F})$, we let $F^+(x, y) \in \mathcal{F}$ denote a set in $\mathcal{F}$ such that

$$x(F \cap F^+(x, y)) \geq y(F \cap F^+(x, y)), \ \ \forall F \in \mathcal{F},$$

we set $F^-(x, y) = S \setminus F^+(x, y)$, we define $x \vee y \in \mathcal{M}(S, \mathcal{F})$ by

$$x \vee y(F) = x(F \cap F^+(x, y)) + y(F \cap F^-(x, y)),$$

and we define $x \wedge y \in \mathcal{M}(S, \mathcal{F})$ by

$$x \wedge y(F) = x(F \cap F^-(x, y)) + y(F \cap F^+(x, y)).$$

If $x \in \mathcal{M}(S, \mathcal{F})$ we define $x^+ \in \mathcal{Q}(S, \mathcal{F})$ by $x^+ = x \vee 0$ where thus $0 \in \mathcal{M}(S, \mathcal{F})$ denotes the $0 - measure$, we define $x^- \in \mathcal{Q}(S, \mathcal{F})$ by $x^- = (-x) \vee 0$, and we define $|x| \in \mathcal{Q}(S, \mathcal{F})$ by $|x| = x^+ + x^-$.

For $x \in \mathcal{M}(S, \mathcal{F})$ we also define $||x||$ by $||x|| = |x|(S)$. It is of course well-known that $|| \cdot ||$ satisfies the properties of a norm and therefore $\mathcal{M}(S, \mathcal{F})$ can be regarded as a normed vector space.

The following inequality, which will be of use to us later, is easily proved by using the triangle inequality.

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

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

For $x, y \in \mathcal{M}(S, \mathcal{F})$ we call $||x - y||$ the total variation between x and y. As is well-known, if we define $\delta_{TV} : \mathcal{M}(S, \mathcal{F}) \times \mathcal{M}(S, \mathcal{F}) \to [0, \infty)$ by $\delta_{TV}(x, y) = ||x - y||$ then δ_{TV} determines a metric on $\mathcal{M}(S, \mathcal{F})$. We call δ_{TV} the *total variation metric.*

For $r > 0$ we define $\mathcal{Q}^r(S, \mathcal{F}) = \{x \in \mathcal{Q}(S, \mathcal{F}) : ||x|| = r\}$. The following well-known and easily proved inequality will also be of use to us.

Lemma 2.2 *Let $r > 0$, let $x, y \in \mathcal{Q}^r(S, \mathcal{F})$ and let $u \in B[S, \mathcal{F}]$. Then*

$$|\langle u, x\rangle - \langle u, y\rangle| \leq osc(u)(1/2)||x - y||. \ \square$$

Next, let $\delta_0 : S \times S \to [0, \infty)$ be a metric on $(S, \mathcal{F})$. We then always assume implicitly that the $\sigma - algebra$ $\mathcal{F}$ is the Borel-field induced by the metric δ_0. Usually we then write $(S, \mathcal{F}, \delta_0)$ instead of $(S, \mathcal{F})$ and sometimes we write (S, δ_0) instead. We let $C[S]$ or $C[S, \mathcal{F}]$ denote the set of real, bounded, continuous functions on $(S, \mathcal{F})$. For $u \in C[S]$ we define

$$\gamma(u) = \sup\{\frac{|u(s) - u(t)|}{\delta_0(s, t)} : s, t \in S, \ s \neq t\},$$

define $Lip[S] = \{u \in C[S] : \gamma(u) < \infty\}$ and $Lip_1[S] = \{u \in Lip[S] : \gamma(u) \leq 1\}$.

Next, let $\lambda \in \mathcal{Q}^\infty(S, \mathcal{F})$. We let $\mathcal{L}^1_\lambda[S]$ denote the set

$$\{f \in B_u[S] : \int_S |f(s)|\lambda(ds) < \infty\}.$$

If a measure $x \in \mathcal{Q}(S, \mathcal{F})$ is such that there exists a function $f \in \mathcal{L}^1_\lambda[S]$ such that for all $F \in \mathcal{F}$

$$x(F) = \int_F f(s)\lambda(ds),$$

then we say that x has a *density* f with respect to the *base measure* λ, we say that $x \in \mathcal{Q}_\lambda(S, \mathcal{F})$ and we call f a *representative* of x. If also $x \in \mathcal{P}(S, \mathcal{F})$, then we write $x \in \mathcal{P}_\lambda(S, \mathcal{F})$. If $x, y \in \mathcal{Q}_\lambda(S, \mathcal{F})$ and f, g are representatives of x and y respectively, then it is well-known that

$$||x - y|| = \int_S |f(s) - g(s)|\lambda(ds).$$

Furthermore, if $(S, \mathcal{F})$ is a complete, separable, measurable space then it is well-known that also $(\mathcal{Q}_\lambda(S, \mathcal{F}), \delta_{TV})$ is a complete, separable, measurable space. (See e.g. the book [8] by R Dudley.)

Next let $(S_1, \mathcal{F}_1)$ and $(S_2, \mathcal{F}_2)$ be two given measurable spaces. A mapping

$$K : S_1 \times \mathcal{F}_2 \to [0, \infty)$$

will in this paper be called a **transition function** (tr.f) from $(S_1, \mathcal{F}_1)$ to $(S_2, \mathcal{F}_2)$ if

1) $K(s, \cdot)$ belongs to $\mathcal{Q}(S_2, \mathcal{F}_2)$ for all $s \in S_1$,

2) $K(\cdot, F)$ belongs to $B[S_1, \mathcal{F}_1]$ for all $F \in \mathcal{F}_2$.
We denote the set of all tr.fs from $(S_1, \mathcal{F}_1)$ to $(S_2, \mathcal{F}_2)$ by $\mathcal{TQ}((S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2))$.

If a tr.f K from $(S_1, \mathcal{F}_1)$ to $(S_2, \mathcal{F}_2)$ is such that $K(s, \cdot) \in \mathcal{P}(S_2, \mathcal{F}_2)$, $\forall s \in S_1$, then we call K a **transition probability function** (tr.pr.f) from $(S_1, \mathcal{F}_1)$ to $(S_2, \mathcal{F}_2)$ and we denote the set of all tr.pr.fs from $(S_1, \mathcal{F}_1)$ to $(S_2, \mathcal{F}_2)$ by $\mathcal{TP}((S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2))$. We often use the letter P to denote a tr.pr.f.

If K is a tr.f from $(S_1, \mathcal{F}_1)$ to $(S_1, \mathcal{F}_1)$ we say that K is tr.f. *on* $(S_1, \mathcal{F}_1)$ and we write $K \in \mathcal{TQ}((S_1, \mathcal{F}_1))$. Similarly, if P is a tr.pr.f from $(S_1, \mathcal{F}_1)$ to $(S_1, \mathcal{F}_1)$ we say that P is tr.p.f. *on* $(S_1, \mathcal{F}_1)$ and we write $P \in \mathcal{TP}((S_1, \mathcal{F}_1))$. If P is a tr.p.f. on $(S_1, \mathcal{F}_1)$ we call P a **Markov kernel** on $(S_1, \mathcal{F}_1)$.

If $K_1 \in \mathcal{TQ}((S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2))$ and $K_2 \in \mathcal{TQ}((S_2, \mathcal{F}_2), (S_3, \mathcal{F}_3))$ one can define a tr.f $K_1K_2 \in \mathcal{TQ}((S_1, \mathcal{F}_1), (S_3, \mathcal{F}_3))$ by

$$(K_1K_2)(s, F) = \int_{S_2} K_1(s, dt)K_2(t, F), \; s \in S_1, \; F \in \mathcal{F}_3.$$

We call K_1K_2 the **product** of K_1 and K_2.

As is well-known $K_1(K_2K_3) = (K_1K_2)K_3$ (see e.g section 1.1 of the book [23] by D Revuz). More generally, if $K_m \in \mathcal{TQ}((S_m, \mathcal{F}_m), (S_{m+1}, \mathcal{F}_{m+1}))$, $m = 1, 2, ..., n$, we can define

$$K_1K_2...K_m \in \mathcal{Q}((S_1, \mathcal{F}_1), (S_{m+1}, \mathcal{F}_{m+1})), \;\; m = 2, 3, ...n$$

recursively by

$$K_1K_2...K_m = (K_1K_2...K_{m-1})K_m, \; m = 2, 3, ..., n.$$

Associated to a tr.f $K \in \mathcal{TQ}((S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2))$ we can define two linear mappings. We define $T : B[S_2, \mathcal{F}_2] \to B[S_1, \mathcal{F}_1]$ by

$$Tu(s) = \int_{S_2} K(s, dt)u(t).$$

We call T the *transition operator* associated to K.

Now suppose P is a tr.pr.f on $(S, \mathcal{F})$ where $\mathcal{F}$ is a Borel field, and let T denote the associated transition operator. If T is such that

$$u \in C[S] \Rightarrow Tu \in C[S]$$

then the tr.pr.f P is called *Feller continuous*.

The following terminology is not standard and therefore we make a more formal definition.

Definition 2.1 *Suppose $(S, \mathcal{F}, \delta_0)$ is a metric space and suppose $P \in \mathcal{TP}((S, \mathcal{F}))$.*
I . If the associated transition operator T satisfies

$$u \in Lip[S] \Rightarrow Tu \in Lip[S]$$

then we call P **Lipschitz-continuous**.
II. If P is Lipschitz-continuous and also there exists a constant $C > 0$ such that the associated transition operator T satisfies

$$\gamma(T^n u) \leq C\gamma(u), \;\; n = 1, 2, ..., \forall u \in Lip[S]$$

then we call P **Lipschitz equicontinuous**. $\square$

The other linear mapping, $\breve{K}$, associated to a tr.f $K \in \mathcal{TQ}((S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2))$ is a mapping from $\mathcal{Q}(S_1, \mathcal{F}_1)$ to $\mathcal{Q}(S_2, \mathcal{F}_2)$ and is defined by

$$\breve{K}x(F) = \int_S x(ds)K(s, F), \quad F \in \mathcal{F}_2.$$

Instead of $\breve{K}x$ we shall usually write xK.

As is well-known the following relation holds for $u \in B[S_2, \mathcal{F}_2]$, $x \in \mathcal{Q}(S_1, \mathcal{F}_1)$ and $K \in \mathcal{TQ}((S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2))$ (see e.g [23], section 1.1):

$$\langle Tu, x\rangle = \langle u, xK\rangle \ (= \langle u, \breve{K}x\rangle). \tag{2}$$

Next, let $(S_1, \mathcal{F}_1)$ and $(S_2, \mathcal{F}_2)$ be two given measurable sets, and let $\lambda \in \mathcal{Q}^\infty(S_2, \mathcal{F}_2)$. We let $D_\lambda[S_1, S_2]$ denote the subset of $B_u[S_1 \times S_2, \mathcal{F}_1 \otimes \mathcal{F}_2]$ defined by

$$D_\lambda[S_1, S_2] = \tag{3}$$

$$\{f \in B_u[S_1 \times S_2, \mathcal{F}_1 \otimes \mathcal{F}_2] : f \geq 0, \int_{S_2} f(s, t)\lambda(dt) < \infty, \ \forall s \in S_1\}.$$

If $(S_1, \mathcal{F}_1) = (S_2, \mathcal{F}_2)$ we simply write $D_\lambda[S_1]$ instead of $D_\lambda[S_1, S_1]$.

Furthermore, if $K \in \mathcal{TQ}((S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2))$ is such that there exists a function $k \in D_\lambda[S_1, S_2]$ such that for all $s \in S_1$ and $F \in \mathcal{F}_2$

$$K(s, F) = \int_F k(s, t)\lambda(dt), \tag{4}$$

then we say that the tr.f K has a *density kernel* k with respect to the base measure λ and we say that K belongs to the set $\mathcal{TQ}_\lambda((S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2))$. If K is a tr.pr.f we usually write P instead of K, we write p instead of k, we call p a *probability density kernel* and we write $P \in \mathcal{TP}_\lambda((S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2))$. Furthermore, if $(S_1, \mathcal{F}_1) = (S_2, \mathcal{F}_2)$ we simply write $\mathcal{TQ}_\lambda((S_1, \mathcal{F}_1))$ instead of $\mathcal{TQ}_\lambda((S_1, \mathcal{F}_1), (S_1, \mathcal{F}_1))$ and $\mathcal{TP}_\lambda((S_1, \mathcal{F}_1))$ instead of $\mathcal{TP}_\lambda((S_1, \mathcal{F}_1), (S_1, \mathcal{F}_1))$.

If $k \in D_\lambda[S_1, S_2]$ and $K \in \mathcal{TQ}_\lambda((S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2))$ denotes the tr.f defined by (4), we call K the tr.f *determined* by $k \in D_\lambda[S_1, S_2]$.

An important observation is the following:

Proposition 2.1 *Suppose* $K \in \mathcal{TQ}_\lambda((S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2))$ *and that* $k \in D_\lambda[S_1, S_2]$ *is a density kernel of* K. *Then, if* $x \in \mathcal{Q}(S_1, \mathcal{F}_1)$, *it follows that* $xK \in \mathcal{Q}_\lambda(S_2, \mathcal{F}_2)$ *and, if we define* $f_1 : S_2 \to [0, \infty)$ *by*

$$f_1(t) = \int_{S_1} k(s, t)x(ds),$$

then f_1 *is a density of* xK *with respect to the base measure* λ.

Proof. Let $F \in \mathcal{F}_2$. Then

$$xK(F) = \int_{S_1} \int_F k(s, t)x(ds)\lambda(dt) = \int_F f_1(t)\lambda(dt)$$

from which the conclusion of the proposition follows. $\square$.

Next a few notations regarding Markov chains. Let again $(S, \mathcal{F})$ be a measurable space and let $P \in \mathcal{TP}((S, \mathcal{F}))$. If $x \in \mathcal{P}(S, \mathcal{F})$ we denote the Markov chain, which is generated by x and P, by $\{X_{n,x}, \ n = 0, 1, 2, ...\}$. If $x = \delta_s$ we usually write $\{X_n(s), \ n = 0, 1, 2...\}$ instead.

Definition 2.2 *Let $(S, \mathcal{F})$ be a measurable space. Suppose $P \in \mathcal{TP}((S, \mathcal{F}))$. If for any two probabilities x and y in $\mathcal{P}(S, \mathcal{F})$*

$$\lim_{n \to \infty} ||xP^n - yP^n|| = 0,$$

then we say that P is **strongly ergodic**.

If furthermore there exists a measure $\pi \in \mathcal{P}(S, \mathcal{F})$ such that

$$\lim_{n \to \infty} ||xP^n - \pi|| = 0, \ \forall \, x \in \mathcal{P}(S, \mathcal{F})$$

then we say that P is **strongly ergodic** *with* **stationary measure** π. □

Definition 2.3 *Let $(S, \mathcal{F})$ be a measurable space. Suppose $P \in \mathcal{TP}((S, \mathcal{F}))$. If*

$$\lim_{n \to \infty} \sup\{||xP^n - yP^n|| : x, y \in \mathcal{P}(S, \mathcal{F})\} = 0,$$

then we say that P is **uniformly ergodic**.

If furthermore there exists a measure $\pi \in \mathcal{P}(S, \mathcal{F})$ such that

$$\lim_{n \to \infty} \sup\{||xP^n - \pi|| : x \in \mathcal{P}(S, \mathcal{F})\} = 0,$$

then we say that P is **uniformly ergodic** *with* **stationary measure** π.□

Before our next definition we define

$$C_{uniform}[S, \mathcal{F}] = \{u \in C[S, \mathcal{F}] \ : \ u \ uniformly \ continuous \ \}.$$

Definition 2.4 *Let $(S, \mathcal{F}, \delta_0)$ be a measurable space and let $\mathcal{F}$ be the Borel field generated by the metric δ_0. Suppose $P \in \mathcal{TP}((S, \mathcal{F}))$. Let T denote the transition operator associated to P.*

If

$$\lim_{n \to \infty} |T^n u(s) - T^n u(t)| = 0, \ \forall \, u \in C_{uniform}[S, \mathcal{F}] \ \text{and} \ \forall \, s, t \in S$$

then we say P is **weakly ergodic**.

If furthermore there exists a measure $\pi \in \mathcal{P}(S, \mathcal{F})$ such that

$$\lim_{n \to \infty} |T^n u(s) - \langle u, \pi \rangle| = 0, \ \forall \, u \in C[S] \ \text{and} \ \forall \, s \in S$$

then we say P is **weakly ergodic** *with* **stationary measure** π. □

We end this long section about notations with another simple but important observation.

Proposition 2.2 *Suppose $P \in \mathcal{TP}_\lambda((S, \mathcal{F}))$, that $\pi \in \mathcal{P}(S, \mathcal{F})$ and that $\pi = \pi P$. Then $\pi \in \mathcal{P}_\lambda(S, \mathcal{F})$.*

Proof. Follows immediately from Proposition 2.1. □

3 Partitions of transition probability functions.

In [16] we introduced the notion "*partition of a transition probability matrix*". In this section shall generalise this notion by defining partitions of transition probability function.

Definition 3.1 *Let $(S_1, \mathcal{F}_1)$ and $(S_2, \mathcal{F}_2)$ be two measurable spaces and let P be a tr.pr.f from $(S_1, \mathcal{F}_1)$ to $(S_2, \mathcal{F}_2)$. Let $(A, \mathcal{A})$ be yet another measurable space and let*

$$M \in \mathcal{TP}((S_1, \mathcal{F}_1), (S_2 \times A, \mathcal{F}_2 \otimes \mathcal{A}))$$

be a transition probability function from $(S_1, \mathcal{F}_1)$ to $(S_2 \times A, \mathcal{F}_2 \otimes \mathcal{A})$. If

$$M(s_1, F_2, A) = P(s_1, F_2), \ \forall s_1 \in S_1, \ \ \forall F_2 \in \mathcal{F}_2,$$

then we call the tr.pr.f M a **partition** *of the tr.pr.f P.*

Furthermore, if there exists a measure $\lambda \in \mathcal{Q}^\infty(S_2, \mathcal{F}_2)$, a probability density kernel $p \in D_\lambda[S_1, S_2]$ such that

$$P(s, F_2) = \int_{F_2} p(s,t)\lambda(dt), \ \forall s \in S_1, \ \forall F_2 \in \mathcal{F}_2,$$

a measure $\tau \in \mathcal{Q}^\infty(A, \mathcal{A})$ and a measurable function $m : S_1 \times S_2 \times A \to [0, \infty)$ such that

$$M(s, F, B) = \int_F \int_B m(s,t,a)\lambda(dt)\tau(da)$$

then we denote the partition M by (m, τ), we call m a density of M and call m a **partition function**. □

Next let us present some simple facts about partitions of tr.pr.fs which we state without proof.

1) Let $M_1 \in \mathcal{TP}((S_1, \mathcal{F}_1), (S_2 \times A_1, \mathcal{F}_2 \otimes \mathcal{A}_1))$ be a partition of $P_1 \in \mathcal{TP}((S_1, \mathcal{F}_1), (S_2, \mathcal{F}_2))$ and let $M_2 \in \mathcal{TP}((S_2, \mathcal{F}_3), (S_3 \times A_2, \mathcal{F}_3 \otimes \mathcal{A}_2))$ be a partition of $P_2 \in \mathcal{TP}((S_2, \mathcal{F}_2), (S_3, \mathcal{F}_3))$. Define $M_3 \in \mathcal{TP}((S_1, \mathcal{F}_1), (S_3 \times A_1 \times A_2, \mathcal{F}_3 \otimes \mathcal{A}_1 \otimes \mathcal{A}_2))$ by

$$M_3(s_1, F_3, B_1 \times B_2) = \int_{s_3 \in F_3} \int_{S_2} M_1(s_1, ds_2, B_1) M_2(s_2, ds_3, B_2).$$

Then M_3 is a partition of $P_1 P_2$. We call M_3 the *product* of M_1 and M_2.

2) Let M_i be a partition of P_i for $i = 1, 2, 3$ and suppose that the product $P_1 P_2 P_3$ is well-defined. Then

$$(M_1 M_2) M_3 = (M_1) M_2 M_3.$$

This implies that if P is a tr.pr.f on $(S, \mathcal{F})$ and $M \in \mathcal{TP}(S, \mathcal{F}), (S \times A, \mathcal{F} \otimes \mathcal{A}))$ is a partition of P, then, for $n = 2, 3, ...$, the *product* $M^n \in \mathcal{TP}(S, \mathcal{F}), (S \times A^n, \mathcal{F} \otimes \mathcal{A}^n))$ is well-defined and is a partition of P^n.

We shall soon see the connection between partitions of tr.pr.fs and hidden Markov models.

Remark. Recall that a measure on the product space of two measurable spaces is determined by knowing the values on the so called *rectangular sets*. □

4 Hidden Markov Models and Random Mappings

Let $(S, \mathcal{F})$ and $(A, \mathcal{A})$ be two 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 $\Lambda \in \mathcal{TP}((S \times A, \mathcal{F} \otimes \mathcal{A}))$. 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 $\{(X_{n,\xi}, Y_{n,\xi}), n = 0, 1, 2, ...\}$.

In this section we shall consider two special classes of bivariate Markov chains.

Definition 4.1 *Let $(S, \mathcal{F})$ and $(A, \mathcal{A})$ be two measurable spaces, let $M \in \mathcal{TP}((S, \mathcal{F}), (S \times A, \mathcal{F} \otimes \mathcal{A}))$ and define $P \in \mathcal{TP}((S, \mathcal{F}))$ by*

$$P(s, F) = M(s, F \times A).$$

Then we call M a **Hidden Markov Model kernel** *(***HMM-kernel***) and we call P the* **Markov kernel** *associated to the HMM-kernel M.*

Remark. In other words, if P is a Markov kernel on $(S, \mathcal{F})$, then any partition of P determines a HMM-kernel. □

In the next definition we present our definition of a Hidden Markov Model.

Definition 4.2 *Let $(S, \mathcal{F})$ and $(A, \mathcal{A})$ be two measurable spaces, let P be a tr.pr.f on $(S, \mathcal{F})$ and let $M \in \mathcal{TP}((S \times A, \mathcal{F} \otimes \mathcal{A}))$ be a partition of P. The set*

$$\mathcal{H} = \{(S, \mathcal{F}), P, (A, \mathcal{A}), M\} \quad (5)$$

is called a **Hidden Markov Model (HMM)**. *We call $(S, \mathcal{F})$ the* **state space** *and call $(A, \mathcal{A})$ the* **observation space**.

Remark. Since the tr.pr.f P is determined by M we could have excluded P in the expression of the right hand side of (5). We have included P for sake of clarity. □

Definition 4.3 *Let*

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

be a HMM, and define $\Lambda \in \mathcal{TP}((S \times A, \mathcal{F} \otimes \mathcal{A}))$ by

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

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

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

Remark 1. Our definition of a HMM is similar to the definition given in [5].
Remark 2. It is for sake of convenience that we start the observation sequence with $n = 1$ instead of $n = 0$. □

In our next definition we introduce a smaller class of HMMs.

Definition 4.4 *Let* $(S, \mathcal{F}, \delta_0)$ *and* $(A, \mathcal{A}, \varrho)$ *be two complete, separable, metric spaces and let*

$$\mathcal{H} = \{(S, \mathcal{F}, \delta_0), P, (A, \mathcal{A}, \varrho), M\}$$

be a HMM. If there exists a measure $\lambda \in \mathcal{Q}^\infty(S, \mathcal{F})$*, a probability density kernel* $p \in D_\lambda[S]$ *such that*

$$P(s, F) = \int_F p(s,t)\lambda(dt), \ \forall s \in S, \ \forall F \in \mathcal{F},$$

a measure $\tau \in \mathcal{Q}^\infty(A, \mathcal{A})$ *and a measurable function* $m : S \times S \times A \to [0, \infty)$ *such that*

$$M(s, F, B) = \int_F \int_B m(s,t,a)\lambda(dt)\tau(da)$$

then we call $\mathcal{H}$ *a* **HMM with densities***, and we denote the HMM by*

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

We call λ *and* τ **base measures**. □

Remark 1. Our concept "HMM with densities" is a minor generalisation of the concept "fully dominated HMM" as defined in [5], Definition 2.2.3.
Remark 2. See Section 2 for the definition of $D_\lambda[S]$. □

We now present three simple examples.

Example 4.1 *Let* $(S, \mathcal{F})$ *be a measurable set. Let* $\{S_i, i \in \mathcal{I}\}$ *be a denumerable set of disjoint subsets of* S *such that* $\cup_{i \in \mathcal{I}} S_i = S$ *and such that* $S_i \in \mathcal{F}, i \in \mathcal{I}$*, and let* $P \in \mathcal{TP}((S, \mathcal{F}))$*. Set* $A = \mathcal{I}$ *and let* $\mathcal{A}$ *denote all subsets of* $\mathcal{I}$*. Define* $M : S \times \mathcal{F} \times \mathcal{A} \to [0, 1]$ *simply by*

$$M(s, F, B) = \sum_{a \in B} P(s, F \cap S_a)$$

Then clearly M **is a partition** *of* P. □

Example 4.2 (Standard hidden Markov model. The denumerable case.) *Let* S *and* A *be denumerable sets and let* $\mathcal{F}$ *and* $\mathcal{A}$ *denote all subsets of* S *and* A *respectively.*

Let $p : S \times S \to [0, 1]$ *be a transition probability matrix (tr.pr.m) on* S *and let* $r : S \times A \to [0, 1]$ *be a tr.pr.m from* S *to* A*. Define* $M : S \times \mathcal{F} \times \mathcal{A} \to [0, 1]$ *simply by*

$$M(s, F, B) = \sum_{t \in F} \sum_{a \in B} p(s,t) r(t, a)$$

and let $P \in \mathcal{TP}((S, \mathcal{F}))$ *be defined by*

$$p(s, F) = \sum_{t \in F} p(s, t).$$

Again it is clear that M **is a partition** *of* P. □

Our third example can be considered as the standard HMM as defined in [5], Definition 2.2.1.

Example 4.3 (Standard hidden Markov model. The general case.) *Let* $(S, \mathcal{F})$ *and* $(A, \mathcal{A})$ *be two measurable sets. Let* $P \in \mathcal{TP}((S, \mathcal{F}))$ *and* $R \in \mathcal{TP}((S, \mathcal{F}), (A, \mathcal{A}))$*. Define* $M : S \times \mathcal{F} \times \mathcal{A} \to [0, 1]$ *simply by*

$$M(s, F \times B) = \int_F R(y, B)P(s, dy).$$

Again it is clear that M *is a HMM-kernel and that* M *is a* **partition** *of the Markov kernel* P*.* □

We shall next consider another special kind of bivariate Markov chains.

Definition 4.5 *Let* $(S, \mathcal{F})$ *and* $(A, \mathcal{A})$ *be two measurable sets. Let* $h \in B[S \times A, S]$ *be a measurable function from* $S \times A$ *to* S *and let* $Q \in \mathcal{TP}((S, \mathcal{F}), (A, \mathcal{A}))$ *be a tr.pr.f from* $(S, \mathcal{F})$ *to* $(A, \mathcal{A})$*. We call the 4-tuple*

$$\mathcal{R} = \{(S, \mathcal{F}), (A, \mathcal{A}), h, Q\}$$

a **random mapping***. We call* h *the* **response function***, we call* Q *the* **index probability***, we call* $(S, \mathcal{F})$ *the* **state space** *and we call* $(A, \mathcal{A})$ *the* **index space***.* □

Definition 4.6 *Let*

$$\mathcal{R} = \{(S, \mathcal{F}), (A, \mathcal{A}), h, Q\}$$

be a random mapping. For $s \in S$ *and* $F \in \mathcal{F}$*, we define* $h^{-1}(s, F) \in \mathcal{A}$ *by* $\{a \in A : h(s, a) \in F\}$*. We call the tr.pr.f* $M \in \mathcal{TP}((S, \mathcal{F}), (S \times A, \mathcal{F} \otimes \mathcal{A}))$ *defined by*

$$M(s, F \times B) = Q(s, h^{-1}(s, F) \cap B) \tag{6}$$

the **HMM-kernel induced** *by the random mapping* $\mathcal{R}$*, we call the tr.pr.f* $P \in \mathcal{TP}((S, \mathcal{F}))$ *defined by*

$$P(s, F) = Q(s, h^{-1}(s, F)) \; (= M(s, F \times A)) \tag{7}$$

the **Markov kernel induced** *by the random mapping* $\mathcal{R}$ *and we call the hidden Markov model*

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

where P *and* M *are defined by (7) and (6) respectively, the* **HMM induced by the random mapping** $\mathcal{R}$*.*

Furthermore, if $\{X_{n,x}, n = 0, 1, 2, ...\}$ *and* $\{Y_{n,x}, n = 1, 2, ...\}$ *denote the hidden Markov chain and the observation sequence generated by the HMM* $\mathcal{H}$ *induced by the random mapping* $\mathcal{R}$ *and the initial distribution* $x \in K$*, we call* $\{X_{n,x}, n = 0, 1, 2, ...\}$ *the* **state sequence** *and* $\{Y_{n,x}, n = 1, 2, ...\}$ *the* **index sequence** *generated by the random mapping* $\mathcal{R}$*.* □

In case a random mapping $\mathcal{R} = \{(S, \mathcal{F}), (A, \mathcal{A}), h, Q\}$ is such that Q has a density, that is if there exist a $\sigma - finite$ measure τ on $(A, \mathcal{A})$ and a function $q : S \times A \to [0, \infty)$ such that $q \in D_\tau[S, A]$ and such that

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

we usually denote the random mapping by $\mathcal{R} = \{(S,\mathcal{F}),(A,\mathcal{A}),h,(q,\tau)\}$.
Remark. The classical name for the 4-tuple

$$\{(S,\mathcal{F}),(A,\mathcal{A}),h,Q\}$$

is *random system with complete connections.* (See e.g the book [10] by M Iosifescu and R Theodorescu or the book [11] by M Iosefescu and S Grigorescu.) Another classical name is *learning model.* (See e.g the book [20] by F Norman.) A later terminology, introduced by M.Barnsley and coworkers, is *iterated function system with place-dependent probabilities* (see e.g [2]). The terminology *random mapping* is inspired by the terminology used in the paper [7] by P Diaconis and D Freedman. In learning model theory the index space is called the *event space* and the index sequence $\{Y_{n,x}, n = 1,2,...\}$ is called the *event sequence.* (See e.g [20].) $\Box$

The motivation for introducing the concept random mapping is that there is a strong connection between the theory on filtering processes and the theory on iterations of random mappings, which we shall see later in this paper. (See Section 6.)

The study of random mappings has a long history (see e.g [10], [20], [14],[11]); here we shall just present a few basic facts.

First suppose $\mathcal{R} = \{(S,\mathcal{F}),(A,\mathcal{A}),h,Q\}$ is a random mapping and let $P \in \mathcal{TP}((S,\mathcal{F}))$ be the Markov kernel induced by $\mathcal{R}$. Let $T : B[S,\mathcal{F}] \to B[S,\mathcal{F}]$ be the transition operator associated to the Markov kernel P. We have

$$Tu(s) = \int_S u(t)P(s,dt) = \int_A u(h(s,a))Q(s,da).$$

In order to state some further relations we need some further notations. Thus let $\{(S,\mathcal{F}),(A,\mathcal{A}),h,Q\}$ be a random mapping. We set $A^1 = A$ and define A^n iteratively by

$$A^{n+1} = A^1 \times A^n,\ n = 1,2,... .$$

Similarly, we set $\mathcal{A}^1 = \mathcal{A}$ and define $\mathcal{A}^n$ iteratively by

$$\mathcal{A}^{n+1} = \mathcal{A}^n \otimes \mathcal{A}.$$

For $(a_1,a_2,...,a_n) \in A^n$ we use the notation

$$a^n = (a_1,a_2,...,a_n)$$

and we write

$$B^n = B_1 \times B_2 \times ... \times B_n,\ \ if\ B_m \in \mathcal{A},\ \ m = 1,2,...,n.$$

We define $h^n : S \times A^n \to K, n = 1,2,..$ by first defining $h^1 = h$, and then defining h^n iteratively by

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

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

$$Q^{n+1}(s,B' \times B_{n+1}) = \int_{B'}\int_{B_{n+1}} Q^n(s,da^n)Q(h^n(s,a^n),da_{n+1}).$$

It is well-known (see [11]) and easily proved that h^n is measurable for each positive integer n and that $Q^n \in \mathcal{TP}((S,\mathcal{F})(A^n,\mathcal{A}^n))$. This implies that also $\{(S,\mathcal{F}),(A^n,\mathcal{A}^n),h^n,Q^n\}$ is a random mapping for each positive integer n.

If $P \in \mathcal{TP}((S,\mathcal{F}))$ is the Markov kernel induced by the random mapping $\{(S,\mathcal{F}),(A,\mathcal{A}),h,Q\}$ and $P^{(n)}$ denotes the Markov kernel induced by the random mapping $\{(S,\mathcal{F}),(A^n,\mathcal{A}^n),h^n,Q^n\}$, then it is easily proved that

$$P^n = P^{(n)},\ n = 1,2,\dots .$$

Furthermore, if $u \in B[S,\mathcal{F}]$ and $s \in S$, then

$$\int_S u(t)P^n(s,dt) = E[u(X_n(s))] = E[u(h^n(s,Y^n(s))] =$$

$$\int_{A^n} u(h^n(s,a^n))Q^n(s,da^n) = T^n u(s)$$

where of course $Y^n(s) = (Y_1(s),Y_2(s),...,Y_n(s))$. We also have that for $n = 1,2,...$

$$X_n(s) = h^n(s,Y^n(s)),\ s \in S,$$

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

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

$$q^{n+1}(s,a^{n+1}) = q^n(s,a^n)q(h^n(s,a^n),a_{n+1}),\ n = 1,2,...,$$

and then we can express $Q^n \in \mathcal{TP}((S,\mathcal{F})(A^n,\mathcal{A}^n))$, for n=1,2,..., by

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

where $\tau^n = \tau \otimes \tau \otimes ... \otimes \tau$ (n times). We call $h^n : S \times A^n \to S$ the *nth composition* of $h : S \times A \to S$, we call $Q^n \in \mathcal{TP}((S,\mathcal{F})(A^n,\mathcal{A}^n))$ the *nth composition* of $Q \in \mathcal{TP}((S,\mathcal{F})(A,\mathcal{A}))$ and we call $\{(S,\mathcal{F}),(A^n,\mathcal{A}^n),h^n,Q^n\}$ the *nth composition* of $\{(S,\mathcal{F}),(A,\mathcal{A}),h,Q\}$. In case Q has a density q with respect to the base measure τ we denote the nth composition of $\{(S,\mathcal{F}),(A,\mathcal{A}),h,(q,\tau)\}$ by $\{(S,\mathcal{F}),(A^n,\mathcal{A}^n),h^n,(q^n,\tau^n)\}$.

5 The filter kernel associated to an HMM with densities.

From now on we shall mainly restrict our analysis to HMMs with densities, and we shall assume that the measurable space $(S,\mathcal{F})$ is a complete, separable, metric space with metric δ_0.

Thus let

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

be a HMM with densities. As above let $\mathcal{P}_\lambda(S,\mathcal{F})$ denote the subset of $\mathcal{P}(S,\mathcal{F})$ consisting of all probability measures which have a density with respect to a σ-finite measure λ. To simplify notations we shall usually write

$$K = \mathcal{P}_\lambda(S,\mathcal{F}).$$

Next, let δ_{TV} be the metric determined by the total variation on $\mathcal{P}_\lambda(S,\mathcal{F})$, and let $\mathcal{E}$ be the $\sigma - algebra$ on $\mathcal{P}_\lambda(S,\mathcal{F})$ generated by δ_{TV}. In agreement with our notations introduced above, we let $\mathcal{P}(K,\mathcal{E})$ denote the set of probability measures on $(K,\mathcal{E})$, we let $\mathcal{Q}(K,\mathcal{E})$ denote the set of finite, nonnegative measures on $(K,\mathcal{E})$, we let $B[K,\mathcal{E}]$ denote the set of real, bounded, measurable functions on $(K,\mathcal{E})$ and we let $C[K,\mathcal{E}]$ denote the set of real, bounded, continuous functions on $(K,\mathcal{E})$.

Next let us for each $a \in A$ define $m_a \in D_\lambda[S]$ simply by

$$m_a(s,t) = m(s,t,a)$$

and let $M_a \in \mathcal{TQ}((S,\mathcal{F}))$ be the tr.f on $(S,\mathcal{F})$ determined by the density kernel m_a that is

$$M_a(s,F) = \int_F m_a(s,t)\lambda(dt), = (\int_F m(s,t,a)\lambda(dt)\,), \quad s \in S,\ F \in \mathcal{F}. \tag{8}$$

(See (3) for the definition of $D_\lambda[S]$.) As we showed in Section 2 (see Proposition 2.1) the tr.f M_a induces a mapping $\breve{M}_a$ from $\mathcal{Q}_\lambda(S,\mathcal{E})$ to $\mathcal{Q}_\lambda(S,\mathcal{E})$. As we mentioned above we shall usually write xM_a instead of $\breve{M}_a x$.

We now define $\overline{M} : \mathcal{Q}_\lambda(S,\mathcal{F}) \times A \to \mathcal{Q}_\lambda(S,\mathcal{F})$ by

$$\overline{M}(x,a) = xM_a. \tag{9}$$

In order to simplify the proof of certain measurability properties for various quantities below, we now make the following definition.

Definition 5.1 *Let*

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

be a HMM with densities. Let $\overline{M} : \mathcal{Q}_\lambda(S,\mathcal{F}) \times A \to \mathcal{Q}_\lambda(S,\mathcal{F})$ *be defined by (9) and (8). If* $\overline{M} : \mathcal{Q}_\lambda(S,\mathcal{F}) \times A \to \mathcal{Q}_\lambda(S,\mathcal{F})$ *is a* **continuous** *function then we say that the partition* (m,τ) *is* **regular**.

Proposition 5.1 *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 two hidden Markov models with densities and regular partitions, and with the same state space and the same base measure* λ*. Define* $m^{1,2} : S \times S \times A_1 \times A_2 \to [0,\infty)$ *by*

$$m^{1,2}(s,t,a_1,a_2) = \int_S m_1(s,\sigma,a_1)m_2(\sigma,t,a_2)\lambda(d\sigma)$$

and define $\tau^{1,2}$ *as the product measure* $\tau_1 \otimes \tau_2$*. Then* $(m^{1,2},\tau^{1,2})$ *is the product of the partitions* (m_1,τ) *and* (m_2,τ_2) *and furthermore the partition* $(m^{1,2},\tau^{1,2})$ *is also regular.* □

We give a proof for sake of completeness.
Proof. Set $M_1 = (m_1,\tau_1)$, set $M_2 = (m_2,\tau_2)$ and let $M_3 = M_1M_2$ denote the product of the partitions M_1 and M_2. By definition

$$M_3(s,F,B_1,B_2) = \int_{\sigma\in S}\int_{t\in F} M_1(s,d\sigma,B_1)M_2(\sigma,dt,B_2) =$$

$$\int_{\sigma\in S}\int_{a_1\in B_1}\int_{t\in F}\int_{a_2\in B_2} m_1(s,\sigma,a_1)m_2(\sigma,t,a_2)\lambda(d\sigma)\lambda(dt)\tau_1(da_1)\tau_2(da_2) =$$

$$\int_{t\in F}\int_{a_1\in B_1}\int_{a_2\in B_2} m^{1,2}(s,t,a_1,a_2)\lambda(dt)\tau^{1,2}(da_1,da_2)$$

from which follows that $M_3 = (m^{1,2},\tau^{1,2})$ and thereby we have shown that $(m^{1,2},\tau^{1,2})$ is the product of the partitions (m_1,τ) and (m_2,τ_2).

To prove that $m^{1,2}$ is regular we proceed as follows. Define $\overline{M}_1 : \mathcal{Q}_\lambda(S,\mathcal{F})\times A_1 \to \mathcal{Q}_\lambda(S,\mathcal{F})$, $\overline{M}_2 : \mathcal{Q}_\lambda(S,\mathcal{F})\times A_2 \to \mathcal{Q}_\lambda(S,\mathcal{F})$ and $\overline{M}_3 : \mathcal{Q}_\lambda(S,\mathcal{F})\times A_1\times A_2 \to \mathcal{Q}_\lambda(S,\mathcal{F})$ by

$$\overline{M}_1(x,a_1)(F) = \int_{t\in F}\int_{s\in S} m_1(s,t,a_1)x(ds)\lambda(dt),\quad F\in\mathcal{F}$$

$$\overline{M}_2(x,a_2)(F) = \int_{t\in F}\int_{s\in S} m_2(s,t,a_2)x(ds)\lambda(dt),\quad F\in\mathcal{F}$$

and

$$\overline{M}_3(x,a_1,a_2)(F) = \int_{t\in F}\int_{s\in S} m^{1,2}(s,t,a_1,a_2)x(ds)\lambda(dt),\quad F\in\mathcal{F}$$

respectively. Since

$$m^{1,2}(s,t,a_1,a_2) = \int_S m_1(s,\sigma,a_1)m_2(\sigma,t,a_2)\lambda(d\sigma)$$

and

$$\int_F\int_S m_1(s,t,a_1)x(ds)\lambda(dt) = xM_{a_1}(F)$$

we find that

$$\overline{M}_3(x,a_1,a_2)(F) =$$

$$\int_{t\in F}\int_{s\in S}\int_{\sigma\in S} m_1(s,\sigma,a_1)m_2(\sigma,t,a_2)\lambda(d\sigma)x(ds)\lambda(dt) =$$

$$\int_F m_2(\sigma,t,a_2)xM_{a_1}(d\sigma)\lambda(dt).$$

Hence

$$\overline{M}_3(x,a_1,a_2) = \overline{M}_2(xM_{a_1},a_2) = \overline{M}_2(\overline{M}_1(x,a_1),a_2)$$

because of (9), and since both $\overline{M}_2$ and $\overline{M}_1$ are continuous it follows that also $\overline{M}_3$ is continuous. □

We next prove that if the observation space is denumerable and the partition function is bounded then the partition is regular.

Proposition 5.2 *Let*

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

be a HMM with densities. Suppose that A is denumerable, that the metric ϱ is the discrete metric and assume also that the partition function m is bounded. Then the partition (m,τ) is regular.

Proof. Since ϱ is the discrete metric it suffices to prove that for each $a \in A$ $\breve{M}_a : \mathcal{Q}_\lambda(S,\mathcal{F}) \to \mathcal{Q}_\lambda(S,\mathcal{F})$ is continuous in the topology induced by the total variation metric.

Thus let $\epsilon > 0$ be given. Since the partition function m is assumed to be bounded we can find a constant C say, such that $m(s,t,a) < C$, $\forall s,t \in S, \forall a \in A$.

Now let $x, y \in \mathcal{Q}_\lambda(S,\mathcal{F})$ and let f and g be representatives of x and y with respect to the base measure λ. We find

$$||xM_a - yM_a|| = |\int_S \int_S (f(s) - g(s))m(s,t,a)\lambda(ds)\lambda(dt)| \le$$

$$C \int_S \int_S |f(s) - g(s)|\lambda(ds)\lambda(dt)| = C||x - y||.$$

Hence if $||x - y|| < \epsilon/C$ it follows that $||xM_a - yM_a|| < \epsilon$. □

Remark. Another, less trivial, result is presented in Section 18. (See Example 18.2 and Proposition 18.1.)

Our next aim is to introduce a notion for HMMs with densities and regular partitions, which we call the **filter kernel**.

Thus, as usual, let

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

be a HMM with densities and assume that the partition (m,τ) is regular. We now define the set $KA^+ \subset K \times A$ by

$$KA^+ = \{(x,a) \in K \times A : ||xM_a|| > 0\}$$

and for each $x \in K$ we let A_x^+ be the set

$$A_x^+ = \{a \in A : ||xM_a|| > 0\}.$$

Since $||xM_a|| = ||\overline{M}(x,a)||$ and $\overline{M}(x,a)$ is a continuous function, it follows that KA^+ is an open set, and also that $A_x^+ \subset A$ is an open set for each $x \in K$. In particular $KA^+ \in \mathcal{E} \otimes \mathcal{A}$ is measurable, and also $A_x^+ \in \mathcal{A}$ for each $x \in K$. Furthermore, if $E_0 \in \mathcal{E}$ is an **open** set and we define $KA^+(E_0)$ as the subset of KA^+ defined by

$$\{(x,a) \in KA^+ : xM_a/||xM_a|| \in E_0\},$$

then it follows from the continuity of the map $\overline{M}(\cdot,\cdot)$ and Lemma 2.1 that $KA^+(E_0)$ is an open set.

We now define the tr.pr.f $\mathbf{P}$ on $(K,\mathcal{E})$ by

$$\mathbf{P}(x,E) = \int_{A_x^+} I_E(\frac{xM_a}{||xM_a||})||xM_a||\tau(da) \tag{10}$$

and we define $\mathbf{T} : B[K,\mathcal{E}] \to B[K,\mathcal{E}]$ by

$$\mathbf{T}u(x) = \int_{A_x^+} u(\frac{xM_a}{||xM_a||})||xM_a||\tau(da) \tag{11}$$

That $\mathbf{P}(x,\cdot)$ is a probability measure in $\mathcal{P}(K,\mathcal{E})$ for every $x \in K$ is easily proved and that $\mathbf{P}(\cdot,E)$ is $\mathcal{E} - measurable$ for each open $E \in \mathcal{E}$ follows from the fact that the set $\{(x,a) \in KA^+ : xM_a/||xM_a|| \in E_0\}$ is an open set if E_0 is open. That $\mathbf{P}(\cdot, E)$ is $\mathcal{E}-measurable$ for each $E \in \mathcal{E}$ follows then easily from the fact that the set $\mathcal{B} = \{E \subset K : \mathbf{P}(\cdot, E)\ measurable\}$ is a $\sigma - algebra$ and contains all open sets.

That $\mathbf{T}$ is the transition operator associated to $\mathbf{P}$ is evident from (10) and (11).

Definition 5.2 *Let* $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ *be a HMM with densities such that* (m, τ) *is regular, and let* $\mathbf{P}$ *be defined by (10). We call* $\mathbf{P}$ *the* **filter kernel** *induced by the HMM* $\mathcal{H}$ *or simply the* **filter kernel** *induced by the* **partition** (m, τ).

If $\mu \in \mathcal{P}(K, \mathcal{E})$ *we call* $\{Z_{n,\mu},\ n = 0, 1, 2, ...\}$ *the* **filtering process** *generated by the HMM* $\mathcal{H}$ *and the initial distribution* μ *or more simply the* **filtering process** *induced by the partition* (m, τ) *and the initial distribution* μ.

If the initial distribution is the Dirac-measure at $x \in K$ *we write* $\{Z_n(x),\ n = 0, 1, 2, ...\}$ *instead of* $\{Z_{n,\delta_x},\ n = 0, 1, 2, ...\}$.

To emphasize the dependence of the HMM $\mathcal{H}$ *we may write* $\mathbf{P}_{\mathcal{H}}$ *instead of just* $\mathbf{P}$. *We use a similar notation for the associated transition operator.* □

The following lemma is easy to prove but yet quite important.

Lemma 5.1 *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 two hidden Markov models with densities such that both* (m_1, τ_1) *and* (m_2, τ_2) *are regular. Let* $(m^{1,2}, \tau^{1,2})$ *be the product of their partitions* (m_1, τ_1) *and* (m_2, τ_2), *let* p_3 *denote the density of the Markov kernel determined by* $(m^{1,2}, \tau^{1,2})$ *and let* $\mathcal{H}_3$ *be the HMM defined by*

$$\mathcal{H}_3 = \{(S, \mathcal{F}, \delta_0), (p_3, \lambda), (A^{1,2}, \mathcal{A}^{1,2}), (m^{1,2}, \tau^{1,2})\}.$$

Let $\mathbf{P}_{\mathcal{H}_1}, \mathbf{P}_{\mathcal{H}_2}, \mathbf{P}_{\mathcal{H}_3}$ *denote the* **filter kernels** *induced by* $\mathcal{H}_1$,$\mathcal{H}_2$ *and* $\mathcal{H}_3$ *respectively and let* $\mathbf{T}_{\mathcal{H}_1}, \mathbf{T}_{\mathcal{H}_2}, \mathbf{T}_{\mathcal{H}_3}$ *denote the associated transition operators.*

Then a)

$$\mathbf{P}_{\mathcal{H}_1}\mathbf{P}_{\mathcal{H}_2} = \mathbf{P}_{\mathcal{H}_3}$$

and b)

$$\mathbf{T}_{\mathcal{H}_1}\mathbf{T}_{\mathcal{H}_2} = \mathbf{T}_{\mathcal{H}_3}.\ \square$$

Proof. The equality in a) follows from the equality in b) if one uses the identity (2) of Section 2.

To prove b) let $u \in B[K, \mathcal{E}]$, and set $u_2 = \mathbf{T}_{\mathcal{H}_2} u$. From (11) we find that

$$u_2(x) = \int_{A^+_{2,x}} u(\frac{xM_{a_2}}{||xM_{a_2}||})||xM_{a_2}||\tau_2(da_2).$$

Hence

$$\mathbf{T}_{\mathcal{H}_1}\mathbf{T}_{\mathcal{H}_2}u(x) = \int_{A^+_{1,x}} u_2(\frac{xM_{a_1}}{||xM_{a_1}||})||xM_{a_1}||\tau_1(da_1) =$$

$$\int_{A^+_{1,x}} \int_{A^+_{2,y(a_1)}} u(\frac{(\frac{xM_{a_1}}{||xM_{a_1}||}M_{a_2})}{||\frac{xM_{a_1}}{||xM_{a_1}||}M_{a_2}||})||\frac{xM_{a_1}}{||xM_{a_1}||}M_{a_2}||\tau(da_2)||xM_{a_1}||\tau(da_1) =$$

$$\int_{A^+_{1,x}} \int_{A^+_{2,y(a_1)}} u(\frac{xM_{a_1}M_{a_2}}{||xM_{a_1}M_{a_2}||})||xM_{a_1}M_{a_2}||\tau(da_2)\tau(da_1)$$

where thus $y(a_1)$ and $A^+_{2,y(a_1)}$ are defined by

$$y(a_1) = xM_{a_1}/||xM_{a_1}||, \;\; a_1 \in A^+_{1,x}$$

and

$$A^+_{2,y(a_1)} = \{a_2 \in A_2 : ||y(a_1)M_{a_2}|| > 0\}$$

respectively.

It is easily checked that the set

$$B(x) = \{(a_1, a_2) \in A_1 \times A_2 : ||xM_{a_1}M_{a_2}|| > 0\}$$

satisfies

$$B(x) = \{(a_1, a_2)\} \in A_1 \times A_2 : a_1 \in A_{1,x} \textit{ and } a_2 \in A_{2,y(a_1)}\}.$$

Hence

$$\mathbf{T}_{\mathcal{H}_1}\mathbf{T}_{\mathcal{H}_2}u(x) = \int_{B(x)} u(\frac{xM_{a_1}M_{a_2}}{||M_{a_1}M_{a_2}||})||xM_{a_1}M_{a_2}||\tau^2(da_1, da_2) = \mathbf{T}_{\mathcal{H}_3}u(x). \;\; \square$$

The following result is an immediate corollary of the previous lemma.

Corollary 5.1 *Let $\mathcal{H}_n = \{(S, \mathcal{F}, \delta_0), (p_n, \lambda), (A_n, \mathcal{A}_n, \varrho_n), (m_n, \tau_n)\}$, $n = 1, 2, ..., N$ be a sequence of HMMs with densities having the same state space $(S, \mathcal{F})$ and such that the partitions $(m_n, \tau_n), n = 1, 2, ...N$ are regular. For $n = 1, 2, ..., N$ let let $\mathbf{P}_n$ denote the* **filter kernel** *induced by $\mathcal{H}_n$ and let $\mathbf{T}_n$ be the associated transition operator.*

Set $(m^1, \tau^1) = (m, \tau)$ and, for $n = 2, 3, ..., N$, let (m^n, τ^n) be the product of the partitions $(m_n, \tau_n), n = 1, 2, ...N$ defined recursively by

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

and

$$\tau^{n+1} = \tau^n \otimes \tau_{n+1},$$

and let p^n denote the density kernel of the Markov kernel determined by the partition (m^n, τ^n). From Proposition 5.1 we know that (m^n, τ^n) is also regular.

Let $\mathcal{H}^N$ denote the HMM defined by

$$\mathcal{H}^N = \{(S, \mathcal{F}, \delta_0), (p^N, \lambda), (A^N, \mathcal{A}^N), (m^N, \tau^N)\}, \qquad (12)$$

let $\mathbf{P}^N$ denote the **filter kernel** *induced by $\mathcal{H}^N$ and let $\mathbf{T}^N$ be the associated transition operator.*

For $(a_1, a_2, ..., a_N) \in \mathcal{A}^N$ we write $M_{a_1}M_{a_2}...M_{a_N} = M^N_{a^N}$.

Then a)

$$\mathbf{P}_1\mathbf{P}_2...\mathbf{P}_N(x, E) = \mathbf{P}^N(x, E) =$$

$$\int_{\{a^N:||xM^N_{a^N}||>0\}} I_E(\frac{xM^N_{a^N}}{||xM^N_{a^N}||})||xM^N_{a^N}||\tau^N(da^N), \; \forall x \in K, \; \forall E \in \mathcal{E},$$

b)

$$\mathbf{T}_1\mathbf{T}_2...\mathbf{T}_N u(x) = \mathbf{T}^N u(x) =$$

$$\int_{\{a^n:||xM^N_{a^N}||>0\}} u(\frac{xM^N_{a^N}}{||xM^N_{a^N}||})||xM^N_{a^N}||\tau^N(da^N), \; \forall x \in K, \; \forall u \in B[K, \mathcal{E}]. \; \square$$

We end this section recalling the notion of weak ergodicity.

Definition 5.3 *Let* $\mathcal{H} = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$ *be a HMM with densities such that the partition* (m,τ) *is regular, let* $\mathbf{P}$ *be the filter kernel induced by* $\mathcal{H}$ *and let* $\mathbf{T}$ *be the transition operator associated to* $\mathbf{P}$*. We say that the filter kernel* $\mathbf{P}$ *is* **weakly ergodic** *if*

$$\lim_{n\to\infty} |\mathbf{T}^n u(x) - \mathbf{T}^n u(y)| = 0$$

for all x, y *in* K *and all functions* $u \in C[K,\mathcal{E}]$ *which are uniformly continuous.*

If furthermore there exists a measure $\nu \in \mathcal{P}(K,\mathcal{E})$ *such that*

$$\lim_{n\to\infty} \langle u, \mu\mathbf{P}^n\rangle = \langle u,\nu\rangle,\ \forall\mu \in \mathcal{P}(K,\mathcal{E}) \ \text{ and } \forall u \in C[K,\mathcal{E}]$$

then we say that $\mathbf{P}$ *is* **weakly ergodic with stationary measure** ν*.*

6 The random mapping associated to a Hidden Markov Model with densities.

Let again

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

be a HMM with densities, such that the partition (m,τ) is regular. As before, set $K = \mathcal{P}_\lambda(S,\mathcal{F})$, let δ_{TV} denote the "total variation metric" on K and let $\mathcal{E}$ denote the $\sigma - algebra$ generated by δ_{TV}.

Let $x \in \mathcal{Q}_\lambda(S,\mathcal{F})$. As before, we write

$$||xM_a|| = \int_S\int_S m(s,t,a)x(ds)\lambda(dt),$$

and for each $a \in A$ we let $xM_a \in \mathcal{Q}(S,\mathcal{F})$ be defined by

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

Also as above, we define $\overline{M} : \mathcal{Q}_\lambda(S,\mathcal{F}) \times A \to \mathcal{Q}_\lambda(S,\mathcal{F})$ by $\overline{M}(x,a) = xM_a$. Since the partition (m,τ) is regular it follows by definition that $\overline{M} : K \times A \to K$ is a continuous function.

We now define $g : K \times A \to [0,\infty)$ by

$$g(x,a) = ||xM_a|| \tag{13}$$

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

$$h(x,a) = xM_a/||xM_a||,\ \ if\ \ ||xM_a|| > 0 \tag{14}$$

$$h(x,a) = x,\ \ if\ \ ||xM_a|| = 0. \tag{15}$$

Now, since $g(x,a) = ||\overline{M}(x,a)||$, and the mapping $\overline{M} : \mathcal{Q}_\lambda(S,\mathcal{F}) \times A \to \mathcal{Q}_\lambda(S,\mathcal{F})$ is continuous, it follows that the mapping $g : K \times A \to [0,\infty)$ is also continuous, and therefore in particular g is measurable. Furthermore,

$$\int_A g(x,a)\tau(da) = 1, \forall x \in K$$

since

$$\int_A ||xM_a||\tau(da) = \int_A \int_S \int_S m(s,t,a)x(ds)\lambda(dt)\tau(da) =$$

$$= \int_S \int_S m(s,t,a)\tau(da)\lambda(dt)x(ds) = \int_S \int_S p(s,t)\lambda(dt)x(ds)) = \int_S x(ds) = 1.$$

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

$$G(x,B) = \int_B g(x,a)\tau(da) \tag{16}$$

it is clear that $G(x,A) = 1$ for all $x \in K$ and that $G(\cdot, B)$ is measurable for each $B \in \mathcal{A}$; therefore $G : K \times \mathcal{A} \to [0,1]$ defines a tr.pr.f from $(K,\mathcal{E})$ to $(A,\mathcal{A})$. Furthermore, using the inequality (2.1) and the fact that the set $\{(x,a) \in K \times A : ||xM_a|| = 0\}$ is a closed set it is not difficult to prove that $h : K \times A \to K$ is *measurable*, and therefore $\{(K,\mathcal{E}),(A,\mathcal{A}),h,(g,\tau)\}$ defines a **random mapping**.

Definition 6.1 *We call the random mapping $\{(K,\mathcal{E}),(A,\mathcal{A}),h,G\}$ defined by (14), (15), (13) and (16) the* **random mapping associated to the HMM** $\{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$.

Next, let $\mathbf{Q}$ denote the Markov kernel induced by the random mapping $\{(K,\mathcal{E}),(A,\mathcal{A}),h,G\}$ (see formula (7) of Definition 4.6) and let $\mathbf{P}$ denote the filter kernel induced by the HMM $\{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$ (see Definition 5.2).

Observation 6.1

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

Proof.

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

$$\int_{\{a:h(x,a)\in E\}} ||xM_a||\tau(da) = \int_{A_x^+} I_E(xM_a/||xM_a)||xM_a||\tau(da) = \mathbf{P}(x,E). \ \Box$$

Having made this observation we now make the following definition which partly replaces Definition 5.2.

Definition 6.2 *Let $\mathcal{H} = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A}),(m,\tau)\}$ be a HMM with density such that the partition (m,τ) is regular, and let $\{(K,\mathcal{E}),(A,\mathcal{A}),h,G\}$ be the* **random mapping** *associated to $\mathcal{H}$. (See (13), (16), (14), and (15).)*

For $\mu \in \mathcal{P}(K,\mathcal{E})$, let $\{Y_{n,\mu}, n = 1,2,...\}$ be the index sequence associated to the random mapping $\{(K,\mathcal{E}),(A,\mathcal{A}),h,G\}$ and let $\{Z_{n,\mu}, n = 0,1,2,...\}$ be the state sequence associated to the random mapping $\{(K,\mathcal{E}),(A,\mathcal{A}),h,G\}$. (The state sequence and the index sequence associated to a random mapping were defined in Definition 4.5.)

Let $\mathbf{P} \in \mathcal{TP}((K,\mathcal{E}))$ denote the Markov kernel induced by the random mapping $\{(K,\mathcal{E}),(A,\mathcal{A}),h,G\}$.

We now introduce the following terminology:
1) The Markov chain $Z_{n,\mu}, n = 0,1,2,...$ with values in $(K,\mathcal{E})$ is called the **filtering process** *associated to the HMM $\mathcal{H}$.*

2) The sequence $\{Y_{n,\mu}, n = 1, 2, ...\}$ *is called the* **observation sequence** *associated to the HMM* $\mathcal{H}$.
3) The Markov kernel $\mathbf{P} \in \mathcal{TP}((K, \mathcal{E}))$ *is called the* **filter kernel**.

In case $\mu = \delta_x$ *for some* $x \in K$ *we write* $Z_n(x)$ *instead of* Z_{n,δ_x} *and* $Y_n(x)$ *instead of* Y_{n,δ_x}. $\square$

Remark. That HMM induces a random mapping is by no means a new observation. Already in the paper [4] Blackwell proves a theorem for random mappings which he applies to the filtering process he is considering. (A finite state space and an observation sequence which is determined by a lumping function.) In section 2.3.3 of the book [10] the connection between partially observed Markov chains and random mappings is described and in the book [11] this connection is mentioned at several places. In the paper [12] from 1973 it is proved that the filtering process converges in distribution with *geometric convergence rate* in case the state space is finite and the tr.pr.m P has strictly positive elements by showing that the associated random mapping is a so called "distance diminishing model". (See [20], chapter 2.) In a recent paper by C Anton Popescu (see [1]) a similar result is proved. The connection between filtering processes and random mappings is also emphasized in [13]. $\square$

Now some further notations similar to those in the previous section. Set

$$A^1 = A,\ A^{n+1} = A^1 \times A^n,\ n = 1, 2, ...$$

For $(a_1, a_2, ..., a_n) \in A^n$ we use the notation

$$a^n = (a_1, a_2, ..., a_n)$$

and consequently we also write

$$Y_\mu^n = (Y_{1,\mu}, Y_{2,\mu}, ..., Y_{n,\mu}).$$

We define $h^n : K \times A^n \to K, n = 1, 2, ...$ by first defining $h^1 = h$, where thus h is defined by (16) and (14), and then defining h^n iteratively by

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

We define $g^n : S \times A^n \to [0, \infty), n = 1, 2, ...$ iteratively by $g^1 = g$ and

$$g^{n+1}(s, a^{n+1}) = g^n(s, a^n)g(h^n(s, a^n), a_{n+1}).$$

Recall that g is defined by (13).

As before we denote the *n-product measure* of $\tau \in (A, \mathcal{A})$ by τ^n. For $(a_1, a_2, ..., a_n) \in \mathcal{A}^n$ we write $M_{a_1}M_{a_2}...M_{a_n} = M^n_{a^n}$.

Clearly $\{(K, \mathcal{E}), (A^n, \mathcal{A}^n), h^n, (g^n, \tau^n)\}$ is the $n : th$ composition of the random mapping $\{(K, \mathcal{E}), (A, \mathcal{A}), h, (g, \tau)\}$.

Due to the special form of both $h : K \times A \to K$ and $g : K \times A \to [0, \infty)$ the following relations hold.

Proposition 6.1 *Let* $\mathcal{H}= \{(S, \mathcal{F}, \delta_0\}, (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$, *be a HMM with densities such that the partition* (m, τ) *is regular. Let* $\{(K, \mathcal{E}), (A, \mathcal{A}), h, (g, \tau)\}$ *denote the random mapping associated to the HMM* $\mathcal{H}$, *and for* $n = 1, 2, ...$ *let* $\{(K, \mathcal{E}), (A^n, \mathcal{A}^n), h^n, (g^n, \tau^n))\}$ *be the nth composition of the random mapping* $\{(K, \mathcal{E}), (A, \mathcal{A}), h, (g, \tau)\}$.

Then a)

$$g^n(x,a^n) = ||xM^n_{a^n}||, \ \forall x \in K, \forall a^n \in A^n,$$

b)

$$h^n(x,a^n) = xM^n_{a^n}/||xM^n_{a^n}||, \ \ if \ \ ||xM^n_{a^n}|| > 0.$$

Proof. See Lemma 5.1 and its proof. □

We end this section with some formulas for the state sequence and the observation sequence of an HMM with densities and regular partition.

Thus, let $\{Z_n(x), n = 0,1,2,...\}$ denote the filter process and let $\{Y_n(x), n = 1,2,...\}$ denote the observation sequence associated to a HMM $\mathcal{H} = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$ and the initial distribution $\delta_x \in \mathcal{P}(K,\mathcal{E})$, and let $\mathcal{H} = \{(K,\mathcal{E}),(A,\mathcal{A}),h,G\}$ be the associated random mapping. Then

$$Z_n(x) = h^n(x, Y^n(x)), \ \ n = 1,2,... \, .$$

Furthermore, if we let $\mathbf{P}$ denote the filter kernel induced by $\mathcal{H}$ and let $\mathbf{T} : B[K,\mathcal{E}] \to B[K,\mathcal{E}]$ denote the transition operator associated to $\mathbf{P}$, then

$$\mathbf{T}u(x) = \int_K u(y)\mathbf{P}(x,dy) = E[u(h(x,Y_1(x)))] = \int_{A^+_x} u(h(x,a))g(x,a)\tau(da) =$$

$$\int_{A^+_x} u(\frac{xM_a}{||xM_a||})||xM_a||\tau(da) \tag{17}$$

and, for $n = 2,3,...$, we have

$$\mathbf{T}^n u(x) = \int_K u(y)\mathbf{P}^n(x,dy) = E[u(h^n(x,Y^n(x)))] =$$

$$\int_{\{a^n:||xM^n_{a^n}||>0\}} u(h^n(x,a^n))g^n(x,a^n)\tau^n(da^n) =$$

$$\int_{\{a^n:||xM^n_{a^n}||>0\}} u(\frac{xM^n_{a^n}}{||xM^n_{a^n}||})||xM^n_{a^n}||\tau^n(da^n) \tag{18}$$

where the last equality is a consequence of Proposition 6.1.

7 The Kantorovich distance on the space $\mathcal{P}(\mathcal{P}_\lambda(S))$

As above let $K = \mathcal{P}_\lambda(S,\mathcal{F})$, let δ_{TV} be the metric on K determined by the total variation, and let $\mathcal{E}$ be the $\sigma - algebra$ on K generated by δ_{TV}.

If μ and ν belong to $\mathcal{P}(K,\mathcal{E})$, we let $\mathcal{P}(K^2;\mu,\nu)$ denote the set of probability measures in $\mathcal{P}(K^2,\mathcal{E}^2)$ defined by

$$\mathcal{P}(K^2;\mu,\nu) = \{\tilde{\mu} \in \mathcal{P}(K^2) : \tilde{\mu}(E \times K) = \mu(E), \ \ \tilde{\mu}(K \times E) = \nu(E), \ \forall E \in \mathcal{E}\}.$$

The **Kantorovich distance** $d_K(\mu,\nu)$ between μ and ν is defined as

$$d_K(\mu,\nu) = \inf\{\int_{K^2} \delta_{TV}(x,y)\tilde{\mu}(dx,dy) : \tilde{\mu} \in \mathcal{P}(K^2;\mu,\nu)\}. \tag{19}$$

Since $0 \le \delta_{TV}(x,y) \le 2$ for $x,y \in K$ it is clear that $d_K(\mu,\nu)$ is well-defined.

From the so called Kantorovich-Rubenstein theorem (see [8], Theorem 11.8.2) it follows that the Kantorovich distance d_K can also be defined by

$$d_K(\mu,\nu) = \sup\{\int_K u(x)\mu(dx) - \int_K u(x)\nu(dx) : u \in Lip_1[K]\ \}, \qquad (20)$$

where thus

$$Lip_1[K] = \{u \in C[K] : \sup\{\frac{|u(x)-u(y)|}{\delta_{TV}(x,y)} : x,y \in K,\ \delta_{TV}(x,y) > 0\} \le 1\}.$$

From this representation it is easily proved that the Kantorovich distance d_K is in fact a **metric** on $\mathcal{P}(K,\mathcal{E})$.

It is well-known that the topology on $\mathcal{P}(K,\mathcal{E})$ induced by the metric d_K is equivalent to the weak topology (see e.g [8], chapter 11).

8 The barycenter

An important notion within the theory of filtering processes is the notion **barycenter**. This concept was introduced into the theory of filtering processes by Kunita. (See [18].)

Let $(S,\mathcal{F},\delta_0)$ be a complete, separable, metric space, and let $\lambda \in \mathcal{Q}^\infty(S,\mathcal{F})$. As before, let $K = \mathcal{P}_\lambda(S,\mathcal{F})$, let δ_{TV} be the metric on K determined by the total variation, let $\mathcal{E}$ be the $\sigma - algebra$ on K generated by δ_{TV} and let $\mathcal{P}(K,\mathcal{E})$ denote the set of probability measures on $(K,\mathcal{E})$.

Now let $\mu \in \mathcal{P}(K,\mathcal{E})$. The **barycenter** of μ, which we denote by $\bar{b}(\mu)$, is an element in K which is defined as follows: First, for each $F \in \mathcal{S}$ we let $I_F : S \to \{0,1\}$ denote the indicator function of F defined as usual by

$$I_F(s) = 1,\ \ if\ \ s \in F,\ \ I_F(s) = 0,\ \ if\ \ s \notin F.$$

For each $F \in \mathcal{F}$ we also define a mapping $U_F : K \to [0,1]$ by

$$U_F(x) = \langle I_F, x\rangle.$$

Since $|\langle I_F,x\rangle - \langle I_F,y\rangle| \le ||x-y||/2$, it is clear that $U_F \in C[K,\mathcal{E}]$ and in particular $U_F \in B[K,\mathcal{E}]$. Therefore, we can define a mapping $\bar{b}(\mu) : \mathcal{F} \to [0,1]$ by

$$\bar{b}(\mu)(F) = \langle U_F,\mu\rangle\ \ (= \int_K \langle I_F,x\rangle\mu(dx)).$$

Since

$$\bar{b}(\mu)(S) = \int_K \langle I_S,x\rangle\mu(dx) = \int_K \mu(dx) = 1,$$

$$\bar{b}(\mu)(S\setminus F)) = \int_K \langle I_{S\setminus F},x\rangle\mu(dx) =$$

$$\int_K \langle I_S,x\rangle\mu(dx) - \int_K \langle I_F,x\rangle\mu(dx) = 1 - \bar{b}(\mu)(F)),\ \ \forall F \in \mathcal{F}$$

and

$$\bar{b}(\mu)(\cup_{i=1}^\infty F_i)) = \int_K \langle\sum_{i=1}^\infty I_{F_i},x\rangle\mu(dx) =$$

$$\sum_{i=1}^{\infty}(\int_K \langle I_{F_i}, x\rangle \mu(dx) = \sum_{i=1}^{\infty} \bar{b}(\mu)(F_i),$$

if $F_i \in \mathcal{F}, i = 1, 2, , ...$ and $F_i \cap F_j = \emptyset \; if \;\; i \neq j$, it is clear that $\bar{b}(\mu) \in \mathcal{P}(S, \mathcal{F})$. Moreover, since $\lambda(F) = 0$ implies that $U_F(x) = \int_F x(ds) = x(F) = 0$ for all $x \in K$, it also follows that $\lambda(F) = 0$ implies that $\bar{b}(\mu)(F) = 0$. Hence $\bar{b}(\mu) \in K$.

We denote the set of all probability measures in $\mathcal{P}(K, \mathcal{E})$ with barycenter equal to x by $\mathcal{P}(K|x)$.

The following theorem is essentially due to Kunita. (See [18].)

Theorem 8.1 *Let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a HMM with densities such that the partition (m, τ) is regular. (Recall that in the definition of a HMM with densities we assume that both $(S, \mathcal{F}, \delta_0)$ and $(A, \mathcal{A}, \varrho)$ are complete, separable, metric spaces.) Let $P \in \mathcal{TF}((S, \mathcal{F}))$ be the Markov kernel determined by (p, λ). As usual, let $K = \mathcal{P}_\lambda(S, \mathcal{F})$, let δ_{TV} denote the total variation metric, let $\mathcal{E}$ be the $\sigma-$algebra determined by the total variation metric δ_{TV} and let $\mathbf{P} \in \mathcal{TP}((K, \mathcal{E}))$ be the filter kernel induced by the HMM $\mathcal{H}$.*

Then for all $x \in K$

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

Proof. Let $F \in \mathcal{F}$ and let $\mathbf{T}$ denote the transition operator associated to the filter kernel $\mathbf{P}$. From the definition of the barycenter we find

$$\bar{b}(x\mathbf{P})(F) = \langle U_F, \delta_x \mathbf{P}\rangle = \langle \mathbf{T}U_F, \delta_x\rangle = \mathbf{T}U_F(x) =$$

$$\int_{A_x{}^+} U_F\left(\frac{xM_a}{||xM_a||}\right)||xM_a||\tau(da) = \int_{A_x{}^+} \langle I_F, \frac{xM_a}{||xM_a||}\rangle ||xM_a||\tau(da) =$$

$$\int_{A_x{}^+} \langle I_F, xM_a\rangle \tau(da) = \int_{A_x{}^+} \int_F xM_a(dt)\tau(da) =$$

$$\int_{A_x{}^+} \int_F \int_S m(s, t, a)x(ds)\lambda(dt)\tau(da) =$$

$$\int_F \int_S \int_{A_x{}^+} m(s, t, a)\tau(da)x(ds)\lambda(dt) = \int_F \int_S p(s, t)x(ds)\lambda(dt) =$$

$$\int_F xP(dt) = xP(F)$$

from which the conclusion of the theorem follows for $n = 1$. That the conclusion also holds for $n \geq 2$ follows from Corollary 5.1.

9 A few words on couplings

Definition 9.1 *Let $(K, \mathcal{E})$ be a measurable space. Let $r > 0$ and let $\mu, \nu \in \mathcal{Q}^r(K, \mathcal{E})$. A measure $\tilde{\mu} \in \mathcal{Q}^r(K^2, \mathcal{E}^2)$ is called a* **coupling** *of μ and ν if*

$$\tilde{\mu}(E \times K) = \mu(E), \; \forall E \in \mathcal{E}$$

$$\tilde{\mu}(K \times E) = \nu(E), \; \forall E \in \mathcal{E}. \; \Box$$

We shall only use two specific couplings.

Definition 9.2 *Let* $(K, \mathcal{E})$ *be a measurable space. Let* $r > 0$ *and let* $\mu, \nu \in \mathcal{Q}^r(K, \mathcal{E})$. *The measure* $\tilde{\mu} \in \mathcal{Q}^r(K^2, \mathcal{E}^2)$ *defined by*

$$\tilde{\mu}(E_1 \times E_2) = (1/r)\mu(E_1)\nu(E_2),\ \forall E_1, E_2 \in \mathcal{E} \tag{21}$$

is called the **trivial coupling** *of* μ *and* ν.

Definition 9.3 *Let* $(K, \mathcal{E})$ *be a measurable space. Let* $r > 0$ *and let* $\mu, \nu \in \mathcal{Q}^r(K, \mathcal{E})$. *Let* $E^+ \in \mathcal{E}$ *denote a set in* $\mathcal{E}$ *such that*

$$\mu(E \cap E^+) \geq \nu(E \cap E^+),\ \ \forall E \in \mathcal{E},$$

and set $E^- = K \setminus E^+$. *(Recall that* $\mu(E^+) - \nu(E^+) = (1/2)||\mu - \nu||$.*)*

The measure $\tilde{\mu} \in \mathcal{Q}^r(K^2, \mathcal{E}^2)$ *defined by*

$$\tilde{\mu}(E_1 \times E_2) = \mu(E_1 \cap E_2 \cap E^-) + \nu(E_1 \cap E_2 \cap E^+) +$$

$$(\mu(E_1 \cap E^+) - \nu(E_1 \cap E^+))(\nu(E_2 \cap E^-) - \mu(E_2 \cap E^-))/(\mu(E^+) - \nu(E^+)), \tag{22}$$

where the second term is omitted if $\mu = \nu$, *is called the* **Vaserstein coupling** *of* μ *and* ν.

Remark. In the paper [24] L Vaserstein introduced the coupling defined by (22) for probabilities defined on a denumerable set. □

That the measures defined by (21) and (22) are couplings of μ and ν is easily checked. We shall usually denote the Vaserstein coupling by using a V as subscript.

We shall have use of the following well-known fact regarding the Vaserstein coupling.

Lemma 9.1 *Let* $(K, \mathcal{E})$ *be a complete, separable, metric space. Let* $r > 0$ *and let* $\mu, \nu \in \mathcal{Q}^r(K, \mathcal{E})$, *let* $\tilde{\mu}_V \in \mathcal{Q}^r((K^2, \mathcal{E}^2))$ *be the Vaserstein coupling and let* $D = \{(x, y) \in K \times K : x = y\}$. *Then*

$$\tilde{\mu}_V(D) = r - ||x - y||/2.\ \square$$

For a proof of the lemma when $r = 1$ see e.g [19], Section 1.5, Theorem 5.2.

10 The Vaserstein coupling of Markov kernels with densities.

Lat $(K, \mathcal{E})$ and $(A, \mathcal{A})$ be two measurable spaces and let $Q \in \mathcal{TP}((K, \mathcal{E}), (A, \mathcal{A}))$ be a tr.pr.f from $(K, \mathcal{E})$ to $(A, \mathcal{A})$.

Let ψ be a positive $\sigma - finite$ measure on $(A, \mathcal{A})$ and suppose that $q \in D_\psi[K, A]$ is a probability density kernel of $Q \in \mathcal{TP}((K, \mathcal{E}), (A, \mathcal{A}))$. As usual let $K^2 = K \times K$, $A^2 = A \times A$, $\mathcal{E}^2 = \mathcal{E} \otimes \mathcal{E}$ and $\mathcal{A}^2 = \mathcal{A} \otimes \mathcal{A}$.

We shall now define a tr.pr.f $\tilde{Q}_V \in \mathcal{PT}((K^2, \mathcal{E}^2), (A^2, \mathcal{A}^2))$ as follows.

We first define $\hat{q} : K^2 \times A \to [0, \infty)$ by

$$\hat{q}(x, y, a) = \min\{q(x, a), q(y, a)\},$$

we define $f : K^2 \times A \to [0, \infty)$ and $g : K^2 \times A \to [0, \infty)$ by

$$f(x, y, a) = q(x, a) - \hat{q}(x, y, a)$$

and

$$g(x, y, a) = q(y, a) - \hat{q}(x, y, a)$$

respectively, and we define

$$\Delta(x, y) = \int |q(x, a) - q(y, a)|\psi(da).$$

To define $\tilde{Q}_V \in \mathcal{PT}((K^2, \mathcal{E}^2)$ it suffices to specify $\tilde{Q}_V(x, y, \tilde{B})$ for sets $\tilde{B}$ which are rectangular. Thus let $(x, y) \in K^2$ and $B_1, B_2 \in \mathcal{A}$. Set $D = B_1 \cap B_2$. We now define

$$\tilde{Q}_V((x, y), B_1 \times B_2) =$$

$$\int_D \hat{q}(x, y, a)\psi(da) + \int_{B_2}\int_{B_1} f(x, y, a)g(x, y, b)\psi(da)\psi(db)/\Delta(x, y)$$

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

That $\tilde{Q}_V((x, y), \cdot))$ is a coupling of $Q(x, \cdot)$ and $Q(y, \cdot)$ for every $x, y \in K$ is easy to check. We present the arguments for sake of completeness.

Thus, let $x, y \in K$. Define

$$A^+ = \{a \in \mathcal{A} : q(x, a) \geq q(y, a)\}$$

and $A^- = A \setminus A^+$. From the definition now follows, if $\Delta(x, y) > 0$ and $B \in \mathcal{A}$, that

$$\tilde{Q}_V((x, y), B \times A)) =$$

$$\int_{B\cap A^+} q(y, a)\psi(da) + \int_{B\cap A^-} q(x, a)\psi(da) +$$

$$\int_{A^1}\int_{B\cap A^+} (q(x, a) - q(y, a))(q(y, b) - q(x, b)\psi(da)\psi(db)/\Delta(x, y) =$$

$$\int_{B\cap A^+} q(y, a)\psi(da) + \int_{B\cap A^-} q(x, a)\psi(da) +$$

$$\int_{B\cap A^+} (q(x, a) - q(y, a))\psi(da) = \int_B q(x, a)\psi(da) = Q(x, B).$$

If $\Delta(x, y) = 0$ then clearly

$$\tilde{Q}_V((x, y), B \times A)) = \int_B q(x, a)\psi(da) = Q(x, B).$$

In the same way one can show that for all $x, y \in K$ and $B \in \mathcal{A}$

$$\tilde{Q}_V((x, y), A \times B)) = Q(y, B).$$

That furthermore $\tilde{Q}_V((\cdot, \cdot), B_1 \times B_2)$ is measurable for all $B_1, B_2 \in \mathcal{A}$ follows from the fact that all integrands are measurable, and therefore it follows that $\tilde{Q}_V((\cdot, \cdot), \tilde{B})$ is measurable for all $\tilde{B} \in \mathcal{A} \otimes \mathcal{A}$.

We call $\tilde{Q}_V$, as defined above, the **Vaserstein coupling** of the Markov kernel Q.

11 An ergodic theorem.

We now return to hidden Markov models and present the main theorem of this paper.

Theorem 11.1 *Let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a HMM with densities such that (m, τ) is regular. Let $P \in \mathcal{TF}((S, \mathcal{F}))$ be the Markov kernel determined by (p, λ). As usual, let $K = \mathcal{P}_\lambda(S, \mathcal{F})$, let δ_{TV} denote the total variation metric, let $\mathcal{E}$ be the $\sigma-$ algebra determined by the total variation metric δ_{TV} and let $\mathbf{P} \in \mathcal{TP}((K, \mathcal{E}))$ be the filter kernel induced by the HMM $\mathcal{H}$.*

Suppose now that the Markov kernel P is **strongly ergodic** *with stationary measure π. Suppose also that the filter kernel $\mathbf{P}$ satisfies the following condition:* **Condition E***:* To every ρ there exist an integer N and a number α such that for any two measures μ and ν with barycenters equal to π there exists a coupling $\tilde{\mu}_N$, say, of $\mu\mathbf{P}^N$ and $\nu\mathbf{P}^N$ such that if we set $D_\rho = \{(x, y) \in K^2 : \delta_{TV}(x, y) < \rho\}$ then

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

Then the conclusion is that the filter kernel $\mathbf{P}$ is **weakly ergodic**.

If furthermore, either

1) there exists a measure $\mu \in \mathcal{P}(K, \mathcal{E})$ which is **invariant** *with respect to $\mathbf{P}$ or*

2) there exists an element $x_0 \in \mathcal{P}_\lambda(S, \mathcal{F})$ such that the sequence

$$\{\mathbf{P}(x_0, \cdot), n = 1, 2, ...\}$$

is a **tight** *sequence, or*

3) the Markov kernel $P \in \mathcal{TF}((S, \mathcal{F}))$ is **uniformly ergodic***,*

- then there exists a unique invariant measure $\mu \in \mathcal{P}(K, \mathcal{E})$ such that $\mathbf{P}$ is **weakly ergodic** *with* **stationary measure** *μ.*

Remark. In the paper [16] we proved that, if the state space S is a denumerable set and $x \in K$, then the set of probability measures on $(K, \mathcal{E})$ with barycenter equal to x is a *tight* set. Therefore, in this case $\{\mathbf{P}(\pi, \cdot), n = 1, 2, ...\}$ is a tight sequence. We believe that also under the hypotheses of Theorem 11.1 the sequence $\{\mathbf{P}(\pi, \cdot), n = 1, 2, ...\}$ is a *tight* sequence; therefore we believe that also under the hypotheses of Theorem 11.1 the conclusion should be that the filtering process is weakly ergodic *with a stationary measure.*

12 Auxiliary theorems for Markov chains in complete, separable, bounded, metric spaces.

In this section we shall state two limit theorems for Markov chains in complete, separable, bounded, metric spaces. In this section $(K, \mathcal{E})$ will denote an arbitrary complete, separable, metric space, with metric δ and $\sigma - algebra\ \mathcal{E}$.

Let $Q : K \times \mathcal{E} \to [0, 1]$ be a tr.pr.f on $(K, \mathcal{E})$ and let $T : B[K] \to B[K]$ denote the **transition operator** associated to Q defined as usual by $Tu(x) = \int_K u(y)Q(x, dy)$. We define $T^0u(x) = u(x) = \int_K u(y)Q^0(x, dy)$. Recall that

$$osc(T^{n+1}u) \leq osc(T^n u), \quad n = 0, 1, 2, ..., \quad u \in B[K], \tag{23}$$

since T is an "averaging" operator.

When stating and proving the forthcoming auxiliary theorem we shall use the notion *shrinking property*, a property we introduced in [16].

Definition 12.1 *Let Q be a tr.pr.f on $(K, \mathcal{E})$, and let T be the associated transition operator. If for every $\rho > 0$ there exists a number α, $0 < \alpha < 1$ such that for every nonempty, compact set $E \subset K$, every $\eta > 0$ and every $\kappa > 0$, there exist an integer N and another nonempty, compact set $F \subset K$ such that, if the integer $n \geq N$, then for all $u \in Lip[K]$*

$$osc_E(T^n u) \leq \eta\gamma(u) + \kappa osc(u) + \alpha\rho\gamma(u) + (1-\alpha) osc_F(T^{n-N}u),$$

then we say that Q has the **shrinking property**. *We call α a* **shrinking number** *associated to ρ.*

Theorem 12.1 *Let $(K, \mathcal{E})$ be a complete, separable, bounded, metric space with metric δ, let Q be a tr.p.f on $(K, \mathcal{E})$ and suppose that Q has the* **shrinking property**. *Suppose also that Q is* **Lipschitz equicontinuous**. *Then Q is* **weakly ergodic**.

Furthermore, if either
1) there exists an invariant probability measure or
2) there exists $x^* \in K$ such that $\{Q^n(x^*, \cdot),\ n = 1, 2....\}$ is a tight sequence,
then there exists a unique invariant probability measure ν such that Q is **weakly ergodic** *with* **stationary measure** *ν.*

Proof. Set

$$D = \sup\{\delta(x, y) : x, y \in K\}.$$

Since K is assumed to be bounded, $D < \infty$. Now define a new metric δ_1, say, on K by $\delta_1 = (2/D)\delta$. Then clearly the topology induced by δ_1 is equal to the topology induced by δ and therefore it is clear that we may assume from the beginning that $D = 2$. As usual we define $Lip_1[K] = \{u \in Lip[K] : \gamma(u) \leq 1\}$.

Next let d_K denote the Kantorovich distance on the set $\mathcal{P}(K, \mathcal{E})$ consisting of all probability measures on $(K, \mathcal{E})$ that is

$$d_K(\mu, \nu) = \inf\{\int_{K^2} \delta(x, y)\tilde{\mu}(dx, dy) : \tilde{\mu} \in \mathcal{P}(K^2; \mu, \nu)\}$$

where $\mathcal{P}(K^2; \mu, \nu)$ denotes the set of probability measures in $\mathcal{P}(K^2, \mathcal{E}^2)$ defined by

$$\mathcal{P}(K^2; \mu, \nu) = \{\tilde{\mu} \in \mathcal{P}(K^2) : \tilde{\mu}(E \times K) = \mu(E),\ \ \tilde{\mu}(K \times E) = \nu(E), \forall E \in \mathcal{E}\}.$$

Since $D = 2$ it follows from the theorem of Kantorovich and Rubenstein (see [8], Thm 11.8.2) that

$$d_K(\mu, \nu) = \sup\{\int u(x)\mu(dx) - \int u(x)\nu(dx) : u \in Lip_1[K]\} =$$

$$\sup\{\int u(x)\mu(dx) - \int u(x)\nu(dx) : u \in Lip_1[K], ||u|| \leq 1\}.$$

We shall next prove that for all $x, y \in K$ and all $u \in Lip_1[K]$,

$$\lim_{n\to\infty} |\int_K u(z)Q^n(x, dz) - \int_K u(z)Q^n(y, dz)| = 0. \quad (24)$$

Thus, let $\epsilon > 0$ be given. In order to prove (24) we shall show, that for any two elements $x, y \in K$ and any function $u \in Lip_1[K]$, we can find an integer N, depending on x and y but not on u, such that

$$|\int_K u(z)Q^n(x,dz) - \int_K u(z)Q^n(y,dz)| < 6\epsilon, \tag{25}$$

for all $n \geq N$.

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 $\kappa = \kappa_1 = \epsilon/2$, we can find an integer N_1 and a compact set E_1 such that, if $n \geq N_1$, then

$$|\int_K u(z)Q^n(x,dz) - \int_K u(z)Q^n(y,dz)| \leq \eta_1 + 2\kappa_1 + \alpha\epsilon +$$

$$(1-\alpha) \sup_{z_1,z_2 \in E_1} |\int_K u(z)Q^{n-N_1}(z_1,dz) - \int_K u(z)Q^{n-N_1}(z_2,dz)|$$

where we have used the fact that $\gamma(u) \leq 1$, that $osc(u) \leq 2$ since $\sup\{\delta(z_1, z_2) : z_1, z_2 \in K\} \leq 2$ and $\gamma(u) \leq 1$, and the fact that we have chosen $\rho = \epsilon$.

We now choose $M = \min\{m : (1-\alpha)^m < \epsilon\}$. For $i = 2, 3, ..., M$ we define the numbers η_i by

$$\eta_i = \epsilon/2^i,$$

the numbers κ_i by

$$\kappa_i = \epsilon/2^i$$

and having defined the compact sets E_i, for $i = 1, 2, ..., j-1$, and the integers N_i, for $i = 1, 2, ..., j-1$, it follows from the shrinking property that we can find a compact set E_j and an integer N_j such that

$$\sup_{z_1,z_2 \in E_{j-1}} |T^m u(z_1) - T^m u(z_2)| \leq \eta_j + 2\kappa_j + \alpha\rho +$$

$$(1-\alpha) \sup_{z_1,z_2 \in E_j} |T^{m-N_j} u(z_1) - T^{m-N_j} u(z_2)| \tag{26}$$

if $m \geq N_j$. By using (26) repeatedly it follows that if the integer n satisfies $n \geq N_1 + N_2 + ... + N_j$ then

$$|T^n u(x) - T^n u(y)| \leq \epsilon/2 + 2\epsilon/2 + \alpha\epsilon +$$

$$(1-\alpha) \sup_{z_1,z_2 \in E_1} |T^{n-N_1} u(z_1) - T^{m-N_1} u(z_2)| \leq$$

$$\sum_{i=1}^{j} \epsilon/2^i + 2\sum_{i=1}^{j} \epsilon/2^i + \epsilon\alpha(1 + (1-\alpha) + (1-\alpha)^2 + ... + (1-\alpha)^{j-1}) +$$

$$(1-\alpha)^j \sup_{z_1,z_2 \in E_j} |T^{n-(N_1+N_2+...+N_j)}u(z_1) - T^{n-(N_1+N_2+...+N_j)}u(z_2)|.$$

In particular, if $j = M$ and the integer n satisfies $n \geq N_1 + N_2 + ... + N_M$, then

$$|T^n u(x) - T^n u(y)| \leq$$

$$\sum_{i=1}^{M} \epsilon/2^i + 2\sum_{i=1}^{M} \epsilon/2^i + \epsilon\alpha(1 + (1-\alpha) + (1-\alpha)^2 + ... + (1-\alpha)^{M-1}) +$$

$$(1-\alpha)^M \sup_{z_1,z_2 \in E_M} |T^{n-N}u(z_1) - T^{n-N}u(z_2)|$$

where $N = N_1 + N_2 + ... + N_M$, and by using the simply proved and well-known fact that $osc(Tu) \leq osc(u)$ for any transition operator T, and the fact that

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

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 M is chosen in such a way that

$$(1-\alpha)^M < \epsilon$$

it follows that

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

if $n \geq N$. Hence (25) holds if $n \geq N$, from which it follows that (24) is satisfied, if x, y in K and $u \in Lip_1[K]$.

Since (24) holds for all $u \in Lip_1[K]$, it also holds for all $u \in Lip[K]$ and since $Lip[K]$ is dense in $C_{uniform}[K]$ in the supremum norm topology, it follows that (24) holds for all $u \in C_{uniform}[K]$ and all $x, y \in K$ and hence Q is **weakly ergodic**. Thereby the first part of Theorem 12.1 is proved.

Next using the estimate (25) and the fact that Q is assumed to be Lipschitz equicontinuous we shall prove:

Lemma 12.1 *Let as before $(K, \mathcal{E})$ be a complete, separable, bounded, metric space with metric δ, let Q be a tr.p.f on $(K, \mathcal{E})$ and suppose that Q has the* **shrinking property**. *Suppose also that Q is* **Lipschitz equicontinuous**. *Then, to every nonempty, compact set $E \in \mathcal{E}$ and every $\epsilon > 0$, we can find an integer N such that, for any function $u \in Lip_1[K]$,*

$$\sup_{x,y \in E} |\int_K u(z)Q^n(x,dz) - \int_K u(z)Q^n(y,dz)| \leq \epsilon, \tag{27}$$

for all $n \geq N$.

Proof of Lemma 12.1. Let $E \in \mathcal{E}$ and ϵ be given, where E is a nonempty, compact set. Since we have assumed that Q has the Lipschitz equicontinuity property, there exists a constant C such that for all $n \geq 1$

$$|\int_K u(z)Q^n(x,dz) - \int_K u(z)Q^n(y,dz)| \leq C\delta(x,y)\gamma(u) \tag{28}$$

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

Next, set $\epsilon_1 = \epsilon/3C$. Since E is compact we can find a finite set $\Psi = \{x_i, i = 1, 2, ..., M\}$ consisting of M elements such that for every $x \in E$

$$\inf\{\delta(x, x_i) : x_i \in \Psi\} < \epsilon_1.$$

Further, let $x_i, x_j \in \Psi$ be two arbitrary elements. From (24) it follows, that we can find an integer N_{x_i,x_j} such that, if $n \geq N_{x_i,x_j}$,

$$|\int_K u(z)Q^n(x_i, dz) - \int_K u(z)Q^n(x_j, dz)| < \epsilon/3.$$

Therefore, if we define $N = \max\{N_{x_i,x_j} : (x_i, x_j) \in \Psi \times \Psi, x_i \neq x_j\}$ it follows that

$$|\int_K u(z)Q^n(x_i, dz) - \int_K u(z)Q^n(x_j, dz)| < \epsilon/3$$

if $n \geq N$ and $x_i, x_j \in \Psi$ and $\gamma(u) \leq 1$.

Now, let $x, y \in E$ be chosen arbitrary. We can then find an element x_i, say, belonging to Ψ, such that $\delta(x, x_i) < \epsilon_1$ and an element x_j such that $\delta(y, x_j) < \epsilon_1$. Now let $u \in Lip_1[K]$. Using the triangle inequality, (28) and that $\epsilon_1 = \epsilon/3C$ we now find that if $n \geq N$ then

$$|\int_K u(z)Q^n(x, dz) - \int_K u(z)Q^n(y, dz)| \leq$$

$$|\int_K u(z)Q^n(x, dz) - \int_K u(z)Q^n(x_i, dz)| +$$

$$|\int_K u(z)Q^n(x_i, dz) - \int_K u(z)Q^n(x_j, dz)| +$$

$$|\int_K u(z)Q^n(x_j, dz) - \int_K u(z)Q^n(y, dz)| <$$

$$C\delta(x, x_i) + \epsilon/3 + C\delta(x_j, y) = C\epsilon_1 + \epsilon/3 + C\epsilon_1 = \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.$$

Hence

$$\sup\{|\int_K u(z)Q^n(x, dz) - \int_K u(z)Q^n(y, dz)| : x, y \in E, u \in Lip_1[K]\} \leq \epsilon$$

if $n \geq N$ and thereby the lemma is proved. $\square$.

We shall now prove the remaining part of Theorem 12.1.

We first assume that the tr.pr.f Q has an invariant measure ν say. Thus $\nu = \nu Q$. We shall now show, that under the hypotheses of the theorem, ν is the *only* invariant measure.

Suppose the contrary and thus assume that also $\mu = \mu Q$ and that $\mu \neq \nu$. Since $Lip[K, \mathcal{E}]$ is a measure determining set of functions there exists a function $u \in Lip_1[K]$ satisfying $||u|| \leq 1$ such that

$$a = \langle u, \mu\rangle - \langle u, \nu\rangle > 0.$$

Next, by choosing the compact set C sufficiently large, it is clear that for every nonnegative integer $n = 1, 2, ...,$ we have

$$a = \int_K T^n u(x)\mu(dx) - \int_K T^n u(y)\nu(dy) <$$

$$a/4 + \int_C \int_C |T^n u(x) - T^n u(y)| \mu(dx)\nu(dy)$$

and then from Lemma 12.1 we can conclude that

$$\int_C \int_C |T^n u(x) - T^n u(y)| \mu(dx)\nu(dy) < a/4,$$

if n is sufficiently large and hence $a < a/2$. Thereby we have obtained a contradiction.

To prove that

$$\lim_{n\to\infty} \int_K u(y)Q^n(x,dy) = \int_K u(y)\nu(dy) \tag{29}$$

if $u \in Lip_1[K]$ is also easy. Since ν is invariant we find

$$|\int_K u(y)Q^n(x,dy) - \int_K u(y)\nu(dy)| = |T^n u(x) - \int_K T^n u(y)\nu(dy)| =$$

$$|\int_K (T^n u(x) - T^n u(y))\nu(dy)| \leq$$

$$\int_K |T^n u(x) - T^n u(y)|\nu(dy). \tag{30}$$

Now let $\epsilon > 0$ be given and choose the compact set C so large that $x \in C$ and $\nu(C) > 1 - \epsilon$. From Lemma 12.1 it follows that we can choose an integer N so large that

$$\sup_{z_1,z_2 \in C} |T^n u(z_1) - T^n u(z_2)| < \epsilon,\ \forall n \geq N. \tag{31}$$

By using the inequalities (30) and (31) it now clearly follows that for any $\epsilon > 0$ and $x \in K$ we can find an integer N such that

$$|\int_K u(y)Q^n(x,dy) - \int_K u(y)\nu(dy)| < \epsilon(1-\epsilon) + \epsilon osc(u) \leq \epsilon + 2\epsilon = 3\epsilon$$

if $n \geq N$ from which follows that (29) holds for all $u \in Lip_1[K]$ and all $x \in K$.

Since the set $Lip_1[K]$ is dense in $C_{uniform}[K]$, it follows easily that (29) holds for all $u \in C_{uniform}[K]$ and all $x \in K$, and by applying Theorem 6.1, Chapter I of [21] it follows that (29) holds for all $u \in C[K]$ and consequently Q is weakly ergodic with stationary measure ν.

Next, instead of assuming that Q has an invariant measure let us assume that there exists an element $x^* \in K$ such that $\{Q^n(x^*,\cdot), n = 1,2....\}$ is a tight sequence. Using standard arguments we shall now show that there exists an invariant probability measure with respect to Q. Once this is proven the proof of Theorem 12.1 is completed.

As usual let T denote the transition operator associated to Q. For $n = 1,2,...$ we define $T^{(n)} : B[K] \to B[K]$ by

$$T^{(n)} = (1/n) \sum_{k=1}^{n} T^n$$

and we define $Q^{(n)} : K \times \mathcal{E} \to [0,1]$ by

$$Q^{(n)} = (1/n) \sum_{k=1}^{n} Q^n.$$

Now, since $\{Q^n(x^*, \cdot), n = 1, 2....\}$ is a tight sequence, we can extract a subsequence $n_j, j = 1, 2, ...$ such that $\{Q^{n_j}(x^*, \cdot), , j = 1, 2, ...\}$ converges weakly towards a probability measure ν say, and, since the *Cesaro limit* is equal to the ordinary limit if the latter exists, it also follows that $\{Q^{(n_j)}(x^*, \cdot), j = 1, 2, ...\}$ converges weakly towards the probability measure ν. Hence

$$\lim_{j \to \infty} T^{(n_j)} u(x^*) = \langle u, \nu \rangle \tag{32}$$

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

Now assume that $u \in Lip[K]$. We first prove that

$$\langle u, \nu Q \rangle = \langle u, \nu \rangle. \tag{33}$$

By considering the sequence $\{T^{(n_j+1)} u(x^*), j = 1, 2, ...\}$ we find on the one hand that

$$T^{(n_j+1)} u(x^*) = \langle T^{(n_j+1)} u, \delta_{x^*} \rangle =$$

$$(n_j/(n_j+1)) \langle T^{(n_j)} u, \delta_{x^*} \rangle + (1/(n_j+1)) \langle T^{n_j+1} u, \delta_{x^*} \rangle.$$

From (32) if follows that the first term tends towards $\langle u, \nu \rangle$ as j tends to infinity, and, since $\langle T^{n_j+1} u, \delta_{x^*} \rangle$ is bounded, the second term tends to zero. Hence

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

On the other hand we also find that

$$T^{(n_j+1)} u(x^*) = \langle T^{(n_j+1)} u, \delta_{x^*} \rangle =$$

$$(1/(n_j+1)) T u(x^*) + 1/(n_j+1) \sum_{k=2}^{n_j+1} \langle T^k u, \delta_{x^*} \rangle. \tag{34}$$

The first term obviously tends to 0 as j tends to infinity. To prove that the second term tends to $\langle Tu, \nu \rangle$ as j tends to infinity we do as follows. Set $Tu = u_1$. Since Q is Lipschitz continuous and $u \in Lip[K]$, it follows that $u_1 \in Lip[K]$. Now, considering the second term in (34), we find that

$$1/(n_j+1) \sum_{k=2}^{n_j+1} \langle T^k u, \delta_{x^*} \rangle = 1/(n_j+1) \sum_{k=1}^{n_j} \langle T^k u_1, \delta_{x^*} \rangle =$$

$$n_j/(n_j+1) \langle T^{(n_j)} u_1, \delta_{x^*} \rangle,$$

which tends to $\langle u_1, \nu \rangle = \langle Tu, \nu \rangle = \langle u, \nu Q \rangle$ as j tends to infinity. Hence, if $u \in Lip[K]$ (33) holds, and since the set of Lipschitz continuous functions is measure determining it follows that (33) holds for $u \in C[K]$ which was what we wanted to prove. Thereby Theorem 12.1 is proved. □

We shall next state another auxiliary theorem which is based on a property similar to the shrinking property.

Definition 12.2 *Let Q be a tr.pr.f on $(K,\mathcal{E})$, and let T be the associated transition operator. If for every $\rho > 0$ there exists a number α, $0 < \alpha < 1$ and an integer N such that if the integer $n \geq N$ then for all $u \in Lip[K]$*

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

then we say that Q has the **strong shrinking property***. We call α a* **shrinking number** *associated to ρ.*

Theorem 12.2 *Let $(K,\mathcal{E})$ be a complete, separable,* **bounded***, metric space with metric δ, let Q be a tr.p.f on $(K,\mathcal{E})$ and suppose that Q has the* **strong shrinking property***. Suppose also that Q is* **Lipschitz equicontinuous***. Then, there exists a probability measure $\nu \in \mathcal{P}(K,\mathcal{E})$ such that Q is* **weakly ergodic** *with* **stationary measure** *ν.*

Proof. Let as usual T denote the transition operator associated to Q. We shall first prove that

$$\lim_{n\to\infty} \sup\{osc(T^n u) : u \in Lip_1[K]\} = 0. \tag{35}$$

Let $\epsilon > 0$ be given. Choose $\rho = \epsilon$. From the strong shrinking property follows that we can find a number $\alpha > 0$ and an integer N such that if $u \in Lip_1[K]$ and $n > N$, then

$$osc(T^n u) \leq \epsilon\alpha + (1-\alpha)osc(T^{n-N}u). \tag{36}$$

Now define

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

where thus $D = \max\{\delta(x,y) : x, y \in K\}$. Then, if $n > NM$, it follows from (36) that

$$osc(T^n u) \leq \epsilon\alpha + (1-\alpha)osc(T^{n-N}u) \leq \epsilon\alpha + (1-\alpha)(\epsilon\alpha + (1-\alpha)osc(T^{n-2N}u)) \leq ... <$$

$$\epsilon\alpha(1/(1-(1-\alpha))) + (1-\alpha)^M D < 2\epsilon$$

if $u \in Lip_1[K]$ and, since ϵ is arbitrarily chosen, (35) follows.

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

$$\sup\{|\int_K u(y)Q^n(x_0,dy) - \int_K u(y)Q^{n+m}(x_0,dy)| : u \in Lip_1[K]\} < \epsilon, \tag{37}$$

if $n \geq N$.

Thus, let $\epsilon > 0$ be given and also the integer $m \geq 1$ be given. Then, if $u \in Lip_1[K]$ we find, for $n = 1, 2, ...$, that

$$|\int_K u(y)Q^n(x_0,dy) - \int_K u(y)Q^{n+m}(x_0,dy)| =$$

$$|\langle u, \delta_{x_0}Q^n\rangle - \langle u, \delta_{x_0}Q^{n+m}\rangle| = |\langle u, \delta_{x_0}Q^n\rangle - \langle u, \delta_{x_0}Q^m Q^n\rangle| =$$

$$|\langle u, \delta_{x_0}Q^n\rangle - \langle u, \nu_{x_0}Q^n\rangle| = |\langle T^n u, \delta_{x_0}\rangle - \langle T^n u, \nu_{x_0}\rangle| =$$

$$|T^n u(x_0) - \int_K T^n u(y)\nu_{x_0}(dy)| \leq \int_K |T^n u(x_0) - T^n u(y)|\nu_{x_0}(dy)$$

where we have defined $\nu_{x_0} \in \mathcal{P}(K,\mathcal{E})$ by $\nu_{x_0} = \delta_{x_0} Q^m$. Hence

$$|T^n u(x_0) - T^{n+m} u(x_0)| \leq \int_K |T^n u(x_0) - T^n(y)| \nu_{x_0}(dy). \tag{38}$$

But from the limit relation (35) it follows that we can find an integer N, which is independent of the integer m, such that for any $u \in Lip_1[K]$ and all $y \in K$

$$|T^n u(x_0) - T^n u(y)| < \epsilon, \ \ n \geq N.$$

Together with the estimate (38) this implies that (37) holds for all $n \geq N$. From the definition of the Kantorovich distance it follows that

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

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

Since we have assumed that $(K,\mathcal{E})$ is a complete, separable, metric space it follows that $(\mathcal{P}(K,\mathcal{E}), \mathcal{G}, d_K)$ is also a complete, separable, metric space, if we let $\mathcal{G}$ denote the Borel field generated by the Kantorovich metric d_K. (See e.g [8], Corollary 11.5.5 and Theorem 11.8.2.)

Therefore it follows that 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 $\lim_{n\to\infty} \sup\{osc(T^n u) : u \in Lip_1[K]\} = 0$ because of (35), it now also follows that

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

which implies that

$$\lim_{n\to\infty} \int_K u(y) Q^n(x,dy) - \int_K u(y)\mu(dy) = 0, \ \ \forall\, x \in K \tag{39}$$

and for all $u \in Lip_1[K]$. But if (39) holds for all $u \in Lip_1[K]$, it also holds for all $u \in Lip[K]$ and, since $Lip[K]$ is dense in $C_{uniform}[K]$, it follows that (39) holds for all $u \in C_{uniform}[K]$ and finally it follows that (39) holds for all $u \in C[K]$ because of Theorem 6.1, Chapter I of [21]. Hence Q is weakly ergodic with stationary measure μ. That μ is the only invariant probability measure of Q is also easily proved by using (35) and (39). □.

13 A result on Lipschitz equicontinuity.

In this section we state and prove an important inequality.

Lemma 13.1 *Let $\mathcal{H} = \{(S,\mathcal{F},\delta_0),(p,\lambda),(A,\mathcal{A},\varrho),(m,\tau)\}$ be a HMM with densities such that the partition (m,τ) is regular. Let $\mathbf{P}$ be the filter kernel induced by the partition (m,τ), and let $\mathbf{T}$ denote the transition operator associated to $\mathbf{P}$. Then*

$$\gamma(\mathbf{T}^n u) \leq 3\gamma(u), \ \ n = 1, 2...., \ \forall u \in Lip[K] \tag{40}$$

and hence $\mathbf{P}$ is both Lipschitz-continuous and Lipschitz equicontinuous.

Proof. The proof is similar to the proof of Lemma 4.3 in [13] and the proof of Lemma 3.5 in [16].

We shall first show that

$$|\mathbf{T}^n u(x) - \mathbf{T}^n u(y)| \leq (||u|| + 2\gamma(u))||x - y|| \tag{41}$$

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

Recall that for $x \in K$ the set A_x^+ is defined as $A_x^+ = \{a : ||xM_a|| > 0\}$ where thus M_a, for each $a \in A$, is the tr.f in $\mathcal{TQ}((S, \mathcal{F}))$ defined by

$$M_a(s, F) = \int_F m(s, t, a)\lambda(dt)$$

and

$$||xM_a|| = \int_S M_a(s, S)x(ds).$$

Recall also that $\mathbf{T} : B[K] \to B[K]$ is defined by

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

Next let us note that if $x, y \in K$ and $a \in A$ then

$$|(||xM_a|| - ||yM_a||)| \leq ||xM_a - yM_a|| = ||(x - y)M_a||. \tag{42}$$

Furthermore, if x and y in K, and f and g in $\mathcal{L}_\lambda^1[S]$ are representatives of x and y respectively, we find from the definition of a partition that

$$\int_A ||(x - y)M_a||\tau(da) = \int_A \int_S |f(s) - g(s)|m(s, S, a)\lambda(ds)\tau(da) =$$

$$\int_S |f(s) - g(s)|p(s, S)\lambda(ds) = \int_S |f(s) - g(s)|\lambda(ds) = ||x - y||. \tag{43}$$

We shall later also have use of the following inequality which is an immediate consequence of Lemma 2.1.

Proposition 13.1 *Let $x, y \in K$ and $a \in A$ be such that $||xM_a|| > 0$ and $||yM_a|| > 0$. Then*

$$\|\frac{xM_a}{\|xM_a\|} - \frac{yM_a}{\|yM_a\|}\| \leq \frac{2\|xM_a - yM_a\|}{\|xM_a\|}. \quad \Box$$

Now let $x, y \in K$ and $u \in Lip[K]$. Define $B \subset A$ by

$$B = \{a \in A : ||xM_a|| > 0, ||yM_a|| > 0\}$$

and define B_1 and B_2 by

$$B_1 = A_x^+ \setminus B, \;\; B_2 = A_y^+ \setminus B.$$

Obviously B, B_1, B_2 are disjoint sets.

We can now express $\mathbf{T}u(x)$ as

$$\mathbf{T}u(x) = \int_B u(\frac{xM_a}{||xM_a||})||xM_a||\tau(da) + \int_{B_1} u(\frac{xM_a}{||xM_a||})||xM_a||\tau(da)$$

and $\mathbf{T}u(y)$ as

$$\mathbf{T}u(y) = \int_B u(\frac{yM_a}{||yM_a||})||yM_a||\tau(da) + \int_{B_2} u(\frac{yM_a}{||yM_a||})||yM_a||\tau(da).$$

Hence

$$|\mathbf{T}u(x) - \mathbf{T}u(y)| \leq$$

$$|\int_B (u(\frac{xM_a}{||xM_a||})||xM_a|| - u(\frac{yM_a}{||yM_a||})||yM_a||)\tau(da)|+$$

$$||u||\int_{B_1} ||xM_a||\tau(da) + ||u||\int_{B_2} ||yM_a||\tau(da).$$

Estimating the first integral we obtain

$$|\int_B (u(\frac{xM_a}{||xM_a||})||xM_a|| - u(\frac{yM_a}{||yM_a||})||yM_a||)\tau(da)| =$$

$$|\int_B (u(\frac{xM_a}{||xM_a||}) - u(\frac{yM_a}{||yM_a||}))||xM_a|| + u(\frac{yM_a}{||yM_a||})(||xM_a|| - ||yM_a||)\tau(da)| \leq$$

$$\gamma(u)\int_B \|\frac{xM_a}{||xM_a||} - \frac{yM_a}{||yM_a||}\| \cdot ||xM_a||\tau(da) + ||u||\int_B ||xM_a - yM_a||\tau(da).$$

Hence, using Proposition 13.1, the inequality (42) and the equality (43), we obtain

$$|\mathbf{T}u(x) - \mathbf{T}u(y)| \leq$$

$$2\gamma(u)\int_B \|xM_a - yM_a\|\tau(da) + ||u||\int_B \|xM_a - yM_a\|\tau(da)+$$

$$||u||\int_{B_1} ||xM_a - yM_a||\tau(da) + ||u||\int_{B_2} ||yM_a - xM_a||\tau(da) \leq$$

$$2\gamma(u)\int_A \|(x-y)M_a\|\tau(da) + ||u||\int_A \|(x-y)M_a\|\tau(da) = (2\gamma(u) + ||u||)||x-y||$$

and thereby the inequality (41) is proved.

From (41) it immediately follows that

$$\gamma(\mathbf{T}u) \leq 2\gamma(u) + ||u|| \tag{44}$$

for all $u \in Lip[K]$. Now, if $\gamma(u) = 0$, then u is a constant which implies that also $\mathbf{T}u$ is constant and then (40) holds trivially. If instead $\gamma(u) > 0$, set $v = u/\gamma(u)$ and define $v_0 = v - osc(v)/2 - \inf\{v(x) : x \in K\}$. Clearly $\gamma(v) = \gamma(v_0) = 1$. Since $\sup\{||x - y|| : x, y \in K\} = 2$, it is also clear that $||v_0|| \leq 1$. Hence, by (44) follows that

$$\gamma(\mathbf{T}v_0) \leq 3$$

and hence

$$\gamma(\mathbf{T}u) \leq 3\gamma(u) \tag{45}$$

since $\gamma(\mathbf{T}u) = \gamma(u)\gamma(\mathbf{T}v_0)$ and thereby the inequality in (40) is proved for $n = 1$.

That the inequality (40) also holds for $n \geq 2$ is an immediate consequence of Corollary 5.1 and the fact that the right hand side of (45) is independent of the choice of the partition (m, τ).

14 On the Kantorovich distance between sets with different barycenters

As usual, let $(S, \mathcal{F})$ be a measurable space with metric δ_0, let λ denote a σ-finite, nonnegative measure on $(S, \mathcal{F})$ and set $K = \mathcal{P}_\lambda(S, \mathcal{F})$. Also, as usual, let δ_{TV} denote the metric on K induced by the total variation and let $\mathcal{E}$ denote the $\sigma - algebra$ generated by δ_{TV}. Instead of writing $\delta_{TV}(x, y)$ we shall in this section 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 $Q^r(K, \mathcal{E})$ 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 $\mathcal{P}(K, \mathcal{E})$ (see Section 7). Recall that the Kantorovich distance on $\mathcal{P}(K, \mathcal{E})$ has two equivalent definitions namely either by the formula (19) or by the formula (20).

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

$$d_K(\mu, \nu) = r d_K(\mu/r, \nu/r), \ \mu, \nu \ \in \mathcal{Q}^r(K, \mathcal{E}). \tag{46}$$

Next, let $\mathcal{P}(K|x)$ denote the set of probability measures on $(K, \mathcal{E})$ for which the barycenter is equal to x. (See Section 8.) For $\mu \in \mathcal{Q}^r(K, \mathcal{E})$ we also define a barycenter $\bar{b}(\mu)$ simply by

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

We let $\mathcal{Q}^r(K|x)$ denote the set of measures in $\mathcal{Q}^r(K, \mathcal{E})$ which have barycenter equal to x.

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

Theorem 14.1 *Let $r > 0$, let $x, y \in K$ and let $\mu \in \mathcal{Q}^r(K|x)$. Then*

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

Proof. In the paper [16] the conclusion of this theorem is proved when the underlying space S is a denumerable set. We shall here essentially copy the proof given in [16], section 5.

Note that if $\mu, \nu \in \mathcal{Q}^r(K, \mathcal{E})$ then

$$d_K(\mu, \nu) = \sup\{\int_K u(y)\mu(dy) - \int_K u(y)\nu(dy) : \ u \in Lip_1[K], ||u|| \leq 1 \ \}$$

since $\sup\{||x - y|| : x, y \in K\} = 2$.

As mentioned above $d_K(\cdot, \cdot)$ determines a metric on $\mathcal{P}(K, \mathcal{E})$ and from (46) follows easily that d_K determines a metric also on $\mathcal{Q}^r(K, \mathcal{E})$ for any $r > 0$. Also in this case we call the metric $d_K(\cdot, \cdot)$ the *Kantorovich distance.*

From the definition (19) of $d_K(\cdot, \cdot)$ it readily follows that

$$d_K(\delta_x, \delta_y) = \delta_{TV}(x, y) = ||x - y||,$$

where thus δ_x and δ_y denote the Dirac measures at x and y respectively.

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

Lemma 14.1 *Let* $\mu, \nu \in \mathcal{Q}^r(K, \mathcal{E})$. *Then*

$$||\bar{b}(\mu) - \bar{b}(\nu)|| \leq d_K(\mu, \nu).$$

Proof. The conclusion of the lemma is trivially true if $\bar{b}(\mu) = \bar{b}(\nu)$. We thus assume that $\bar{b}(\mu) \neq \bar{b}(\nu)$. From the definitions of the Kantorovich distance in $\mathcal{Q}^r(K, \mathcal{E})$ and the barycenter of a measure in $\mathcal{Q}^r(K, \mathcal{E})$, it is easily seen, that it suffices to prove the inequality if $\mu, \nu \in \mathcal{P}(K, \mathcal{E})$.

Thus, let $\mu, \nu \in \mathcal{P}(K, \mathcal{E})$ and set

$$x = \bar{b}(\mu), \quad and \quad y = \bar{b}(\nu).$$

Let $F_0 \in \mathcal{F}$ be such that

$$x(F_0 \cap F) \geq y(F_0 \cap F), \quad \forall F \in \mathcal{F},$$

and define the function $J_{F_0} : S \to [-1, 1]$ by

$$J_{F_0}(s) = 1, \quad if \quad s \in F_0$$

$$J_{F_0}(s) = -1 \quad if \quad s \in S \setminus F_0.$$

Hence

$$J_{F_0} = I_{F_0} - I_{S \setminus F_0}, \tag{47}$$

where thus I_{F_0} denotes the indicator function of the set F_0.

Next, define $v \in B[K, \mathcal{E}]$ by

$$v(z) = \langle J_{F_0}, z \rangle.$$

Since $osc(J_{F_0}) \leq 2$, it follows from Proposition 2.2, that

$$|v(z_1) - v(z_2)| = |\langle J_{F_0}, z_1 \rangle - \langle J_{F_0}, z_2 \rangle| \leq osc(J_{F_0})||z_1 - z_2||/2 \leq ||z_1 - z_2||$$

and hence $v \in Lip_1[K, \mathcal{E}]$. From the definition of the Kantorovich distance it then follows, that

$$d_K(\mu, \nu) \geq |\int_K v(z)\mu(dz) - \int_K v(z)\nu(dz)| \tag{48}$$

and from the definition of the barycenter and (47), it then follows, that

$$|\int_K v(z)\mu(dz) - \int_K v(z)\nu(dz)| = |\int_K \langle J_{F_0}, z \rangle \mu(dz) - \int_K \langle J_{F_0}, z \rangle \nu(dz)| =$$

$$|\int_K (\langle I_{F_0}, z \rangle - \langle I_{S \setminus F_0}, z \rangle)\mu(dz) - \int_K (\langle I_{F_0}, z \rangle - \langle I_{S \setminus F_0}, z \rangle)\nu(dz)| =$$

$$|\langle I_{F_0}, \bar{b}(\mu) \rangle - \langle I_{S \setminus F_0}, \bar{b}(\mu) \rangle - \langle I_{F_0}, \bar{b}(\nu) \rangle + \langle I_{S \setminus F_0}, \bar{b}(\nu) \rangle| =$$

$$|x(F_0) - y(F_0) + y(S \setminus F_0) - x(S \setminus F_0)| = ||x - y|| = ||\bar{b}(\mu) - \bar{b}(\nu)||,$$

which together with (48) implies that

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

which was what we wanted to prove. $\square$

We now continue our proof of Theorem 14.1 by proving that, if the measure $\mu \in \mathcal{Q}(K,\mathcal{E})$ is a weighted finite sum of Dirac measures, then for every $y \in K$ we can find a measure $\nu \in \mathcal{Q}(K,\mathcal{E})$ such that $\mu(K) = \nu(K)$, $\overline{b}(\nu) = y$ and

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

As usual, if ξ denotes an arbitrary element in K, we denote by δ_ξ the Dirac measure at ξ.

Lemma 14.2 *Let N be a positive integer and let $\xi_k,\ k = 1,2,...,N$ be elements in K. Let $\beta_k > 0,\ k = 1,2,...,N$, define the measure $\varphi \in \mathcal{Q}(K,\mathcal{E})$ by*

$$\varphi = \sum_{k=1}^{N} \beta_k \delta_{\xi_k}$$

and define the element $a \in \mathcal{Q}_\lambda(S,\mathcal{F})$ by

$$a = \sum_{k=1}^{N} \beta_k \xi_k.$$

Let $b \in \mathcal{Q}_\lambda(S,\mathcal{F})$ be an element satisfying

$$||b|| = ||a||.$$

Then, there exist elements $\zeta_k,\ \ k = 1,2,...,N,$ in K such that

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

and such that, if we define

$$\Psi = \sum_{k=1}^{N} \beta_k \delta_{\zeta_k},$$

then

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

Proof. Let us first note that, if $\psi \in \mathcal{Q}(K,\mathcal{E})$ is defined by

$$\psi = \sum_{k=1}^{N} \beta_k \delta_{\zeta_k},$$

where $\beta_k,\ k = 1,2,...,N$ are real positive numbers and $\zeta_k,\ \ k = 1,2,...,N,$ belong to K, then

$$\overline{b}(\psi) = \sum \beta_k \zeta_k. \tag{49}$$

This follows easily 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 $\zeta_1, \zeta_2, ..., \zeta_N$ denote an arbitrary set of elements 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

$$\tilde{\varphi}(\{(\xi_k, \zeta_k)\}) = \beta_k,\ k = 1, 2, ..., N.$$

Then clearly

$$\tilde{\varphi}(A \times K) = \varphi(A),\ \forall A \in \mathcal{E},$$

and

$$\tilde{\varphi}(K \times A) = \theta(A),\ \forall A \in \mathcal{E},$$

from which follows that the Kantorovich distance $d_K(\varphi, \theta)$ satisfies

$$d_K(\varphi, \theta) \leq \sum_{k=1}^{N} \beta_k ||\xi_k - \zeta_k||, \tag{50}$$

since clearly

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

By combining (50) and (49) with Lemma 14.1, it follows, that in order to prove Lemma 14.2, it suffices to find probability measures $\zeta_k,\ \ k = 1, 2, ..., N$ belonging to K, such that

$$b = \sum_{k=1}^{N} \beta_k \zeta_k \tag{51}$$

and also

$$\sum_{k=1}^{N} \beta_k ||\xi_k - \zeta_k|| = ||a - b||. \tag{52}$$

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 we want it to be. The case when $b = a$ is also trivial. Just take $\zeta_k = \xi_k,\ \ k = 1, 2, ..., N$. In the remaining part of the proof we therefore assume that $a \neq b$.

We shall now prove, that we can find probability measures $\zeta_k,\ \ k = 1, 2, ..., N,\ \in K$ such that (51) and (52) hold, by induction. Thus, let us assume, that, if $N = M - 1$, where $M \geq 2$, then, if $a = \sum_{k=1}^{N} \beta_k \xi_k$ where $\beta_k > 0,\ k = 1, 2, ...N$, and $\xi_k \in K, k = 1, 2, ..., N$, and $b \in \mathcal{Q}_\lambda(S, \mathcal{F})$ satisfies $||a|| = ||b||$, we can find $\zeta_k, k = 1, 2, ..., N$ in K such that (51) and (52) hold.

Now, let $N = M$, set $r = ||a||$, let $\beta_k > 0,\ k = 1, 2, ...M$, let $\xi_k \in K, k = 1, 2, ..., M$ and let $b \in \mathcal{Q}_\lambda^r(S, \mathcal{F})$. Define

$$a = \sum_{k=1}^{M} \beta_k \xi_k.$$

Our aim is thus to find elements $\zeta_k, k = 1, 2, ..., M$ in K, such that

$$b = \sum_{k=1}^{M} \beta_k \zeta_k \tag{53}$$

and also

$$\sum_{k=1}^{M} \beta_k ||\xi_k - \zeta_k|| = ||a - b||. \tag{54}$$

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

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

Let us also define $a_1 \in \mathcal{Q}_\lambda(S, \mathcal{F})$ by

$$a_1 = \sum_{k=1}^{M-1} \beta_k \xi_k.$$

Clearly $||a_1|| = ||a|| - \beta_M$.

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

$$b_1 = b - \beta_M \zeta_M \tag{55}$$

then

1)

$$b_1 \in \mathcal{Q}_\lambda(S, \mathcal{F}) \tag{56}$$

2)

$$||a - b|| = ||a_1 - b_1|| + \beta_M ||\xi_M - \zeta_M||. \tag{57}$$

From (56) and the definition of b_1 it then follows that $||b_1|| = ||b|| - \beta_M = ||a|| - \beta_M = ||a_1||$ and then, using the induction hypothesis, it follows that we can find probability measures $\zeta_k, k = 1, 2, ..., M - 1$ such that

$$b_1 = \sum_{k=1}^{M-1} \beta_k \zeta_k$$

and

$$\sum_{k=1}^{M-1} \beta_k ||\xi_k - \zeta_k|| = ||a_1 - b_1|| \tag{58}$$

and consequently, by using (57) and (58), it follows that

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

and hence (53) and (54) hold with $N = M$.

It thus remains to determine a vector $\zeta_M \in K$ such that, if we define b_1 by (55), then (56) and (57) hold. In order to do this we proceed as follows.

First, let the set $F_1 \in \mathcal{F}$ be such that

$$a(F_1 \cap F) \geq b(F_1 \cap F),\ \forall F \in \mathcal{F}$$

and define

$$F_2 = S \setminus F_1.$$

Let us write

$$\mathcal{F}_1 = \{F \in \mathcal{F} : F \subset F_1\}$$

and

$$\mathcal{F}_2 = \{F \in \mathcal{F} : F \subset F_2\}.$$

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

$$c(F) = ((a - a_1) \wedge (a - b))(F \cap F_1),\ F \in \mathcal{F} \tag{59}$$

and set

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

We now define ζ_M as follows:

$$\zeta_M(F) = \xi_M(F) - c(F)/\beta_M,\ \ if\ F \in \mathcal{F}_1$$

$$\zeta_M(F) = \xi_M(F) + (\Delta_0/\Delta)(b(F) - a(F))/\beta_M,\ \ if\ F \in \mathcal{F}_2.$$

We want to show 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 (59)), 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 $\zeta_M \in \mathcal{Q}_\lambda(S, \mathcal{F})$.

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

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

and

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

since $b(F_2) - a(F_2) = \Delta$, we find that it also follows that $\zeta_M(S) = \zeta_M(F_1) + \zeta_M(F_2) = \xi_M(F_1) - \Delta_0/\beta_M + \xi_M(F_2) + \Delta_0/\beta_M = \xi_M(S) = 1$, and hence $\zeta_M \in K$.

We also find that

$$||\xi_M - \zeta_M|| = \xi_M(F_1) - \zeta_M(F_1) + \zeta_M(F_2) - \xi_M(F_2) =$$

$$c(F_1)/\beta_M + c(F_1)/\beta_M = 2\Delta_0/\beta_M. \tag{60}$$

Next, if b_1 is defined by (55), we find, that if $F \in \mathcal{F}_1$, then

$$b_1(F) = b(F) - \beta_M \xi_M(F) + c(F) = b(F) - a(F) + a_1(F) + c(F) =$$

$$b(F) + a_1(F) + ((a - a_1) \wedge (a - b))(F) - a(F)$$

$$= b(F) + a_1(F) - (a_1 \vee b)(F) \geq 0,$$

and, if $F \in \mathcal{F}_2$, then

$$b_1(F) = b(F) - \beta_M \xi_M(F) - (b(F) - a(F))\Delta_0/\Delta$$

$$\geq b(F) - a(F) + a_1(F) - (b(F) - a(F) \geq a_1(F).$$

Hence (56) is satisfied.

It thus remains to show that (57) 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

$$||a_1 - b_1|| = a_1(F_1) - b_1(F_1) + b_1(F_2) - a_1(F_2) =$$

$$a(F_1) - \beta_M \xi_M(F_1) - b(F_1) + \beta_M \xi_M(F_1) + c(F_1) +$$

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

$$a(F_1) - b(F_1) + \Delta_0 + b(F_2) - a(F_2) + \Delta_0 = 2\Delta + 2\Delta_0$$

and since $||a - b|| = 2\Delta$ and $\beta_M ||\xi_M - \zeta_M|| = 2\Delta_0$ because of (60), the equality (57) holds and thereby the proof of the lemma is completed. $\square$

Using Lemma 14.2 and Lemma 14.1 it is now easy to conclude the proof of Theorem 14.1. Thus let $x, y \in K$ and suppose $\mu \in \mathcal{Q}^r(K|x)$. What we want to prove is that to every $\epsilon > 0$ we can find a measure $\nu \in \mathcal{Q}^r(K|y)$ such that

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

Thus, let $\epsilon > 0$ be given. 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 thus $\xi_k, k = 1, 2, ...N$ belong to K and $\beta_k > 0$ for $k = 1, 2, ..., N$. From Lemma 14.1 now follows, that we have

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

and from Lemma 14.2 follows, that we can find a measure $\nu \in \mathcal{Q}^r(K|y)$, such that

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

From the triangle inequality now follows, that

$$d_K(\mu, \nu) \leq d_K(\mu, \mu_1) + d_K(\mu_1, \nu) < \epsilon/2 + ||\bar{b}(\mu_1) - y|| \leq$$

$$\epsilon/2 + ||\bar{b}(\mu_1) - x|| + ||x - y|| \leq \epsilon/2 + \epsilon/2 + ||x - y||.$$

Hence,

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

and thereby Theorem 14.1 is proved. $\square$

15 Verifying the shrinking property.

From Theorem 12.1 it follows, that in order to prove the first part of Theorem 11.1, it suffices to prove, that the filter kernel **P** is Lipschitz equicontinuous and has the shrinking property.

That **P** is Lipschitz equicontinuous follows from Lemma 13.1. It thus remains to verify, that **P** has the shrinking property. This we can do by using arguments similar to those given in [16], section 8.

Thus, let $\rho > 0$ be given. What we have to do is to prove, that we can find a number $\alpha > 0$, such that for each nonempty, compact set $E \in \mathcal{E}$, each $\eta > 0$ and each $\kappa > 0$, we can find an integer N and another nonempty, compact set $F \in \mathcal{E}$, such that for each $u \in Lip[K]$

$$osc_E(\mathbf{T}^n u) \leq \eta\gamma(u) + \kappa osc(u) + \alpha\rho\gamma(u) + (1-\alpha)osc_F(\mathbf{T}^{n-N}u). \qquad (61)$$

Now, let also E, η and κ be given, where thus $E \in \mathcal{E}$ is a nonempty, compact set, $\eta > 0$ and $\kappa > 0$. Since E is a compact set, we can find a finite set $\mathcal{M} = \{x_i,\ i = 1, 2, ...M\}$ of points belonging to K, such that for every $x \in E$ there exists an element $x_j \in \mathcal{M}$ such that $\delta_{TV}(x, x_j) < \eta/12$.

Next, let x_i and x_j be two arbitrary elements in the set $\mathcal{M}$. We shall first prove that we can find a number α depending on ρ, an integer N depending on the compact set E and on ρ, and a compact set $F = F(x_i, x_j)$ belonging to $\mathcal{E}$ depending on x_i and x_j such that, for each $u \in Lip[K]$,

$$|\mathbf{T}^n u(x_i) - \mathbf{T}^n u(x_j)| \leq$$

$$\eta\gamma(u)/2 + \kappa osc(u) + \alpha\rho\gamma(u) + (1-\alpha)osc_{F(x_i,x_j)}(\mathbf{T}^{n-N}u) \qquad (62)$$

if $n \geq N$.

This we shall do in three steps. In the first step we choose N_1 so large that, if $n \geq N_1$, then the barycenter of $\mathbf{P}^n(x_i, \cdot)$ as well as the barycenter of $\mathbf{P}^n(x_j, \cdot)$ are close to the stationary measure π. This we can do because of Lemma 8.1. After that we use Theorem 14.1 to find two measures belonging to $\mathcal{P}(K|\pi)$ which are close to the measures $\mathbf{P}^n(x_i, \cdot)$ and $\mathbf{P}^n(x_j, \cdot)$ respectively thereby preparing us for the use of Condition E.

In the second step we use Condition E in order to determine a number $\alpha > 0$. This number will only depend on ρ and not on x_i or x_j. In this second step we also introduce another integer N_2 (only dependent of α) and a compact set $F = F(x_i, x_j)$ depending on κ, α, x_i and x_j.

In the third step we make the necessary estimations in order to verify that (62) holds if $n \geq N_1 + N_2$.

Step 1. Since the Markov kernel P determined by the HMM $\mathcal{H}$ is strongly ergodic with stationary measure π, it follows that for every $z \in E$ we can find an integer $N_0(z)$ such that

$$||zP^n - \pi|| < \eta/12,\ if\ n > N_0(z).$$

Since $||z_1P^n - z_2P^n|| \leq ||z_1 - z_2||$ for all $n \geq 1$ and all $z_1, z_2 \in K$ and also E is assumed to be a non-zero compact set, it follows, that we can find an integer N_1, such that

$$\sup_{z\in E} ||zP^n - \pi|| < \eta/12,\ if\ n \geq N_1.$$

Now, let $x \in \mathcal{M}$ and $y \in \mathcal{M}$ be given. Set

$$\mu_{n,x}(\cdot) = \mathbf{P}^n(x, \cdot), \ \ and \ \ \mu_{n,y}(\cdot) = \mathbf{P}^n(y, \cdot), \ \ n = 1, 2, ... \, .$$

From Lemma 8.1 it follows, that $\bar{b}(\mu_{n,x}) = xP^n$ and that $\bar{b}(\mu_{n,y}) = yP^n$. Therefore, if $n \geq N_1$, we conclude that

$$||\bar{b}(\mu_{n,x}) - \pi|| < \eta/12 \ \ and \ \ ||\bar{b}(\mu_{n,y}) - \pi|| < \eta/12.$$

From Theorem 14.1 now follows, that we can find two measures ν_x and ν_y, both in $\mathcal{P}(K|\pi)$, such that, if $u \in Lip[K]$, then

$$|\int_K u(z)\mu_{N_1,x}(dz) - \int_K u(z)\nu_x(dz)| \leq \gamma(u)\eta/12$$

and

$$|\int_K u(z)\mu_{N_1,y}(dz) - \int_K u(z)\nu_y(dz)| \leq \gamma(u)\eta/12.$$

From Lemma 13.1 (the Lipschitz equicontinuity property) we also find that, for $m = 0, 1, 2, ...,$

$$|\int_K \mathbf{T}^m u(z)\mu_{N_1,x}(dz) - \int_K \mathbf{T}^m u(z)\nu_x(dz)| \leq 3\gamma(u)\eta/12 = \gamma(u)\eta/4$$

and that

$$|\int_K \mathbf{T}^m u(z)\mu_{N_1,y}(dz) - \int_K \mathbf{T}^m u(z)\nu_y(dz)| \leq 3\gamma(u)\eta/12 = \gamma(u)\eta/4.$$

Thus, if $n \geq N_1$, we have

$$|\mathbf{T}^n u(x) - \mathbf{T}^n u(y)| \leq \eta\gamma(u)/2+$$

$$|\int_K \mathbf{T}^{n-N_1} u(z)\nu_x(dz) - \int_K \mathbf{T}^{n-N_1} u(z)\nu_y(dz)|. \tag{63}$$

Step 2. Recall that ρ was given. Let us define the set $D \subset K$ by

$$D = \{(z, w) \in K^2 : \delta_{TV}(z, w) < \rho/3\}.$$

Since the filter kernel $\mathbf{P}$ is assumed to satisfy Condition E and both ν_x and ν_y belong to $\mathcal{P}(K|\pi)$, we can find an integer N_2 - independent of x and y -, a constant $\alpha_1 > 0$ and a coupling $\tilde{\nu}^*_{x,y}$ of $\nu_x\mathbf{P}^{N_2}$ and $\nu_y\mathbf{P}^{N_2}$ such that

$$\tilde{\nu}^*_{x,y}(D) \geq \alpha_1. \tag{64}$$

Our choice of shrinking number α associated to the given number ρ will be

$$\alpha = \alpha_1/2.$$

Next, since $(K, \mathcal{E})$ is a complete, separable metric space, we can find a compact set $F(x, y) \in \mathcal{E}$ depending on $x, y \in \mathcal{M}$, such that

$$\tilde{\nu}^*_{x,y}(F(x, y) \times F(x, y)) > \max\{(1 - \alpha_1)/2, (1 - \kappa)\}. \tag{65}$$

From (65) and (64) then follows that

$$\tilde{\nu}^*_{x,y}(\{(z,z') : (z,z') \in D \cap F(x,y) \times F(x,y))) \geq$$

$$\tilde{\nu}^*_{x,y}(D) - (1 - \tilde{\nu}^*_{x,y}(F(x,y) \times F(x,y)) \geq \alpha_1 - \alpha_1/2 = \alpha_1/2 = \alpha. \tag{66}$$

This concludes step 2.

Step 3. Set $N = N_1 + N_2$. To simplify notations we write $F = F_{x,y}$. In this last step we shall estimate

$$|\int_K \mathbf{T}^{n-N_1} u(z)\nu_x(dz) - \int_K \mathbf{T}^{n-N_1} u(z)\nu_y(dz)|$$

when $n \geq N$.

Set $m = n - N$ and $v = \mathbf{T}^m u$. Then $n - N_1 = m + N_2$. Hence

$$|\int_K \mathbf{T}^{n-N_1} u(z)\nu_x(dz) - \int_K \mathbf{T}^{n-N_1} u(z)\nu_y(dz)| =$$

$$|\int_K \mathbf{T}^{m+N_2} u(z)\nu_x(dz) - \int_K \mathbf{T}^{m+N_2} u(z)\nu_y(dz)| = |\langle \mathbf{T}^m u, \nu_x \mathbf{P}^{N_2}\rangle - \langle \mathbf{T}^m u, \nu_y \mathbf{P}^{N_2}\rangle| =$$

$$|\int_{K^{(2)}} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz,dz')| \leq$$

$$|\int_{B_1} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz,dz')| + |\int_{B_2} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz,dz')| +$$

$$|\int_{B_3} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz,dz')|, \tag{67}$$

where

$$B_1 = \{(z,z') \in K^2 : \delta_{TV}(z,z') < \rho/3, z \in F, z' \in F\},$$

$$B_2 = \{(z,z') \in K^2 : \delta_{TV}(z,z') \geq \rho/3, z \in F, z' \in F\}, \; and \; B_3 = K^2 \setminus (B_1 \cup B_2).$$

From (65) follows easily, that

$$|\int_{B_3} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz,dz')| \leq osc(v)(1 - \kappa) \tag{68}$$

and thereby we have an estimate of the third integral occurring on the right hand side of (67).

To estimate the sum of the other two integrals of the right hand side of (67), we first note that

$$|\int_{B_2} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz,dz')| \leq osc_F(v)\tilde{\nu}^*_{x,y}(B_2) \leq osc_F(v)(1 - \tilde{\nu}^*_{x,y}(B_1)).$$

Next let us note, that, if $0 < a \leq b \leq 1$, $\epsilon > 0$ and $\Theta > 0$, then, by elementary calculations, it follows that

$$b \min\{\epsilon, \Theta\} + (1-b)\Theta \leq a\epsilon + (1-a)\Theta. \tag{69}$$

Since $\tilde{\nu}^*_{x,y}(B_1) \geq \alpha$ because of (66), it follows from (69), that

$$|\int_{B_1} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz,dz')| + |\int_{B_2} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz,dz')| \leq$$

$$\min\{\gamma(v)\rho/3, osc_F(v)\} \cdot \tilde{\nu}^*_{x,y}(B_1) + osc_F(v)(1 - \tilde{\nu}^*_{x,y}(B_1) \leq$$

$$\alpha\gamma(v)\rho/3 + (1-\alpha)osc_F(v),$$

which combined with (68) and (67) implies that

$$|\int_K \mathbf{T}^{n-N_1}u(z)\nu_x(dz) - \int_K \mathbf{T}^{n-N_1}u(z)\nu_y(dz)| \leq$$

$$osc(u)\kappa + \alpha\gamma(u)\rho + (1-\alpha)osc_F(\mathbf{T}^{n-N}u), \tag{70}$$

where we also have used the fact that $\gamma(v) \leq 3\gamma(u)$ because of Lemma 13.1.

By combining (70) and (63) it follows, that we can find an integer N, such that to every pair of points x_i and x_j in $\mathcal{M}$, we can find a set $F(x_i, x_j)$, such that (62) is satisfied.

To verify the shrinking property is now easily done by using (62) together with Lemma 13.1. First, let us define

$$F = \cup_{x_i, x_j \in \mathcal{M}} F(x_i, x_j).$$

Then clearly F is also a compact set and from (62) it obviously follows, that for all $u \in Lip[K]$ and for all $x_i, x_j \in \mathcal{M}$, we have

$$|\mathbf{T}^n u(x_i) - \mathbf{T}^n u(x_j)| \leq$$

$$\eta\gamma(u)/2 + \kappa osc(u) + \alpha\rho\gamma(u) + (1-\alpha)osc_F(\mathbf{T}^{n-N}u), \tag{71}$$

if $n \geq N$.

Now, let z_1 and z_2 be two arbitrary elements in the compact set E, choose $x \in \mathcal{M}$ such that $\delta_{TV}(z_1, x) < \eta/12$ and choose $y \in \mathcal{M}$ such that $\delta_{TV}(z_2, y) < \eta/12$. From Lemma 13.1 follows that for any $u \in Lip[K]$ and all integers $n \geq 1$ we have

$$|\mathbf{T}^n u(z_1) - \mathbf{T}^n u(z_2)| \leq$$

$$|\mathbf{T}^n u(z_1) - \mathbf{T}^n u(x)| + |\mathbf{T}^n u(x) - \mathbf{T}^n u(y)| + |\mathbf{T}^n u(y) - \mathbf{T}^n u(z_2)| \leq$$

$$3\gamma(u)\eta/12 + |\mathbf{T}^n u(x) - \mathbf{T}^n u(y)| + 3\gamma(u)\eta/12 =$$

$$\gamma(u)\eta/2 + |\mathbf{T}^n u(x) - \mathbf{T}^n u(y)|.$$

Hence, if $n \geq N$, we can conclude from (71) that

$$|\mathbf{T}^n u(z_1) - \mathbf{T}^n u(z_2)| \leq$$

$$\eta\gamma(u)/2 + \eta\gamma(u)/2 + \kappa osc(u) + \alpha\rho\gamma(u) + (1-\alpha)osc_F(T^{n-N}u) =$$

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

and, since z_1, z_2 are arbitrary points in E, (61) follows.

Thereby, we have proved, that, if Condition E is satisfied, then the filter kernel $\mathbf{P}$ is weakly ergodic. Moreover, if $\mathbf{P}$ has an invariant measure or if there exists an element $x_0 \in \mathcal{P}_\lambda(S, \mathcal{F})$ such that the sequence $\{\mathbf{P}(x_0, \cdot), n = 1, 2, ...\}$ is a tight sequence, then, by again applying Theorem 12.1, it also follows that there exists a unique invariant probability measure $\nu \in \mathcal{P}(K, \mathcal{E})$ such that $\mathbf{P}$ is weakly ergodic with stationary measure ν.

It thus remains to show, that, if the Markov kernel P is uniformly ergodic, then also in this case there exists a unique, invariant probability measure $\nu \in \mathcal{P}(K, \mathcal{E})$, such that $\mathbf{P}$ is weakly ergodic with stationary measure ν.

Since $\mathbf{P}$ is Lipschitz equicontinuous this follows from auxiliary Theorem 12.2, if we verify that in this case the filter kernel $\mathbf{P}$ not only has the shrinking property but also the *strong* shrinking property.

In order to prove this we proceed in a similar way as when proving the shrinking property.

Thus, let $\rho > 0$ be given. What we have to do is to show, that we can find a positive number $\alpha > 0$ and an integer N such that for all $u \in Lip[K]$

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

Set $\rho_1 = \rho/6$. From Condition E follows, that we can find a real number $\alpha > 0$ and an integer N_1 such that for any two measures μ and ν with barycenters equal to π, there exists a coupling $\tilde{\mu}_{N_1}$, say, of $\mu\mathbf{P}^{N_1}$ and $\nu\mathbf{P}^{N_1}$, such that

$$\tilde{\mu}_{N_1}(D_{\rho_1}) \geq \alpha,$$

where D_{ρ_1} denotes the set $\{(x, y) \in K^2 : \delta_{TV}(x, y) < \rho_1\}$.

Next, since we have assumed that the Markov kernel P is uniformly ergodic, we can find an integer N_0 such that

$$||zP^{N_0} - \pi|| < \alpha\rho/12$$

for all $z \in K$.

Now let x and y be two arbitrary probability measures in K. Set

$$\mu_{n,x}(\cdot) = \mathbf{P}^n(x, \cdot), \;\; and \;\; \mu_{n,y}(\cdot) = \mathbf{P}^n(y, \cdot), \;\; n = 1, 2, \dots .$$

From Lemma 8.1 it follows that the barycenter $\bar{b}(\mu_{n,x})$ satisfies $\bar{b}(\mu_{n,x}) = xP^n$ and that $\bar{b}(\mu_{n,y}) = yP^n$. Therefore, if $n \geq N_0$, we conclude that

$$||\bar{b}(\mu_{n,x}) - \pi|| < \alpha\rho/12 \;\; and \;\; ||\bar{b}(\mu_{n,y}) - \pi|| < \alpha\rho/12.$$

From Theorem 14.1 now follows that we can find two measures ν_x and ν_y, both in $\mathcal{P}(K|\pi)$, such that, if $u \in Lip[K]$, then

$$|\int_K u(z)\mu_{N_0,x}(dz) - \int_K u(z)\nu_x(dz)| < \gamma(u)\alpha\rho/12$$

and

$$|\int_K u(z)\mu_{N_0,y}(dz) - \int_K u(z)\nu_y(dz)| < \gamma(u)\alpha\rho/12.$$

By using Lemma 13.1 (the Lipschitz equicontinuity property), it therefore follows that for $m = 0, 1, 2, ...$

$$|\int_K \mathbf{T}^m u(z)\mu_{N_0,x}(dz) - \int_K \mathbf{T}^m u(z)\nu_x(dz)| \leq 3\gamma(u)\alpha\rho/12 = \gamma(u)\alpha\rho/4$$

and that

$$|\int_K \mathbf{T}^m u(z)\mu_{N_0,y}(dz) - \int_K \mathbf{T}^m u(z)\nu_y(dz)| \leq 3\gamma(u)\alpha\rho/12 = \gamma(u)\alpha\rho/4.$$

Thus, if $n \geq N_0$, we have

$$|\mathbf{T}^n u(x) - \mathbf{T}^n u(y)| \leq$$

$$\gamma(u)\alpha\rho/2 + |\int_K \mathbf{T}^{n-N_0} u(z)\nu_x(dz) - \int_K \mathbf{T}^{n-N_0} u(z)\nu_y(dz)|. \tag{72}$$

Since the filter kernel $\mathbf{P}$ is assumed to satisfy Condition E and both ν_x and ν_y belong to $\mathcal{P}(K|\pi)$, it follows from the definition of the number $\alpha > 0$ and the definition of the integer N_1 made above, that we can find a coupling $\tilde{\nu}^*_{x,y}$ of $\nu_x \mathbf{P}^{N_1}$ and $\nu_y \mathbf{P}^{N_1}$ such that

$$\tilde{\nu}^*_{x,y}(D_{\rho_1}) \geq \alpha.$$

Therefore, if $m \geq N_1$ and we denote $D_{\rho_1} = D$, we find

$$|\int_K \mathbf{T}^m u(z)\nu_x(dz) - \int_K \mathbf{T}^m u(z)\nu_y(dz)| =$$

$$|\int_K \mathbf{T}^{m-N_1} u(z)\nu_x \mathbf{P}^{N_1}(dz) - \int_K \mathbf{T}^{m-N_1} u(z)\nu_y \mathbf{P}^{N_1}(dz)| =$$

$$|\int_{K^2} (\mathbf{T}^{m-N_1} u(z) - \mathbf{T}^{m-N_1} u(w))\tilde{\nu}^*_{x,y}(dzdw)| =$$

$$|\int_D (\mathbf{T}^{m-N_1} u(z) - \mathbf{T}^{m-N_1} u(w))\tilde{\nu}^*_{x,y}(dzdw) +$$

$$\int_{K^2 \setminus D} (\mathbf{T}^{m-N_1} u(z) - \mathbf{T}^{m-N_1} u(w))\tilde{\nu}^*_{x,y}(dzdw)| \leq$$

$$\alpha\gamma(\mathbf{T}^{m-N_1} u)\rho_1 + (1-\alpha)osc(\mathbf{T}^{m-N_1} u) \leq$$

$$\alpha\gamma(u)3\rho_1 + (1-\alpha)osc(\mathbf{T}^{m-N_1} u) =$$

$$\alpha\gamma(u)\rho/2 + (1-\alpha)osc(\mathbf{T}^{m-N_1} u). \tag{73}$$

Hence, if we set $N = N_0 + N_1$, we conclude from (72) and (73) that

$$|\mathbf{T}^n u(x) - \mathbf{T}^n u(y)| \leq$$

$$\gamma(u)\alpha\rho/2 + |\int_K \mathbf{T}^{n-N_1} u(x) - \int_K \mathbf{T}^{n-N_1} u(y)| \leq$$

$$\alpha\gamma(u)\rho/2 + \alpha\gamma(u)\rho/2 + (1-\alpha)osc(\mathbf{T}^{n-N_1-N_0} u) =$$

$$\alpha\gamma(u)\rho + (1-\alpha)osc(\mathbf{T}^{m-N} u),$$

and, since x, y were arbitrary points in K, it follows that

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

which was what we wanted to prove. Thereby, Theorem 11.1 is proved. □

16 Contracting kernels.

Our next aim is to give some conditions that will imply that the filter kernel induced by a HMM with densities will satisfy Condition E.

In this section we shall state and prove a theorem in which an estimate for a class of nonnegative kernels is obtained. The theorem is similar to Lemma 6.2 of [13]. The essential ideas goes - at least - back to the paper [9] by Furstenberg and Kesten.

Let as usual $(S, \mathcal{F})$ be a measurable space and let λ be a positive $\sigma - finite$ measure on $(S, \mathcal{F})$.

Definition 16.1 *Let $k \in D_\lambda[S]$. We say that k has* **rectangular support** *if there exist $F \in \mathcal{F}$ and $G \in \mathcal{F}$ such that*

$$\lambda(F) > 0 \; and \;\; \lambda(G) > 0,$$

$$k(s,t) > 0, \;(s,t) \in F \times G$$

and

$$k(s,t) = 0, \;(s,t) \notin F \times G.$$

Theorem 16.1 *Let $k_m, \; m = 1,2,...,n, \;\; n \geq 1$ be density kernels belonging to $D_\lambda[S]$ having rectangular supports $F_m \times G_m, \; m = 1,2,...,n,$ where thus $\lambda(F_m)\lambda(G_m) > 0, \; m = 1,2,...,n.$*

As usual, 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 set

$$K^m = K_1 K_2 ..., K_m, \; m = 1,2,...,n.$$

Now, suppose that there exist numbers $d_m, D_m, m = 1,2,...,n,$ such that, for $1 \leq m \leq n$,

$$0 < d_m \leq \inf\{k_m(s,t) : (s,t) \in F_m \times G_m\} \tag{74}$$

and

$$\sup\{k_m(s,t) : (s,t) \in F_m \times G_m\} \leq D_m. \tag{75}$$

Suppose also, that

$$K^n(s,S) > 0 \tag{76}$$

for all $s \in F_1$.

Then, if $x, y \in \mathcal{Q}(S,\mathcal{F})$ are such that $x(F_1) > 0$ and also $y(F_1) > 0$ it follows that if $n = 1$ or $n = 2$, then

$$||\frac{xK^n}{||xK^n||} - \frac{yK^n}{||yK^n||}|| \leq 2(1 - (d_1/D_1)),$$

and, if $n \geq 3$, then

$$||\frac{xK^n}{||xK^n||} - \frac{yK^n}{||yK^n||}|| \leq 2\prod_{m=1}^{n-1}(1 - (d_m/D_m)^2).$$

In order to prove this theorem we shall first prove the following lemma.

Lemma 16.1 *Let $k_m, K_m, K^m,\ m = 1,2,...,n,\ \ n \geq 1$ be defined - and have the same properties - as in Theorem 16.1.*

a) Firstly, let $n = 1$. Then, for all $s_1, s_2 \in F_1$ and all $E \in \mathcal{F}$,

$$|\frac{K^1(s_1,E)}{K^1(s_1,G_1)} - \frac{K^1(s_2,E)}{K^1(s_2,G_1)}| \leq 1 - (d_1/D_1). \tag{77}$$

b) Secondly, let $n = 2$. Then, for all $s_1, s_2 \in F_1$ and all $E \in \mathcal{F}$,

$$|\frac{K^2(s_1,E)}{K^2(s_1,G_2)} - \frac{K^2(s_2,E)}{K^2(s_2,G_2)}| \leq 1 - (d_1/D_1). \tag{78}$$

c) Thirdly, let $n \geq 3$. Then, for all $s_1, s_2 \in F_1$ and all $E \in \mathcal{F}$,

$$|\frac{K^n(s_1,E)}{K^n(s_1,G_n)} - \frac{K^n(s_2,E)}{K^n(s_2,G_n)}| \leq \prod_{m=1}^{n-1}(1 - (d_m/D_m)^2). \tag{79}$$

Proof of Lemma. Define $u_m : S \times S \to [0,\infty),\ m = 1,2,...,n$, recursively by $u_1 = k_1$ and

$$u_{m+1}(s,t) = \int_F u_m(s,\tau)k_{m+1}(\tau,t)\lambda(d\tau).$$

Clearly $u_m,\ \ m = 1,2,...,n$, are density kernels and $K^m,\ \ m = 1,2,...,n$ are the associated tr.fs.

Since all the density kernels $k_m, m = 1,2,...,n$ are assumed to have rectangular supports, it is easy to prove that also $u_m, m = 1,2,...,n$ have rectangular supports, and since we assume that $K^n(s,S) > 0$ if $s \in F_1$ (see (76)), it follows, that

$$u_m(s,G_m) > 0,\ if\ m = 1,2,...,n,\ and\ s \in F_1,$$

and that the support of u_m is equal to $F_1 \times G_m$. For $m = 1,2,...,n$, we let $\mathcal{G}_m$ be defined by

$$\mathcal{G}_m = \{E \in \mathcal{F} : E \subset G_m\},$$

and we define $U_m : S \times \mathcal{F} \to [0,\infty)$ by

$$U_m(s,E) = \int_S u_m(s,t)\lambda(dt).$$

Clearly $U_m = K^m$. We use U_m instead of K^m in order to have notations similar to the notations used in [13]. If $s \in F_1$ and $E \in \mathcal{F}$ we write

$$U_m^*(s,E) = \frac{U_m(s,E \cap G_m)}{U_m(s,G_m)}.$$

We shall first verify that (77) holds. Since $U_1(s,E) = \int_E k_1(s,t)\lambda(dt)$ it follows from (74) and (75) that

$$d_1\lambda(E) \leq U_1(s,E) \leq D_1\lambda(E),\ \ s \in F_1, E \in \mathcal{G}_1.$$

Hence, if $s \in F_1$ and $E \in \mathcal{G}_1$ we find that

$$U_1^*(s,E) = \frac{U_1(s,E)}{U_1(s,G_1)} \geq \frac{(d_1/D_1)\lambda(E)}{\lambda(G_1)}.$$

This implies that

$$||U_1^*(s_1,\cdot)-U_1^*(s_2,\cdot)||=$$
$$||(U_1^*(s_1,\cdot)-(d_1/D_1)\lambda^*)-(U_1^*(s_2,\cdot)-(d_1/D_1)\lambda^*)||\le 2(1-(d_1/D_1)),$$

if $s_1,s_2\in F_1$, where thus $\lambda^*\in\mathcal{P}(S,\mathcal{F})$ is defined by

$$\lambda^*(E)=\frac{\lambda(E\cap G_1)}{\lambda(G_1)}.$$

Hence, if $s_1,s_2\in F_1$,

$$\sup_{E\in\mathcal{F}}|U_1^*(s_1,E)-U_1^*(s_2,E)|=(1/2)||U_1^*(s_1,\cdot)-U_1^*(s_2,\cdot)||\le$$
$$(1/2)2(1-(d_1/D_1))=1-(d_1/D_1).$$

Hence (77) holds.

Next, let us note, that since

$$U_2(s,E)=\int_E\int_{G_1}k_1(s,\sigma)k_2(\sigma,t)\lambda(d\sigma)\lambda(dt),$$

it follows from (74) and (75), that, for $s\in F_1$ and $E\in\mathcal{G}_2$,

$$U_2(s,E)\ge d_1\int_{F_2\cap G_1}K_2(\sigma,E)\lambda(d\sigma)$$

and that

$$U_2(s,E)\le D_1\int_{F_2\cap G_1}K_2(\sigma,E)\lambda(d\sigma).$$

Hence,

$$U_2^*(s,E)\ge(d_1/D_1)\mu^*(E),\ \forall s\in F_1,\ \forall E\in\mathcal{G}_2,$$

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

$$\mu^*(E)=\int_{F_2\cap G_1}K_2(\sigma,E\cap G_2)\lambda(d\sigma)/\int_{F_2\cap G_1}K_2(\sigma,G_2)\lambda(d\sigma).$$

Hence, by arguments similar to those used for the case $n=1$, we can conclude, that if $s_1,s_2\in F_1$ then

$$||U_2^*(s_1,\cdot)-U_2^*(s_2,\cdot)||=$$
$$||(U_2^*(s_1,\cdot)-(d_1/D_1)\mu^*)-(U_2^*(s_2,\cdot)-(d_1/D_1)\mu^*)||\le 2(1-(d_1/D_1)),$$

from which follows that

$$\sup_{E\in\mathcal{F}}|U_2^*(s_1,E)-U_2^*(s_2,E)|=(1/2)||U_2^*(s_1,\cdot)-U_2^*(s_2,\cdot)||\le$$
$$(1/2)2(1-(d_1/D_1))=1-(d_1/D_1)$$

if $s_1,s_2\in F_1$. Hence (78) holds.

Now, let $n\ge 3$. Define $v:F_2\times S\to[0,\infty)$ by

$$v(s,t)=\int_{G_2}\int_{G_3}...\int_{G_{n-1}}k_2(s,\sigma_2)k_3(\sigma_2,\sigma_3)...k_n(\sigma_{n-1},t)\lambda(d\sigma_2)\lambda(d\sigma_3)...\lambda(d\sigma_{n-1})$$

and define $V : F_2 \times \mathcal{F} \to [0, \infty)$ by

$$V(s,E) = \int_E v(s,t)\lambda(dt).$$

Then, for $s \in F_1$ and $E \in \mathcal{F}$, we can express $U_n(s,E)$ as

$$U_n(s,E) = \int_{G_1} k_1(s,\sigma)V(\sigma,E)\lambda(d\sigma).$$

Hence, if $s \in F_1$,

$$U_n(s,E) \geq d_1 \int_{G_1 \cap F_2} V(\sigma,E)\lambda(d\sigma)$$

and

$$U_n(s,E) \leq D_1 \int_{G_1 \cap F_2} V(\sigma,E)\lambda(d\sigma).$$

Hence, for $s_1, s_2 \in F_1$ and $E \in \mathcal{F}$, we find that

$$(d_1/D_1)U_n(s_2,E) \leq U_n(s_1,E) \leq (D_1/d_1)U_n(s_2,E),$$

and in particular we have

$$(d_1/D_1)U_n(s_2,G_n) \leq U_n(s_1,G_n) \leq (D_1/d_1)U_n(s_2,G_n), \ s_1, s_2 \in F_1. \tag{80}$$

Now in order to prove (79) we shall use the same trick as used in the classical paper [9]. Let $s_1 \in F_1$ and $E \in \mathcal{F}$. We then find that

$$U_n^*(s_1,E) = U_n(s_1,E)/U_n(s_1,G_n) =$$

$$\int_{G_1} k_1(s_1,\sigma)(\frac{V(\sigma,E)}{U_n(s_1,G_n)})\lambda(d\sigma) =$$

$$\int_{G_1} k_1(s_1,\sigma)(\frac{V(\sigma,G_n)}{U_n(s_1,G_n)})(\frac{V(\sigma,E)}{V(\sigma,G_n)})\lambda(d\sigma) =$$

$$\int_{G_1} f(\sigma)(\frac{V(\sigma,E)}{V(\sigma,G_n)})\lambda(d\sigma),$$

where thus

$$f(\sigma) = k_1(s_1,\sigma)\frac{V(\sigma,G_n)}{U_n(s_1,G_n)}.$$

Clearly,

$$\int_{G_1} f(\sigma)\lambda(d\sigma) = 1,$$

since

$$\int_{G_1} k_1(s_1,\sigma)V(\sigma,G_n)\lambda(d\sigma) = U_n(s_1,G_n).$$

Similarly, if we let $s_2 \in F_1$ satisfy $s_2 \neq s_1$, we obtain

$$U_n^*(s_2,E) = U_n(s_2,E)/U_n(s_2,G_n) =$$

$$\int_{G_1} k_1(s_2,\sigma)(\frac{V(\sigma,G_n)}{U_n(s_2,G_n)})(\frac{V(\sigma,E)}{V(\sigma,G_n)})\lambda(d\sigma) =$$

$$\int_{G_1} g(\sigma)(\frac{V(\sigma, E)}{V(\sigma, G_n)})\lambda(d\sigma),$$

where thus now

$$g(\sigma) = k_1(s_2, \sigma)\frac{V(\sigma, G_n)}{U_n(s_2, G_n)}$$

and we again have

$$\int_{G_1} g(\sigma)\lambda(d\sigma) = 1.$$

Let us now define $\alpha \in \mathcal{P}(G_1, \mathcal{G}_1)$ and $\beta \in \mathcal{P}(G_1, \mathcal{G}_1)$ by

$$\alpha(E) = \int_E f(\sigma)\lambda(d\sigma),\ E \in \mathcal{G}_1$$

and

$$\beta(E) = \int_E g(\sigma)\lambda(d\sigma),\ E \in \mathcal{G}_1.$$

From the hypotheses (74) and (75) it clearly follows, that

$$(d_1/D_1)k_1(s_2, t) \le k_1(s_1, t) \le k_1(s_2, t)(D_1/d_1),\ \forall s_1, s_2 \in F_1,\ \forall t \in G_1, \quad (81)$$

and from (81) and (80) we find

$$f(\sigma) = k_1(s_1, \sigma)V(\sigma, G_n)/U_n(s_1, G_n) \ge$$

$$\frac{(d_1/D_1)k_1(s_2, \sigma)V(\sigma, G_n)}{(D_1/d_1)U(s_2, G_n)} = (d_1/D_1)^2 g(\sigma)$$

from which it follows that

$$\alpha(E) \ge (d_1/D_1)^2\beta(E),\ E \in \mathcal{G}_1.$$

Hence

$$||\alpha - \beta|| = ||\alpha - (d_1/D_1)^2\beta - (1 - (d_1/D_1)^2)\beta|| \le$$

$$1 - (d_1/D_1)^2 + 1 - (d_1/D_1)^2 = 2(1 - (d_1/D_1)^2).$$

Therefore, if $s_1, s_2 \in F_1$ and $E \in \mathcal{F}$, we obtain, by using Lemma 2.2, that

$$|U_n^*(s_1, E) - U_n^*(s_2, E)| =$$

$$|\int_{F_1} \frac{V(\sigma, E)}{V(\sigma, G_n)}\alpha(d\sigma) - \int_{F_1} \frac{V(\sigma, E)}{V(\sigma, G_n)}\beta(d\sigma)| \le$$

$$\sup\{|\frac{V(t_1, E)}{V(t_1, G_n)} - \frac{V(t_2, E)}{V(t_2, G_n)}| : t_1, t_2 \in F_2\}(1/2)||\alpha - \beta|| \le$$

$$\sup\{|\frac{V(t_1, E)}{V(t_1, G_n)} - \frac{V(t_2, E)}{V(t_2, G_n)}| : t_1, t_2 \in F_2\}(1 - (d_1/D_1)^2). \quad (82)$$

By induction it now follows, that

$$|U_n^*(s_1, E) - U_n^*(s_2, E)| \le$$

$$(1 - (d_1/D_1)^2)(1 - (d_2/D_2)^2)...(1 - (d_{n-2}/D_{N-2})^2)(1 - (d_{n-1}/D_{N-1}), \quad (83)$$

where thus the first $n-2$ factors of the right hand side of (83) follow by applying (82) repeatedly, and the last factor follows from (78). Hence (79) holds and thereby the lemma is proved. □

We now prove Theorem 16.1. Thus, let $x, y \in \mathcal{Q}(S, \mathcal{F})$ be such that both $x(F_1) > 0$ and $y(F_1) > 0$. We write $K^n = U$. What we want to prove is that, if $n = 1$ or $n = 2$, then

$$||\frac{xU}{||xU||} - \frac{yU}{||yU||}|| \leq 2(1 - (d_1/D_1)^2)$$

and, if $n \geq 3$, then

$$||\frac{xU}{||xU||} - \frac{yU}{||yU||}|| \leq 2\prod_{m=1}^{n-1}(1 - (d_m/D_m)^2).$$

Let $E \in \mathcal{F}$. Then $xU(E)/||xU||$ can be written

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

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

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

where thus

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

Evidently $\alpha \in \mathcal{P}(F_1, \mathcal{F}_1)$.

In a similar manner we can write

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

where thus $\beta \in \mathcal{P}(F_1, \mathcal{F}_1)$ is defined by

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

Hence, by again using Lemma 2.2, we find

$$|\frac{xU(E)}{||xU||} - \frac{yU(E)}{||yU||}| =$$

$$|\int_{F_1} \frac{U(s,E)}{U(s,G_n)}\alpha(ds) - \int_{F_1} \frac{U(s,E)}{U(s,G_n)}\beta(ds)| \leq$$

$$\sup\{\frac{U(s_1,E)}{U(s_1,G_n)} - \frac{U(s_2,E)}{U(s_2,G_n)} : s_1, s_2 \in F_1\}(1/2)||\alpha - \beta||.$$

Since $||\alpha - \beta|| \leq 2$, it follows from Lemma 16.1, that

$$|\frac{xU(E)}{||xU||} - \frac{yU(E)}{||yU||}| \leq (1 - (d_1/D_1),$$

if $n = 1$ or $n = 2$, and

$$|\frac{xU(E)}{||xU||} - \frac{yU(E)}{||yU||}| \le \prod_{m=1}^{n-1}(1-(d_m/D_m)^2),$$

if $n \ge 3$. Since the right hand sides are independent of $E \in \mathcal{G}_n$ it follows that

$$||\frac{xU}{||xU||} - \frac{yU}{||yU||}|| \le 2(1-(d_1/D_1)$$

if $n = 1$ or $n = 2$ and

$$||\frac{xU}{||xU||} - \frac{yU}{||yU||}|| \le 2\prod_{m=1}^{n-1}(1-(d_m/D_m)^2)$$

if $n \ge 3$.

Thereby Theorem 16.1 is proved. □

17 On Condition E

Our aim in this section is to find conditions on a HMM with densities such that Condition E holds.

We shall first prove that Condition B introduced in the paper [16] implies Condition E. Thus, let $\mathcal{H} = \{(S,\mathcal{F}), P, (A,\mathcal{A}), M\}$ be a HMM for which both the state space S and the observation space A are denumerable sets and the $\sigma - algebras$ $\mathcal{F}$ and $\mathcal{A}$ are defined as all the subsets of S and A respectively. If we define δ_0 as the discrete metric on S, ϱ as the discrete metric on A, let λ denote the counting measure on $(S,\mathcal{F})$, let τ denote the counting measure on $(A,\mathcal{A})$, define $m : S \times S \times A \to [0,1]$ by

$$m(s,t,a) = M(s,t,\{a\})$$

and define $p : S \times S \to [0,1]$ by

$$p(s,t) = \sum_{a \in A} m(s,t,a),$$

then the HMM $\mathcal{H}$ can also be defined by the set

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

which clearly is a HMM with densities. Since A is denumerable and $m : S \times S \times A \to [0,1]$ is bounded, we also note that the partition (m,τ) is regular because of Proposition 5.2. We also suppose that the hidden Markov chain of the HMM is strongly ergodic with stationary measure π and for simplicity we identify the state s_i with its index i.

As usual let $K = \mathcal{P}(S,\mathcal{F})$, let $\mathcal{E}$ denote the $\sigma - field$ generated by the total variation norm, let $\mathcal{P}(K|\pi)$ be the set of all probability measures on $(K,\mathcal{E})$ with barycenter equal to π. For $a \in A$ we define the $S \times S$ matrix $M(a)$ by

$$(M(a))_{(i,j)} = m(i,j,a),\ \forall i,j \in S$$

where thus $(M(a))_{(i,j)}$ denotes the *i-j-th* element of $M(a)$. For $i \in S$ we let e_i denote the probability vector in K satisfying $(e_i)_i = 1$.

In [16], Section 1.6, the following condition was introduced.

Condition B. *For every $\rho > 0$ there exists an element $i_0 \in S$ such that if $C \subset K$ is a compact set satisfying*

$$\mu(C \cap \{x : (x)_{i_0} > (\pi)_{i_0}/2\}) \geq (\pi)_{i_0}/3, \quad \forall\, \mu \in \mathcal{P}(K|\pi), \tag{84}$$

then we can find an integer N, and a sequence $\{a_1, a_2, ...a_N\}$ of elements in A, such that, if we set

$$\mathbf{M}(\mathbf{a^N}) = M(a_1)M(a_2)...M(a_N),$$

then

$$||e^{i_0}\mathbf{M}(\mathbf{a^N})|| > 0$$

and if $x \in C \cap \{x : (x)_{i_0} > (\pi)_{i_0}/2\}$ then also

$$||\frac{x\mathbf{M}(\mathbf{a^N})}{||x\mathbf{M}(\mathbf{a^N})||} - \frac{e^{i_0}\mathbf{M}(\mathbf{a^N})}{||e^{i_0}\mathbf{M}(\mathbf{a^N})||}|| < \rho. \ \square$$

Remark. That there exists a compact set C, such that (84) holds, follows from Proposition 5.7 of [16] .

Proposition 17.1 *Let $\mathcal{H} = \{(S, \mathcal{F}), P, (A, \mathcal{A}), M\}$ be a HMM for which both the state space S and the observation space A are denumerable and such that the hidden Markov chain is strongly ergodic with stationary measure π. Suppose Condition B is satisfied. Then Condition E is satisfied.*

Proof. Let $\rho > 0$, From Condition B follows,that we can find an element $i_0 \in S$ such, if $C \subset K$ is a compact set satisfying (84), then we can find an integer N and a sequence $a_1, a_2, ...a_N$, such that, if

$$\mathbf{M}(\mathbf{a^N}) = M(a_1)M(a_2)...M(a_N),$$

then

$$||e^{i_0}\mathbf{M}(\mathbf{a^N})|| > 0,$$

and, if $x \in C \cap \{x : (x)_{i_0} > (\pi)_{i_0}/2\}$, then also

$$||\frac{x\mathbf{M}(\mathbf{a^N})}{||x\mathbf{M}(\mathbf{a^N})||} - \frac{e^{i_0}\mathbf{M}(\mathbf{a^N})}{||e^{i_0}\mathbf{M}(\mathbf{a^N})||}|| < \rho/2. \tag{85}$$

For simplicity, let us set

$$E_{i_0} = \{x \in K : (x)_{i_0} > (\pi)_{i_0}/2\}. \tag{86}$$

Now, let μ and ν belong to $\mathcal{P}(K|\pi)$ and let $\tilde{\mu}_0$ denote the product measure of μ and ν. Then it follows from (84), that

$$\tilde{\mu}_0((x, y) \in (C \cap E_{i_0}) \times (C \cap E_{i_0})) \geq ((\pi)_{i_0}/3)^2.$$

Now simply let $\tilde{\mu}_N$ be the product measure of $\mu\mathbf{P}^N$ and $\nu\mathbf{P}^N$, and set

$$\beta = ||e_{i_0}\mathbf{M}(\mathbf{a^N})||. \tag{87}$$

From (85) and the triangle inequality it follows that, if

$$(x, y) \in (C \cap E_{i_0}) \times (C \cap E_{i_0}),$$

then

$$||\frac{x\mathbf{M}(\mathbf{a^N})}{||x\mathbf{M}(\mathbf{a^N})} - \frac{y\mathbf{M}(\mathbf{a^N})}{||y\mathbf{M}(\mathbf{a^N})||}|| < \rho.$$

Furthermore, if $x \in E_{i_0}$ we conclude from (87) and (86), that

$$||x\mathbf{M}(\mathbf{a^N})|| \geq (\pi_{i_0}/2)\beta.$$

Hence, if as usual we set $D_\rho = \{(x, y) \in K^2 : ||x - y|| < \rho\}$, we find that

$$\tilde{\mu}_N(D_\rho) \geq (\pi_{i_0}/3)^2(\pi_{i_0}/2)^2\beta^2$$

and hence, if we set

$$\alpha = (\pi_{i_0}/3)^2(\pi_{i_0}/2)^2\beta^2,$$

we see that Condition E is satisfied. □

We have just showed that Condition E is a fairly straight forward generalisation of Condition B in [16]. Our next aim is to describe a condition which can be considered as a generalisation of Condition B1 of [16]. (See again [16], Section 1.6.)

Thus let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a HMM with densities such that the partition (m, τ) is regular, and let as usual $K = \mathcal{P}_\lambda(S, \mathcal{F})$, let δ_{TV} denote the total variation metric, let $\mathcal{E}$ be the $\sigma - algebra$ determined by the total variation metric δ_{TV} and let $\mathbf{P} \in \mathcal{TP}((K, \mathcal{E}))$ be the filter kernel induced by the HMM $\mathcal{H}$. Let $\{(S, \mathcal{F}), (A, \mathcal{A}), h, G\}$ be the random mapping associated to $\mathcal{H}$ as defined in Definition 6.1 of Section 6.

Now, let us for each $F \in \mathcal{F}$ define

$$E(F) = \{x \in K : x(F) > \pi(F)/2\}. \tag{88}$$

Clearly $E(F)$ is an open set.

Next let us define $K^2A^{n+}(F)$ by

$$\{(x, y, a^n) \in E(F) \times E(F) \times A^n : ||xM^n_{a^n}|| \cdot ||yM^n_{a^n}|| > 0\}.$$

Clearly $K^2A^{n+}(F)$ is also an open set, since $||xM^n_{a^n}||$ is a continuous function in both variables.

Further, let again $F \in \mathcal{F}$, let $\epsilon > 0$, and let n be a positive integer. Let $\{(S, \mathcal{F}), (A^n, \mathcal{A}^n), h^n, G^n\}$ denote the n:th composition of the random mapping $\{(S, \mathcal{F}), (A, \mathcal{A}), h, G\}$. We now define $K^2A^{n+}(F, \epsilon)$ as the subset of $K^2A^{n+}(F)$ such that

$$||h^n(x, a^n) - h^n(y, a^n)|| < \epsilon.$$

Since $h^n(x, a^n)$ is continuous in both variables on the set $\{(x, a^n) : ||xM^n_{a^n}|| > 0\}$ it follows that also the set $K^2A^{n+}(F, \epsilon)$ is an open set and hence measurable.

Next, for each $(x, y) \in E(F) \times E(F)$, we define

$$A^{n+}_{x,y}(F, \epsilon) = \{a^n \in A^n : (x, y, a^n) \in K^2A^{n+}(F, \epsilon)\}.$$

Since the set $A^{n+}_{x,y}(F, \epsilon)$ is a section of $K^2A^{n+}(F, \epsilon)$, it belongs to $\mathcal{A}^n$.

The following condition is not only a generalisation of Condition B1 in [16] but also in essence a generalisation of the *"rank one condition"* introduced by Kochman and Reeds in the paper [17]. Recall that $E(F)$ is defined by (88).

Condition 17.1 *To every $\rho > 0$ we can find an integer N, a number $\beta > 0$ and a set $F_0 \in \mathcal{F}$ satisfying $\pi(F_0) > 0$ such that for every pair $(x, y) \in E(F_0) \times E(F_0)$*

$$G^N(x, A^{N+}_{x,y}(F_0, \rho)) \geq \beta. \tag{89}$$

Proposition 17.2 *Let $\mathcal{H} = \{(S, \mathcal{F}, \delta_0\}, (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be a HMM with densities such that the partition (m, τ) is regular, and suppose that underlying Markov chain is* **strongly ergodic with stationary measure** *π. Suppose Condition 17.1 holds. Then Condition E holds.*

Proof. We shall first prove the following lemma which is similar to Lemma 5.7 of [16].

Lemma 17.1 *Let $\mu \in \mathcal{P}(K|\pi)$, and suppose that $F \in \mathcal{F}$ is such that $\pi(F) > 0$. Let $E = \{x \in K : x(F) > \pi(F)/2)\}$. Then*

$$\mu(E) > \pi(F)/2.$$

Proof. Since $\mu \in \mathcal{P}(K|\pi)$ we have $\int_K \langle I_F, x\rangle \mu(dx) = \pi(F)$. Hence

$$\pi(F) = \int_E \langle I_F, x\rangle \mu(dx) + \int_{K\setminus E} \langle I_F, x\rangle \mu(dx) =$$

$$\int_E \int_F x(ds)\mu(dx) + \int_{K\setminus E} \int_F x(ds)\mu(dx) \leq$$

$$\leq \mu(E) + (1 - \mu(E))\pi(F)/2.$$

Hence

$$\mu(E)(1 - \pi(F)/2) \geq \pi(F)/2$$

and hence

$$\mu(E) > \pi(F)/2$$

which was what we wanted to prove. □

To finish the proof of Proposition 17.2 is now rather easy. Let $\rho > 0$ be given, and let the set $F_0 \in \mathcal{F}$, the integer N and the number β be such that (89) is satisfied. Let $D_\rho = \{(z_1, z_2) \in K^2 : \delta_{TV}(z_1, z_2) < \rho\}$ and let $\mathbf{P}$ denote the filter kernel induced by the HMM $\mathcal{H}$. What we want to show is, that we can find an integer N_1 and a number α, such that for any $\mu, \nu \in \mathcal{P}(K|\pi)$, we can find a coupling $\tilde{\mu}_{N_1}$ of $\mu\mathbf{P}^{N_1}$ and $\nu\mathbf{P}^{N_1}$, such that

$$\tilde{\mu}_{N_1}(D_\rho) \geq \alpha.$$

We shall choose $N_1 = N$ and we shall choose

$$\alpha = \beta(\pi(F_0)/2)^2. \tag{90}$$

It thus remains to define a coupling.

Let μ and ν belong to $\mathcal{P}(K|\pi)$. Our aim is to construct a coupling $\tilde{\mu}_N$ of $\mu\mathbf{P}^N$ and $\mu\mathbf{P}^N$ such that

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

where thus α is defined by (90). We shall do this by constructing a random mapping

$$\{(K^2, \mathcal{E}^2), (A^{2N}, \mathcal{A}^{2N}), \tilde{h}_N, \tilde{G}_N\}$$

for which the state space is $(K^2,\mathcal{E}^2)$ and the index space is $(A^{2N},\mathcal{A}^{2N})$, as follows:

1) We let $\tilde{G}_N : K^2\times\mathcal{A}^{2N}\to[0,1]$ be the Vaserstein coupling of the tr.pr.f $G^N : K\times\mathcal{A}^N\to[0,1]$. (See Section 10 for the definition of the Vaserstein coupling of a Markov kernel with density.)
2) We define $\tilde{h}_N : K^2\times A^{2N}\to K^2$ simply by

$$\tilde{h}_N((x,y),(a^N,b^N)) = (h^N(x,a^N),h^N(y,b^N)).$$

Since $\tilde{G}_N : K^2\times\mathcal{A}^{2N}\to[0,1]$ belongs to $\mathcal{TP}((K^2,\mathcal{E}^2),(A^{2N},\mathcal{A}^{2N}))$ and obviously $\tilde{h}_N : K^2\times A^{2N}\to K^2$ is measurable, since each of its components is measurable, it is clear that

$$\{(K^2,\mathcal{E}^2),(A^{2N}\,\mathcal{A}^{2N}),\tilde{h}_N,\tilde{G}_N\}$$

constitutes a random mapping.

Next, if we let $\tilde{\mathbf{P}}_N$ denote the tr.pr.f associated to the random mapping

$$\{(K^2,\mathcal{E}^2),(A^{2N},\mathcal{A}^{2N}),\tilde{h}_N,\tilde{G}_N\}$$

it is easily checked, that for all $x,y\in K$,

$$\tilde{\mathbf{P}}_N((x,y),E\times K)) = \mathbf{P}^N(x,E),\ \ \forall E\in\mathcal{E}$$

and that

$$\tilde{\mathbf{P}}_N((x,y),K\times E)) = \mathbf{P}^N(y,E),\ \ \forall E\in\mathcal{E}.$$

Therefore, if we let $\tilde{\mu}_0\in\mathcal{P}(K^2,\mathcal{E}^2)$ denote the product measure $\mu\otimes\nu$ of μ and ν, it follows easily, that

$$\tilde{\mu}_0\tilde{\mathbf{P}}_N$$

is a coupling of $\mu\mathbf{P}^N$ and $\nu\mathbf{P}^N$.

Now let $E_0=\{x\in K : x(F_0)>\pi(F_0)/2\}$. From Lemma 17.1 we know, that

$$\mu(E_0)>\pi(F_0)/2\ \ \textit{and}\ \ \nu(E_0)>\pi(F_0)/2. \tag{91}$$

Furthermore, since by assumption $G^N(x,A^{N+}_{x,y}(F_0,\rho))\geq\beta$ for all $(x,y)\in E_0\times E_0$, and also $(A^N,\mathcal{A}^N)$ is a complete, separable, metric space, it follows from the definition of the Vaserstein coupling and Lemma 9.1, that

$$\tilde{G}_N((x,y),\tilde{D}_N(A^{N+}_{x,y}(F_0,\rho)))\geq\beta \tag{92}$$

for all $x,y\in E_0$, where thus

$$\tilde{D}_N(A^{N+}_{x,y}(F_0,\rho)) = \{(a^N,b^N) : a^N=b^N,\ a^N\in A^{N+}_{x,y}(F_0,\rho))\}.$$

But if $x,y\in E_0$ and $a^N\in A^{N+}_{x,y}(F_0,\rho)$, then

$$||h^N(x,a^N)-h^N(y,a^N)|<\rho.$$

This together with (92), (90) and the inequalities in (91) implies now that

$$\tilde{\mu}_0\tilde{\mathbf{P}}_N(D_\rho)\geq(\pi(F_0)/2)^2\beta=\alpha.$$

Thereby Proposition 17.2 is proved.□

Remark. Instead of using the Vaserstein coupling of $G_N(x)$ and $G_N(y)$ we could have used the trivial coupling. Our choice of α would then have been $\alpha = (\pi(F_0)/2)^2\beta^2$. □

Our next condition is a generalisation of a condition called Condition A, first introduced in the paper [13]. To simplify a comparison with Theorem 9.3 of [16] we recapitulate the definition of Condition A and the notion "localizing" introduced in [16].

Definition 17.1 *Let $\mathcal{H} = \{(S, \mathcal{F}), P, (A, \mathcal{A}), M\}$ be a HMM such that the state space and the observation space are denumerable. If there exist an integer N and a sequence $a_1, a_2, ..., a_N$ such that*

1)

$$M^N_{a^N} \neq 0,$$

2)

$$(M^N_{a^N})_{i_1,j_1} > 0 \text{ and } (M^N_{a^N})_{i_2,j_2} > 0$$

$$\Rightarrow$$

$$(M^N_{a^N})_{i_1,j_2} > 0 \text{ and } (M^N_{a^N})_{i_2,j_1} > 0,$$

then we say that **Condition A** *holds.*

If furthermore the set

$$S' = \{j : (M^N_{a^N})_{i,j} > 0, \text{ some } i \in S\ \}$$

is a **finite** *set, then we say that the partition M is* **localizing**. □

Our next condition reads as follows. As usual we are considering a HMM $\mathcal{H}= \{(S, \mathcal{F}, \delta_0\}, (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ with densities such that the partition (m, τ) is regular, and such that the hidden Markov chain is strongly ergodic with stationary measure π.

Condition 17.2 *There exists a set $F_0 \in \mathcal{F}$, a positive integer N_0 and a set $B_0 \in \mathcal{A}^{N_0}$ such that*

1)

$$\lambda(F_0) > 0,$$

2)

$$\pi(F_0) > 0,$$

3)

$$\tau^{N_0}(B_0) > 0$$

4) for all $s \in S$ and $a^{N_0} \in B_0$

$$m(s, t, a^{N_0}) = 0, \ \ if \ \ t \notin F_0$$

5) there exist positive numbers d_0 and D_0 such that for every $a^{N_0} \in B_0$ there exists a subset $F_1(a^{N_0}) \in \mathcal{F}$ such that

(a)

$$d_0 \leq m(s, t, a^{N_0}) \leq D_0, \ \forall (s, t) \in F_0 \times F_1(a^{N_0})$$

(b)

$$m(s,t,a^{N_0}) = 0, \ \ \forall (s,t) \in F_0 \times (F_0 \setminus F_1(a^{N_0})),$$

6) there exists a number $\beta_0 > 0$ such that for all $s \in F_0$

$$\int_{F_0} \int_B m(s,t,a^{N_0})\lambda(dt)\tau^N(da^{N_0}) \geq \beta_0 \tau^{N_0}(B), \ \forall\, B \in \mathcal{A}^{N_0} \text{ such that } B \subset B_0.$$

Remark. It is easily verified that, if 1) $\mathcal{H}= \{(S,\mathcal{F},\delta_0\}, (p,\lambda), (A,\mathcal{A},\varrho), (m,\tau)\}$ have denumerable state space and observation space, 2) Condition A is satisfied and 3) the partition M is localizing, then Condition 17.2 is satisfied. □

Proposition 17.3 *Let $\mathcal{H}= \{(S,\mathcal{F},\delta_0\}, (p,\lambda), (A,\mathcal{A},\varrho), (m,\tau)\}$ be a HMM with densities such that the partition (m,τ) is regular, and suppose that the underlying Markov chain is* **strongly ergodic with stationary measure** π. *Suppose Condition 17.2 holds. Then Condition E is satisfied.*

Proof. Let P denote the Markov kernel determined by (p,λ). For $n = 2,3,...$, let $p^n : S \times S \to [0,\infty)$ denote a density kernel of P^n, and let

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

be defined by formula (12) of Corollary 5.1

Since the stationary measure π for the hidden Markov chain associated to the HMM $\mathcal{H}$ is also a stationary measure for the hidden Markov chain associated to the HMM $\mathcal{H}^n$ for any integer $n \geq 2$, it follows trivially that, if Condition E is satisfied for the HMM $\mathcal{H}^n$ for some $n \geq 2$, then Condition E is also satisfied for $\mathcal{H}$. Therefore, it suffices to prove the lemma, when we in Condition 17.2 assume that the integer $N_0 = 1$. This assumption simplifies the notations somewhat.

Next, let

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

denote the random mapping associated to the HMM $\mathcal{H}$, and let $\mathbf{P}$ denote the filter kernel induced by the HMM $\mathcal{H}$. Recall that $\mathbf{P}$ is equal to the Markov kernel induced by the random mapping $\mathcal{R}$.

Let us first assume that $x' \in K$ is such that

$$x'(F_0) = 1$$

where thus F_0 is the set introduced in Condition 17.2. Then from hypothesis 4) and hypothesis 6) of Condition 17.2 it follows that for all $B \subset B_0$ such that $B \in \mathcal{A}$

$$G(x',B) = \int_S \int_S \int_B m(s,t,a)\tau(da)\lambda(dt)x'(ds) =$$

$$\int_{F_0} (\int_{F_0} \int_B m(s,t,a)\tau(da)\lambda(dt))x'(ds) \geq \int_{F_0} \beta_0\tau(B)x'(ds) = \beta_0\tau(B).$$

Therefore, for an arbitrary $x \in K$, it follows that

$$G(x,B) \geq x(F_0)\beta_0\tau(B). \tag{93}$$

Next let $\tilde{G}_V : K^2 \times \mathcal{A}^2 \to [0,1]$ denote the Vaserstein coupling of the tr.pr.f $G : K \times \mathcal{A} \to [0,1]$ (see Section 10), and define $\tilde{h} : K^2 \times A^2 \to K^2$ by

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

Clearly $\tilde{h}$ is measurable since each of its components is measurable. Therefore the set

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

is a random mapping with state space $(K^2, \mathcal{E}^2)$ and index space $(A^2, \mathcal{A}^2)$.

Now, let $\tilde{\mathbf{P}} : K^2 \times \mathcal{E}^2 \to [0, 1]$ denote the Markov kernel associated to the random mapping $\tilde{\mathcal{R}}$. From the definition of $\tilde{G}_V$ and $\tilde{h}$ it is easily checked that $\tilde{\mathbf{P}}((x, y), (\cdot, \cdot))$ is a coupling of $\mathbf{P}(x, \cdot)$ and $\mathbf{P}(y, \cdot)$ for any two elements x and y in $(K, \mathcal{E})$. Therefore, if μ and ν are probability measures in $\mathcal{P}(K, \mathcal{E})$ and $\tilde{\mu}$ denotes the product measure of μ and ν, it follows that $\tilde{\mu}\tilde{\mathbf{P}}^n$ is a coupling of $\mu\mathbf{P}^n$ and $\nu\mathbf{P}^n$ for $n = 1, 2, \ldots$.

Now , let $\rho > 0$ be given. What we shall prove is, that we can find an integer N and a constant $\alpha > 0$, such that, for any two probability measures μ and ν which belong to $\mathcal{P}(K|\pi)$, the coupling $\tilde{\mu}\tilde{\mathbf{P}}^N$ of $\mu\mathbf{P}^N$ and $\nu\mathbf{P}^N$ is such that

$$\tilde{\mu}\tilde{\mathbf{P}}^N(D_\rho) \geq \alpha, \tag{94}$$

where thus $\tilde{\mu} = \mu \otimes \nu$ and as above $D_\rho = \{(z_1, z_2) \in K^2 : ||z_1 - z_2|| < \rho\}$.

We choose the integer N as

$$N = \min\{n \geq 4 : 2(1 - (d_0/D_0)^2)^{n-2} < \rho\}, \tag{95}$$

and we chose $\alpha > 0$ such that

$$\alpha = (\pi(F_0)/2)^3(\beta_0\tau(B_0))^N \tag{96}$$

where thus F_0, β_0 and B_0 are the parameters introduced in Condition 17.2.

Next, let us define $\tilde{D}_{B_0^n}$, for $n = 1, 2, ...,$ by

$$\tilde{D}_{B_0^n} = \{(a^n, b^n) \in A^n \times A^n : a^n = b^n,\ a_m \in B_0, m = 1, 2, ..., n\}.$$

As before let

$$E_0 = \{x \in K : x(F_0) > \pi(F_0)/2\}.$$

From the inequality (93) we know that if x and y belong to E_0, then both $G(x, B) > (\pi(F_0)/2)\beta_0\tau(B)$ and $G(y, B) > (\pi(F_0)/2)\beta_0\tau(B)$ if $B \subset B_0$ and $B \in \mathcal{A}$. This implies that if $x, y \in E_0$ then

$$||G(x, \cdot) - G(y, \cdot)|| \leq 2(1 - (\pi(F_0)/2)\beta_0\tau(B_0)).$$

Therefore, since we have assumed that $(A, \mathcal{A}, \varrho)$ is a complete, separable, metric space, it follows by Lemma 9.1, that

$$\tilde{G}_V((x, y), \tilde{D}_{B_0^1}) \geq (\pi(F_0)/2)\beta_0\tau(B_0), \quad \forall x, y \in E_0.$$

But if $a \in B_0$, then, because of hypothesis 4) of Condition 17.2, it follows, that the probability measure $x' = h(x, a) \in K$ is such that $x'(F_0) = 1$, and it also follows that the probability measure $y' = h(y, a) \in K$ is such that $y'(F_0) = 1$.

From the Markov property it now follows that if $x, y \in E_0$ then

$$\tilde{G}_V^2((x, y), \tilde{D}_{B^2}) \geq (\pi(F_0)/2)(\beta_0\tau(B_0))^2.$$

Generally, since $h^m(x, a^{1,m})$ and $h^m(y, a^{1,m})$ are such that their supports belong to F_0 for $m = 2, 3, \ldots$ if $a^{1,m} \in B_0^m$, we find that, if $a^{1,m} \in B_0^m$, then it follows, that

$$\tilde{G}_V^n((x, y), \tilde{D}_{B_0^n}) \geq (\pi(F_0)/2)(\beta_0\tau(B_0))^n,\ \ \forall x, y \in E_0\ . \tag{97}$$

Now, let N be defined by (95), let $(a^N, b^N) \in \tilde{D}_{B_0^N}$ and let us consider $\tilde{h}^N((x,y),(a^N,b^N))$ when $x, y \in E_0$. Since $(a^N, b^N) \in \tilde{D}_{B_0^N}$ it is clear that

$$\tilde{h}^N((x,y),(a^N,b^N)) = \tilde{h}^N((x,y),(a^N,a^N)) = (h^N(x,a^n), h^N(y,a^N)).$$

Next set $x' = h(x, a_1)$ and $y' = h(y, a_1)$. From hypothesis 4) of Condition 17.2 it follows that $x'(F_0) = 1$ and that also $y'(F_0) = 1$.

For $m = 2, 3, ...,$ we now define $k_m : S \times S \to [0, \infty)$ by

$$k_m(s,t) = m(s,t,a_m).$$

From hypothesis 4) of Condition 17.2 we know that $k_m(s,t) = 0$ if $t \notin F_0$, and furthermore, if we restrict k_m to the set $F_0 \times F_0$, then, from condition 5) of Condition 17.2, it follows that k_m has a rectangular support.

Let $\mathcal{F}_0$ be defined by

$$\mathcal{F}_0 = \{F \in \mathcal{F} : F \subset F_0\}.$$

From Theorem 16.1 it follows, that if we define $K_m : F_0 \times \mathcal{F}_0 \to [0, \infty)$ by

$$K_m(s,F) = \int_F k_m(s,t)\lambda(dt),\ m = 2,3,...,N,$$

we assume that x and y belong to E_0 and set $x' = h(x, a_1)$ and $y' = h(y, a_1)$, then it follows that

$$||h^N(x,a^N) - h^N(y,a^N)|| = ||h^{N-1}(x',a^{2,N}) - h^{N-1}(y',a^{2,N})|| =$$

$$||\frac{x'K_2K_3...K_N}{||xK_2K_3...K_N||} - \frac{y'K_2K_3...K_N}{||y'K_2K_3...K_N||}|| \leq$$

$$2\prod_{m=2}^{N-1}(1-(d_0/D_0)^2) = 2(1-(d_0/D_0)^2)^{N-2} < \rho.$$

Hence, if $(a^N, b^N) \in \tilde{D}_{B_0^N}$ and $x, y \in E_0$ then

$$\tilde{h}^N((x,y),(a^N,b^N)) \in D_\rho.$$

Hence, by the inequality (97), we can conclude that

$$\tilde{\mathbf{P}}((x,y),D_\rho) \geq (\pi(F_0)/2)(\beta_0\tau(B_0))^N. \tag{98}$$

Finally, recalling Lemma 17.1, we know that both $\mu(E_0) > \pi(F_0)/2$ and $\nu(E_0) > \pi(F_0)/2$. Hence, from (98) and (96) it follows that

$$\tilde{\mu}\tilde{\mathbf{P}}^N(D_\rho) = \int_K\int_K \tilde{\mathbf{P}}^N(x,y,D_\rho)\mu(dx)\nu(dy) \geq$$

$$\int_{E_0}\int_{E_0} \tilde{\mathbf{P}}^N(x,y,D_\rho)\mu(dx)\nu(dy) \geq (\pi(F_0)/2)(\beta_0\tau(B_0))^N\mu(E_0)\nu(E_0) \geq$$

$$(\pi(F_0)/2)^3(\beta_0\tau(B_0))^N = \alpha$$

and thereby (94) is proved. □

18 Examples.

Our first example has Example 4.1 as its starting point.

Example 18.1 *Let* $(S, \mathcal{F}, \delta_0)$ *be a complete, separable, metric space, let* λ *be a* $\sigma - finite$ *measure on* $(S, \mathcal{F})$*, let* $P \in \mathcal{TP}_\lambda((S, \mathcal{F}))$ *and let* $p : S \times S \to [0, \infty)$ *be a density kernel of* P *such that*

$$\sup\{p(s,t) : s, t \in S\} < \infty.$$

Let A *be a denumerable set, let* $\{S_a, a \in A\}$ *be a denumerable set of disjoint subsets of* S *such that 1)* $\cup_{a \in A} S_a = S$*, 2)* $S_a \in \mathcal{F}$*,* $\forall a \in A$ *and*
3) $\lambda(S_a) > 0$*,* $\forall a \in A$*.*

Let ϱ *denote the discrete metric on* A*, let* $\mathcal{A}$ *denote all subsets of* A*, let* τ *denote the counting measure on* $(A, \mathcal{A})$*, and define* $m : S \times S \times A \to [0, \infty)$ *simply by*

$$m(s,t,a) = p(s,t), \; if \; t \in S_a;$$

$$m(s,t,a) = 0, \; if \; t \notin S_a.$$

Since obviously

$$\sum_{a \in A} m(s,t,a) = p(s,t)$$

we see that (m, τ) *is a partition of the density kernel* p*. Hence*

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

constitutes a HMM with densities. Moreover, since A *is a denumerable set and* $p : S \times S \to [0, \infty)$ *is assumed to be uniformly bounded, it follows from Proposition 5.2 that the partition* (m, τ) *is* regular*. (End of example.)*

Theorem 18.1 *Let* $\mathcal{H}_1 = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ *be the HMM defined as in Example 18.1. Suppose that*
a) the hidden Markov chain determined by the tr.pr.f P *is* **strongly ergodic** *with stationary measure* π*;*
b) there exists a set S_{a_0} *where* $a_0 \in A$ *and* $\pi(S_{a_0}) > 0$*, an integer* N*, an element* $b^N = (b_1, b_2, ..., b_N) \in A^N$*, two positive numbers* d_0, D_0 *satisfying* $d_0 \leq D_0$ *and a number* $\beta > 0$ *such that*
(i)

$$b_N = a_0,$$

(ii)

$$d_1 \leq m^N(s,t,b^N) \leq D_1, \; \forall (s,t) \in S_{a_0} \times S_{a_0}$$

and (iii)

$$\inf_{s \in S_{a_0}} \int_{S_{a_0}} m^N(s,t,b^N) \lambda(dt) \geq \beta.$$

Then the filtering process generated by the HMM is weakly ergodic.

If furthermore the underlying hidden Markov chain is **uniformly ergodic** *then there exists a measure* $\mu \in \mathcal{P}(K, \mathcal{E})$ *such that the filtering process is weakly ergodic with stationary measure* μ*.* $\square$

Remark. As usual $K = \mathcal{P}_\lambda(S, \mathcal{F})$ and $\mathcal{E}$ is the $\sigma - algebra$ on K generated by the total variation metric. $\Box$

Proof. It suffices to verify that the HMM $\mathcal{H}_1$ satisfies the hypotheses of Condition 17.2.

Set $F_0 = S_{a_0}$. Then clearly, since $\lambda(S_{a_0}) > 0$ and $\pi(S_{a_0}) > 0$, we obviously have $\lambda(F_0) > 0$ and $\pi(F_0) > 0$. Hence 1) and 2) of Condition 17.2 are satisfied with this choice of F_0.

Next set $N_0 = N$ and let $B_0 \subset \mathcal{A}^{N_0}$ be the set defined by $B_0 = \{b^N\}$. Since τ is the counting measure on A it follows that $\tau^{N_0}(B_0) = 1 > 0$ and hence hypothesis 3) of Condition 17.2 is satisfied.

Furthermore, since $b_N = a_0$, it is clear that if $t \notin F_0$ then $m(s, t, a^{N_0}) = 0$ for all $s \in S$ and hence hypothesis 4) of Condition 17.2 is satisfied.

In order to verify condition 5) of 17.2 we first need to define a set $F_1(b^{N_0})$. We simply define $F_1(b^{N_0}) = F_0$. From hypotheses (ii) follows that hypothesis 5a) of Condition 17.2 is satisfied with this choice of $F_1(b^{N_0})$ and that hypothesis 5b) of Condition 17.2 is satisfied follows again from the fact that $b_N = a_0$.

Finally from hypothesis (iii) of the theorem, it follows easily that if we define $\beta_0 = \beta\lambda(S_{a_0})$, where thus β is the constant occurring in hypothesis (iii), then hypothesis 6) of Condition 17.2 is satisfied with this choice of β_0.

Hence Condition 17.2 is satisfied and hence Condition E is satisfied. The conclusion then follows from Theorem 11.1. $\Box$

Our second example has Example 4.3 as its starting point.

Example 18.2 *Let $(S, \mathcal{F}, \delta_0)$ and $(A, \mathcal{A}, \varrho)$ be two complete, separable, metric spaces and let λ be a $\sigma - finite$ measure on $(S, \mathcal{F})$. Let $p \in D_\lambda[S]$ and $q \in D_\tau[S, A]$ be two probability density kernels. (See (3) in Section 2 for the definition of $D_\lambda[S]$ and $D_\tau[S, A]$.)*

Define $m : S \times S \times A \to [0, \infty)$ by

$$m(s, t, a) = p(s, t)q(t, a).$$

Then clearly (m, τ) is a partition of the probability density kernel p and, if we set

$$\mathcal{H}_2 = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\},$$

then clearly $\mathcal{H}_2$ is a HMM.

In order to guarantee that the partition is regular, we shall next make a few assumptions about the densities p and q.

We first introduce the notation

$$S_+(a) = \{t \in S : q(t, a) > 0\}.$$

Now, regarding the probability density function $q : S \times A \to [0, \infty)$ we shall first assume that it satisfies the following **boundedness** *condition:*
(i): There exists a positive constant C_0 such that

$$\sup\{q(t, a) : (t, a) \in S \times A\} \leq C_0.$$

Secondly we assume that
(ii)

$$\lambda(S_+(a)) > 0, \ \forall a \in A.$$

We also assume that the density function q has two **continuity** *properties. Our first continuity assumption reads as follows:*
(iii) For every $\epsilon > 0$ we can find an $\eta > 0$ such that if $\varrho(a, b) < \eta$ then

$$\lambda(S_+(a) \,\Delta\, S_+(b)) < \epsilon.$$

Here

$$S_+(a) \,\Delta\, S_+(b) = (S_+(a) \setminus S_+(b)) \cup (S_+(b) \setminus S_+(a)).$$

Our second continuity assumption reads as follows:
(iv) For every $\epsilon > 0$ we can find an $\eta > 0$ such that if $\varrho(a, b) < \eta$ then

$$|q(t, a) - q(t, b)| < \epsilon, \ \forall t \in S_+(a) \cap S_+(b).$$

Regarding the density function p we make the following assumption:
(v): there exists a constant C_1 such that

$$\sup\{p(s, t) : \ s, t \in S\} \leq C_1.$$

(End of example.)

Proposition 18.1 *Let $m : S \times S \times A \to [0, \infty)$ be defined by*

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

and suppose that p and q have the properties described in Example 18.2. Then the partition (m, τ) is regular. (See Definition 5.1 for the definition of a regular partition.)

Proof. Define $\overline{M} : \mathcal{Q}_\lambda(S, \mathcal{F}) \times A \to \mathcal{Q}_\lambda(S, \mathcal{F})$ by

$$\overline{M}(x, a)(F) = \int_S \int_F p(s, t)q(t, a)\lambda(dt)x(ds).$$

What we need to prove is that $\overline{M} : \mathcal{Q}_\lambda(S, \mathcal{F}) \times A \to \mathcal{Q}_\lambda(S, \mathcal{F})$ is a *continuous* function.

We first prove that $\overline{M}$ is a continuous function in the first variable. Thus, let $a \in A$ be fixed but arbitrary, and let $\epsilon > 0$ be given. Let $x, y \in \mathcal{Q}_\lambda(S, \mathcal{F})$, and let f and g be representatives of x and y respectively. Now, since

$$||\overline{M}(x, a) - \overline{M}(y, a)|| \leq \int_S (\int_S p(s, t)q(t, a)\lambda(dt))|f(s) - g(s)|\lambda(ds) \leq$$

$$C_0 \int_S |f(s) - g(s)|\lambda(ds) = C_0||x - y||$$

we see that, if $||x-y|| < \epsilon/C_0$, then $||\overline{M}(x, a)-\overline{M}(y, a)|| < \epsilon$ from which follows, that $\overline{M} : \mathcal{Q}_\lambda(S, \mathcal{F}) \times A \to \mathcal{Q}_\lambda(S, \mathcal{F})$ is a continuous function in the first variable.

In order to prove that $\overline{M}(x, \cdot)$ is a continuous function for each $x \in \mathcal{Q}_\lambda(S, \mathcal{F})$, we first consider the case, when $x \in \mathcal{P}_\lambda(S, \mathcal{F})$.

Thus, let $x \in \mathcal{P}_\lambda(S, \mathcal{F})$ be fixed but arbitrary chosen and let $f : S \to [0, \infty)$ be a representative of x. Define $f_1 : S \to [0, \infty)$ by

$$f_1(t) = \int_S p(s, t)f(s)\lambda(ds) = \int_S p(s, t)x(ds).$$

Since we have assumed that $\sup\{p(s,t): \; s,t \in S\} \leq C_1$ (see hypothesis (v)), it follows that

$$\sup_{t \in S} f_1(t) \leq C_1. \tag{99}$$

Now, let $F \in \mathcal{F}$ and let $a \in A$. We then find that

$$\overline{M}(x,a)(F) = \int_F (\int_S f(s)p(s,t)q(t,a)\lambda(ds))\lambda(dt) =$$

$$\int_F f_1(t)q(t,a)\lambda(dt).$$

Hence, if $a, b \in A$ we find that

$$||\overline{M}(x,a) - \overline{M}(x,b)|| = \int_S f_1(t)|q(t,a) - q(t,b)|\lambda(dt) \leq$$

$$\int_{S_+(a)\cap S_+(b)} f_1(t)|q(t,a) - q(t,b)|\lambda(dt) + \int_{S_+(a)\setminus S_+(b)} f_1(t)q(t,a)\lambda(dt) +$$

$$\int_{S_+(b)\setminus S_+(a)} f_1(t)q(t,b)\lambda(dt). \tag{100}$$

Now, let $\epsilon > 0$ be given. Then, first choose η_1 so small that

$$\lambda(S_+(a)\Delta S_+(b)) < (1/2)\epsilon/C_0C_1, \tag{101}$$

if $\varrho(a,b) < \eta_1$. This is possible to do because of hypothesis (ii). Next choose η_2 so small that

$$|q(t,a) - q(t,b)| < \epsilon/2, \; \forall t \in S_+(a) \cap S_+(b), \tag{102}$$

if $\varrho(a,b) < \eta_2$. Therefore, if we define $\eta = \min\{\eta_1, \eta_2\}$, it follows from (100), (99), (101) and (102) that, if $\varrho(a,b) < \eta$, then

$$||\overline{M}(x,a) - \overline{M}(x,b)|| \leq$$

$$\epsilon/2 \int_{S_+(a)\cap S_+(b)} f_1(t)\lambda(dt) + C_1C_0 \int_{S_+(a)\Delta S_+(b)} \lambda(dt) \leq$$

$$\epsilon/2 + C_1C_0(1/2)\epsilon/C_0C_1 = \epsilon,$$

from which follows, that $\overline{M} : \mathcal{P}_\lambda(S,\mathcal{F}) \times A \to \mathcal{Q}_\lambda(S,\mathcal{F})$ is a **continuous** function. But since $\overline{M}(x,a) = ||x||\overline{M}(\frac{x}{||x||}, a)$, it clearly follows that also $\overline{M} : \mathcal{Q}_\lambda(S,\mathcal{F}) \times A \to \mathcal{Q}_\lambda(S,\mathcal{F})$ is a **continuous** function. Thereby, the proof of the proposition is completed. □

Theorem 18.2 *Let $\mathcal{H}_2 = \{(S,\mathcal{F},\delta_0), (p,\lambda), (A,\mathcal{A},\varrho), (m,\tau)\}$ be the HMM defined as in Example 18.2. Suppose that*
a) the hidden Markov chain determined by the tr.pr.f P is **strongly ergodic** *with stationary measure π;*
b) there exists a set $F_0 \in \mathcal{F}$ and a set $B_0 \in \mathcal{A}$ such that
(i)

$$\lambda(F_0) > 0,$$

(ii)

$$\pi(F_0) > 0,$$

(iii)

$$\tau(B_0) > 0.$$

Furthermore, for each $a \in B_0$, let the set $F_1(a)$ be defined by

$$F_1(a) = \{t : q(t,a) > 0\},$$

and assume that
(iv)

$$F_1(a) \subset F_0,$$

(v) there exists $\beta_1 > 0$ such that for all $a \in B_0$

$$\lambda(F_1(a)) \geq \beta_1,$$

(vi) there exists a constant c_0 such that for each $a \in B_0$

$$\{t : 0 < q(t,a) < c_0\} = \emptyset,$$

(vii) there exists a constant $c_1 > 0$ such that

$$p(s,t) \geq c_1,\ (s,t) \in F_0 \times F_0.$$

Then the filtering process generated by the HMM is weakly ergodic.
If furthermore, the underlying hidden Markov chain is **uniformly ergodic***, then there exists a measure $\mu \in \mathcal{P}(K, \mathcal{E})$, such that the filtering process is weakly ergodic with stationary measure μ.*

Proof. It suffices to verify that the HMM $\mathcal{H}_2$ satisfies the hypotheses of Condition 17.2.

Since $\lambda(F_0) > 0$ and $\pi(F_0) > 0$ because of hypotheses (i) and (ii), it follows that 1) and 2) of Condition 17.2 is satisfied with this choice of F_0.

Next, set $N_0 = 1$. From hypothesis (iii) then follows that 3) of Condition 17.2 is satisfied if we let B_0 be as in the theorem and $N_0 = 1$.

Since $q(t,a) = 0$ if $a \in B_0$ and $t \notin F_0$ because of hypothesis (iv), it follows that $m(s,t,a) = p(s,t)q(t,a) = 0$ for all $(s,a) \in S \times B_0$, if $t \notin F_0$. Hence 4) of Condition 17.2 is satisfied.

Next, let $a \in B_0$ and let as above $F_1(a)$ be defined by

$$F_1(a) = \{t : q(t,a) > 0\}.$$

From hypothesis (iv) we know that $F_1(a) \subset F_0$.

From hypothesis (v) follows, that, if $t \in F_1(a)$, then $q(t,a) \geq c_0$ and from the assumptions we have made in Example 18.2 regarding the HMM $\mathcal{H}_2$, we also know that $q(t,a) \leq C_0$.

From hypothesis (vi) we also know that $p(s,t) \geq c_1$ for all $(s,t) \in F_0 \times F_0$ and from the assumptions made in Example 18.2, we also know that $p(s,t) \leq C_1$ for all $s,t \in S$. Hence, if we define $d_0 = c_1c_0$ and $D_0 = C_0C_1$ and recall that $m(s,t,a) = p(s,t)q(t,a)$, we find that

$$d_0 \leq m(s,t,a) \leq D_0,\ \ if\ \ (s,t) \in F_0 \times F_1(a)$$

and that

$$m(s,t,a) = 0, \;\; if \;\; s \in F_0 \;\; and \;\; t \notin F_1(a).$$

Hence, also condition 5) of Condition 17.2 is satisfied.

It remains to verify condition 6) of Condition 17.2. Let $B \subset B_0$ and let $s \in F_0$. We find:

$$\int_{F_0} \int_B m(s,t,a)\lambda(dt)\tau(da) = \int_B \int_{F_1(a)} p(s,t)q(t,a)\lambda(dt)\tau(da) \geq d_0\beta_1 \int_B \tau(da).$$

Hence, if we define $\beta_0 = d_0\beta_1$, we find that also condition 6) of Condition 17.2 is satisfied. Hence Condition 17.2 is satisfied and hence Condition E is satisfied because of Proposition 17.3. The conclusion then follows from Theorem 11.1. □

We shall end this section with a simple and elementary example, which shows, that even, if the HMM has densities with regular partition and is such that the hidden Markov chain is uniformly ergodic, then the filtering process may not be weakly ergodic. The filtering process may even be periodic. The example is a straight forward modification of Example 11.1 in [16], which in its turn is an extension of the example presented in [13], section 10. Unfortunately our presentation is somewhat tedious.

Example 18.3 *Let $S = [-4, 4]$, let $\delta_0 : S \times S \to [0, 8]$ be the Euclidean metric and let $\mathcal{F}$ be the Borel field generated by the metric δ_0. We set $S_1 = [-4, -3)$, $S_2 = [-3, -2)$, $S_3 = [-2, -1)$, $S_4 = [-1, 0]$ $S_5 = [0, 1)$, $S_6 = [1, 2)$, $S_7 = [2, 3)$, $S_8 = [3, 4]$ and we let λ denote the Borel-Lebesque measure on $(S, \mathcal{F})$.*

We now define a probability density kernel $p : S \times S \to [0, \infty)$ as follows:

1. *If $s \in S_1$ then $p(s,t) = 1/2$, $if\ t \in S_1 \cup S_5$ and $p(s,t) = 0\ if\ t \notin S_1 \cup S_5$;*
2. *if $s \in S_2$ then $p(s,t) = 1/2$, $if\ t \in S_2 \cup S_6$ and $p(s,t) = 0\ if\ t \notin S_2 \cup S_6$;*
3. *if $s \in S_3$ then $p(s,t) = 1/2$, $if\ t \in S_4 \cup S_8$ and $p(s,t) = 0\ if\ t \notin S_4 \cup S_8$;*
4. *if $s \in S_4$ then $p(s,t) = 1/2$, $if\ t \in S_3 \cup S_7$ and $p(s,t) = 0\ if\ t \notin S_3 \cup S_7$;*
5. *if $s \in S_5$ then $p(s,t) = 1/2$, $if\ t \in S_1 \cup S_8$ and $p(s,t) = 0\ if\ t \notin S_1 \cup S_8$;*
6. *if $s \in S_6$ then $p(s,t) = 1/2$, $if\ t \in S_2 \cup S_7$ and $p(s,t) = 0\ if\ t \notin S_2 \cup S_7$;*
7. *if $s \in S_7$ then $p(s,t) = 1/2$, $if\ t \in S_4 \cup S_5$ and $p(s,t) = 0\ if\ t \notin S_4 \cup S_5$;*
8. *if $s \in S_8$ then $p(s,t) = 1/2$, $if\ t \in S_3 \cup S_6$, and $p(s,t) = 0\ if\ t \notin S_3 \cup S_6$.*

Let $P \in \mathcal{TP}((S, \mathcal{F})$ be the tr.pr.f determined by p. It is easily seen that the Markov chain generated by P is uniformly ergodic and has the **uniform measure** *on S as stationary distribution.*

Next, let $A = \{a, b\}$, and define the partition function $m : S \times S \times A \to [0, \infty)$ simply by
$m(s,t,a) = p(s,t), \;\; if \;\; -4 \leq t < 0,$
$m(s,t,a) = 0, \;\; if\ 0 \leq t \leq 4,$
$m(s,t,b) = 0, \;\; if \;\; -4 \leq t < 0,$
$m(s,t,b) = p(s,t), \; if\ 0 \leq t \leq 4.$

Let τ be the counting measure on A, let $\mathcal{A}$ be the subsets of A and let ϱ be the discrete measure on A.

Let $\mathcal{H}_3$ denote the HMM with densities defined by Let

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

(End of Example 18.3.)

Since A is finite and the tr.p.f p is bounded, the partition (m, τ) of $\mathcal{H}_3$ is regular.

Proposition 18.2 *Let $\mathcal{H}_3 = \{(S, \mathcal{F}, \delta_0), (p, \lambda), (A, \mathcal{A}, \varrho), (m, \tau)\}$ be defined as in Example 18.3. As usual, let $K = \mathcal{P}_\lambda(S, \mathcal{F})$, let δ_{TV} be the total variation metric on K and let $\mathcal{E}$ be the $\sigma-algebra$ generated by δ_{TV}. Let $0 < \alpha < 1/2$. Let $x_0 \in K$ be defined by*
$x_0(F) = \alpha\lambda(F)/2,\ \ if\ F \subset S_1$ or $F \subset S_3$;
$x_0(F) = (1-\alpha)\lambda(F)/2,\ \ if\ F \subset S_2$ or $F \subset S_4$;
$x_0(F) = 0,\ \ if\ F \subset S_5 \cup S_6 \cup S_7 \cup S_8$.
*Let **P** be the filter kernel induced by the HMM $\mathcal{H}_3$, and let $\{Z_n(x_0), n = 0, 1, 2, ...\}$ denote the Markov chain generated by **P** and the starting point $x_0 \in K$.*

*Then, the Markov chain $\{Z_n(x_0), n = 0, 1, 2, ...\}$ is a **periodic** Markov chain on the finite set $K_1 = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\} \subset K$ defined as follows:*

1. *$x_1 = x_0$;*

2. *x_2 is defined by*
   $x_2(F) = \alpha\lambda(F)/2,\ if\ F \subset S_1$ or $F \subset S_4$,
   $x_2(F) = (1-\alpha)\lambda(F)/2\ if\ F \subset S_2$ or $F \subset S_3$,
   $x_2(F) = 0\ if\ F \subset S_5 \cup S_6 \cup S_7 \cup S_8$

3. *x_3 is defined by*
   $x_3(F) = \alpha\lambda(F)/2,\ if\ F \subset S_2$ or $F \subset S_4$,
   $x_3(F) = (1-\alpha)\lambda(F)/2\ if\ F \subset S_1$ or $F \subset S_3$,
   $x_3(F) = 0\ if\ F \subset S_5 \cup S_6 \cup S_7 \cup S_8$;

4. *x_4 is defined by*
   $x_4(F) = \alpha\lambda(F)/2,\ if\ F \subset S_2$ or $F \subset S_3$,
   $x_4(F) = (1-\alpha)\lambda(F)/2\ if\ F \subset S_1$ or $F \subset S_4$,
   $x_4(F) = 0\ if\ F \subset S_5 \cup S_6 \cup S_7 \cup S_8$;

5. *x_5 is defined by*
   $x_5(F) = 0\ if\ F \subset S_1 \cup S_2 \cup S_3 \cup S_4$,
   $x_5(F) = \alpha\lambda(F)/2,\ if\ F \subset S_5$ or $F \subset S_7$,
   $x_5(F) = (1-\alpha)\lambda(F)/2\ if\ F \subset S_6$ or $F \subset S_8$;

6. *x_6 is defined by*
   $x_6(F) = 0\ if\ F \subset S_1 \cup S_2 \cup S_3 \cup S_4$,
   $x_6(F) = \alpha\lambda(F)/2,\ if\ F \subset S_5$ or $F \subset S_8$,
   $x_6(F) = (1-\alpha)\lambda(F)/2\ if\ F \subset S_6$ or $F \subset S_7$;

7. *x_7 is defined by*
   $x_7(F) = 0\ if\ F \subset S_1 \cup S_2 \cup S_3 \cup S_4$,
   $x_7(F) = \alpha\lambda(F)/2,\ if\ F \subset S_6$ or $F \subset S_8$,
   $x_7(F) = (1-\alpha)\lambda(F)/2\ if\ F \subset S_5$ or $F \subset S_7$;

8. x_8 *is defined by*
$x_8(F) = 0 \; if \; F \subset S_1 \cup S_2 \cup S_3 \cup S_4$,
$x_8(F) = \alpha\lambda(F)/2, \; if \; F \subset S_6 \; or \; F \subset S_7$,
$x_8(F) = (1-\alpha)\lambda(F)/2 \; if \; F \subset S_5 \; or \; F \subset S_8$.

Proof. Let $\{(K,\mathcal{F}),(A,\mathcal{A}),h,(g,\tau)\}$ be the random mapping associated to the HMM $\mathcal{H}_3$. We then find that

$$g(x_1,a) = ||x_1M(a)|| = \int_{-4}^{0}(1/2)x_1(ds) = 1/2.$$

Similarly we find that $g(x_1,b) = 1/2$ and generally we find that
$g(x_i,a) = 1/2, \; i = 1,2,...,8 \; and \; \; g(x_i,b) = 1/2, \; \; i = 1,2,...,8$.

Regarding the function $h : K \times A \to K$ we find that, if $F \subset S_1$, then

$$h(x_1,a)(F) = x_1M(a)(F)/||x_1M(a)|| = 2x_1M(a)(F) =$$

$$2\int_{-4}^{-3}\int_F (1/2)dtx_1(ds) + 2\int_{-3}^{-2}\int_F 0\cdot x_1(ds)+$$

$$2\int_{-2}^{-1}\int_F 0\cdot dtx_1(ds) + 2\int_{-1}^{-0}\int_F 0\cdot x_1(ds) =$$

$$2\int_{-4}^{-3}\int_F (1/2)dtx_1(ds) = 2(1/2)\lambda(F)\alpha/2\int_{-4}^{-3}\lambda(ds) = \alpha\lambda(F)/2.$$

If instead $F \subset S_2$, we find in a similar way that

$$h(x_1,a)(F) = (1-\alpha)\lambda(F)/2,$$

if $F \subset S_3$ we find

$$h(x_1,a)(F) = (1-\alpha)\lambda(F)/2,$$

and if $F \subset S_4$ we find

$$h(x_1,a)(F) = \alpha\lambda(F)/2.$$

Hence $h(x_1,a) = x_2$. Generally we find: $h(x_1,a) = x_2, \;\; h(x_1,b) = x_6, h(x_2,a) = x_1, \;\; h(x_2,b) = x_5, h(x_3,a) = x_4, \;\; h(x_3,b) = x_8, h(x_4,a) = x_3, \;\; h(x_4,b) = x_7, h(x_5,a) = x_2, \;\; h(x_1,b) = x_6, h(x_6,a) = x_1, \;\; h(x_6,b) = x_7, h(x_7,a) = x_4, \;\; h(x_7,b) = x_8, h(x_8,a) = x_3, \;\; h(x_8,b) = x_5$.

Next, if we as usual let $\mathbf{P}$ denote the filter kernel induced by the HMM under consideration, it follows from the equalities above, that $\mathbf{P}$ satisfies the following equalities:

$$\mathbf{P}(x_1,\{x_i\}) = 1/2, \; if \; i = 2,6; \quad \mathbf{P}(x_2,\{x_i\}) = 1/2, \; if \; i = 1,5;$$

$$\mathbf{P}(x_3,\{x_i\}) = 1/2, \; if \; i = 4,8; \quad \mathbf{P}(x_4,\{x_i\}) = 1/2, \; if \; i = 3,7;$$

$$\mathbf{P}(x_5,\{x_i\}) = 1/2, \; if \; i = 2,6; \quad \mathbf{P}(x_6,\{x_i\}) = 1/2, \; if \; i = 1,7;$$

$$\mathbf{P}(x_7,\{x_i\}) = 1/2, \; if \; i = 4,8; \quad \mathbf{P}(x_8,\{x_i\}) = 1/2, \; if \; i = 3,5;$$

$$\mathbf{P}(x_i,\{x_j\}) = 0 \; otherwise.$$

Therefore, if we define the 8×8 matrix $P_{special}$ by

$$(P_{special})_{i,j} = \mathbf{P}(x_i,\{x_j\}), \;\; 1 \le i,j \le 8$$

we find that the Markov chain generated by $P_{special}$ is irreducible and has *period* 2. Consequently $\{\mathbf{P}^n(x_0,\cdot), n = 1, 2, ...\}$ can not converge in distribution in spite of the fact that the hidden Markov chain associated to the HMM is uniformly ergodic. Furthermore, note, that if we define the measure $\mu_{x_0} \in \mathcal{P}(K, \mathcal{E})$ by

$$\mu_{x_0}(\{x_i\}) = 1/8, \; i = 1, 2, ..., 8,$$

then

$$\mu_{x_0}\mathbf{P} = \mu_{x_0}. \; \square$$

References

[1] C. Anton Popescu, "Applications of Dependence with Complete Connections to Hidden Markov Models", *Bulletin of the Transilvania University of Brasov, Series III: Mathematics, Information, Physics*, **5**, 2012, Special Issue: Proceedings of the Seventh Congress of Romanian Mathematicians, pp 23-32, (2012).

[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, Prob Stat*, **24**, 367-394, (1988).

[3] L.E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains", *Ann of Math Stat*, **37**, 164-71, (1966).

[4] D. Blackwell, "The entropy of functions of finite-state Markov chains", *Trans First Prague Conf Inf Th.*, Prague, (1957).

[5] O. Cappé, E. Moulines and T. Ryden, *Inference in Hidden Markov Models*, Springer Series in Statistics, Springer, 2005.

[6] P. Chigansky and R. van Handel, "A complete solution to Blackwell's unique ergodicity problem for hidden Markov chains',' *Ann Appl Prob*, **20**, 2318-45, (2010).

[7] P. Diaconis and D. Freedman, "Iterated random functions", *SIAM Rev*, **41**, 45-76, (1989).

[8] R.M. Dudley, *Real analysis and probability*, Cambridge University Press, 2002.

[9] H. Furstenberg and H. Kesten, "Products of random matrices", *Ann Math Stat*, **31**, pp 457-469, (1960)

[10] M. Iosifescu and R. Theodorescu, *Random Processes and Learning*, Springer, 1969.

[11] M. Iosifescu and S. Grigorescu, *Dependence with complete connections and its applications*, Cambridge University Press, Cambridge, 1990.

[12] T. Kaijser, "A limit theorem for the conditional distributions of Markov chains with incomplete state information", Report LiH-Mat-R-73-7, *Department of Mathematics, Linköping University*, (1973).

[13] T. Kaijser, "A limit theorem for partially observed Markov chains", *Ann of Prob*, **3**, 677-96, (1975).

[14] T. Kaijser, "On a new contraction condition for random systems with complete connections',' *Rev. Roumaine Math. Pure Appl.*, **26**, 1075 - 1117, (1981).

[15] T. Kaijser, "On Markov chains induced by partitioned transition probability matrices", *arXiv:0907.4502,* (2009).

[16] T. Kaijser, "On Markov chains induced by partitioned transition probability matrices", *Acta Math Sinica*, **20**, 441-476, (2011).

[17] F. Kochman and J. Reeds, "A simple proof of Kaijser's unique ergodicity result for hidden Markov α-chains", *Ann Appl Prob*, **16**, 1805 - 15, (2006).

[18] H. Kunita, "Asymptotic behaviour of the nonlinear filtering errors of Markov Processes", *J Multivariate Analysis* , **1**, 365-93, (1971)

[19] T. Lindwall, *Lectures on the coupling method*, Wiley and Sons, 1992.

[20] F. Norman, *Markov Processes and Learning Models*, Academic Press, New York, 1972.

[21] K.R. Parthasarathy, *Probability measures on metric spaces*, Academic Press, New York 1967

[22] L.R. Rabiner and B.H. Juang , "An introduction to Hidden Markov Models", *IEEE, ASSP Magazine* (1986).

[23] D. Revuz, *Markov chains*, North-Holland/American Elsevier, 1975.

[24] L. Vaserstein, "Markov processes over denumerable products of spaces describing large systems of automata", *Problems of Information Transmission*, **5**, vpp 64-73, (1969).