On Markov chains induced by partitioned transition probability matrices.

Thomas Kaijser

Department of Mathematics and Information Coding Group,

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

Abstract

Let S be a denumerable state space and let P be a transition probability matrix on S. If a denumerable set $\mathcal{M}$ of nonnegative matrices is such that the sum of the matrices is equal to P, then we call $\mathcal{M}$ a *partition* of P.

Let K denote the set of probability vectors on S. To every partition $\mathcal{M}$ of P we can associate a transition probability function $\mathbf{P}_{\mathcal{M}}$ on K defined in such a way that if $p \in K$ and $M \in \mathcal{M}$ are such that $||pM|| > 0$, then, with probability $||pM||$, the vector p is transferred to the vector $pM/||pM||$. Here $|| \cdot ||$ denotes the $l_1 - norm$.

In this paper we investigate convergence in distribution for Markov chains generated by transition probability functions induced by partitions of transition probability matrices.

An important application of the convergence results obtained is to filtering processes of partially observed Markov chains.

Keywords: Markov chains on nonlocally compact spaces, convergence in distribution, conditional state distribution, filtering processes, functions of Markov chains, partially observed Markov chains, hidden Markov chains, partitions, equicontinuity, entropy, Blackwell's entropy formula, Kantorovich metric, products of random matrices, barycenter, subrectangular matrices.

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

1 Introduction

1.1. The transition probability function $\mathbf{P}_{\mathcal{M}}$. Let S be a denumerable set and let P be a transition probability matrix (tr.pr.m) on S. A denumerable set $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ of nonnegative $S \times S$ matrices such that $\sum_{w \in \mathcal{W}} M(w) = P$ will in this paper be called a **partition** of P. We denote the set of all tr.pr.ms on S by $PM(S \times S)$.

Next, let K denote the set of probability vectors on S, let $|| \cdot ||$ denote the $l_1 - norm$, define a metric δ on K by $\delta(x, y) = ||x - y||$ and let $\mathcal{E}$ denote the Borel field induced by δ. Let $\mathcal{P}(K)$ denote the set of probability measures on $(K, \mathcal{E})$. An element x in K will be considered as a row vector. We denote the

$i:th$ coordinate of a generic vector x by $(x)_i$ and we denote the $i,j-th$ element of a generic matrix M by $(M)_{i,j}$.

To every partition $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ of $P \in PM(S \times S)$ one can define a transition probability function (tr.pr.f) $\mathbf{P}_{\mathcal{M}} : K \times \mathcal{E} \to [0,1]$ on $(K, \mathcal{E})$ by

$$\mathbf{P}_{\mathcal{M}}(x,B) = \sum_{w \in \mathcal{W}_{\mathcal{M}}(x,B)} ||xM(w)||, \ \ x \in K, \ \ B \in \mathcal{E} \tag{1}$$

where

$$\mathcal{W}_{\mathcal{M}}(x,B) = \{w \in \mathcal{W} : ||xM(w)|| > 0, xM(w)/||xM(w)|| \in B\}.$$

That $\mathbf{P}_{\mathcal{M}}(x,\cdot)$ is a probability measure for each $x \in K$ is easily proved and that $\mathbf{P}_{\mathcal{M}}(\cdot, B)$ is Borel-measurable for each $B \in \mathcal{E}$ can be proved by fairly standard arguments. The details will be given in the last section.

Let $C[K]$ denote the set of real, bounded, continuous function on $(K, \mathcal{E})$. From the definition of $\mathbf{P}_{\mathcal{M}}(x,\cdot)$ it is easily seen that

$$\int_K u(y)\mathbf{P}_{\mathcal{M}}(x,dy) = \sum_{w \in W_{\mathcal{M}}(x)} u(xM(w)/||xM(w)||)||xM(w)||$$

where

$$W_{\mathcal{M}}(x) = \{w \in \mathcal{W} : ||xM(w)|| > 0\}. \tag{2}$$

Next, let $\mathbf{P}^n_{\mathcal{M}}(\cdot,\cdot)$ denote the n-step tr.pr.f defined recursively by

$$\mathbf{P}^1_{\mathcal{M}}(x,B) = \mathbf{P}_{\mathcal{M}}(x,B), \quad x \in K, \ \ B \in \mathcal{E}$$

$$\mathbf{P}^{n+1}_{\mathcal{M}}(x,B) = \int_K \mathbf{P}^n_{\mathcal{M}}(y,B)\mathbf{P}_{\mathcal{M}}(x,dy) \ \ x \in K, \ \ B \in \mathcal{E}, \ \ n = 1,2,\dots .$$

If the tr.pr.f $\mathbf{P}_{\mathcal{M}}(\cdot,\cdot)$ is such that there exists a probability measure $\mu \in \mathcal{P}(K)$ such that

$$\lim_{n\to\infty} \int_K u(y)\mathbf{P}^n_{\mathcal{M}}(x,dy) = \int_K u(y)\mu(dy), \ \forall u \in C[K], \ \forall x \in K$$

then we say that $\mathbf{P}_{\mathcal{M}}(\cdot,\cdot)$ is **asymptotically stable**.

The main purpose of this paper is to give a *sufficient condition* for asymptotic stability of $\mathbf{P}_{\mathcal{M}}(\cdot,\cdot)$ when the tr.pr.m P on S is irreducible, aperiodic and positively recurrent.

1.2. Motivation. The interrelationship with the filtering process. Let $(S, \mathcal{S})$ be a measurable space, and let $P : S \times \mathcal{S} \to [0,1]$ be a tr.pr.f on $(S, \mathcal{S})$. Let $(A, \mathcal{A})$ be another measurable space and let $R : S \times \mathcal{A} \to [0,1]$ be a tr.pr.f from $(S, \mathcal{S})$ to $(A, \mathcal{A})$. Let $\mathcal{P}(S)$ denote the set of probability measures on $(S, \mathcal{S})$. It is well-known (see e.g [11] or [24] for the case when S and A are finite sets) that one to each $p_0 \in \mathcal{P}(S)$ can define two *stochastic processes* $\{X_0, X_1, X_2, X_3, \dots\}$ and $\{Y_1, Y_2, Y_3, \dots\}$ taking values in S and A respectively, such that

1) $\{X_n, n = 0,1,2,...\}$ is a Markov chain with tr.pr.f P and initial distribution p_0;

2) for $n = 1,2,3,...$ and $D \in \mathcal{S}$

$$Pr[X_{n+1} \in D | X_0 = i_0, X_1 = i_1, Y_1 = a_1, X_2 = i_2, Y_2 = a_2, \dots, X_n = i, Y_n = a_n] =$$

$$Pr[X_{n+1} \in D | X_n = i] = P(i, D)$$

3) for $n = 1, 2, 3, ...$ and $B \in \mathcal{A}$

$$Pr[Y_n \in B | X_0 = i_0, X_1 = i_1, Y_1 = a_1, X_2 = i_2, Y_2 = a_2, \dots, X_n = i] =$$

$$Pr[Y_n \in B | X_n = i] = R(i, B).$$

The Markov chain $\{X_n\}$ is often called the *hidden* process, whereas the process $\{Y_n\}$ is often called the *observation* process.

Next let $Z_n, n = 1, 2, ...$ denote the *conditional distribution* of X_n given the observations $Y_1, Y_2, ..., Y_n$. We call Z_n the *conditional state distribution*. The process $\{Z_n, n = 1, 2, ...\}$ is often called the *filtering process*, in particular in engineering literature.

For more than half a century scientists have worked on the problem to draw conclusions about the hidden process $\{X_n\}$ from the observed process $\{Y_n\}$, under various assumptions about the spaces $(S, \mathcal{S})$ and $(A, \mathcal{A})$ and the tr.pr.fs $P : S \times \mathcal{S} \to [0, 1]$ and $R : S \times \mathcal{A} \to [0, 1]$. Also much work has been done when both the hidden process and the observed process are time-continuous processes.

The following is a short list of natural problems:
1) *Use the observations* $Y_1, Y_2, ...$ *to estimate the parameters of the tr.pr.f* P *and the tr.pr.f* R. (An early paper regarding this problem for the case when both the space S and A are finite is the paper [2] by L.Blum and T.Petrie.)
2) *Use the observations* $Y_1, Y_2, ..., Y_n$ *to give an estimation of both* X_n *and* X_{n+1}. (When both the hidden process and the observed process are linear Gaussian processes this problem falls into to the theory called *Kalman filter theory*.)
3) *How does* Z_n *depend on the initial distribution* p_0*?* If Z_n becomes insensitive to the initial distribution p_0 of X_0, this property is often called the *stability property* of the filtering problem (see e.g [14]) or simply the *forgetting of initial distribution property* (see e.g [5], chapter 4). This problem has been studied actively in the last few years, and we refer to the recent paper [14] by van Händel for more information about this important problem. In [14] references to other work can be found some of which have a detailed reference list.
4) *Does the filtering process* $\{Z_n\}$ *converge in distribution and if so is the limit distribution independent of the initial distribution* p_0*?* This problem is in the center of this paper.

It seems fair to say that the first three problems are important from the practical point of view. Roughly speaking problem 1) deals with the problem of how to determine the most accurate model, problem 2) deals with the problem of how to make the "best" estimation given the observations and problem 3) deals with the problem of how dependent our "best" estimation is with respect to initial assumptions. The fourth problem is perhaps mainly of theoretical interest but since this problem is related to the computation of the entropy rate of the observation process it is perhaps fair to also consider problem 4) as being of importance from a practical point of view.

One of the first papers within this field of probability theory is the paper [4] from 1957 by D. Blackwell. He considered the case when S is a finite set, implying that the $\{X_n\}-process$ is an ordinary finite Markov chain with finite-dimensional tr.pr.m P, and he assumed that the $\{Y_n\}-process$ was determined by a "lumping" function $g : S \to A$ such that $Y_n = g(X_n)$. The main result in

[4] is a formula for the entropy rate of the $\{Y_n\} - process$, a formula based on the existence of a stationary measure for the filtering process Z_n. Blackwell also showed that the filtering process itself is a *Markov chain* and proved the existence of a unique stationary measure for the filtering process when the tr.pr.m $P \in P(S \times S)$ has "*nearly identical rows and no element which is very small*".

A classical paper in this field of probability is the paper [20] form 1971 by H.Kunita. Kunita assumes that the hidden process is a time-continuous Markov process on a compact, separable Hausdorff space, and that the observed process is also a time-continuous processes taking values in R^n and such that it can be described as a stochastic integral based on the Wiener process. Among other things also the question regarding convergence in distribution of the filtering process is considered in [20].

Another rather early paper in this field is the paper [16] from 1975. In [16] it is assumed that both the space S and the space A are finite and that the observation process is determined by a lumping function $g : S \to A$. It was proved that the distributions of the filtering process converge in distribution towards a unique limit distribution independent of the initial distribution p_0 when P is irreducible and aperiodic and a rather mild extra condition holds. (See Condition A of [16].) In [16] an example was given that shows that there need not be a unique limit measure for the filtering process even if the Markov chain $\{X_n\}$ is irreducible and aperiodic.

In the paper [19] from 2006 by F.Kochman and J.Reeds, the result obtained in [16] was generalized and the proof simplified. Condition A of [16] was replaced by a *"rank 1 condition"* for certain matrix products that arise naturally when studying the filtering process for denumerable sets S and A.

The same year the paper [15] by T.Holliday, P.Glynn and A.Goldsmith was published in which also a convergence theorem regarding the distributions of the filtering process is proved for the case when both the sets S and A are finite sets. The result in [15] is proved under a positivity condition for the tr.pr.f (tr.pr.m) from S to A, which is stronger than Condition A introduced in [16].

In the paper [8] from 2005 by G.B.DiMasi and L.Stettner the authors consider the filtering process when the set $(S, \mathcal{S})$ is a complete, separable, metrizable space. They also require that a positivity condition shall be satisfied, when proving that the distributions of the filtering process converge in distribution to a unique limit measure. They use the so called Hilbert norm to measure distances between measures. One drawback with the Hilbert norm is that it is equal to infinity if the supports of the two measures under consideration are different, and it seems that it is for this reason the authors of [8] need to make certain uniform, positivity assumptions.

Although a few papers have been published in the last decade, roughly speaking, the convergence problem for the distributions of filtering process (problem 4 above) has not been considered very often in the literature, in particular if one compares with the number of papers dealing with the so called stability problem for the filtering process. One reason could be due to the fact that, if the space $(S, \mathcal{S})$ on which the hidden process takes its values is *not compact*, then the set of probability measures on the set of probability measures on $(S, \mathcal{S})$ is a rather complicated set. Already if S is a denumerable set and we measure the distance between two probability vectors on S by the $l_1 - norm$ then the set of of probability vectors on S will be a nonlocally compact set, and therefore the Markov chain associated to the filtering process will be a Markov chain on a nonlocally compact space. Moreover, if both the set S and the set A are denumerable, then in general the tr.pr.f of the filtering process will not satisfy a *Doeblin condition* nor be ψ*-irreducible.* (See e.g [9] or [5] for the definitions of the Doeblin condition and ψ-irreducibility.) In fact if we consider two filtering processes having different initial distributions then, in a generic situation, the supports of the distributions of the two filtering processes will be *non-overlapping*, which causes some technical complications.

We shall now show how the filtering processes described above are interrelated with Markov chains generated by tr.pr.fs induced by partitions of tr.pr.ms, when the spaces S and A are denumerable.

Thus, let now S denote a denumerable space and let P be a tr.pr.m on S. Let A be another denumerable space, let $R = \{(R)_{i,a} : i \in S, a \in A\}$ be a tr.pr.m. from S to A and let $p_0 \in \mathcal{P}(S)$. As described above (see also [24]), let

$$\{X_0, X_1, X_2, X_3, ... \} \ and \ \{Y_1, Y_2, Y_3, ... \}$$

be the two stochastic processes taking values in S and A respectively, that are generated by p_0, P and R.

For $n = 1, 2, ...$ and $i \in S$ let

$$Z_{n,i} = Pr[X_n = i|Y_1, Y_2, ..., Y_n] \tag{3}$$

and set

$$Z_n = (Z_{n,i}, \ \ i \in S). \tag{4}$$

Next let us for each $a \in A$ define a $S \times S$ matrix $M(a)$ by

$$(M(a))_{i,j} = (P)_{i,j}(R)_{j,a}, \ \ i \in S, \ j \in S \ . \tag{5}$$

It is easily checked that

$$\sum_{a \in A} M(a) = P$$

and hence if we set $\mathcal{M}_0 = \{M(a) : a \in A\}$ then $\mathcal{M}_0$ is a partition of P.

The interesting fact with this partition is the fact that the distribution of the conditional state distribution Z_n is given by $\mathbf{P}^n_{\mathcal{M}_0}(p_0, \cdot)$ that is

$$Pr[Z_n \in B] = \mathbf{P}^n_{\mathcal{M}_0}(p_0, B), \ B \in \mathcal{E}, ; n = 1, 2, ...$$

a relation which in principal was proved already in [1]. (See also [26], [16] and in particular [19].) Hence in order to prove convergence in distribution of the distributions of the filtering process it suffices to prove convergence in distribution of the Markov chain generated by the tr.pr.f $\mathbf{P}_{\mathcal{M}_0}(\cdot, \cdot)$ induced by the partition $\mathcal{M}_0 = \{M(a) : a \in A\}$ of P, where thus a matrix $M(a)$ of $\mathcal{M}_0$ is defined by (5).

The main application of the convergence result regarding Markov chains generated by tr.pr.fs induced by partitions of tr.pr.ms, proved in this paper, is thus to the filtering process $\{Z_n, n = 1, 2, ...\}$ for denumerable spaces S and A. We also use the convergence result to generalize Blackwell's entropy formula for functions of Markov chains from finite state spaces to denumerable state spaces.

1.3. The main theorem. Let S be a denumerable set and let as above K denote the set of probability vectors on S that is

$$K = \{x = ((x)_i, i \in S) : \sum_{i \in S} (x)_i = 1,\ (x)_i \geq 0\}. \tag{6}$$

Let M be a $S \times S$ matrix. We define the norm $||M||$ by

$$||M|| = \sup\{||xM|| : ||x|| = 1, x \in \mathtt{R}^S\}, \tag{7}$$

where thus $||\cdot||$ denotes the $l_1 - norm$ and $\mathtt{R}$ the real numbers.

Let $\mathcal{U}$ denote the set of $S - dimensional$ vectors specified by

$$\mathcal{U} = \{u = ((u)_i, i \in S) : u_i \geq 0, \quad and \quad \sup\{u_i : i \in S\} = \ 1 \ \}$$

and let $\mathtt{W}$ denote the set of $S \times S$ matrices defined by

$$\mathtt{W} = \{W = u^c v : u \in \mathcal{U}, v \in K\}$$

where u^c denotes the transpose of u. Note that if $W \in \mathtt{W}$ then $||W|| = 1$ since $0 \leq \sum_{j \in S} u_i v_j = u_i \leq 1$ for all $i \in S$ and $\sup_i u_i = 1$. We call an element in $\mathtt{W}$ a *nonnegative rank 1 matrix of norm 1.*

Next, let P be a tr.p.m on S and let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ be a partition of P. If $\{w_1, w_2, ..., w_m\}$ is a finite sequence of elements in $\mathcal{W}$ we use the notations

$$\mathbf{w^m} = \{w_1, w_2, ...w_m\},$$

and

$$\mathbf{M}(\mathbf{w^m}) = M(w_1)M(w_2)...M(w_m).$$

Our first condition is a rather straight forward generalization of a "rank 1 condition" introduced in [19] for the case when the state space S is a finite set. (See [19] page 1807.) Here and throughout this paper we let $e^i,\ i \in S$ denote the vector in K defined by

$$(e^i)_i = 1. \tag{8}$$

Condition B1. *There exists a nonnegative rank 1 matrix $W = u^c v$ of norm 1, a sequence of integers $\{n_1, n_2, ...\}$ and a sequence $\{\mathbf{w_j^{n_j}}, j = 1, 2, ...\}$ of sequences $\mathbf{w_j^{n_j}} = \{w_{1,j}, w_{2,j}, ..., w_{n_j,j}\} \in \mathcal{W}^{n_j}$, such that $||\mathbf{M}(\mathbf{w_j^{n_j}})|| > 0, \quad j = 1, 2, ...$ and such that for all $i \in S$*

$$\lim_{j \to \infty} ||e^i \mathbf{M}(\mathbf{w_j^{n_j}}) / ||\mathbf{M}(\mathbf{w_j^{n_j}})|| - e^i W|| = 0.$$

It is not difficult to prove that if the state space S is finite and the partition is determined by an "observation matrix" R (see (5)), then Condition B1 is equivalent to the " rank 1 condition" introduced in [19].

In order to define our next condition we first need to introduce the notion **barycenter**. The barycenter of a measure $\mu \in \mathcal{P}(K)$ is defined as that vector $\overline{b}(\mu) \in K$ whose $i : th$ coordinate $(\overline{b}(\mu))_i$ is defined by

$$(\overline{b}(\mu))_i = \int_K (x)_i \mu(dx). \tag{9}$$

That the vector $\overline{b}(\mu)$ belongs to K follows from the fact that the set K is convex. We let $\mathcal{P}(K|q)$ denote the subset of $\mathcal{P}(K)$ such that each $\mu \in \mathcal{P}(K|q)$ has barycenter equal to q.

We are now ready to introduce the condition under which the main theorem of this paper is proved. Let S be a denumerable set, let P be a tr.p.m on S such that P is irreducible, aperiodic and positively recurrent, let π denote the unique probability vector in K such that $\pi P = \pi$ and let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ be a partition of P. Since P is irreducible and positively recurrent it follows that $(\pi)_i > 0$, $\forall i \in S$.

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*

$$\mu(C \cap \{x : (x)_{i_0} > (\pi)_{i_0}/2\}) \geq (\pi)_{i_0}/3, \quad \forall\, \mu \in \mathcal{P}(K|\pi), \tag{10}$$

then we can find an integer N, and a sequence $\{w_1, w_2, ...w_N\}$ of elements in $\mathcal{W}$, such that, if we set

$$\mathbf{M}(\mathbf{w^N}) = M(w_1)M(w_2)...M(w_N),$$

then

$$||e^{i_0}\mathbf{M}(\mathbf{w^N})|| > 0$$

and if $x \in C \cap \{x : (x)_{i_0} > (\pi)_{i_0}/2\}$ then also

$$||(x\mathbf{M}(\mathbf{w^N})/||x\mathbf{M}(\mathbf{w^N})|| - e^{i_0}\mathbf{M}(\mathbf{w^N})/||e^{i_0}\mathbf{M}(\mathbf{w^N})||)|| < \rho.$$

That there exists a compact set C such that (10) holds is proved in section 4.

It is not very difficult to prove that Condition B1 implies Condition B when P is irreducible, aperiodic and positively recurrent, a fact we shall prove in section 9.

The main theorem of this paper reads as follows:

Theorem 1.1 *Let S be a denumerable set, let $P \in PM(S \times S)$ be irreducible, aperiodic and positively recurrent, let $\pi \in K$ satisfy $\pi P = \pi$ and let $\mathcal{M}$ be a partition of P. Suppose also that Condition B holds. Then $\mathbf{P}_{\mathcal{M}}$ is asymptotically stable.*

1.5. Some remarks about the proof of the main theorem. In this subsection we first introduce some further notations and concepts and state a few obvious but important properties for partitions.

As usual let S denote a denumerable set, let K denote the set of probability vectors on S and let $\mathcal{E}$ be the Borel field induced by the $l_1 - metric$ on K. Let $F[K]$ denote the set of real, bounded functions on K, let $B[K]$ denote the set of real, bounded, $\mathcal{E} - measurable$ functions on K, for $u \in C[K]$ define $\gamma(u) = \sup\{|u(x) - u(y)|/||x - y|| : x, y \in K,\ x \neq y\}$, let $Lip[K] = \{u \in C[K] : \gamma(u) < \infty\}$ and $Lip_1[K] = \{u \in Lip[K] : \gamma(u) \leq 1\}$.

Now let $P \in PM(S \times S)$, let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ be a partition of P and let $\mathbf{P}_{\mathcal{M}}$ denote the tr.pr.f on $(K, \mathcal{E})$ induced by $\mathcal{M}$. The mapping $T_{\mathcal{M}} : F[K] \to F[K]$ defined by

$$T_{\mathcal{M}}u(x) = \sum_{w \in \mathcal{W}(x)} u(xM(w)/||xM(w)||)||xM(w)||,\ x \in K$$

where as above $\mathcal{W}_{\mathcal{M}}(x) = \{w \in \mathcal{W} : ||xM(w)|| > 0\}$ (see (2)), will be called the **transition operator** induced by $\mathcal{M}$. For $u \in F[K]$ it is clear from the definition of $\mathbf{P}_{\mathcal{M}}(\cdot,\cdot)$ (see (1)) that

$$T_{\mathcal{M}}u(x) = \sum_{w \in \mathcal{W}(x)} u(xM(w)/||xM(w)||)||xM(w)|| = \int_K u(y)\mathbf{P}_{\mathcal{M}}(x, dy).$$

The mapping $\breve{P}_{\mathcal{M}} : \mathcal{P}(K) \to \mathcal{P}(K)$ defined by

$$\breve{P}_{\mathcal{M}}\mu(B) = \int_K \mathbf{P}_{\mathcal{M}}(x, B)\mu(dx), \ B \in \mathcal{E}$$

will be called the **transition probability operator** induced by $\mathcal{M}$.

For $u \in B[K]$ and $\mu \in \mathcal{P}(K)$ we shall - when convenient - write

$$\langle u, \mu \rangle = \int_K u(x)\mu(dx).$$

It is well-known that

$$\langle T_{\mathcal{M}}u, \mu \rangle = \langle u, \breve{P}_{\mathcal{M}}\mu \rangle. \tag{11}$$

(See e.g. [25], chapter 1, section 1.)

A crucial property for partitions of matrices is the following. Let P_1 and P_2 be two tr.pr.ms in $PM(S \times S)$, let $\mathcal{M}_1 = \{M(w_1) : w_1 \in \mathcal{W}_1\}$ be a partition of P_1 and let $\mathcal{M}_2 = \{M(w_2) : w_2 \in \mathcal{W}_2\}$ be a partition of P_2. Define the set

$$\mathcal{M}_3 = \{M(w_1)M(w_2) : w_1 \in \mathcal{W}_1, w_2 \in \mathcal{W}_2\}. \tag{12}$$

Then $\mathcal{M}_3$ is a partition of the matrix P_1P_2. The proof is elementary. We call $\mathcal{M}_3$ the *product* of $\mathcal{M}_1$ and $\mathcal{M}_2$ and we write $\mathcal{M}_3 = \mathcal{M}_1 \cdot \mathcal{M}_2$. It is also elementary to show that if we have three partitions $\mathcal{M}_1, \mathcal{M}_2$ and $\mathcal{M}_3$ with matrices of the same format then

$$(\mathcal{M}_1 \cdot \mathcal{M}_2) \cdot \mathcal{M}_3 = \mathcal{M}_1 \cdot (\mathcal{M}_2 \cdot \mathcal{M}_3). \tag{13}$$

Therefore, if $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ is a partition of P, then $\mathcal{M}^n$ is well-defined and is a partition of P^n.

Another obvious relation is

$$\sum_{w \in \mathcal{W}} ||xM(w)|| = 1, \ x \in K \tag{14}$$

if $\{M(w) : w \in \mathcal{W}\}$ is a partition of a tr.pr.m P since $\sum_{w \in \mathcal{W}} ||xM(w)|| = ||xP|| = 1$ if $x \in K$.

Now some remarks about the proof of Theorem 1.1. When proving asymptotic stability for a tr.pr.f Q on some general metric space $(\mathcal{K}, \mathcal{A})$ one strategy is to first verify that there exists a point $x^* \in \mathcal{K}$ such that the sequence $\{Q^n(x^*, \cdot), n = 1, 2, ...\}$ of probability measures is a *tight sequence* which together with *Feller continuity* implies that there exists an invariant measure for Q, that is a measure ν such that

$$\int_{\mathcal{K}} Q(x, B)\nu(dx) = \nu(B), \ \ \forall B \in \mathcal{A}.$$

Then, by assuming some kind of *contact condition, recurrence condition* or *contraction condition*, one proves that if the function u belongs to a sufficiently large set of functions, then

$$\lim_{n\to\infty} |T^n u(x) - T^n u(y)| = 0,\ x, y \in \mathcal{K}, \tag{15}$$

where thus T denotes the transition operator associated to the tr.pr.f Q. Finally one uses this limit result to prove both 1) that Q has only one invariant measure and 2) that Q is asymptotically stable.

This is precisely the strategy that we shall use in this paper. In order to prove that there exists a point $x^* \in K$ such that the sequence $\{\mathbf{P}^n_{\mathcal{M}}(x^*, \cdot), n = 1, 2, ...\}$ of probability measures is a tight sequence we shall need two important facts. The first fact is that for each $q \in K$ the set $\mathcal{P}(K|q)$ of probability measures with equal barycenter is a *tight set* of probability measures; this is proved in section 4. The second fact we need is that, if $\pi \in K$ satisfies $\pi P = \pi$ and $\mu \in \mathcal{P}(K|\pi)$, then it follows that $\check{P}_{\mathcal{M}}\mu \in \mathcal{P}(K|\pi)$ for *any* partition $\mathcal{M}$ of P; this is proved in section 5. From these two facts it follows immediately that the sequence $\{\mathbf{P}^n_{\mathcal{M}}(\pi, \cdot), n = 1, 2, ...\}$ is a tight sequence if $\pi P = \pi$.

To prove that (15) is satisfied we first of all use a universal inequality for the transition operator $T_{\mathcal{M}}$ induced by a partition $\mathcal{M}$, which reads as follows:

$$|T_{\mathcal{M}} u(x) - T_{\mathcal{M}} u(y)| \le 3\gamma(u)||x - y||,\ \ x, y \in K,\ u \in Lip[K]. \tag{16}$$

In principal this inequality was already proved in [16], but the universal character of this inequality was not observed in [16].

From (16) follows easily that the sequence $\{T^n_{\mathcal{M}} u,\ n = 1, 2, ...\}$ is an equicontinuous sequence when $u \in Lip[K]$. Now if the set $(K, \mathcal{E})$ would have been a compact set then, as is well-known and easily proved, in order to prove (15) it is enough to verify that to every $\rho > 0$ there exists an integer N and a number $\alpha > 0$ such that for any two initial points x and y in K

$$Pr[\delta(X_N(x), \tilde{X}_N(y)) < \rho] \ge \alpha \tag{17}$$

where thus $\{X_n(x), n = 0, 1, 2, ...\}$ and $\{\tilde{X}_n(y), n = 0, 1, 2, ...\}$ are two independent Markov chains generated by $P_{\mathcal{M}}(\cdot, \cdot)$ the former starting at x the latter at y.

Since, in our case, the space $(K, \mathcal{E})$ is a non-compact space if the set S is denumerable and infinite, one can not in general expect that one can find an integer N and a number $\alpha > 0$ such that (17) holds for *all* $x, y \in K$. For this reason we have introduced a notion we call the *shrinking property* which is defined as follows:

The shrinking property. *Let Q be a tr.pr.f on a metric space $(\mathcal{K}, \mathcal{A})$ where $\mathcal{A}$ is the Borel field generated by a metric ς. Let $Lip[\mathcal{K}]$ denote the set of Lipschitz-functions on $(\mathcal{K}, \mathcal{A})$, 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 $A \subset \mathcal{K}$, every $\eta > 0$ and every $\kappa > 0$, we can find an integer N and another nonempty, compact set $B \subset \mathcal{K}$ such that, if the integer $n \ge N$, then for all $u \in Lip[\mathcal{K}]$*

$$\sup_{x,y\in A} |T^n u(x) - T^n u(y)| \le \eta\gamma(u) + \kappa osc(u) + \alpha\rho\gamma(u)+$$

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

then we say that Q has the shrinking property. We call α a **shrinking number** *associated to ρ.*
Note that a shrinking number α only depends on which ρ is chosen, whereas the integer N and the new nonempty compact set B depend on the initial choice of the nonempty compact set A and depend also on how small we have chosen the numbers η and κ; roughly speaking, a large compact set A, a small η and a small κ require a large integer N and a large compact set B.

Now by showing that the tr.pr.f $\mathbf{P}_{\mathcal{M}}$ has the shrinking property, when Condition B is satisfied, and using an auxiliary theorem for Markov chains having the shrinking property proved in section 6, we are able to prove (15) and then it is a simple matter to prove asymptotic stability, and hence Theorem 1.1.

1.6. Exceptional cases. One consequence of asymptotic stability is that there only exists one *invariant measure*. Therefore if $\mathbf{P}_{\mathcal{M}}$ fulfills the hypotheses of Theorem 1.1 then the equation

$$\int_K \mathbf{P}_{\mathcal{M}}(dx,B)\mu(dx)=\mu(B),\quad \forall B\in\mathcal{E} \tag{18}$$

has a unique solution in $\mathcal{P}(K)$. In the paper [4] Blackwell conjectured that the equation (18) has a unique solution if S is finite, $P\in PM(S\times S)$ is indecomposable and the partition is determined by a lumping function on S. However, there are counterexamples to this conjecture and one such counterexample was presented in [16]. In fact, already in 1974, H.Kesten constructed an example, not published before, which shows that the tr.pr.f $\mathbf{P}_{\mathcal{M}}$ can in fact even be **periodic** ([18]). In section 8 we present this counterexample.

In section 8 we also state and prove a theorem with hypotheses that guarantee that $\mathbf{P}_{\mathcal{M}}$ is *not* asymptotically stable. To state the theorem we need two further notations, $K(x,\mathcal{M})$ and $K_{S'}$. Let $\mathcal{M}=\{M(w):w\in\mathcal{W}\}$ be an arbitrary partition of a tr.pr.m. Recall that $\mathcal{W}_{\mathcal{M}}(x)$ is defined by (2). For each $x\in K$ we define

$$K(x,\mathcal{M})=\{y\in K: y=xM(w)/||xM(w)||\ \ \textit{some}\ \ w\in\mathcal{W}_{\mathcal{M}}(x)\}.$$

Let S be a denumerable set and $S'\subset S$. We define

$$K_{S'}=\{x\in K:(x)_i=0,\ i\notin S'\}.$$

Theorem 1.2 *Let S be a denumerable set, let $P\in PM(S\times S)$ and let $\mathcal{M}=\{M(w):w\in\mathcal{W}\}$ be a partition of P. Now suppose that there exists a subset $S'\subset S$ consisting of at least two elements, such that*
1) for every $x\in K_{S'}$ the set $\cup_{n=1}^{\infty}K(x,\mathcal{M}^n)$ consists of isolated points,
2) if both x and y are in $K_{S'}$ then $\mathcal{W}_{\mathcal{M}^n}(x)=\mathcal{W}_{\mathcal{M}^n}(y),\ \ n=1,2,....$
and
3) if x and y in $K_{S'}$, $n\geq 1$ and $\mathbf{w^n}\in\mathcal{W}_{\mathcal{M}^n}(x)$ then

$$||(x\mathbf{M}(\mathbf{w^n})/||x\mathbf{M}(\mathbf{w^n})||-y\mathbf{M}(\mathbf{w^n})/||y\mathbf{M}(\mathbf{w^n})||)||=||x-y||.$$

If these conditions are fulfilled then $\mathbf{P}_{\mathcal{M}}$ is **not** *asymptotically stable.*

Remark. If we could prove that the hypotheses of Theorem 1.2 are also necessary in order for $\mathbf{P}_{\mathcal{M}}$ to be a tr.pr.f which is *not asymptotically stable* when $P \in PM(S \times S)$ is irreducible, aperiodic and positively recurrent, we would be able to formulate an *easily checked* criterion for deciding whether a tr.pr.f $\mathbf{P}_{\mathcal{M}}$ induced by a partition $\mathcal{M}$ of a tr.pr.m P is asymptotically stable or not. $\Box$

Conjecture 1.1 *Let S be a denumerable set, let $P \in PM(S \times S)$ be irreducible, aperiodic and positively recurrent, let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ be a partition of P and let $\mathbf{P}_{\mathcal{M}}$ be the tr.pr.f induced by $\mathcal{M}$. Then either the hypotheses of Theorem 1.1 or the hypotheses of Theorem 1.2 are satisfied.*

In section 8 we also describe a class of tr.pr.ms and partitions for which the hypotheses of Theorem 1.2 are fulfilled, and show that both Kesten's example and the counterexample in [16] belong to this class.

1.7. Blackwell's entropy formula. Let S be a denumerable set, let $P \in PM(S \times S)$ be irreducible, aperiodic and positively recurrent, let $\{X_0, X_1, X_2, ...\}$ denote a Markov chain with tr.pr.m P and initial distribution π where π satisfies $\pi P = \pi$, let $g : S \to A$ be a "lumping" function on S, let $\mathcal{M} = \{M(a) : a \in A\}$ be the partition of P defined by

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

$$(M(a))_{i,j} = 0 \quad if \quad g(j) \neq a,$$

and define $Y_n,\ n = 0, 1, 2, ...$ by $Y_n = g(X_n), n = 0, 1, 2, ...$ In the paper [4] it was assumed that S is finite, and it was shown that the entropy rate of the $\{Y_n\}$-process is given by

$$\sum_{a \in A} \int_K h(||yM(a)||)\mu(dy)$$

where

$$h(t) = -(1/\ln 2)t\ln(t), \quad if \ \ 0 < t \leq 1, \qquad and \quad h(0) = 0,$$

and where μ is an invariant measure associated to the tr.pr.f $\mathbf{P}_{\mathcal{M}}$ induced by the partition $\mathcal{M}$. In section 13 we generalize the entropy result obtained by Blackwell to Markov chains on denumerable state spaces. By using *convexity properties* proved in section 11 we can also give lower and upper bounds for the entropy.

1.8. The plan of the paper. In section 2 we prove the rather obvious fact that, if S is a denumerable set, P_1 and P_2 belong to $PM(S \times S)$, $\mathcal{M}_1$ is a partition of P_1 and $\mathcal{M}_2$ is a partition of P_2, then

$$\check{P}_{\mathcal{M}_1 \mathcal{M}_2} = \check{P}_{\mathcal{M}_2} \check{P}_{\mathcal{M}_1}$$

from which follows that for any partition $\mathcal{M}$

$$\mathbf{P}^n_{\mathcal{M}} = \mathbf{P}_{\mathcal{M}^n}, \quad n = 1, 2, ...$$

which is a very useful relation.

In section 3 we first prove the universal inequality (16). From this inequality it readily follows that for any *any partition* $\mathcal{M}$

$$\gamma(T^n_{\mathcal{M}}) \leq 3\gamma(u), \quad n = 1, 2, ... \quad \forall u \in Lip[K]. \tag{19}$$

We also introduce the well-known Kantorovich metric for probability measures in $\mathcal{P}(K)$, and we present an example which shows that the number 3 on the right hand side of the inequality (19) can not be decreased to a number less than 2.

In section 4 we consider probability measures with equal **barycenters**. We prove that for any $q \in K$ the set $\mathcal{P}(K|q)$ is a tight set in $\mathcal{P}(K)$ and we also prove that the Kantorovich distance between a measure $\mu \in \mathcal{P}(K|p)$ and the set $\mathcal{P}(K|q)$ is equal to $||q - p||$.

In section 5, we prove that if $P \in PM(S \times S)$ is irreducible, aperiodic and positively recurrent, and $\pi P = \pi$, then it follows that if μ belongs to $\mathcal{P}(K|\pi)$ then $\breve{P}_{\mathcal{M}}\mu$ also belongs to $\mathcal{P}(K|\pi)$. In section 5 we also prove that the barycenter of $\mathbf{P}^n_{\mathcal{M}}(x, \cdot)$ is equal to xP^n.

In section 6 we prove an auxiliary theorem for a Markov chain on a complete, separable, metric space for which its tr.pr.f has the shrinking property. In section 7 we use this auxiliary theorem to prove Theorem 1.1, and in section 8 we give some examples for which asymptotic stability does not hold and prove Theorem 1.2.

In section 9 we focus on Condition B and first prove that Condition B1 implies Condition B. Furthermore we prove that, if the partition $\mathcal{M}$ is such that, 1) sooner or later the Markov chain of probability vectors generated by $\mathbf{P}_{\mathcal{M}}$ at some moment will have a finite support with positive probability, and 2) a condition similar to Condition A introduced in [16] is satisfied, then it follows that Condition B is satisfied.

In section 10, since we have not been able to find a simple criterion from which Condition B follows, we present a *concrete random walk example* on the integers for which Condition B is satisfied. The partition in this example consists of just two matrices, $M(1)$ and $M(2)$, such that $(M(1))_{i,j} = (P)_{i,j}$ if j is odd and zero otherwise, where P denotes the tr.pr.m governing the random walk.

In section 11 we consider *convex functions* on the set K. Let $C_{convex}[K]$ denote the set of all continuous, bounded, convex functions on K. If $u \in C_{convex}[K]$ is such that it can be obtained as

$$u = \sup\{v_n : n \in \mathcal{N}\}$$

where $\mathcal{N}$ is an arbitrary index set and each v_n is an affine function on K such that

$$v_n(x) = xa^c_n + b_n$$

where $a_n \in l^\infty(S)$, y^c denotes the transpose of a vector y and b_n is a real number, then we say that $u \in C'_{convex}[K]$. (Here $l^\infty(S) = \{a = ((a)_i, i \in S) : \sup_{i\in S} |(a)_i| < \infty\}$.)

In section 11 we prove that a transition operator $T_{\mathcal{M}}$, induced by a partition $\mathcal{M}$, maps $C'_{convex}[K]$ into $C_{convex}[K]$. We also show that if $\mathcal{M}$ is a partition of a tr.pr.m P and the vector π in K satisfies $\pi P = \pi$ then, if $u \in C'_{convex}[K]$, we have the following string of inequalities for $n = 1, 2, ...$:

$$\langle u, \breve{P}^n_{\mathcal{M}}\delta_q\rangle \leq \langle u, \breve{P}^{n+1}_{\mathcal{M}}\delta_q\rangle \leq \langle u, \breve{P}^{n+1}_{\mathcal{M}}\psi_q\rangle \leq \langle u, \breve{P}^n_{\mathcal{M}}\psi_q\rangle. \tag{20}$$

Here and throughout this paper, for each $x \in K$ we let δ_x denote the measure in $\mathcal{P}(K)$ which is defined by

$$\delta_x(\{x\}) = 1,$$

and we let ψ_x be the measure in $\mathcal{P}(K)$ which is defined by

$$\psi_x(\{e^i\}) = (x)_i, \ \ i \in S.$$

The inequalities in (20) are reminiscent of results obtained by Kunita in [20].

In section 12 we introduce a *martingale* by reversing the order of the matrix multiplication.

In section 13, as mentioned above, we generalize *Blackwell's formula for the entropy rate of functions of Markov chains.* We also use the inequalities (20) to obtain upper and lower bounds of the entropy rate.

In section 14 finally, we give a proof of the rather intuitive fact that $\mathbf{P}_{\mathcal{M}}$, as defined in subsection 1.1, is in fact a transition probability function.

2 Further notations and some multiplication properties.

Let S be a denumerable set, let $\mathcal{W}$ denote another denumerable set, and let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ denote a denumerable set of nonnegative, $S \times S$ matrices such that

$$\sum_{j \in S} \sum_{w \in \mathcal{W}} (M(w))_{i,j} = 1, \ \ \forall i \in S. \tag{21}$$

We let $\mathcal{G}(S)$ denote the family of denumerable sets of nonnegative, $S \times S$ matrices for which each set $\mathcal{M} = \{M(w), w \in \mathcal{W}\} \in \mathcal{G}(S)$ is such that (21) holds.

For each $\mathcal{M} \in \mathcal{G}(S)$ we can - of course - associate a matrix $P \in PM(S \times S)$ defined by

$$(P)_{i,j} = \sum_{w \in \mathcal{W}} (M(w))_{i,j}, \ \ i, j \in S,$$

and by definition it follows that $\mathcal{M}$ is a partition of P. Therefore we call an element $\mathcal{M} \in \mathcal{G}(S)$ a partition.

Throughout the rest of the paper if $\mathcal{M} \in \mathcal{G}(S)$ then P denotes the tr.pr.m associated to $\mathcal{M}$, and $\mathbf{P}_{\mathcal{M}}$, $T_{\mathcal{M}}$ and $\breve{P}_{\mathcal{M}}$ will denote the transition probability function, the transition operator and the transition probability operator induced by $\mathcal{M}$ as defined in the introduction. Also, if a denumerable set S is given, the capital letter K always refers to the set defined by (6).

A vector $x \in K$ will be called *positive* if $(x)_i > 0, \ \forall i \in S$ and we set $K^+ = \{x \in K : x \ \textit{positive}\}$. If $\pi \in K^+$, we let $\mathcal{G}_\pi(S)$ denote the subset of $\mathcal{G}(S)$ consisting of those partitions $\mathcal{M}$ for which the associated tr.pr.m P is irreducible, aperiodic and positively recurrent satisfying $\pi P = \pi$.

From the multiplication properties of partitions it is clear that $\mathcal{G}(S)$ is a semigroup, and since the set consisting of just the unit matrix belongs to $\mathcal{G}(S)$, it is a semigroup with unit. Clearly $\mathcal{G}_\pi(S)$ is also a semigroup for each $\pi \in K^+$.

We also define $\mathcal{G}'(S) = \cup_{\{\pi \in K^+\}} \mathcal{G}_\pi(S)$, and let $\mathcal{G}^{',\mathcal{Z}}(S)$ denote the subset of $\mathcal{G}'(S)$ consisting of those partitions $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ for which

$$\sup_{x \in K} \sum_{w \in \mathcal{W}_{\mathcal{M}}(x)} -(\ln ||xM(w)||) \cdot ||xM(w)|| < \infty. \tag{22}$$

Next, let $n \geq 2$ and consider the product $\mathcal{M}_1\mathcal{M}_2...\mathcal{M}_n$ where $\mathcal{M}_i = \{M_i(w_i) : w_i \in \mathcal{W}_i\}, i = 1, 2, ..., n$ belongs to $\mathcal{G}(S)$. We denote the product by $\mathcal{M}^n$, we set $\mathcal{W}^n = \prod_{i=1}^n \mathcal{W}_i$, and as before we let $\mathbf{w^n}$ denote an element in $\mathcal{W}^n$ and we let $\mathbf{M(w^n)}$ denote an element in $\mathcal{M}^n$; thus $\mathcal{M}^n = \{\mathbf{M(w^n)} : \mathbf{w^n} \in \mathcal{W}^n\}$. For $x \in K$ we let

$$\mathcal{W}^n_{\mathcal{M}^n}(x) = \{\mathbf{w^n} \in \mathcal{W}^n : ||x\mathbf{M(w^n)}|| > 0\}.$$

(Compare (2).)

The following trivial *scaling property* for matrix products will be used frequently. We omit the proof.

Lemma 2.1 *Let A and B denote two matrices (not necessarily of finite dimension) and assume that AB is well defined. Let x be a row vector and assume also that*
1) xA is well defined, 2) $0 < ||xA|| < \infty$ and 3) $0 < ||xAB|| < \infty$. Then

$$xAB/||xAB|| = (xA/||xA||)B/||(xA/||xA||)B||. \tag{23}$$

As above, let $\mathtt{R}$ denote the real numbers, and let $\mathtt{R}^S$ denote the set of S-dimensional vectors with real coordinates, for $x \in \mathtt{R}^S$ define $||x|| = \sum_{i \in S} |(x)_i|$ and let $l_1(S)$ denote the set $\{x \in \mathtt{R}^S : ||x|| < \infty\}$. Let $\mathbf{0}$ denote the zero vector in $\mathtt{R}^S$ and let $l_1^{'}(S) = l_1(S) \setminus \{\mathbf{0}\}$. Let $\overline{K} = \{x \in l_1(S) : ||x|| = 1\}$. We now denote the projection map from $l_1^{'}(S)$ to $\overline{K}$ by $[\cdot]$. Thus if $y \in l_1^{'}(S)$ then

$$[y] = y/||y||. \tag{24}$$

With this notation, when $x \in \mathtt{R}^S \setminus \{\mathbf{0}\}$ we can write the formula (23) as

$$[xAB] = [[xA]B]. \tag{25}$$

Next some further notations. Let $x \in K$, let $\mathcal{M}_i \in \mathcal{G}(S)$, $i = 1, 2, ...$ and, for $n = 1, 2, ...$, let $\mathcal{M}^n = \{\mathbf{M(w^n)} : \mathbf{w^n} \in \mathcal{W}^n\}$. Let N be an integer ≥ 1, let $1 \leq n \leq N$, and let $x \in K$. If $(w_1, w_2, ...w_N) \in \mathcal{W}^N$ is such that $\mathbf{w}^n \in \mathcal{W}^n_{\mathcal{M}^n}(x)$, $n = 1, 2, ..., N$, we let $x_n(\mathbf{w^n})$, for $n = 1, 2, ..., N$, be defined by

$$x_n(\mathbf{w^n}) = [x\mathbf{M(w^n)}] = x\mathbf{M(w^n)}/||x\mathbf{M(w^n)}||.$$

For $n = 1$ we write $x_1(w_1) = x_1(\mathbf{w^1})$. By using the scaling property it is easy to conclude that

$$\mathcal{W}^2_{\mathcal{M}^2}(x) = \{(w_1, w_2) \in \mathcal{W}^2 : w_1 \in \mathcal{W}_{\mathcal{M}_1}(x) \textit{ and } w_2 \in \mathcal{W}_{\mathcal{M}_2}(x_1(w_1))\}. \tag{26}$$

Next we define the function $g : [0, 1] \to \mathtt{R}$ by

$$g(t) = -(1/\ln(2)) \ln(t), \quad if \;\; t > 0 \;\; and \;\; g(0) = 0.$$

We let $h : [0, 1] \to [0, 1/(e \ln(2))]$ be defined by

$$h(t) = g(t) \cdot t.$$

It is well-known and easily proved that $h : [0, 1] \to [0, 1/(e \ln(2))]$ is a continuous and concave function on $[0, 1]$. For $\mathcal{M} = \{M(w) : w \in \mathcal{W}\} \in \mathcal{G}^{',\mathcal{Z}}(S)$ we set

$$H_{\mathcal{M}}(x) = \sum_{w \in \mathcal{W}} g(||xM(w)||)||xM(w)|| = \sum_{w \in \mathcal{W}} h(||xM(w)||). \tag{27}$$

Before we state our next theorem let us note that if $\mathcal{M}_1 = \{M_1(w_1) : w_1 \in \mathcal{W}_1\} \in \mathcal{G}(S)$ and $\mathcal{M}_2 = \{M_2(w_2) : w_2 \in \mathcal{W}_2\} \in \mathcal{G}(S)$ then

$$\sum_{w_2 \in \mathcal{W}_2} ||xM_1(w_1)M_2(w_2)|| = ||xM_1(w_1)||. \tag{28}$$

because of (14) and the scaling property (23).

Theorem 2.1 *Let $\mathcal{M}_1,\ \mathcal{M}_2\ \in \mathcal{G}(S)$. Then:*
a)

$$\breve{P}_{\mathcal{M}_2}\breve{P}_{\mathcal{M}_1} = \breve{P}_{\mathcal{M}_1\mathcal{M}_2}; \tag{29}$$

b) Suppose $\mathcal{M}_1 \in \mathcal{G}^{',\mathcal{Z}}(S)$ and $\mathcal{M}_2 \in \mathcal{G}^{',\mathcal{Z}}(S)$. Then $\mathcal{M}_1\mathcal{M}_2 \in \mathcal{G}^{',\mathcal{Z}}(S)$, and

$$H_{\mathcal{M}_1\mathcal{M}_2}(x) = H_{\mathcal{M}_1}(x) + \sum_{w_2 \in \mathcal{W}_2} \int_K h(||yM_2(w_2)||)\breve{P}_{\mathcal{M}_1}\delta_x(dy). \tag{30}$$

Proof. Since

$$\langle u, \breve{P}_{\mathcal{M}_2}\breve{P}_{\mathcal{M}_1}\mu\rangle = \langle T_{\mathcal{M}_1}T_{\mathcal{M}_2}u, \mu\rangle$$

and

$$\langle u, \breve{P}_{\mathcal{M}_1\mathcal{M}_2}\mu\rangle = \langle T_{\mathcal{M}_1\mathcal{M}_2}u, \mu\rangle$$

if $u \in C[K]$ and $\mu \in \mathcal{P}(K)$ because of (11), in order to prove a) it therefore suffices to prove that if $u \in C[K]$ and $x \in K$ then

$$T_{\mathcal{M}_1\mathcal{M}_2}u(x) = T_{\mathcal{M}_1}T_{\mathcal{M}_2}u(x). \tag{31}$$

Clearly

$$T_{\mathcal{M}_1\mathcal{M}_2}u(x) = \sum_{(w_1,w_2) \in \mathcal{W}^2_{\mathcal{M}^2}(x)} u(x_2((w_1, w_2))) \cdot ||xM_1(w_1)M_2(w_2)||,$$

and by using the scaling property (25) and the relation (26) we find that

$$\sum_{(w_1,w_2) \in \mathcal{W}^2_{\mathcal{M}^2}(x)} u(x_2((w_1, w_2))) \cdot ||xM_1(w_1)M_2(w_2)|| =$$

$$\sum_{w_1 \in \mathcal{W}_{\mathcal{M}_1}(x)} \sum_{w_2 \in \mathcal{W}_{\mathcal{M}_2}(x_1(w_1))} u([x_1(w_1)M_2(w_2)]) \cdot ||x_1(w_1)M_2(w_2)|| \cdot ||xM_1(w_1)|| =$$

$$\sum_{w_1 \in \mathcal{W}_{\mathcal{M}_1}(x)} T_{\mathcal{M}_2}u(x_1(w_1))(||xM_1(w_1)||) = T_{\mathcal{M}_1}T_{\mathcal{M}_2}u(x).$$

Hence (31) holds and hence a) is proved.

The proof of b) is similar. We first assume that $\mathcal{M}_1,\ \mathcal{M}_2,\ \mathcal{M}_1\mathcal{M}_2 \in \mathcal{G}^{',\mathcal{Z}}(S)$. By definition

$$H_{\mathcal{M}_1\mathcal{M}_2}(x) = \sum_{(w_1,w_2) \in \mathcal{W}^2_{\mathcal{M}_1\mathcal{M}_2}(x)} g(||xM(w_1)M(w_2)||) \cdot ||xM_1(w_1)M_2(w_2)||.$$

Now from the scaling property and the additivity property of the function g we find, for $x \in W_{\mathcal{M}_1}(x)$, that

$$g(||xM_1(w_1)M_2(w_2)||) \cdot ||xM_1(w_1)M_2(w_2)|| =$$
$$g(||x_1(w_1)M_2(w_2)|| \cdot ||xM_1(w_1)||) \cdot ||xM_1(w_1)M_2(w_2)|| =$$
$$(g(||x_1(w_1)M_2(w_2)||) + g(||xM_1(w_1)||)) \cdot ||xM_1(w_1)M_2(w_2)||.$$

Hence

$$H_{\mathcal{M}_1\mathcal{M}_2}(x) = \sum_{(w_1,w_2)\in\mathcal{W}^2_{\mathcal{M}_1\mathcal{M}_2}(x)} g(||x_1(w_1)M_2(w_2)||) \cdot ||xM_1(w_1)M_2(w_2)|| +$$
$$\sum_{(w_1,w_2)\in\mathcal{W}^2_{\mathcal{M}_1\mathcal{M}_2}(x)} g(||xM_1(w_1)||)) \cdot ||xM_1(w_1)M_2(w_2)||.$$

The second term is equal to $H_{\mathcal{M}_1}(x)$ because of (28). Furthermore by using the scaling property again we find that

$$g(||x_1(w_1)M_2(w_2)||) \cdot ||xM_1(w_1)M_2(w_2)|| =$$
$$g(||x_1(w_1)M_2(w_2)||) \cdot ||x_1(w_1)M_2(w_2)|| \cdot ||xM_1(w_1)||$$

if $x \in W_{\mathcal{M}_1}(x)$, and then, using also (26), we find

$$\sum_{(w_1,w_2)\in\mathcal{W}^2_{\mathcal{M}_1\mathcal{M}_2}(x)} g(||x_1(w_1)M_2(w_2)||) \cdot ||xM_1(w_1)M_2(w_2)|| =$$
$$\sum_{(w_1,w_2)\in\mathcal{W}^2_{\mathcal{M}_1\mathcal{M}_2}(x)} g(||x_1(w_1)M_2(w_2)||) \cdot ||x_1(w_1)M_2(w_2)|| \cdot ||xM_1(w_1)|| =$$
$$\sum_{w_1\in\mathcal{W}_{\mathcal{M}_1}(x)} ||xM_1(w_1)|| \sum_{w_2\in\mathcal{W}_{\mathcal{M}_2}(x_1(w_1))} g(||x_1(w_1)M_2(w_2)||) \cdot ||x_1(w_1)M_2(w_2)|| =$$
$$\sum_{w_1\in\mathcal{W}_{\mathcal{M}_1}(x)} ||xM_1(w_1)|| \sum_{w_2\in\mathcal{W}_2} h(||x_1(w_1)M_2(w_2)||) =$$
$$\sum_{w_2\in\mathcal{W}_2} \sum_{w_1\in\mathcal{W}_{\mathcal{M}_1}(x)} ||xM_1(w_1)|| h(||x_1(w_1)M_2(w_2)||).$$

From the definition of $\mathbf{P}_{\mathcal{M}}$ it then follows that

$$\sum_{w_2\in\mathcal{W}_2} \sum_{w_1\in\mathcal{W}_{\mathcal{M}_1}(x)} ||xM_1(w_1)|| h(||x_1(w_1)M_2(w_2)||) =$$
$$\sum_{w_2\in\mathcal{W}_2} \int_K h(yM_2(w_2))\mathbf{P}_{\mathcal{M}_1}(x, dy).$$

Hence

$$H_{\mathcal{M}_1\mathcal{M}_2}(x) = H_{\mathcal{M}_1}(x) + \sum_{w_2\in\mathcal{W}_2} \int_K h(yM_2(w_2))\mathbf{P}_{\mathcal{M}_1}(x, dy).$$

Finally since $\breve{P}_{\mathcal{M}_1}\delta_x(\cdot) = \mathbf{P}_{\mathcal{M}_1}(x, \cdot)$ we can conclude that (30) is true.

In order to prove that $\mathcal{M}_1\mathcal{M}_2 \in \mathcal{G}^{',\mathcal{Z}}(S)$, if both $\mathcal{M}_1 \in \mathcal{G}^{',\mathcal{Z}}(S)$ and $\mathcal{M}_2 \in \mathcal{G}^{',\mathcal{Z}}(S)$, we only have to follow the reasoning above backwards. $\square$

Part b) of Theorem 2.1 will be used in section 13 to prove entropy results.

Corollary 2.1 *Let S be a denumerable set and let $\mathcal{M}$ be a partition of $P \in PM(S \times S)$. Then, for each positive integer n,*

$$\breve{P}^n_{\mathcal{M}} = \breve{P}_{\mathcal{M}^n}, \tag{32}$$

and

$$T^n_{\mathcal{M}} = T_{\mathcal{M}^n}. \tag{33}$$

Proof. Relation (32) follows from (29) by induction, and (33) follows from (32) and (11). □

3 A universal inequality.

The following inequality, which we formulate as a theorem, is in principal proved in [16], section 4. If $u \in F[K]$, we define $||u|| = \sup_{x \in K} |u(x)|$.

Theorem 3.1 *Let S be a denumerable set and suppose $\mathcal{M} \in \mathcal{G}(S)$. Then,*

$$u \in Lip[K] \;\Rightarrow T_{\mathcal{M}}u \in Lip[K] \tag{34}$$

and

$$\gamma(T_{\mathcal{M}}u) \leq (||u|| + 2\gamma(u)). \tag{35}$$

Proof. We shall follow the proof of Lemma 4.3 in [16] closely. To simplify notations we shall write T instead of $T_{\mathcal{M}}$.

In order to prove (34) and (35) it suffices to prove that for any $u \in Lip[K]$ and any two elements $x, y \in K$

$$|Tu(x) - Tu(y)| \leq (||u|| + 2\gamma(u)) \cdot ||x - y||. \tag{36}$$

Let $S_1 = \{i : (x)_i > 0,\ (y)_i > 0\}$, $S_2 = \{i : (x)_i > 0,\ (y)_i = 0\}$, $S_3 = \{i : (x)_i = 0,\ (y)_i > 0\}$, let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ and

$$\mathcal{W}_i = \{w \in \mathcal{W} : ||e^i M(w)|| > 0\}.$$

Using the fact that for an arbitrary $z \in K$ and arbitrary $w \in \mathcal{W}$, we have

$$||zM(w)|| = \sum_{i \in S} (z)_i \cdot ||e^i M(w)||,$$

we obtain, by using the triangle inequality and (14), that

$$|Tu(x) - Tu(y)| =$$

$$|\{ \sum_{i \in S_1} (x)_i \sum_{w \in \mathcal{W}_i} u(x_1(w)) \cdot ||e^i M(w)|| + \sum_{i \in S_2} (x)_i \sum_{w \in \mathcal{W}_i} u(x_1(w)) \cdot ||e^i M(w)|| -$$

$$\sum_{i \in S_1} (y)_i \sum_{w \in \mathcal{W}_i} u(y_1(w)) \cdot ||e^i M(w)|| - \sum_{i \in S_3} (y)_i \sum_{w \in \mathcal{W}_i} u(y_1(w)) \cdot ||e^i M(w)|| \;\}| \leq$$

$$\sum_{i \in S} |(x)_i - (y)_i| \cdot ||u|| \sum_{w \in \mathcal{W}} ||e^i M(w)|| +$$

$$| \sum_{i \in S_1} (y)_i \sum_{w \in \mathcal{W}_i} u(x_1(w)) \cdot ||e^i M(w)|| - \sum_{i \in S_1} (y)_i \sum_{w \in \mathcal{W}_i} u(y_1(w)) \cdot ||e^i M(w)|| \, | \leq$$

$$||x - y|| \cdot ||u|| + \gamma(u) \sum_{i \in S_1} (y)_i \sum_{w \in \mathcal{W}_i} ||x_1(w) - y_1(w)|| \cdot ||e^i M(w)||. \tag{37}$$

In order to prove that the last term is less than $2\gamma(u)||x - y||$ we shall use the following inequality. *Let a and b be two nonzero vectors in a normed vector space. Then*

$$||(a/||a|| - b/||b||)|| = ||(a/||a|| - a/||b|| + a/||b|| - b/||b||)|| \leq$$

$$|(||a|| - ||b||)|/||b|| + ||a - b||/||b|| \leq 2||a - b||/||b||$$

and hence

$$||(a/||a|| - b/||b||)|| \leq 2||a - b||/||b||. \tag{38}$$

Using (38) we find that

$$||x_1(w) - y_1(w)|| \leq 2||xM(w) - yM(w)||/||yM(w)|| \tag{39}$$

if $||xM(w)|| \cdot ||yM(w)|| > 0$, and by using (39), the triangle inequality and change of summation order, we obtain

$$\gamma(u) \sum_{i \in S_1} (y)_i \sum_{w \in \mathcal{W}_i} ||x_1(w) - y_1(w)|| \cdot ||e^i M(w)|| \leq$$

$$\gamma(u) \sum_{i \in S_1} (y)_i \sum_{w \in \mathcal{W}_i} (2||xM(w) - yM(w)||/||yM(w)||) \cdot ||e^i M(w)|| \leq$$

$$2\gamma(u) \sum_{w \in \mathcal{W}} ||xM(w) - yM(w)|| \sum_{i \in S} (y)_i ||e^i M(w)||/||yM(w)|| =$$

$$2\gamma(u) \sum_{w \in \mathcal{W}} ||xM(w) - yM(w)|| \leq 2\gamma(u) \sum_{w \in \mathcal{W}} \sum_{i \in \mathcal{I}} |x_i - y_i| \sum_{j \in \mathcal{I}} (M(w))_{i,j} =$$

$$2\gamma(u) \sum_{i \in \mathcal{I}} |x_i - y_i| \sum_{j \in \mathcal{I}} \sum_{w \in \mathcal{W}} (M(w))_{i,j} =$$

$$2\gamma(u) \sum_{i \in \mathcal{I}} |x_i - y_i| \sum_{j \in \mathcal{I}} (P)_{i,j} =$$

$$2\gamma(u) \sum_{i \in \mathcal{I}} |x_i - y_i| = 2\gamma(u)||x - y||$$

which combined with (37) implies (36). □

Corollary 3.1 *Let S be a denumerable set, let $\mathcal{M} \in \mathcal{G}(S)$ and $u \in Lip[K]$. Then*

$$\gamma(T_{\mathcal{M}} u) \leq 3\gamma(u). \tag{40}$$

Proof. Let $u \in Lip[K]$, set $v = u/\gamma(u)$, set $osc(v) = \sup\{v(x) - v(y) : x, y \in K\}$ and define $v_0 = v - osc(v)/2 - \inf\{v(x) : x \in K\}$. Since $\sup\{||x - y|| : x, y \in K\} = 2$ and $\gamma(v_0) = \gamma(v) \leq 1$ it is clear that $||v_0|| \leq 1$. Hence by Theorem 3.1 follows that $\gamma(T_{\mathcal{M}} v_0) \leq 3$ and hence $\gamma(T_{\mathcal{M}} u) = \gamma(u)\gamma(T_{\mathcal{M}} v) = \gamma(u)\gamma(T_{\mathcal{M}} v_0) \leq 3\gamma(u)$. □

Corollary 3.2 *Let S be a denumerable set, let $\mathcal{M} \in \mathcal{G}(S)$ and let $u \in Lip[K]$. Then*

$$\gamma(T^n_{\mathcal{M}}u) \leq 3\gamma(u), \quad n = 1, 2, \dots \ . \tag{41}$$

Proof. Let $u \in Lip[K]$. From Corollary 2.1 follows that $\gamma(T^n_{\mathcal{M}}u) = \gamma(T_{\mathcal{M}^n}u)$ and then (41) follows from Corollary 3.1. □

Next let $\mathcal{Q}(K)$ denote the set of nonnegative, finite, Borel measures on $(K, \mathcal{E})$ with positive total mass. For $\mu \in \mathcal{Q}(K)$ we write $||\mu|| = \mu(K)$. For $r > 0$ we define $\mathcal{Q}_r(K) = \{\mu \in \mathcal{Q}(K) : ||\mu|| = r\}$. If both $\mu, \nu \in \mathcal{Q}_r(K)$ we define - for any $r > 0$ -

$$d_K(\mu, \nu) = \sup\{\int_K u(y)\mu(dy) - \int_K u(y)\nu(dy) : \ u \in Lip_1[K] \ \}. \tag{42}$$

Note that if $\mu, \nu \in \mathcal{Q}_r(K)$ then $\mu/r, \nu/r \in \mathcal{P}(K)$ and

$$d_K(\mu, \nu) = ||\mu|| d_K(\mu/||\mu||, \nu/||\mu||). \tag{43}$$

Note also that

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

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

That $d_K(\cdot, \cdot)$ determines a metric on $\mathcal{P}(K)$ is well-known, (see e.g. [10], Chapter 11, section 3,) and from (43) follows that d_K determines a metric also on $\mathcal{Q}_r(K)$ for any $r > 0$. We shall call the metric $d_K(\cdot, \cdot)$ the *Kantorovich metric.*

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

$$d_K(\delta_x, \delta_y) = \delta(x, y) = ||x - y||.$$

If $\mathcal{P}' \subset \mathcal{P}(K)$ and $\mu \in \mathcal{P}(K)$ we define

$$d_K(\mu, \mathcal{P}') = \inf\{d_K(\mu, \nu) : \nu \in \mathcal{P}'\}.$$

We shall later have use for the following inequality which follows easily from the definition of $d_K(\cdot, \cdot)$ and the triangle inequality.

Proposition 3.1 *Let $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathcal{Q}(K)$ satisfy $||\mu_1|| = ||\nu_1||$ and $||\mu_2|| = ||\nu_2||$. Let $\alpha > 0$ and $\beta > 0$ be two real numbers. Then*

$$d_K(\alpha\mu_1 + \beta\mu_2, \alpha\nu_1 + \beta\nu_2) \leq \alpha d_K(\mu_1, \nu_1) + \beta d_K(\mu_2, \nu_2).$$

It is well-known that the Kantorovich metric δ_K on $\mathcal{Q}_r(K)$ can be defined in another way. Let $K^2 = K \times K$, let $\mathcal{E}^2 = \mathcal{E} \otimes \mathcal{E}$, let $r > 0$ and let $\mathcal{Q}_r(K \times K)$ denote the set of nonnegative measures on $(K^2, \mathcal{E}^2)$ with total mass equal to r. For any two measures $\mu, \nu \in \mathcal{Q}_r(K)$ we let $\tilde{\mathcal{Q}}_r(\mu, \nu)$ denote the subset of $\mathcal{Q}_r(K \times K)$ consisting of those measures $\tilde{\mu}(dx, dy)$ such that

$$\tilde{\mu}(A, K) = \mu(A), \qquad \forall A \in \mathcal{E},$$

and

$$\tilde{\mu}(K, B) = \nu(B), \qquad \forall B \in \mathcal{E}.$$

Then

$$d_K(\mu,\nu) = \inf\{\int_{K\times K} \delta(x,y)\tilde{\mu}(dx,dy) : \ \tilde{\mu}(dx,dy) \in \tilde{\mathcal{Q}}_r(\mu,\nu)\}. \qquad (44)$$

A proof of the equality between (42) and (44) when $r = 1$ can be found in [10], section 11.8, and the equality between (42) and (44) for $r \neq 1$ follows then by using the relation (43). The proof of the fact that the two definitions of d_K give the same value goes back to *Kantorovich* (see [17]). For a short overview of the Kantorovich metric and some applications, see [27].

Having introduced the metric d_K the following corollary to Corollary 3.2 follows immediately. We omit the proof.

Corollary 3.3 *Let S be a denumerable set, let $P \in PM(S \times S)$ and let $\mathcal{M}$ be a partition of P. Let μ and ν be two arbitrary measures in $\mathcal{P}(K)$. Then*

$$d_K(\check{P}^n_{\mathcal{M}}\,\mu, \check{P}^n_{\mathcal{M}}\,\nu) \leq 3d_K(\mu,\nu) \ \ for \ \ n = 1,2,\ldots . \qquad (45)$$

It is not difficult to construct an example that shows that the constant 3, occurring on the right hand side of formula (45), can not be replaced by a constant strictly less than 2. Take e.g $S = \{1,2,3\}$, let $P = I$, the identity matrix, and let the partition consist of two matrices $M(1), M(2)$ such that the first two columns of P and $M(1)$ are equal and the last column of P is equal to the last column of $M(2)$. Let $x = (1-\epsilon,\epsilon,0)$ and $y = (1-\epsilon,0,\epsilon)$. Then, for $0 < \epsilon \leq 1$, $d_K(\delta_x,\delta_y) = 2\epsilon$ and $d_K(\check{P}_{\mathcal{M}}\delta_y, \check{P}_{\mathcal{M}}\delta_x) = d_K(\delta_x,\delta_y)(2-\epsilon)$.

Conjecture 3.1 *Let S be a denumerable set, let $\mathcal{M} \in \mathcal{G}(S)$ and $\mu,\nu \in \mathcal{P}(K)$. Then*

$$d_K(\check{P}_{\mathcal{M}}\mu, \check{P}_{\mathcal{M}}\nu) \leq 2d_K(\mu,\nu).$$

We end this section by stating another well-known result concerning the metric $d_K(\cdot,\cdot)$. See e.g. [10], Theorem 11.3.3.

Proposition 3.2 *Let $\mu \in \mathcal{P}(K)$ and $\{\mu_n, n = 1,2,...\}$ be a sequence of probability measures in $\mathcal{P}(K)$ such that*

$$\lim_{n\to\infty} d_K(\mu_n,\mu) = 0.$$

Then for each $u \in C[K]$

$$\lim_{n\to\infty} \int_K u(y)\mu_n(dy) = \int_K u(y)\mu(dy).$$

4 On probability measures with equal barycenter.

Throughout this section we shall use the set of positive integers when specifying the elements of a denumerable set. We let $\mathcal{J} = \{1,2,...\}$ either denote a finite sequence of consecutive positive integers starting with the integer 1, or the whole set of positive integers, and we denote an arbitrary denumerable set by

$$S = \{i_j : j \in \mathcal{J}\}.$$

In the introduction we defined the barycenter $\overline{b}(\mu)$ of a probability measure $\mu \in \mathcal{P}(K)$ by the relation (9). We use the same definition for a measure $\mu \in \mathcal{Q}(K)$. Thus if $\mu \in \mathcal{Q}(K)$ we call the vector $\overline{b}(\mu)$ defined by

$$(\overline{b}(\mu))_i = \int_K (y)_i \mu(dy), \quad i \in S$$

the *barycenter* of μ. Also recall that we defined $\mathcal{P}(K|q) = \{\mu \in \mathcal{P}(K) : \overline{b}(\mu) = q\}$.

If the underlying space S is finite, then the space $(K, \mathcal{E})$ is a compact, metric space, and from the general theory regarding probability measures on compact, metric spaces, it is well-known that $\mathcal{P}(K)$ is then also a compact, metric space in the topology determined by the Kantorovich metric. From Corollary 4.2 below it follows easily that $\mathcal{P}(K|q)$ is a closed set, and therefore $\mathcal{P}(K|q)$ is also compact when S is finite, since it is a closed subset of the compact set $\mathcal{P}(K)$.

Next let us prove the following tightness result which is based on a moment condition. Recall that a subset $\mathcal{P}'$ of $\mathcal{P}(K)$ is *tight* if for every $\epsilon > 0$ one can find a compact set $C \subset K$ such that $\mu(C) > 1 - \epsilon$, $\forall \mu \in \mathcal{P}'$.

Proposition 4.1 *Let $S = \{i_j, j = 1, 2, ...\}$ be infinite, let $q \in K$ and suppose that there exists a $\kappa > 0$ such that*

$$\sum_{j=1}^{\infty} (q)_{i_j} j^{1+\kappa} < \infty.$$

Then the set $\mathcal{P}(K|q)$ is a tight set in $\mathcal{P}(K)$.

Proof. Let $\epsilon > 0$ be given and set

$$A = \sum_{j=1}^{\infty} (q)_{i_j} j^{1+\kappa}.$$

Let $E_{j,\epsilon}, \ j = 1, 2, ...$ be defined by

$$E_{j,\epsilon} = \{x \in K : (x)_{i_j} \leq 2A/(j^{1+\kappa}\epsilon)\}.$$

Using the fact that

$$\sum_{j=1}^{\infty} 1/(j^{1+\kappa}) < \infty$$

it is easy to prove that the set

$$E_\epsilon = \cap_{j=1}^{\infty} E_{j,\epsilon}$$

is a compact set for every fixed $\epsilon > 0$. But from Markov's inequality it follows that if $\mu \in \mathcal{P}(K|q)$ then

$$\mu(K \setminus E_{j,\epsilon}) \leq (q)_{i_j} (j^{1+\kappa}\epsilon/2A).$$

Hence

$$\mu(K \setminus (\cap_{j=1}^{\infty} E_{j,\epsilon})) \leq \sum_{j=1}^{\infty} \mu(K \setminus E_{j,\epsilon}) \leq \sum_{j=1}^{\infty} (q)_{i_j} (j^{1+\kappa}\epsilon/2A) = \epsilon/2 < \epsilon$$

and thereby we have proved tightness. $\Box$

The moment condition of Proposition 4.1 is though unnecessary. The following is true.

Theorem 4.1 *Let S be a denumerable and infinite set and let $q \in K$. Then $\mathcal{P}(K|q)$ is tight.*

In order to prove Theorem 4.1 our main goal will be to prove the following theorem.

Theorem 4.2 *For each $q \in K$ the set $\mathcal{P}(K|q)$ is a compact set in the topology induced by the metric $d_K(\cdot,\cdot)$.*

In order to prepare ourselves for a rather short proof of this theorem, we shall first prove two lemmas.

Lemma 4.1 *Let $S = \{i_j, j = 1,2,...\}$ be denumerable, let $r > 0$ and let $\mu,\nu \in \mathcal{Q}_r(K)$. Then*

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

Proof. Set

$$a = \overline{b}(\nu), \quad and \quad b = \overline{b}(\mu).$$

For $i \in S$ define

$$u_i(x) = (x)_i.$$

Clearly $u_i \in Lip_1[K],\ i \in S$, and by definition

$$\int_K u_i(x)\nu(dx) = (a)_i \quad and \quad \int_K u_i(x)\mu(dx) = (b)_i \quad for \quad i \in S.$$

Next define S_+ by

$$S_+ = \{i : (a)_i > 0 \quad or \quad (b)_i > 0\},$$

let n be an arbitrary positive integer, let k be an integer in the interval $1 \leq k \leq 2^n$, and let $\epsilon^{n,k} = (\epsilon_{1,k}, \epsilon_{2,k}, ..., \epsilon_{n,k})$ specify one vector such that $\epsilon_{j,k} = +1$ or $\epsilon_{j,k} = -1,\ j = 1,2,...,n$. There are exactly 2^n such vectors.

Next let $v_{\epsilon^{n,k}}$ specify one of the 2^n possible functions one can obtain by the definition

$$v_{\epsilon^{n,k}} = \sum_{j=1}^{n} \epsilon_{j,k} u_{i_j}.$$

Since $u_i = (x)_i,\ i = 1,2,...$ it follows that if $x, y \in K$ then

$$|v_{\epsilon^{n,k}}(x) - v_{\epsilon^{n,k}}(y)| \leq \sum_{j=1}^{n} |(x)_{i_j} - (y)_{i_j}| \leq ||x - y||$$

and hence $v_{\epsilon^{n,k}} \in Lip_1[K]$ for $1 \leq k \leq 2^n$. Thus

$$\delta_K(\mu,\nu) \geq \max\{\int_{K_r} v_{\epsilon^{n,k}}(x)\mu(dx) - \int_{K_r} v_{\epsilon^{n,k}}(x)\nu(dx) : k = 1,2,...,2^n\} =$$

$$= \sum_{j=1}^{n} |(a)_{i_j} - (b)_{i_j}|.$$

Now, if S_+ is finite, by choosing n large enough, it follows that

$$\delta_K(\mu,\nu) \geq \sum_{i=1}^{n} |(a)_i - (b)_i| = ||a-b|| = ||\overline{b}(\nu) - \overline{b}(\mu)||$$

which was what we wanted to prove. If S_+ is infinite, we can conclude that $\delta_K(\mu,\nu) \geq \sum_{i=1}^{n} |(a)_i - (b)_i|$ and by letting $n \to \infty$, we also can conclude that

$$\delta_K(\mu,\nu) \geq ||a-b|| = ||\overline{b}(\nu) - \overline{b}(\mu)||$$

and thereby Lemma 4.1 is proved. □

Lemma 4.2 *Let S be a denumerable set, and let K denote the probability vectors on S. Let N be a positive integer and let $\xi_k,\ k = 1,2,...,N$ be vectors in K. Let $\beta_k > 0,\ k = 1,2,...,N$, define the measure $\varphi \in \mathcal{Q}(K)$ by*

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

where as usual δ_ξ denotes the probability measure with unit mass at ξ, and define the vector a by

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

Let $b = ((b)_i, i \in S)$ be a vector satisfying $(b)_i \geq 0,\ \ i \in S$, and

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

Then there exist vectors $\zeta_k,\ \ k = 1,2,...,N,$ 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

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

Proof. That the conclusion of the lemma is true when $N = 1$ is easily proved. Simply define

$$\zeta_1 = b/\beta_1 .$$

Since $||b|| = ||a|| = \beta_1$ it is clear that $||\zeta_1|| = 1$ and hence $\zeta_1 \in K$ since $(b)_i \geq 0,\ i \in S$. Now defining $\Psi = \beta_1 \delta_{\zeta_1}$ we find

$$\delta_K(\varphi,\Psi) = \delta_K(\beta_1\delta_{\xi_1}, \beta_1\delta_{\zeta_1}) = \beta_1\delta_K(\delta_{\xi_1}, \delta_{\zeta_1}) = \beta_1||\xi_1 - \zeta_1|| =$$

$$||\beta_1\xi_1 - \beta_1\zeta_1|| = ||a-b||$$

where we have used the fact that $d_K(\delta_x, \delta_y) = ||x-y||$ for any pair of points x and y in K.

The case when $a = b$ is trivial. Just take $\zeta_k = \xi_k, k = 1, 2, ..., N$ and define $\Psi = \varphi$. In the remaining part of the proof we therefore assume that $a \neq b$.

Let us now assume that we have proved the conclusion of the lemma for $N = M - 1$ and let us prove the conclusion for $N = M$, where $M \geq 2$.

We first define the sets S_1, S_2 and S_3 by

$$S_1 = \{i \in S : (a)_i > (b)_i \},$$

$$S_2 = \{i \in S : (a)_i < (b)_i \},$$

and

$$S_3 = \{i \in S : (a)_i = (b)_i \}.$$

Since $a \neq b$ and $||a|| = ||b||$ both the sets S_1 and S_2 are nonempty. Let us set

$$\Delta = \sum_{i \in S_1} ((a)_i - (b)_i).$$

Then clearly

$$\Delta = \sum_{j \in S_2} ((b)_j - (a)_j) = ||a - b||/2.$$

Next let us consider the vector ξ_M. We define

$$R_1 = \{i \in S_1 : (\xi_M)_i > 0\}.$$

Assume first that $R_1 = \emptyset$. Then define the vector a_1, the vector ζ_m and the vector b_1, by

$$a_1 = a - \beta_M \xi_M,$$

$$\zeta_M = \xi_M,$$

and

$$b_1 = b - \beta_M \zeta_M.$$

For $i \in S_1$ we now find that $(a_1)_i = (a)_i \;\; and \;\; (b_1)_i = (b)_i$ since $(\xi_M)_i = 0$ if $i \in S_1$. Therefore, we note in particular that $(b_1)_i \geq 0, \;\; i \in S_1$. Moreover, if $i \in S_2 \cup S_3$ then by using the fact that $(a)_i \leq (b)_i, \;\; \forall i \in S_2 \cup S_3$, using the fact that $\zeta_M = \xi_M$ and using the fact that $(a_1)_i \geq 0, i \in S$ we can conclude that $(b_1)_i \geq 0, \;\; i \in S$. Thus both b, b_1 and ζ_M have nonnegative coordinates, which implies that

$$||b_1|| = ||b|| - \beta_M ||\zeta_M||$$

since the norm $|| \cdot ||$ is defined by the $l_1 - metric$; since furthermore $||a|| = ||b||$ and $||\zeta_M|| = ||\xi_M||$, we find that

$$||b_1|| = ||a|| - \beta_M ||\xi_M||.$$

But since also $0 \leq (a_1)_i \leq (a)_i, \;\; i \in S$, it is clear that we also have

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

Hence

$$||a_1|| = ||b_1||.$$

Since $\beta_k > 0,\ k = 1, 2, ..., M$, it is clear that $||a_1|| > 0$, and hence we also have that $||b_1|| > 0$. Furthermore since $\xi_M = \zeta_M$ it is clear that

$$||a_1 - b_1|| = ||a - \beta_M \xi_M - b + \beta_M \xi_M|| = ||a - b||. \tag{46}$$

From the induction hypothesis now follows that there exists a set of vectors $\{\zeta_k \in K,\ \ k = 1, 2, ..., M-1\}$ such that

$$b_1 = \sum_{k=1}^{M-1} \beta_k \zeta_k$$

and such that if we denote

$$\varphi_1 = \sum_{k=1}^{M-1} \beta_k \delta_{\xi_k}$$

and define

$$\Psi_1 = \sum_{k=1}^{M-1} \beta_k \delta_{\zeta_k}$$

then

$$d_K(\varphi_1, \Psi_1) = ||a_1 - b_1||. \tag{47}$$

Therefore, if we define

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

we find that

$$d_K(\varphi, \Psi) = d_K(\varphi_1 + \beta_M \delta_{\xi_M}, \Psi_1 + \beta_M \delta_{\zeta_M})$$

and from Proposition 3.1 it follows that

$$d_K(\varphi, \Psi) \leq d_K(\varphi_1, \Psi_1) + \beta_M d_K(\delta_{\xi_M}, \delta_{\zeta_M}) = d_K(\varphi_1, \Psi_1) \tag{48}$$

since by definition $\zeta_M = \xi_M$. But from (47) and (46) follows that $d_K(\varphi_1, \Psi_1) = ||a_1 - b_1|| = ||a - b||$ and hence

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

because of (48). But $\overline{b}(\varphi) = a$ and $\overline{b}(\Psi) = b$ and therefore by Lemma 4.1 we also know that $d_K(\varphi, \Psi) \geq ||a - b||$. Hence

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

which was what we wanted to prove.

Next consider the case when the set R_1 is not empty. We shall again use ξ_M to define a new vector ζ_M. Roughly speaking what we shall do is to move as large part as possible from a coordinate $(\xi_M)_i,\ i \in R_1$, to a coordinate $(\xi_M)_j, j \in S_2$.

Hence, we claim that there exist nonnegative numbers $t_{i,j},\ i \in R_1,\ j \in S_2$ with the following properties:

$$\sum_{j \in S_2} t_{i,j} = \min\{\beta_M (\xi_M)_i, ((a)_i - (b)_i)\}, \quad \forall i \in R_1 \tag{49}$$

and

$$\sum_{i \in R_1} t_{i,j} \leq ((b)_j - (a)_j), \quad \forall j \in S_2. \tag{50}$$

That such a set $\{t_{i,j} : i \in R_1, \quad j \in S_2\}$ exists follows from the following two observations:
1)

$$\sum_{j \in S_2} ((b)_j - (a)_j) = \Delta,$$

2)

$$\sum_{i \in R_1} \min\{\beta_M(\xi_M)_i, ((a)_i - (b)_i)\} \leq \sum_{i \in S_1} ((a)_i - (b)_i)\} = \Delta.$$

We now simply define the vector ζ_M by

$$(\zeta_M)_i = (\xi_M)_i - (\beta_M)^{-1} \sum_{j \in S_2} t_{i,j}, \quad i \in R_1 \tag{51}$$

$$(\zeta_M)_j = (\xi_M)_j + (\beta_M)^{-1} \sum_{i \in R_1} t_{i,j}, \quad j \in S_2 \tag{52}$$

$$(\zeta_M)_i = (\xi_M)_i, \quad if \quad i \notin R_1 \cup S_2.$$

Since by definition

$$\sum_{j \in S_2} t_{i,j} \leq \beta_M(\xi_M)_i, \quad \forall i \in R_1$$

it is clear that $(\zeta_M)_i \geq 0$, $\forall i \in R_1$. That also $(\zeta_M)_i \geq 0$ if $i \notin R_1$ follows from the fact that $(\xi_M)_i \geq 0$ for all $i \in S$. Hence

$$(\zeta_M)_i \geq 0, \ \forall i \in S$$

from which we conclude that

$$||\zeta_M|| = \sum_{i \in R_1} (\zeta_M)_i + \sum_{j \in S_2} (\zeta_M)_j + \sum_{i \notin R_1 \cup S_2} (\zeta_M)_j =$$

$$||\xi_M|| - \sum_{i \in R_1} (\beta_M)^{-1} (\sum_{j \in S_2} t_{i,j}) + \sum_{j \in S_2} (\beta_M)^{-1} (\sum_{i \in R_1} t_{i,j}) = ||\xi_M||$$

and hence

$$||\zeta_M|| = ||\xi_M|| = 1.$$

Now define

$$a_1 = a - \beta_M \xi_M = \sum_{k=1}^{M-1} \beta_k \xi_k$$

and

$$b_1 = b - \beta_M \zeta_M.$$

Obviously $(a_1)_i \geq 0$, $\forall i \in S$. Since also $(\xi_M)_i \geq 0$, $\forall i \in S$, it follows, since we are using the $l_1 - norm$, that

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

Next let us investigate whether $(b_1)_i \geq 0$ for all $i \in S$. First suppose $i \in S_1 \setminus R_1$. Then $(\xi_M)_i = 0$ and hence

$$(b_1)_i = (b)_i \geq 0.$$

Next suppose that $i \in R_1$. Then

$$(b_1)_i = (b)_i - \beta_M(\xi_M)_i + \sum_{j \in S_2} t_{i,j}.$$

Now if $\beta_M(\xi_M)_i \leq (a)_i - (b)_i$ then $\sum_{j \in S_2} t_{i,j} = \beta_M(\xi_M)_i$ and hence

$$(b_1)_i = (b)_i \geq 0.$$

If instead $\beta_M(\xi_M)_i > (a)_i - (b)_i$ then $\sum_{j \in S_2} t_{i,j} = (a)_i - (b)_i$ and hence

$$(b_1)_i = (b)_i - \beta_M(\xi_M)_i + (a)_i - (b)_i$$

$$= (a)_i - \beta_M(\xi_M)_i = (a_1)_i \geq 0.$$

If $i \in S_3$ then $(a)_i = (b)_i$ and $(\zeta_M)_i = (\xi_M)_i$ and therefore

$$(b_1)_i = (a_1)_i \geq 0, \ \ \forall i \in S_3.$$

Finally if $j \in S_2$ then

$$(b_1)_j = (b)_j - \beta_M(\zeta_M)_j = (b)_j - \beta_M(\xi_M)_j - \sum_{i \in R_1} t_{i,j} \geq (b)_j - \beta_M(\xi_M)_j - ((b)_j - (a)_j) =$$

$$= (a)_j - \beta_M(\xi_M)_j = (a_1)_j \geq 0.$$

Hence

$$(b_1)_i \geq 0, \quad \forall \, i \in S.$$

Therefore, using the fact that $||b|| = ||a||$ and the fact that $||\zeta_M|| = ||\xi_M||$, it follows that

$$||b_1|| = ||b|| - \beta_M||\zeta_M|| = ||a|| - \beta_M||\xi_M|| = ||a_1||.$$

Let us also compute $||a_1 - b_1||$. If $i \in S_1 \setminus R_1$ then $(\xi_M)_i = 0$ and therefore it follows that

$$(a_1)_i - (b_1)_i = (a)_i - (b)_i > 0, \ \forall \, i \in S_1 \setminus R_1.$$

If $i \in R_1$ then

$$(a_1)_i - (b_1)_i = (a)_i - \beta_M(\xi_M)_i - (b)_i + \beta_M(\xi_M)_i - \sum_{j \in S_2} t_{i,j} = (a)_i - (b)_i - \sum_{j \in S_2} t_{i,j} \geq 0$$

because of (49). If $i \in S_3$ then $(a)_i = (b)_i$ and $(\zeta_M)_i = (\xi_M)_i$ and therefore

$$(a_1)_i - (b_1)_i = 0, \ \forall \, i \in S_3,$$

and if $j \in S_2$ then

$$(b_1)_j - (a_1)_j = (b)_j - \beta_M(\zeta_M)_j - (a)_j + \beta_M(\xi_M)_j =$$

$$(b)_j - (a)_j - \sum_{i \in R_1} t_{i,j} \geq 0$$

because of (50). Hence

$$||a_1 - b_1|| = 2 \sum_{j \in S_2} ((b)_j - (a)_j) - 2 \sum_{i \in R_1,\, j \in S_2} t_{i,j}$$

and since

$$||a - b|| = 2 \sum_{j \in S_2} ((b)_j - (a)_j)$$

and

$$2 \sum_{i \in R_1,\, j \in S_2} t_{i,j} = \beta_M ||\xi_M - \zeta_M||$$

because of (51) and (52) we conclude that

$$||a_1 - b_1|| = ||a - b|| - \beta_m ||\xi_M - \zeta_M||. \tag{53}$$

Now let us set

$$\varphi_1 = \sum_{k=1}^{M-1} \beta_k \delta_{\xi_k}.$$

Since b_1 is such that $(b_1)_i \geq 0, \ \forall i \in S$ and $||b_1|| = ||a_1||$, it follows from the induction hypothesis that there exists a set of vectors $\{\zeta_k \in K, \quad k = 1, 2, ..., M-1\}$ such that

$$b_1 = \sum_{k=1}^{M-1} \beta_k \zeta_k$$

and if we define

$$\Psi_1 = \sum_{k=1}^{M-1} \beta_k \delta_{\zeta_k},$$

then

$$d_K(\varphi_1, \Psi_1) = ||a_1 - b_1||.$$

Since $b = b_1 + \beta_M \zeta_M$ we find that

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

Defining

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

it follows by Proposition 3.1 and (53) that

$$d_K(\varphi, \Psi) \leq d_K(\varphi_1, \Psi_1) + d_K(\beta_M \delta_{\xi_M}, \beta_M \delta_{\zeta_M}) =$$

$$||a_1 - b_1|| + \beta_M ||\xi_M - \zeta_M|| = ||a - b||.$$

Again, because of Lemma 4.1, we also know that

$$d_K(\psi, \Psi) \geq ||a - b||$$

and hence

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

and thereby the induction step is completed and the lemma is proved. □

We now state yet another lemma which is a simple consequence of well-known results from the general theory on probability measures on complete, separable, metric spaces. See e.g [10] or [22], Chapter 2, section 6. We omit the proof.

Lemma 4.3 *For every $\mu \in \mathcal{P}(K)$ and every $\epsilon > 0$ we can find an integer N, a sequence $\{\alpha_k,\ k = 1, 2, ..., N\}$ of real positive numbers satisfying*

$$\sum_{k=1}^{N} \alpha_k = 1,$$

and a sequence $\{x_k, k = 1, 2, ..., N\}$ of elements in K such that if we define

$$\nu = \sum_{k=1}^{N} \alpha_k \delta_{x_k}$$

then

$$d_K(\mu, \nu) < \epsilon.$$

The following two results are simple consequences of Lemma 4.3, Lemma 4.1, Lemma 4.2 and the triangle inequality. We omit the details of the proofs. (When S is finite and consequently K is compact, then the conclusion of Corollary 4.2 below is well-known from the general theory on barycenters. See e.g Proposition 26.4 in [6].)

Corollary 4.1 *Let $q \in K$. For every $\mu \in \mathcal{P}(K|q)$ and every $\epsilon > 0$ we can find an integer N, a sequence $\{\alpha_k, k = 1, 2, ..., N\}$ of real positive numbers satisfying*

$$\sum_{k=1}^{N} \alpha_k = 1,$$

and a sequence $\{x_k, k = 1, 2, ..., N\}$ of elements in K such that if we define

$$\nu = \sum_{k=1}^{N} \alpha_k \delta_{x_k}$$

then

$$\nu \in \mathcal{P}(K|q)$$

and

$$d_K(\mu, \nu) < \epsilon.$$

Corollary 4.2 *Let $\mu \in \mathcal{P}$ and let $q \in K$. Then*

$$d_K(\mu, \mathcal{P}(K|q)) = ||\overline{b}(\mu) - q||.$$

Before we prove Theorem 4.2 we shall state one more lemma that follows from the general theory regarding probabilities on compact, metric spaces.

Lemma 4.4 *Let $K' \subset K$ be a compact set and let $\mathcal{P}(K')$ be defined by*

$$\mathcal{P}(K') = \{\nu \in \mathcal{P}(K) : \nu(K') = 1\}$$

Then for every $\epsilon > 0$ we can find an integer N and a set $\mathcal{R} = \{\nu_1, \nu_2, ..., \nu_N\}$ of probability measures in $\mathcal{P}(K')$ such that for every $\nu \in \mathcal{P}(K')$

$$d_K(\nu, \mathcal{R}) < \epsilon.$$

We shall now prove Theorem 4.2. We repeat its formulation.

Theorem 4.2. For each $q \in K$ the set $\mathcal{P}(K|q)$ is a *compact set* in the topology induced by the metric $d_K(\cdot, \cdot)$.

Proof. If the underlying set S is finite then the set K itself is compact and, as we pointed out in the beginning of this section, the set $\mathcal{P}(K|q)$ is a closed set and therefore, in this situation, $\mathcal{P}(K|q)$ is a compact set.

Thus assume that S is an infinite set. To simplify notations we assume that $S = \mathtt{N}$, the set of positive integers, and it is not difficult to convince oneself that this is no loss of generality. With this choice of S the set K is defined by

$$K = \{x = ((x)_i, i = 1, 2, ...) : (x)_i \geq 0, \sum_{i=1}^{\infty}(x)_i = 1\}.$$

Next let us also note that if q has finite support, that is: *there exists an integer N such that $(q)_i = 0$ if $i > N$*, then it follows that any measure $\nu \in \mathcal{P}(K|q)$ has support in the set $K' = \{x \in K : (x)_i = 0, i > N\}$. Since K' is a compact set it follows, as above, that $\mathcal{P}(K|q)$ is compact if q has finite support.

It remains to consider the case when the vector q does not have finite support. Thus let $q \in K$ be given and assume that for every integer N there exists a number $i > N$ such that $q_i > 0$. Since $(K, \mathcal{E})$ is a closed subset of the space l_1, which is a complete, separable, metric space it follows that $(K, \mathcal{E})$ is also a complete, separable, metric space. Therefore, $\mathcal{P}(K)$ is also a complete, separable, metric space. (From [10] Corollary 11.5.5 it follows that $\mathcal{P}(K)$ is complete, and that $\mathcal{P}(K)$ is separable follows easily from Lemma 4.3.) Therefore in order to prove that the set $\mathcal{P}(K|q)$ is compact in the topology induced by the Kantorovich distance it suffices to prove that the set $\mathcal{P}(K|q)$ is totally bounded.

Thus let ϵ, $0 < \epsilon < 1/2$ be chosen arbitrarily. What we shall do is to show that we can find a set $\mathcal{N} = \{\mu_1, \mu_2, ..., \mu_N\}$ of probability measures in $\mathcal{P}(K|q)$ such that for any $\mu \in \mathcal{P}(K|q)$ the following inequality holds:

$$d_K(\mu, \mathcal{N}) < \epsilon.$$

In order to do this, let is define the number L by

$$L = \max\{m : \sum_{n=m}^{\infty}(q)_i \geq \epsilon/6\}.$$

Next let us define

$$K' = \{x \in K : \sum_{i=1}^{L}(x)_i = 1\}.$$

Clearly K' is a compact subset of $(K, \mathcal{E})$.

Let us also define Δ_q by

$$\Delta_q = \sum_{i=L+1}^{\infty} (q)_i$$

and the vector $q' \in K$ by

$$(q')_i = (q)_i/(1 - \Delta_q),\ if\ 1 \leq i \leq L$$

$$(q')_i = 0\ \ if\ i \geq L + 1.$$

Clearly $q' \in K'$ and $\Delta_q <\ \epsilon/6$. Furthermore

$$||q - q'|| = \sum_{i=1}^{L} ((q)_i/(1 - \Delta_q) - (q)_i) + \Delta_q =$$

$$(\Delta_q/(1 - \Delta_q)) \sum_{i=1}^{L} (q)_i) + \Delta_q = 2\Delta_q < 2\epsilon/6. \tag{54}$$

Now, since K' is a compact set, it follows by Lemma 4.4 that there exist an integer N and a set $\mathcal{R} = \{\nu_1, \nu_2, ..., \nu_N\}$ of probability measures in $\mathcal{P}(K')$ such that for every $\nu \in \mathcal{P}(K'|q')$

$$d_K(\nu, \mathcal{R}) < \epsilon/6. \tag{55}$$

Furthermore, since $||q' - q|| < 2\epsilon/6$ (see (54)), and $||\bar{b}(\nu_j) - q'|| < \epsilon/6$ because of (55) and Lemma 4.1, it follows from Corollary 4.2 that we can find measures $\mu_j \in \mathcal{P}(K|q),\ j = 1, 2, ..., N$ such that

$$d_K(\nu_j, \mu_j) < 3\epsilon/6.$$

Now set $\mathcal{N} = \{\mu_1, \mu_2, ..., \mu_N\}$. We claim that

$$d_K(\mu, \mathcal{N}) < \epsilon$$

for all $\mu \in \mathcal{P}(K|q)$. But this is almost obvious from the way we constructed the set $\mathcal{N}$. For let $\mu \in \mathcal{P}(K|q)$. Since $||q - q'|| < 2\epsilon/6$ it follows from Corollary 4.2 that we can find a probability measure $\nu \in \mathcal{P}(K|q')$ such that

$$d_K(\mu, \nu) < 2\epsilon/6. \tag{56}$$

Then we can first find a measure $\nu^* \in \mathcal{R}$ such that

$$d_K(\nu, \nu^*) < \epsilon/6 \tag{57}$$

and then find a probability measure $\mu^* \in \mathcal{N}$ such that

$$d_K(\mu^*, \nu^*) < 3\epsilon/6. \tag{58}$$

Finally by the triangle inequality and the inequalities (56), (57) and (58) we conclude that

$$d_K(\mu, \mathcal{N}) < \epsilon$$

from which follows that the set $\mathcal{P}(K|q)$ is totally bounded which was what we needed to prove in order to prove that the set $\mathcal{P}(K|q)$ is compact. $\square$

Theorem 4.1. *Let S be infinite and let $q \in K$. Then $\mathcal{P}(K|q)$ is tight.*

Proof. Follows from [10], Theorem 11.5.4. $\square$

We end this section with the following lemma to be used later. For $i \in S$ and $0 < \eta \leq 1$, we define the set $E_i(\eta)$ by

$$E_i(\eta) = \{x \in K : (x)_i \geq \eta\}. \tag{59}$$

Lemma 4.5 *Let S, be denumerable, let $i \in S$, let $q \in K$ and suppose also that $(q)_i > 0$. Then we can find a compact set C such that for all $\mu \in \mathcal{P}(K|q)$*

$$\mu(E_i((q)_i/2) \cap C) \geq (q)_i/3. \tag{60}$$

Proof. Let $\mu \in \mathcal{P}(K|q)$. Since $\int_K (y)_i \mu(dy) = (q)_i$ and $0 \leq (y)_i \leq 1$ if $y \in K$ one easily obtains the estimate

$$\mu(E_i((q)_i/2)) \geq (q)_i/2$$

for all $\mu \in \mathcal{P}(K|q)$. Furthermore, since the set $\mathcal{P}(K|q)$ is tight, we can find a compact set $C \subset K$ such that

$$\mu(C) > 1 - (q)_i/6$$

for all $\mu \in \mathcal{P}(K|q)$. Therefore, if we set

$$B(i) = E_i((q)_i/2),$$

$$B^c(i) = K \setminus E_i((q)_i/2)$$

and

$$C^c = K \setminus C,$$

we obtain

$$\mu(B(i) \cap C) = 1 - \mu(B^c(i) \cup C^c) \geq 1 - \mu(B^c(i)) - \mu(C^c) =$$

$$\mu(B(i)) - (1 - \mu(C)) \geq (q)_i/2 - (q)_i/6 = (q)_i/3. \ \square$$

5 The barycenters of Markov chains induced by partitions.

In this section we prove two more theorems concerning barycenters. The latter together with Theorem 4.1 of the previous section implies a useful tightness result.

Theorem 5.1 *Let S be a denumerable set, let $\mathcal{M} \in \mathcal{G}(S)$ and let $P \in PM(S \times S)$ be the associated tr.pr.m. Then, for all $x \in K$,*

$$\overline{b}(\breve{P}^n_{\mathcal{M}}\delta_x) = xP^n, \ n = 1, 2, ... \tag{61}$$

Proof. For an arbitrary $i \in S$ define $u_i \in C[K]$ by

$$u_i(x) = (x)_i.$$

It then follows that

$$\langle u_i, \breve{P}_{\mathcal{M}}\delta_x\rangle = \langle T_{\mathcal{M}}u_i, \delta_x\rangle = T_{\mathcal{M}}u_i(x) = \sum_{w \in \mathcal{W}_{\mathcal{M}(x)}} u_i(xM(w)/||xM(w)||) \cdot ||xM(w)|| =$$

$$\sum_{w \in \mathcal{W}_{\mathcal{M}(x)}} ((xM(w))_i/||xM(w)||) \cdot ||xM(w)|| = \sum_{w \in \mathcal{W}_{\mathcal{M}(x)}} (xM(w))_i = (xP)_i$$

from which follows that (61) holds for $n = 1$.

That (61) also holds for $n \geq 2$ now follows from the fact that $\breve{P}^n_{\mathcal{M}} = \breve{P}_{\mathcal{M}^n}$ (see Corollary 2.1) and the fact that $\mathcal{M}^n$ is a partition of P^n. □

We shall now prove the following interesting result which we also state as a theorem.

Theorem 5.2 *Let S be a denumerable set, let $\mathcal{M} \in \mathcal{G}(S)$, let P be the associated tr.pr.m, let $\pi \in K$ and suppose that $\pi = \pi P$. Then,*

$$\breve{P}_{\mathcal{M}}\mu \in \mathcal{P}(K|\pi), \;\; \forall \; \mu \in \mathcal{P}(K|\pi).$$

Proof. First assume that $\mu \in \mathcal{P}(K|\pi)$ can be written

$$\mu = \sum_{k=1}^{N} \alpha_k \delta_{y_k} \tag{62}$$

for some integer N, where $\alpha_k > 0, k = 1, 2, ..., N$, $\sum_k \alpha_k = 1$, and $y_k \in K, \; k = 1, 2, ..., N$.

Next, let $i \in S$ be chosen arbitrarily, and let the function $u \in C[K]$ be defined by $u(x) = (x)_i$. Below we shall at a few places write

$$u_1 = T_{\mathcal{M}}u.$$

What we shall prove is that $\langle u, \breve{P}_{\mathcal{M}}\mu\rangle = (\pi)_i$. Using the fact that

$$u(\alpha x + \beta y) = \alpha u(x) + \beta u(y), \;\; x \in K, \; y \in K, \;\; \alpha > 0, \;\; \beta > 0, \;\; \alpha + \beta = 1,$$

and the fact that

$$(\alpha x + \beta y)P = \alpha xP + \beta yP, \;\; x \in K, \; y \in K, \;\; \alpha > 0, \;\; \beta > 0, \;\; \alpha + \beta = 1,$$

we find

$$\langle u, \breve{P}_{\mathcal{M}}\mu\rangle = \langle T_{\mathcal{M}}u, \mu\rangle = \langle u_1, \mu\rangle =$$

$$\int_K u_1(x)\mu(dx) = \sum_{k=1}^{N} \alpha_k u_1(y_k) =$$

$$\sum_{k=1}^{N} \alpha_k \sum_{w \in \mathcal{W}_{\mathcal{M}}(y_k)} u(y_kM(w))/||y_kM(w)||) \cdot ||y_kM(w)|| =$$

$$\sum_{k=1}^{N}\alpha_k\sum_{w\in\mathcal{W}_{\mathcal{M}}(y_k)}(y_kM(w))_i=\sum_{k=1}^{N}\alpha_k\sum_{w\in\mathcal{W}}(y_kM(w))_i=$$

$$\sum_{k=1}^{N}\alpha_k(y_kP)_i=((\sum_{k=1}^{N}\alpha_k y_k))P)_i=(\pi P)_i=(\pi)_i,$$

and thereby we have proved the assertion of the theorem when the measure μ can be written as in (62).

Next let $\mu\in\mathcal{P}(K|\pi)$ be chosen arbitrarily and let also $\epsilon>0$ be chosen arbitrarily. From Corollary 4.1 we know that we can find an integer N, a sequence $\{x_1,x_2,...,x_N\}$ of elements in K, and a sequence $\{\alpha_1,\alpha_2,...,\alpha_N\}$ of positive numbers satisfying $\sum_{j=1}^{N}\alpha_j=1$, such that if we define

$$\nu=\sum_{j=1}^{N}\alpha_j\delta_{x_j}$$

then
1)

$$\nu\in\mathcal{P}(K|\pi)$$

2)

$$d_K(\mu,\nu)<\epsilon/3.$$

Hence by Corollary 3.3 it follows that

$$d_K(T_{\mathcal{M}}\mu,T_{\mathcal{M}}\nu)<\epsilon.$$

Since $\gamma(u)=1$ when u is defined as above it follows that $u\in Lip_1[K]$ and since

$$\langle T_{\mathcal{M}}u,\nu\rangle=(\pi)_i$$

it follows that

$$|\langle T_{\mathcal{M}}u,\mu\rangle-(\pi)_i|<\epsilon$$

and since ϵ is arbitrary it follows that

$$\langle T_{\mathcal{M}}u,\mu\rangle=(\pi)_i.$$

Since also $i\in S$ was chosen arbitrarily, it follows that $\bar{b}(T_{\mathcal{M}}\mu)=\pi$ for all $\mu\in\mathcal{P}(K|\pi)$ which was what we wanted to prove. $\Box$

From Theorem 5.2 and Theorem 4.1 we now immediately obtain the following tightness result.

Theorem 5.3 *Let S be a denumerable set, let $\mathcal{M}\in\mathcal{G}(S)$, let P be the associated tr.pr.m, let $\pi\in K$ and suppose that $\pi P=\pi$. Then, for all $\mu\in\mathcal{P}(K|\pi)$*

$$\{\check{P}^n_{\mathcal{M}}\mu,\ \ n=1,2,...\}$$

is a tight sequence.

Conjecture 5.1 *Let S be denumerable, let $\mathcal{M}\in\mathcal{G}(S)$, let P be the associated tr.pr.m, let $\pi\in K$ and suppose that $\pi P=\pi$. Then*

$$d_K(\check{P}_{\mathcal{M}}\mu,\check{P}_{\mathcal{M}}\nu)\le d_K(\mu,\nu),\ \ \forall\mu,\nu\in\mathcal{P}(K|\pi).$$

6 An auxiliary theorem for Markov chains in complete, separable, metric spaces.

In this section we shall prove a limit theorem for Markov chains in a complete, separable, metric space, which we in the next section shall apply to Markov chains generated by tr.pr.fs induced by partitions of a tr.pr.ms.

In this section $(K, \mathcal{E})$ will denote an arbitrary complete, separable, metric space, with metric δ and $\sigma - algebra\ \mathcal{E}$. Other notations will be the same as before. For $u \in F[K]$ we define

$$osc(u) = \sup\{|u(x) - u(y)| : x \in K,\ y \in K\}.$$

As before $\gamma(u)$ denotes the Lipschitz constant defined by

$$\gamma(u) = \sup\{|u(x) - u(y)|/\delta(x, y) : x \in K,\ y \in K,\ x \neq y\}.$$

Let $Q : K \times \mathcal{E} \to [0, 1]$ be a tr.pr.f on $(K, \mathcal{E})$, and let $Q^n : K \times \mathcal{E} \to [0, 1],\ n = 1, 2, ...$ be the sequence of a tr.pr.fs defined recursively by

$$Q^1(x, B) = Q(x, B), \qquad B \in \mathcal{E}$$

$$Q^{n+1}(x, B) = \int_K Q^n(y, E)Q(x, dy), \ \ B\ \in \mathcal{E}.$$

We 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 $T^0u(x) = u(x)$. Note that

$$osc(T^{n+1}u) \leq osc(T^n u), \quad n = 0, 1, 2, ..., \quad u \in B[K]. \tag{63}$$

If a measure $\mu \in \mathcal{P}(K)$ is such that

$$\int_K Q(x, B)\mu(dx) = \mu(B), \ \ \forall B \in \mathcal{E}$$

then μ is an *invariant* probability measure of Q. If Q is such that

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

the Q is *Feller continuous.* If a set $\mathcal{D} \subset C[K]$ is such that for any two measures $\nu, \mu \in \mathcal{P}(K)$

$$\int_K u(y)\mu(dy) = \int_K u(y)\nu(dy),\ \forall\ u \in \mathcal{D}$$
$$\Longrightarrow$$
$$\int_K u(y)\mu(dy) = \int_K u(y)\nu(dy),\ \forall\ u \in C[K]$$

then we say that the set $\mathcal{D}$ is *measure-determining.*

It is well-known that $Lip[K]$ is measure-determining when $(K, \mathcal{E})$ is a complete, separable, metric space. (From [22], Chapter II, Theorem 6.1, follows that the subset of $C[K]$ consisting of uniformly continuous functions is measure-determining and since one, for any uniformly continuous function u and any $\epsilon > 0$, can find a function $v \in Lip[K]$ such that $||u - v|| < \epsilon$ it follows that $Lip[K]$ is measure-determining.)

When stating and proving the forthcoming theorem we shall use the notion *shrinking property* which we introduced in the introduction. We repeat it here for sake of convenience.

Definition 6.1 *Let Q be a tr.pr.f, and let T be the associated transition operator. If for every $\rho > 0$ there exists a number $\alpha,\ 0 < \alpha < 1$ such that for every nonempty, compact set $A \subset K$, every $\eta > 0$ and every $\kappa > 0$, there exist an integer N and another nonempty, compact set $B \subset K$ such that, if the integer $n \geq N$ then for all $u \in Lip[K]$*

$$\sup_{x,y \in A} |T^n u(x) - T^n u(y)| \leq \eta\gamma(u) + \kappa osc(u) + \alpha\rho\gamma(u) +$$

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

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

We now first prove the following lemma.

Lemma 6.1 *Suppose that the tr.p.f Q has the shrinking property. Then for every nonempty compact set $C \subset K$ and every $u \in Lip[K]$*

$$\lim_{n\to\infty} \sup_{x,y \in C} |\int_K u(z)Q^n(x,dz) - \int_K u(z)Q^n(y,dz)| = 0. \tag{64}$$

Proof. Let C be a given nonempty, compact set. Let also $\epsilon > 0$ be given. In order to prove the lemma it suffices to prove that, for every $u \in Lip[K]$, we can find an integer N such that

$$\sup_{x_0,y_0 \in C} |\int_K u(z)Q^n(x_0,dz) - \int_K u(z)Q^n(y_0,dz)| < 4\epsilon \tag{65}$$

if $n \geq N$.

Thus let $u \in Lip[K]$ be given. Obviously (64) holds if $\gamma(u) = 0$. Therefore we may assume that $\gamma(u) > 0$, which also implies that $osc(u) > 0$.

We now define

$$\rho = \epsilon/\gamma(u).$$

Next let $\alpha > 0$ be a shrinking number associated to ρ. Define the integer M by

$$M = \min\{m : (1-\alpha)^m < \epsilon/osc(u)\}. \tag{66}$$

From the shrinking property it now follows that if we choose $\eta = \eta_1 = \epsilon/(2\gamma(u))$ and choose $\kappa = \kappa_1 = \epsilon/(2osc(u))$, then we can find an integer N_1 and a compact set A_1 such that

$$\sup_{x_0,y_0 \in C} |T^n u(x_0) - T^n u(y_0)| \leq$$

$$\eta_1\gamma(u) + \kappa_1 osc(u) + \alpha\rho\gamma(u) + (1-\alpha) \sup_{x,y \in A_1} |T^{n-N_1}u(x) - T^{n-N_1}u(y)| \leq$$

$$\epsilon/2 + \epsilon/2 + \alpha\epsilon + (1-\alpha) \sup_{x,y \in A_1} |T^{n-N_1}u(x) - T^{n-N_1}u(y)| \quad (67)$$

for $n \geq N_1$.

Next setting $\eta = \eta_2 = (\epsilon/(4C\gamma(u))$ and $\kappa = \kappa_2 = \epsilon/(4osc(u))$, and again using the shrinking property, it follows that we can find an integer N_2 and a new compact set A_2 such that

$$\sup_{x_1,y_1 \in A_1} |T^m u(x_1) - T^m u(y_1)| \leq$$

$$\eta_2\gamma(u) + \kappa_2 osc(u) + \alpha\rho\gamma(u) + (1-\alpha) \sup_{x,y \in A_2} |T^{n-N_2}u(x) - T^{n-N_2}u(y)| \leq$$

$$\epsilon/4 + \epsilon/4 + \alpha\epsilon + (1-\alpha) \sup_{x,y \in A_2} |T^{m-N_2}u(x) - T^{m-N_2}u(y)| \quad (68)$$

for $m \geq N_2$.

Hence, by combining (67) and (68) we find that if $n \geq N_1 + N_2$ and we put $m = n - N_1$ then

$$\sup_{x_0,y_0 \in C} |T^n u(x_0) - T^n u(y_0)| \leq$$

$$\epsilon/2 + \epsilon/2 + \alpha\epsilon + (1-\alpha) \sup_{x,y \in A_1} |T^m u(x) - T^m u(y)| \leq$$

$$\epsilon/2+\epsilon/2+\alpha\epsilon+(1-\alpha)(\epsilon/4+\epsilon/4+\alpha\epsilon+(1-\alpha) \sup_{x,y \in A_2} |T^{m-N_2}u(x)-T^{m-N_2}u(y)|) <$$

$$(\epsilon/2 + \epsilon/4) + (\epsilon/2 + \epsilon/4) + \alpha\epsilon(1 + (1-\alpha))+$$

$$(1-\alpha)^2 \sup_{x,y \in A_2} |T^{n-(N_1+N_2)}u(x) - T^{n-(N_1+N_2)}u(y)|).$$

And in this way we proceed. Thus we define the numbers $\eta_i, i = 1, 2, ..., M$ by

$$\eta_i = \epsilon/((2^i)\gamma(u))$$

the numbers $\kappa_i, i = 1, 2, ..., M$ by

$$\kappa_i = \epsilon/((2^i)osc(u))$$

and having defined the compact sets A_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 define A_j and N_j such that

$$\sup_{x,y \in A_{j-1}} |T^m u(x) - T^m u(y)| \leq \eta_j\gamma(u) + \kappa_j osc(u) + \alpha\rho\gamma(u)+$$

$$(1-\alpha) \sup_{x,y \in A_j} |T^{m-N_j}u(x) - T^{m-N_j}u(y)|$$

if $m \geq N_j$. By induction follows that if the integer n satisfies $n \geq N_1 + N_2 + ... + N_j$ then

$$\sup_{x_0,y_0 \in C} |T^n u(x_0) - T^n u(y_0)| <$$

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

$$(1-\alpha)^j \sup_{x,y\in A_j} |T^{n-(N_1+N_2+...+N_j))}u(x) - T^{n-(N_1+N_2+...+N_j)}u(y)|.$$

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

$$\sup_{x_0,y_0\in C} |T^n u(x_0) - T^n u(y_0)| <$$

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

$$(1-\alpha)^M \sup_{x,y\in A_M} |T^{n-N}u(x) - T^{n-N}u(y)|$$

where $N = N_1 + N_2 + ... + N_M$, and by using (63) and the fact that

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

we find that if $n \geq N$ then

$$\sup_{x_0,y_0\in C} |T^n u(x_0) - T^n u(y_0)| < \epsilon + \epsilon + \epsilon + (1-\alpha)^M osc(u),$$

and since M is chosen in such a way that

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

(see (66)), it follows that

$$\sup_{x_0,y_0\in C} |T^n u(x_0) - T^n u(y_0)| < 4\epsilon$$

if $n \geq N$. Hence (65) holds if $n \geq N$ from which the lemma follows. □

Theorem 6.1 *Suppose that the tr.p.f Q is Feller continuous, that Q has the shrinking property and that there exists a point $x^* \in K$ such that the sequence $\{Q^n(x^*, \cdot), n = 1, 2, ...\}$ is a tight sequence of probability measures. Then Q is asymptotically stable.*

Proof. Since Q is Feller continuous and $\{Q^n(x^*, \cdot), n = 1, 2, ...\}$ is tight it is well-known from the general theory on Markov chains that there exists at least one invariant measure for Q.

That there is only one invariant measure ν, say, is not difficult to prove by contradiction if one uses Lemma 6.1. For suppose both ν and μ are invariant measures for Q. Suppose $||u|| = 1$, and that $a > 0$ where a is defined by

$$a = \int_K u(x)\mu(dx) - \int_K u(x)\nu(dx).$$

Since we have assumed that $(K, \mathcal{E})$ is a complete, separable, metric space both ν and μ are tight measures. (See e.g. [3], Theorem 1.4.) By choosing the compact set C sufficiently large it is clear that for every nonnegative integer $n = 0, 1, 2, ...$ 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 6.1 we can conclude that

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

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

After one has proved the uniqueness, it suffices to prove that

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

for all u in $Lip[K]$, since this set is measure-determining.

To do this let us first prove that

$$\lim_{n\to\infty} |T^{n+1}u(x^*) - T^n u(x^*)| = 0 \tag{69}$$

for $u \in Lip[K]$. Thus, let $u \in Lip[K]$ be given and let also $\epsilon > 0$ be given. Choose the compact set C so large that $(1 - Q(x^*, C))osc(u) < \epsilon/2$. Since

$$|T^{n+1}u(x^*) - T^n u(x^*)| = |\int_K (T^n u(y) - T^n u(x^*))Q(x^*, dy)|$$

it follows that

$$|T^{n+1}u(x^*) - T^n u(x^*)| \leq |\int_C (T^n u(y) - T^n u(x^*))Q(x^*, dy)| + \epsilon/2$$

and since

$$|\int_C (T^n u(y) - T^n u(x^*))Q(x^*, dy)| < \epsilon/2$$

if n sufficiently large because of Lemma 6.1, statement (69) follows.

Then, since $\{Q^n(x^*, \cdot), n = 1, 2, ...\}$ is tight, it follows that there exists a measure μ, say, and an increasing sequence of integers $n_j, j = 1, 2, ...$ such that

$$\lim_{j\to\infty} \int_K u(y)Q^{n_j}(x^*, dy) = \lim_{j\to\infty} T^{n_j} u(x^*) = \int_K u(y)\mu(dy) \tag{70}$$

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

Now let $u \in Lip[K]$ and set $u_1 = Tu$. Since $Tu \in Lip[K]$ if $u \in Lip[K]$, it follows from (70) that

$$\lim_{j\to\infty} \int_K u(y)Q^{n_j+1}(x^*, dy) =$$

$$\lim_{j\to\infty} T^{n_j+1}u(x^*) = \lim_{j\to\infty} T^{n_j} u_1(x^*) = \int_K u_1(y)\mu(dy)$$

which together with (69) and (70) implies that

$$\int_K u(y)\mu(dy) = \int_K Tu(y)\mu(dy).$$

Since $Lip[K]$ is measure-determining it follows that μ is an invariant measure of Q and therefore $\mu = \nu$ since we had proved that ν was the only invariant measure of Q. But since the right hand side of (70) is independent of the subsequence, it follows that we in fact have

$$\lim_{n\to\infty} \int_K u(y)Q^n(x^*, dy) = \lim_{n\to\infty} T^n u(x^*) = \int_K u(y)\nu(dy). \tag{71}$$

Finally let $x \in K$ be chosen arbitrarily. Since

$$\lim_{n\to\infty} T^n u(x^*) = \int_K u(y)\nu(dy),$$

it follows from Lemma 6.1 that also

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

and hence

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

for u in $Lip[K]$, and since this set is measure-determining it follows that Q is asymptotically stable. □

We end this section introducing the following terminology.

Definition 6.2 *Let Q be a tr.pr.f and let T be the associated transition operator. (i) If there exists a positive constant L such that for every $u \in Lip[K]$*

$$|Tu(x) - Tu(y)| \leq L\gamma(u)\delta(x, y),$$

then we say that Q is **Lipschitz continuous***.*
(ii) If there exists a positive constant L such that for every $u \in Lip[K]$

$$|T^n u(x) - T^n u(y)| \leq L\gamma(u)\delta(x, y), \quad n = 0, 1, 2, ...$$

then we say that Q is **Lipschitz equicontinuous***.*

Proposition 6.1 *Let S be a denumerable set, let $\mathcal{M} \in \mathcal{G}(S)$ and let $T_{\mathcal{M}}$ denote the transition operator on K induced by $\mathcal{M}$. Then $T_{\mathcal{M}}$ is both Lipschitz continuous and Lipschitz equicontinuous.*

Proof. Follows immediately from Corollary 3.1 and Corollary 3.2. □

7 The proof of Theorem 1.1.

Let us first repeat the definition of Condition B introduced in the introduction.

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*

$$\mu(C \cap \{x : (x)_{i_0} > (\pi)_{i_0}/2\}) \geq (\pi)_{i_0}/3, \quad \forall\, \mu \in \mathcal{P}(K|\pi),$$

then we can find an integer N, and a sequence $\{w_1, w_2, ...w_N\}$ of elements in $\mathcal{W}$, such that, if we set

$$\mathbf{M}(\mathbf{w^N}) = M(w_1)M(w_2)...M(w_N),$$

then

$$||e^{i_0}\mathbf{M}(\mathbf{w^N})|| > 0$$

and if $x \in C \cap \{x : (x)_{i_0} > (\pi)_{i_0}/2\}$ *then also*

$$||(x\mathbf{M}(\mathbf{w^N})/||x\mathbf{M}(\mathbf{w^N})|| - e^{i_0}\mathbf{M}(\mathbf{w^N})/||e^{i_0}\mathbf{M}(\mathbf{w^N})||)|| < \rho.$$

We now repeat the formulation of Theorem 1.1 using the notations introduced in section 2.

Theorem 1.1 *Let S be a denumerable set, let $\pi \in K$ be positive, let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\} \in \mathcal{G}_\pi(S)$ and suppose that Condition B holds. Then $\mathbf{P}_\mathcal{M}$ is asymptotically stable.*

Proof. As usual we denote the tr.pr.m associated to $\mathcal{M}$ by P, and since $\mathcal{M} \in \mathcal{G}_\pi(S)$ we know that $\pi P = \pi$.

Since $(K, \mathcal{E})$ is a complete, separable, metric space it suffices to prove that $\mathbf{P}_\mathcal{M}$ satisfies the hypotheses of Theorem 6.1.

That $\mathbf{P}_\mathcal{M}$ is Feller continuous follows easily from the fact that $\mathbf{P}_\mathcal{M}$ is Lipschitz continuous (see Proposition 6.1) and that $Lip[K]$ is measure determining. From Theorem 5.3 we conclude that $\{\mathbf{P}^n_\mathcal{M}(\pi, \cdot), n = 1, 2, ...\}$ is a tight sequence since $\mathcal{M} \in \mathcal{G}_\pi(S)$ and $\delta_\pi \in \mathcal{P}(K|\pi)$.

It thus remains to show that $\mathbf{P}_\mathcal{M}$ has the shrinking property. To simplify notations we shall throughout the rest of this proof denote the transition probability function $\mathbf{P}_\mathcal{M}$ by $\mathbf{P}$, denote the transition operator $T_\mathcal{M}$ by T and the transition probability operator $\breve{P}_\mathcal{M}$ by $\breve{P}$. Also recall that we have introduced the notation $[x] = x/||x||$ when $||x|| \neq 0$. (See (25).)

Thus let $\rho > 0$ be given. What we have to do is to show that we can find a number $\alpha > 0$ such that for each nonempty compact set A and each $\eta > 0$ and each $\kappa > 0$ we can find an integer N and another nonempty compact set B such that for each $u \in Lip[K]$

$$\sup_{x,y \in A} |T^n u(x) - T^n u(y)| \leq \eta\gamma(u) + \kappa osc(u) + \alpha\rho\gamma(u) +$$

$$(1 - \alpha) \sup_{z,z' \in B} |T^{n-N} u(z) - T^{n-N} u(z')|. \tag{72}$$

Let A be a given nonempty compact set, and let also $\eta > 0$ and $\kappa > 0$ be given. From Corollary 3.2 we know that for every $u \in Lip[K]$

$$|T^n u(x) - T^n u(y)| \leq 3\gamma(u)\delta(x, y), \quad n = 0, 1, 2, ...$$

Let us next recall, from the general theory of Markov chains, that for any $z \in K$ and any $\epsilon > 0$, we can find an integer N' such that if $n \geq N'$ then

$$||zP^n - \pi|| < \epsilon.$$

(See e.g [21], Chapter 2.) Since the set A given above is compact, it follows that we can find an integer N_1 such that for all $z \in A$

$$||zP^n - \pi|| < \eta/6$$

if $n \geq N_1$.

Now let $x \in A$ and $y \in A$ be given. Set

$$\mu_{n,x}(\cdot) = \mathbf{P}^n(x, \cdot), \quad n = 1, 2, ...$$

and

$$\mu_{n,y}(\cdot) = \mathbf{P}^n(y, \cdot), \quad n = 1, 2, ...$$

From Theorem 5.1 follows that $\overline{b}(\mu_{n,x}) = xP^n$ and that $\overline{b}(\mu_{n,y}) = yP^n$. Therefore, if $n \geq N_1$, where N_1 is defined as above, we conclude that

$$||\overline{b}(\mu_{n,x}) - \pi|| < \eta/6$$

and that

$$||\overline{b}(\mu_{n,y}) - \pi|| < \eta/6.$$

From Corollary 4.2 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

$$|\int_K u(z)\mu_{N_1,x}(dz) - \int_K u(z)\nu_x(dz)| \leq \gamma(u)\eta/6$$

and

$$|\int_K u(z)\mu_{N_1,y}(dz) - \int_K u(z)\nu_y(dz)| \leq \gamma(u)\eta/6.$$

From Corollary 3.2 we also find that for $m = 0, 1, 2, ...$

$$|\int_K T^m u(z)\mu_{N_1,x}(dz) - \int_K T^m u(z)\nu_x(dz)| \leq 3\gamma(u)\eta/6 = \gamma(u)\eta/2$$

and similarly that

$$|\int_K T^m_{\mathcal{M}} u(z)\mu_{N_1,y}(dz) - \int_K T^m_{\mathcal{M}} u(z)\nu_y(dz)| \leq 3\gamma(u)\eta/6 = \gamma(u)\eta/2.$$

Thus if $n \geq N_1$ we have

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

$$|\int_K T^{n-N_1} u(z)\nu_x(dz) - \int_K T^{n-N_1} u(z)\nu_y(dz)|. \tag{73}$$

We shall next define a shrinking coefficient $\alpha > 0$ associated to the given number ρ. To do this we shall use Lemma 4.5 and Condition B.

From Lemma 4.5, it follows that for each $i \in S$ we can find a compact set C_i such that for all $\mu \in \mathcal{P}(K|\pi)$

$$\mu(C_i \cap E_i((\pi)_i/2)) \geq (\pi)_i/3. \tag{74}$$

From Condition B it follows that we can find an element $i_0 \in S$, an integer N_0 and a sequence $\{w_1, w_2, ...w_{N_0}\}$ of elements in $\mathcal{W}$ depending on the set C_{i_0}, such that if we set $M(w_1)M(w_2)...M(w_{N_0}) = \mathbf{M}(\mathbf{w}^{\mathbf{N_0}})$ then

$$||e^{i_0}\mathbf{M}(\mathbf{w}^{\mathbf{N_0}}))|| > 0$$

and

$$||([x\mathbf{M}(\mathbf{w}^{\mathbf{N_0}})] - [e^{i_0}\mathbf{M}(\mathbf{w}^{\mathbf{N_0}})])|| < \rho/6 \quad \forall x \in E_{i_0}((\pi)_{i_0}/2) \cap C_{i_0}. \tag{75}$$

Next let us define α_1 by

$$\alpha_1 = ((\pi)_{i_0}/3) \cdot ((\pi)_{i_0}/2) \cdot ||e^{i_0}\mathbf{M}(\mathbf{w}^{\mathbf{N_0}})||, \tag{76}$$

and let us define

$$\alpha = \alpha_1^2/2.$$

Our aim is thus to verify (72) with this choice of α and with $N = N_1 + N_0$. In order to do this let us first set

$$\nu_x^* = \breve{P}^{N_0}\nu_x,$$

$$\nu_y^* = \breve{P}^{N_0}\nu_y,$$

let us write

$$K^{(2)} = K \times K$$

and

$$\mathcal{E}^{(2)} = \mathcal{E} \otimes \mathcal{E},$$

and let $\tilde{\nu}_{x,y}^*$ denote the product measure on $(K^{(2)}, \mathcal{E}^{(2)})$ determined by ν_x^* and ν_y^*.

Furthermore let us denote

$$q_0 = e^{i_0}\mathbf{M}(\mathbf{w}^{N_0})/||e^{i_0}\mathbf{M}(\mathbf{w}^{N_0})||$$

and

$$D = \{z \in K : \delta(z, q_0) < \rho/6\}.$$

Since $(z)_{i_0} \geq (\pi)_{i_0}/2$ if $z \in E_{i_0}((\pi)_{i_0}/2)$ and

$$||xA|| \geq (x)_i ||e^i A||, \ \ \forall i \in S, \tag{77}$$

if A is a nonnegative matrix and $x \in K$, we conclude from (74), (75), (76) and (77) that

$$\nu_x^*(D) \geq ((\pi)_{i_0}/3) \cdot ((\pi)_{i_0}/2) \cdot ||e^{i_0}\mathbf{M}^{N_0}(\mathbf{w}^{N_0})|| = \alpha_1$$

and also that

$$\nu_y^*(D) \geq ((\pi)_{i_0}/3) \cdot ((\pi)_{i_0}/2) \cdot ||e^{i_0}\mathbf{M}^{N_0}(\mathbf{w}^{N_0})|| = \alpha_1.$$

Since $\nu_x, \nu_y \in \mathcal{P}(K|q)$, $\nu_x^* = \breve{P}^{N_0}\nu_x$ and $\nu_y^* = \breve{P}^{N_0}\nu_y$ it follows from Theorem 5.2 that $\nu_x^*, \nu_y^* \in \mathcal{P}(K|q)$. Since $\mathcal{P}(K|q)$ is a tight set it follows that we can find a compact set B independent of $x, y \in A$ such that

$$\nu_x^*(D \cap B) \geq \alpha_1/\sqrt{2},$$

$$\nu_y^*(D \cap B) \geq \alpha_1/\sqrt{2},$$

and also

$$\nu_x^*(B) \geq 1 - \kappa/2$$
$$\nu_y^*(B) \geq 1 - \kappa/2.$$

Therefore, if we denote

$$B^{(2)} = \{(z,z') \in K^{(2)} : z \in B,\ z' \in B\ \}$$

and set

$$B_1^{(2)} = \{(z,z') \in B^{(2)} : \delta(z,z') < \rho/3\},$$

we can conclude that

$$\tilde{\nu}^*_{x,y}(B_1^{(2)}) \geq \nu^*_x(D \cap B) \cdot \nu^*_y(D \cap B) \geq \alpha_1^2/2 = \alpha, \tag{78}$$

and that

$$\tilde{\nu}^*_{x,y}(B^{(2)}) \geq (1-\kappa/2)^2 > 1-\kappa.$$

Next let us define the sets $B_2^{(2)}$ and $B_3^{(2)}$ by

$$B_2^{(2)} = \{(z,z') \in B^{(2)} :\ \ \delta(z,z') \geq \rho/3\}$$

and

$$B_3^{(2)} = K^{(2)} \setminus B^{(2)}.$$

Obviously

$$B_k^{(2)} \cap B_m^{(2)} = \emptyset,\ \ 1 \leq k < m \leq 3$$

and

$$\bigcup_{i=1}^{3} B_i^{(2)} = K^{(2)}.$$

We shall now estimate

$$|\int_K T^{n-N_1}u(z)\nu_x(dz) - \int_K T^{n-N_1}u(z)\nu_y(dz)|$$

for $n \geq N_1 + N_0 = N$. We set

$$m = n - N \quad \textit{and} \quad v = T^m u. \tag{79}$$

Note first that if $n \geq N$ then

$$|\int_K T^{n-N_1}u(z)\nu_x(dz) - \int_K T^{n-N_1}u(z)\nu_y(dz)| =$$
$$|\langle T^{m+N_0}u, \nu_x\rangle - \langle T^{m+N_0}u, \nu_y\rangle| =$$
$$|\langle T^m u, \nu^*_x\rangle - \langle T^m u, \nu^*_y\rangle| =$$
$$|\langle v, \nu^*_x\rangle - \langle v, \nu^*_y\rangle|. \tag{80}$$

Next let us note that since T is a transition operator it follows that

$$osc(v) = osc(T^m u) \leq osc(u), \tag{81}$$

and from Corollary 3.2 we also find that

$$\gamma(v) = \gamma(T^m u) \leq 3\gamma(u). \tag{82}$$

Since $\tilde{\nu}^*_{x,y}$ is the product measure of ν^*_x and ν^*_y it is clear from (80) that

$$|\int_K T^{n-N_1}u(z)\nu_x(dz) - \int_K T^{n-N_1}u(z)\nu_y(dz)| =$$

$$|\int_{K^{(2)}} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz, dz')|. \tag{83}$$

From the definitions of the sets $B_1^{(2)}$, $B_2^{2)}$ and $B_3^{(2)}$ we then find that

$$|\int_{K^{(2)}} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz, dz')| \leq$$

$$|\int_{B_1^{(2)}} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz, dz')|+$$

$$|\int_{B_2^{(2)}} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz, dz')|+$$

$$|\int_{B_3^{(2)}} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz, dz')|. \tag{84}$$

We have already proved that

$$\tilde{\nu}^*_{x,y}(B_3^{(2)}) < \kappa$$

and hence

$$|\int_{B_3^{(2)}} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz, dz')| \leq osc(v)\kappa. \tag{85}$$

Next let us write

$$\Theta = \sup\{|v(z) - v(z')| : z \in B, z' \in B\}. \tag{86}$$

Then

$$|\int_{B_2^{(2)}} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz, dz')| \leq \Theta\tilde{\nu}^*_{x,y}(B_2^{(2)}) \leq$$

$$\Theta(1 - \tilde{\nu}^*_{x,y}(B_1^{(2)})). \tag{87}$$

But if $0 < \alpha \leq \beta \leq 1$, $\epsilon > 0$ and $\Theta > 0$, then by elementary calculations we find

$$\beta \min\{\epsilon, \Theta\} + (1 - \beta)\Theta \leq \alpha\epsilon + (1 - \alpha)\Theta. \tag{88}$$

Since $\tilde{\nu}^*_{x,y}(B_1^{(2)}) \geq \alpha$ because of (78), it follows from (87) and (88) that

$$|\int_{B_1^{(2)}} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz, dz')|+$$

$$|\int_{B_2^{(2)}} (v(z) - v(z'))\tilde{\nu}^*_{x,y}(dz, dz')| \leq$$

$$\min\{\gamma(v)\rho/3, \Theta\} \cdot \tilde{\nu}^*_{x,y}(B_1^{(2)}) + \Theta(1 - \tilde{\nu}^*_{x,y}(B_1^{(2)})) \leq$$

$$\alpha\gamma(v)\rho/3 + (1 - \alpha)\Theta$$

which combined with (83), (84), (85), (81), and (82), implies that

$$|\int_K T^{n-N_1}u(z)\nu_x(dz) - \int_K T^{n-N_1}u(z)\nu_y(dz)| \leq$$

$$osc(u)\kappa + \alpha\gamma(u)\rho + (1-\alpha)\Theta. \tag{89}$$

Hence, from (73), (89) and the definition of Θ (see (86) and (79)) it finally follows that

$$|T^n u(x) - T^n u(y)| \leq$$
$$\gamma(u)\eta + osc(u)\kappa + \alpha\gamma(u)\rho +$$
$$(1-\alpha)\sup\{|T^{n-N}u(z) - T^{n-N}u(z')| : z \in B, z' \in B\}$$

and since x and y were arbitrary chosen in A, it follows that (72) holds for all $u \in Lip[K]$ and hence the tr.p.f $\mathbf{P}_{\mathcal{M}}$ has the shrinking property which was what we wanted to prove in order to complete the proof of Theorem 1.1. □

8 Exceptional cases.

We first repeat the example presented in [16].

Example 8.1 *([16]) Let* $S = \{1,2,3,4\}$ *and define* $P \in PM(S \times S)$ *by*

$$P = \begin{pmatrix} \star & 0 & \star & 0 \\ 0 & \star & 0 & \star \\ \star & 0 & 0 & \star \\ 0 & \star & \star & 0 \end{pmatrix},$$

where each $\star$ *denotes the value 1/2. Obviously* P *is aperiodic and irreducible. Let* $\mathcal{M} = \{M(1), M(2)\}$ *be a partition of* P *such that 1) the first two columns of* $M(1)$ *and* P *are equal and 2) the last two columns of* $M(2)$ *and* P *are equal. Let* $S' = \{1,2\}$*. It is not difficult to show that the hypotheses of Theorem 1.2 are fulfilled with this choice of* S'*, and therefore* $\mathbf{P}_{\mathcal{M}}$ *is not asymptotically stable.*

Kesten's counterexample to Blackwell's conjecture is as follows.

Example 8.2 *([18]) Let* $S = \{1,2,...,8\}$ *and define* $P \in PM(S \times S)$ *by*

$$P = \begin{pmatrix} \star & 0 & 0 & 0 & \star & 0 & 0 & 0 \\ 0 & \star & 0 & 0 & 0 & \star & 0 & 0 \\ 0 & 0 & 0 & \star & 0 & 0 & 0 & \star \\ 0 & 0 & \star & 0 & 0 & 0 & \star & 0 \\ \star & 0 & 0 & 0 & 0 & 0 & 0 & \star \\ 0 & \star & 0 & 0 & 0 & 0 & \star & 0 \\ 0 & 0 & 0 & \star & \star & 0 & 0 & 0 \\ 0 & 0 & \star & 0 & 0 & \star & 0 & 0 \end{pmatrix},$$

where each $\star$ *denotes the value 1/2. Obviously* P *is aperiodic and irreducible. Let* $\mathcal{M} = \{M(1), M(2)\}$ *be a partition of* P *such that 1) the first four columns of* $M(1)$ *and* P *are equal and 2) the four last columns of* $M(2)$ *and* P *are equal.*

Let $S' = \{1,2,3,4\}$*. It is again not difficult to show that the hypotheses of Theorem 1.2 are fulfilled with this choice of* S'*, and therefore* $\mathbf{P}_{\mathcal{M}}$ *is not asymptotically stable. It is also not difficult to verify that if, for example* $x \in K_{S'}$ *is such that* $0 < (x)_1 = (x)_3 < (x)_2 = (x)_4$*, then* $\mathbf{P}_M(x)$ *is a* **periodic** *Markov chain taking its values in a subset of* K *consisting of just 8 elements.*

We shall now prove Theorem 1.2. We repeat its formulation for convenience, using the notation $x_n(\mathbf{w^n}) = [x\mathbf{M}(\mathbf{w^n})] = x\mathbf{M}(\mathbf{w^n})/||x\mathbf{M}(\mathbf{w^n})||$ introduced in section 2.

Theorem 1.2 *Let S be a denumerable set, let $P \in PM(S \times S)$ and let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ be a partition of P.*

Suppose that there exists a subset $S' \subset S$ consisting of at least two elements, such that
1) for every $x \in K_{S'}$ the set $\cup_{n=1}^{\infty} K(x, \mathcal{M}^n)$ consists of isolated points,
2) if both x and y are in $K_{S'}$ then $\mathcal{W}_{\mathcal{M}^n}(x) = \mathcal{W}_{\mathcal{M}^n}(y)$, $n = 1, 2, ...$
and
3) if x and y in $K_{S'}$, then

$$||x_n(\mathbf{w^n}) - y_n(\mathbf{w^n})|| = ||x - y||, \quad n \geq 1, \ \mathbf{w^n} \in \mathcal{W}_{\mathcal{M}^n}(x).$$

If these conditions are fulfilled then $\mathbf{P}_{\mathcal{M}}$ is **not** *asymptotically stable.*

Proof. Let $S' \subset S$ be as in the hypotheses of the theorem. Let $x \in K_{S'}$ and set

$$K'(x) = \cup_{n=1}^{\infty} K(x, \mathcal{M}^n).$$

Because of hypothesis 1), the set $K'(x)$ consists of isolated points, and because of hypothesis 3) it is not difficult to convince oneself that $K'(x)$ must contain at least two points. Therefore

$$\epsilon_0 = \inf\{||z_1 - z_2|| : z_1, z_2 \in K'(x), \ z_1 \neq z_2\} > 0.$$

Since $x \in K_{S'}$ and S' consists of at least two elements, we can find an element $y \in K_{S'}$ such that $||x - y|| = \epsilon_0/2$.

Now let n denote an arbitrary positive integer, let

$$K_1 = \{x_n(\mathbf{w^n}) : \mathbf{w^n} \in \mathcal{W}_{\mathcal{M}^n}(x)\}$$

and

$$K_2 = \{y_n(\mathbf{w^n}) : \mathbf{w^n} \in \mathcal{W}_{\mathcal{M}^n}(x)\}$$

Since the points in $K'(x)$ are isolated points, it is clear that K_1 is the support of $\mathbf{P}^n_{\mathcal{M}}(x, \cdot)$ and since $\mathcal{W}_{\mathcal{M}^n}(y) = \mathcal{W}_{\mathcal{M}^n}(x)$ because of hypothesis 2) and also the set $\cup_{n=1}^{\infty} K(y, \mathcal{M}^n)$ consists of isolated points, it is clear that K_2 is the support of $\mathbf{P}^n_{\mathcal{M}}(y, \cdot)$. Since

$$\inf\{\delta(z, K_1) : z \in K_2\} = \epsilon_0/2$$

because of hypothesis 3), it therefore follows that the Kantorovich distance

$$d_K(\mathbf{P}^n_{\mathcal{M}}(x, \cdot), \mathbf{P}^n_{\mathcal{M}}(y, \cdot)) \geq \epsilon_0/2 > 0$$

and since n was an arbitrary integer

$$d_K(\mathbf{P}^n_{\mathcal{M}}(x, \cdot), \mathbf{P}^n_{\mathcal{M}}(y, \cdot)) \geq \epsilon_0/2$$

for $n = 1, 2, ...$ which implies that the tr.pr.f $\mathbf{P}^n_{\mathcal{M}}(\cdot, \cdot)$ can not be asymptotically stable. □

We shall next describe a family of tr.pr.ms for which one, for each matrix belonging to the family, can find a partition such that the induced tr.pr.f is *not* asymptotically stable.

Let $\mathcal{I}$ denote a denumerable set, let $d \geq 2$ be an integer, set $I_d = \{1, 2, ..., d\}$, and define the set S as the Cartesian product of $\mathcal{I}$ and I_d, that is

$$S = \{(i, j), i \in \mathcal{I}, j \in I_d\}.$$

Let $A \in PM(\mathcal{I} \times \mathcal{I})$, and let $\mathcal{M} = \{M(w) : \ w \in \mathcal{W}\}$ be a partition of A.

Next let $Perm(d)$ denote the set of $d \times d$ *permutation matrices*. For each $w \in \mathcal{W}$ and each $(i, k) \in \mathcal{I} \times \mathcal{I}$ we now associate a matrix $Q(i, k, w) \in Perm(d)$. We write

$$\mathcal{Q}_{\mathcal{I},\mathcal{W}} = \{Q(i, k, w) : \ (i, j) \in \mathcal{I} \times \mathcal{I}, \ w \in \mathcal{W}\}.$$

We define the set

$$\mathcal{M}' = \{M'(w) : w \in \mathcal{W}\}$$

of $S \times S$ matrices by

$$(M'(w))_{(i,j),(k,m)} = (M(w))_{i,k} \cdot (Q(i, k, w))_{j,m}, \tag{90}$$

and define the $S \times S$ matrix P by

$$(P)_{(i,j),(k,m)} = \sum_{w \in \mathcal{W}} (M'(w))_{(i,j),(k,m)}. \tag{91}$$

Proposition 8.1 *The matrix P belongs to $PM(S \times S)$.*

Proof. Obviously $(P)_{(i,j),(k,m)} \geq 0$. It remains to show that

$$\sum_{k,m} (P)_{(i,j),(k,m)} = 1, \ \ \forall (i, j) \in S.$$

Thus let $(i, j) \in S$ be given. From the definition we know that

$$\sum_{k,m} (P)_{(i,j),(k,m)} = \sum_{k,m} \sum_{w \in \mathcal{W}} (M(w))_{i,k} \cdot (Q(i, k, w))_{j,m}$$

and interchanging the order of summation we obtain

$$\sum_{k,m} (P)_{(i,j),(k,m)} = \sum_{w \in \mathcal{W}} \sum_{k} (M(w))_{i,k} \cdot \sum_{m} (Q(i, k, w))_{j,m}.$$

Since $Q(i, k, w)$ is a permutation matrix it follows that

$$\sum_{m} (Q(i, k, w))_{j,m} = 1.$$

Hence

$$\sum_{k,m} (P)_{(i,j),(k,m)} = \sum_{w \in \mathcal{W}} \sum_{k} (M(w))_{i,k} = \sum_{k} \sum_{w \in \mathcal{W}} (M(w))_{i,k} = \sum_{k} (P)_{i,k} = 1. \ \square$$

We call P the tr.pr.m *generated* by A and $\mathcal{Q}_{\mathcal{I},\mathcal{W}}$.

Corollary 8.1 *The set $\mathcal{M}'$ is a partition of P.*

Proof. Follows from (91). □

Next suppose that the partition $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ of A is such that

$$(M(w))_{i,k} > 0 \;\;\Rightarrow\;\; (M(w))_{i,k_1} = 0, \; if \;\; k_1 \neq k, \;\; \forall M(w) \in \mathcal{M}. \qquad (92)$$

Proposition 8.2 *Let $\mathcal{I}$ denote a denumerable set, let $A \in PM(\mathcal{I} \times \mathcal{I})$ and let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ be a partition of A that satisfies (92). Let $d \geq 2$, set $I_d = \{1, 2, ..., d\}$, and let*

$$\mathcal{Q}_{\mathcal{I},\mathcal{W}} = \{Q(i,k,w) : \; (i,j) \in \mathcal{I} \times \mathcal{I}, \; w \in \mathcal{W}\}.$$

Let $S = \{(i,j), i \in \mathcal{I}, j \in I_d\}$, let $P \in PM(S \times S)$ be the tr.pr.m generated by A and $\mathcal{Q}_{\mathcal{I},\mathcal{W}}$ and let $\mathcal{M}'$ be the partition of P defined by (90).

Then $\mathcal{M}'$ satisfies the hypotheses of Theorem 1.2.

Proof. The proof is based on the following observation. For $i \in \mathcal{I}$, let $S'_i = \{(i,j), j = 1, 2, ...d\}$. Let $x \in K_{S'_i}$ and suppose $||xM'(w)|| > 0$. From (92) and (90) follows that $xM'(w)/||xM'(w)|| \in K_{S'_k}$ where thus k is such that $(M(w))_{i,k} > 0$. Furthermore, if we let x' denote the *d-dimensional* vector defined by $(x')_j = (x)_{i,j}, \; j = 1, 2, ..., d$, set $z = xM'(w)/||xM'(w)||$ and let z' denote the *d-dimensional* vector defined by $(z')_j = (z)_{k,j}, \;\; j = 1, 2, ..., d,$ then $z' = x'Q(i,k,w)$. Since i was arbitrary it now easily follows that the hypotheses of Theorem 1.2 are fulfilled. •

It is easy to show that both Example 8.1 and Example 8.2 can be put into the framework of the class just described. We show this for Example 8.2.

Example 8.3 *We need to define a denumerable set $\mathcal{I}$, an integer d, a tr.pr.m $A \in PM(\mathcal{I} \times \mathcal{I})$, a partition $\mathcal{M}$ of A that satisfies (92), and a set of permutation matrices. We choose $\mathcal{I} = \{1, 2\}$, define the matrix A in the simplest possible way by*

$$(A)_{i,k} = 1/2, \;\; i = 1, 2 \;\; k = 1, 2,$$

and let the partition $\mathcal{M} = \{M(1), M(2)\}$ of A be defined such that the first column of $M(1)$ is equal to the first column of A and the second column of $M(2)$ is equal to the second column of A. Clearly $\mathcal{M}$ satisfies (92).

We choose $d = 4$. The state space S is thus $\{1, 2\} \times \{1, 2, 3, 4\}$ which consists of 8 elements. It remains to determine the set $\mathcal{Q}_{\mathcal{I},\mathcal{W}}$ of permutation matrices. We choose the permutation matrices independent of $w \in \mathcal{W}$. Therefore we denote the permutation matrix $Q(i,j,w)$ by $Q(i,j)$. We thus have to determine four permutation matrices $Q(i,k))$, $\;\; i = 1, 2 \;\; k = 1, 2,$ of format 4×4. We define $Q(1,1)$ by

$$(Q(1,1))_{j,j} = 1, \;\; j = 1, 2$$

$$(Q(1,1))_{3,4} = (Q(1,1))_{4,3} = 1,$$

and

$$(Q(1,1))_{j,m} = 0 \;\; otherwise,$$

We let the matrices $Q(1,2)$ and $Q(2,1)$ be defined in the same way as $Q(1,1)$ and finally we define the matrix $Q(2,2)$ by

$$(Q(2,2))_{1,4} = (Q(2,2))_{2,3} = (Q(2,2))_{3,1} = (Q(2,2))_{4,2} = 1$$

$$(Q(2,2))_{j,m} = 0 \;\; otherwise.$$

It is easily checked that the matrix P generated by A and $\mathcal{Q}_{\mathcal{I},\mathcal{W}}$ is aperiodic and irreducible.

The reason that for some (most) initial values the tr.pr.f $\mathbf{P}_{\mathcal{M}}(x,\cdot)$ *is a* **periodic** *Markov chain is due to the fact that all permutation matrices in the set $\mathcal{Q}_{\mathcal{I},\mathcal{W}}$ correspond to* **odd permutations**.

Our last example can be considered as a separate class of examples.

Example 8.4 *Suppose that S is finite set of size $d \geq 2$, and that $P \in PM(S \times S)$ is* **doubly stochastic** *(the transpose of P does also belong to $PM(S \times S)$). It is well-known that each such matrix P can be written as $\sum_{w=1}^{N} a_w Q_w$ where $a_w > 0,\ w = 1,2,...N$, $\sum_{w=1}^{N} a_w = 1$ and each $Q_w, w = 1,2,...N$ is a permutation matrix. If we now simply define $\mathcal{M} = \{a_w Q_w : w = 1,2,...,N\}$ it is not difficult to prove that the hypotheses of Theorem 1.2 are fulfilled.*

Furthermore if we define $\mathcal{I} = \{1\}$, let A denote the 1×1 matrix whose only element is equal to one, redefine $\mathcal{M}$ by $\mathcal{M} = \{a_w A :\ w = 1,2,...,N\}$, let d be the same as before, define $\mathcal{Q}_{\mathcal{I}} = \{Q_w,\ w = 1,2,...N\}$, let $\mathcal{M}'$ be defined by (90) with $M(w) = a_w A$, and let P be defined by (91), we see that this example belongs to the class considered above.

9 On Condition B.

In this section we shall first prove that Condition B1 (see subsection 1.5) implies Condition B. Then we shall make a precise statement of the condition introduced in [19] which we have called the *"rank 1 condition"*, and we prove that if the set S is finite then this "rank 1 condition" implies Condition B1.

We also introduce a condition which we call Condition P, which is adapted to Perron's classical theorem regarding matrices with positive elements (see e.g [13], vol II, Theorem 8.1), and prove that Condition P implies Condition B1.

A condition which seems easy to check in practice, when the space S is finite, is Condition A, introduced in [16]. In the last part of this section we prove a result based on a condition which can be regarded as a slight generalization of Condition A of [16].

Proposition 9.1 *Let S be a denumerable set, let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\} \in \mathcal{G}'(S)$, let $P \in PM(S \times S)$ be the associated tr.p.m, and let π be the positive vector in K that satisfies $\pi P = \pi$. Then, if Condition B1 is satisfied, it follows that Condition B is also satisfied.*

Proof. In order to prove Proposition 9.1 we shall first prove the following lemma in which we formulate some simple inequalities for matrices approaching a matrix in the set `W`. (For the definition of the set `W` see subsection 1.3.)

Lemma 9.1 *Let $u \in \mathcal{U}$, $v \in K$, and define $W = u^c v$. Let i_0 be such that*

$$(u)_{i_0} > 0.$$

Let $\{W_n, n = 1,2,...\}$ be a sequence of matrices of the same format as W, and assume that
1) for $n = 1,2,...$

$$||W_n|| = 1,$$

2)

$$\lim_{n\to\infty} ||e^i(W_n - W)|| = 0,\ i \in S.$$

Then
1) to every $\epsilon > 0$ and every nonempty compact set $C \subset K$ there exists an integer $N = N_{\epsilon,C}$ such that if $x \in C$ then for all integers $n \geq N$

$$|(||xW_n|| - ||xW||)| < \epsilon;$$

2) to every η, $0 < \eta < 1$, there exists an integer N_η such that if $x \in K$ is such that

$$(x)_{i_0} \geq \eta$$

then for all integers $n \geq N_\eta$

$$||xW_n|| > (u)_{i_0}\eta/2\ ;$$

3) to every η, $0 < \eta < 1$ and every nonempty compact set $C \subset K$ and every $\gamma > 0$, there exists an integer $N = N_{\gamma,C,\eta}$ such that if $x \in C$ and also $(x)_{i_0} \geq \eta$ then $||xW_n|| > 0$ for all integers $n \geq N$ and furthermore

$$||xW_n/||xW_n|| - v|| < \gamma;$$

4) to every η, $0 < \eta < 1$ and every nonempty compact set $C \subset K$ and every $\gamma > 0$, there exists an integer $N = N_{\gamma,C,\eta}$ such that if $x \in C$ and $y \in C$ are such that $(x)_{i_0} \geq \eta$ and $(y)_{i_0} \geq \eta$ then $||xW_n|| > 0$ and $||yW_n|| > 0$ for all integers $n \geq N$, and furthermore

$$||(xW_n/||xW_n|| - yW_n/||yW_n||)|| < \gamma.$$

Proof of Lemma 9.1. Let $\epsilon > 0$ and the nonempty compact set $C \subset K$ be given. Since C is compact and K is a complete, separable, metric space we can find a set $\mathcal{N}$ of finitely many points $x_1, x_2, ..., x_M$ such that for any $x \in C$ there exists an element $x_j \in \mathcal{N}$ such that

$$||x_j - x|| < \epsilon/3. \tag{93}$$

From hypotheses 1) and 2), and the fact that

$$\sum_{i\in S} (x_j)_i = 1 < \infty, \quad \forall\, x_j \in \mathcal{N}$$

it follows that we can find an integer $N = N_{\epsilon,c}$ such that if $n \geq N$ then

$$\max\{||x_j(W_n - W)|| : x_j \in \mathcal{N}\} < \epsilon/3. \tag{94}$$

Since $||W_n|| = 1$ for $n = 1, 2....$, we conclude from (93) and (94), that if $x \in C$ and $x_j \in \mathcal{N}$ satisfies $||x_j - x|| < \epsilon/3$ then

$$|(||xW_n|| - ||xW||)| \leq ||x(W_n - W)|| \leq$$

$$||(x - x_j)(W_n - W)|| + ||x_j(W_n - W)|| \leq 2||x - x_j|| + \epsilon/3 < \epsilon,$$

if $n \geq N$, which proves part 1) of the lemma.

Next consider $||xW_n||$ when $x \in K$ is such that $(x)_{i_0} \geq \eta$. We have

$$||xW_n|| \geq (x)_{i_0}||e_{i_0}W_n|| = (x)_{i_0}||e_{i_0}W_n - e_{i_0}W + e_{i_0}W|| \geq$$

$$(x)_{i_0}||e_{i_0}W|| - (x)_{i_0}||e_{i_0}W_n - e_{i_0}W||.$$

But

$$(x)_{i_0}||e_{i_0}W|| \geq \eta u_{i_0}(\sum_{k \in S} v_k) = \eta u_{i_0}$$

and

$$(x)_{i_0}||e_{i_0}W_n - e_{i_0}W|| < (u)_{i_0}\eta/2$$

if n is sufficiently large because of hypothesis 2) of the lemma. Hence

$$||xW_n|| \geq (x)_{i_0}||e_{i_0}W|| - (x)_{i_0}||e_{i_0}W_n - e_{i_0}W|| >$$

$$\eta(u)_{i_0} - (u)_{i_0}\eta/2 = (u)_{i_0}\eta/2$$

if n is sufficiently large and thereby we have proved part 2) of the lemma.

To prove part 3) of the lemma, let $\gamma > 0$ be given, let $\eta,\ 0 < \eta < 1$ be given, and let the nonempty compact set C be given. Let $x \in C$ be such that

$$(x)_{i_0} \geq \eta.$$

Now for any $y \in K$ for which $||yW|| > 0$ we have

$$yW/||yW|| = yu^c v/||yu^c v|| = (y, u^c)v/||(y, u^c)v|| = v/||v|| = v.$$

Since $(x)_{i_0} \geq \eta > 0$ and $u_{i_0} > 0$, we clearly have $||xW|| \geq \eta u_{i_0} > 0$ and hence

$$xW/||xW|| = v.$$

Next let N_0 be so large that $||xW_n|| > (u)_{i_0}\eta/2$ for all $x \in K$ for which $(x)_{i_0} \geq \eta$. That we can find such a number follows from part 2) of this lemma, which we just proved. Now by using inequality (38) of Section 3 we can conclude that

$$||(xW_n/||xW_n||) - v|| = ||(xW_n/||xW_n|| - xW/||xW||)|| \leq$$

$$2||xW_n - xW||/\max\{||xW_n||, ||xW||\} \leq 2||xW_n - xW||(2/(u)_{i_0}\eta) \qquad (95)$$

if $n > N_0$, if $x \in C$ and if $(x)_{i_0} \geq \eta$. From part 1) of this lemma then follows that we can choose an integer $N_1 \geq N_0$ so large that if $n \geq N_1$ then

$$||xW_n - xW|| < \gamma \cdot ((u)_{i_0}\eta/4). \qquad (96)$$

Hence, by combining (95) and (96) we can conclude that

$$||(xW_n/||xW_n|| - xW/||xW||)|| < \gamma$$

if $x \in C,\ (x)_{i_0} \geq \eta$, and $n \geq N_1$ and thereby we have proved part 3) of the lemma.

Part 4) finally follows trivially from part 3) of the lemma and the triangle inequality. $\square$

We now continue the proof of Proposition 9.1. Our aim is to verify Condition B. (See Definition 7.1.) Thus let $\rho > 0$ be given. What we shall prove is that we can find an element $i_0 \in S$, such that if C is a compact set such that

$$\mu(C \cap E_{i_0}((\pi)_{i_0}/2)) \geq (\pi)_{i_0}/3 \tag{97}$$

for all $\mu \in \mathcal{P}(K|\pi)$, then we can find an integer N and an element $\mathbf{w^N} \in \mathcal{W}^N$ such that

$$||e^{i_0}\mathbf{M}(\mathbf{w^N})|| > 0 \tag{98}$$

and

$$||[x\mathbf{M}(\mathbf{w^N})] - [e^{i_0}\mathbf{M}(\mathbf{w^N})]|| < \rho, \qquad \forall x \in E_{i_0}((\pi)_{i_0}/2) \cap C. \tag{99}$$

Since Condition B1 is satisfied there exist a vector $u \in \mathcal{U}$, a vector $v \in K$, a sequence of integers $\{n_1, n_2, ...\}$, and a sequence $\{\mathbf{w_j^{n_j}},\ j = 1, 2, ...\}$ of elements in $\mathcal{W}^{n_j}$ respectively, such that $||\mathbf{M}(\mathbf{w_j^{n_j}})|| > 0,\ j = 1, 2, ...$ and such that if we define $W = u^c v$ then for all $i \in S$

$$\lim_{j\to\infty} ||(e^i\mathbf{M}(\mathbf{w_j^{n_j}})/||\mathbf{M}(\mathbf{w_j^{n_j}})||) - e^iW|| = 0. \tag{100}$$

Let us choose $i_0 \in S$ such that $(u)_{i_0} > 0$ and let C be a compact set such that (97) holds for all $\mu \in \mathcal{P}(K|\pi)$. Since $(u)_{i_0} > 0$ it follows that $||e^{i_0}W|| = (u)_{i_0} > 0$. By (100) then follows that $||\mathbf{M}(\mathbf{w_j^{n_j}})|| > 0,\ j = 1, 2, ...$ if we let the enumeration $n_1, n_2, ...$ start with a sufficiently large n_1. Since obviously

$$||(\mathbf{M}(\mathbf{w_j^{n_j}})/||\mathbf{M}(\mathbf{w_j^{n_j}})||)|| = 1$$

and $C \cup \{e^{i_0}\}$ is a compact set, it follows from assertion 4) of Lemma 9.1 that if j is sufficiently large then

$$||[x\mathbf{M}(\mathbf{w_j^{n_j}})] - [e^{i_0}\mathbf{M}(\mathbf{w_j^{n_j}})]|| < \rho$$

$$\forall x \in E_{i_0}((\pi)_{i_0}/2) \cap C$$

and hence (99) holds which was what we wanted to prove. □

The following condition was introduced in [19] by Kochman and Reeds for the case when S is finite. We call it Condition KR.

Condition KR. *Let S be a finite set, let $P \in PM(S \times S)$, let A be another finite set and let R be a transition probability matrix from S to A. Let $\mathcal{M} = \{(M(a), a \in A\}$ be a partition of P determined by*

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

(As we pointed out in subsection 1.2 it is easily checked that $\mathcal{M} = \{(M(a), a \in A\}$ is a partition of P.)

Define

$$\mathcal{M}^* = \cup_{n=1}^{\infty}\mathcal{M}^n,$$

$$\mathcal{C} = \{C = \alpha M : \alpha > 0, M \in \mathcal{M}^*\}$$

and let the set $\overline{\mathcal{C}}$ be defined as the closure of $\mathcal{C}$ under the usual topology in $\mathtt{R}^{S\times S}$.

If $\overline{\mathcal{C}}$ contains a matrix of **rank 1** *then we say that Condition KR is satisfied.*

Corollary 9.1 *Let S be a finite set , let P be a tr.p.m in $PM(S \times S)$ which is irreducible and aperiodic and let $\mathcal{M}$ be a partition of P. Then Condition KR implies Condition B.*

Proof. Because of Proposition 9.1 it suffices to prove that Condition KR implies Condition B1. In order to do this first note that if W is a finite-dimensional, square, nonnegative $S \times S$ matrix, then W can be written $W = u^c v$ where u is a nonnegative $S-dimensional$ vector satisfying $\max(u)_i = 1$, and where $v \in K$.

Suppose now that Condition KR is satisfied. The we can find a nonnegative matrix W of rank 1, a sequence of integers $n_1, n_2, ...$, a sequence $\mathbf{w_j^{(n_j)}}, \ \ j = 1, 2, ...$ of elements in $\mathcal{W}^{n_j}, j = 1, 2, ...$ and a sequence $\lambda_j, \ j = 1, 2, ...$ of real numbers, such that

$$\lim_{j \to \infty} ||\mathbf{M}(\mathbf{w_j^{(n_j)}})/\lambda_j - W|| = 0. \tag{101}$$

By using the triangle inequality, the norm inequality $|(||A|| - ||B||)| \leq ||A - B||$, and the fact that $||W|| = 1$, it is easily proved that we then also have

$$\lim_{j \to \infty} ||\mathbf{M}(\mathbf{w_j^{(n_j)}})/||\mathbf{M}(\mathbf{w_j^{(n_j)}})|| - W|| = 0 \tag{102}$$

from which Condition B1 follows, since for any two matrices of the same format $||xA - xB|| \leq ||x||(||A - B||)$. $\square$

Next two more notations. Let S be a denumerable set. For an arbitrary matrix $M = \{(M)_{i,j}, \ i \in S, \ j \in S\}$ we define

$$S_1(M) = \{i \in S : (M)_{i,j} > 0, \ \ some \ \ j \in S\} \tag{103}$$

$$S_2(M) = \{j \in S : (M)_{i,j} > 0, \ \ some \ \ i \in S\}. \tag{104}$$

What makes the case when S is finite somewhat easier to handle is that one can use Perron's theorem for finite dimensional matrices with positive elements (see e.g [13], vol II, Theorem 8.1) in order to verify Condition KR. The following condition is adapted to Perron's theorem when S is denumerable.

Condition P. *Let S be a denumerable set, let $P \in PM(S \times S)$ and let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ be a partition of P. If there exist an integer N and an element $\mathbf{w^N} \in \mathcal{W}^n$, such that the matrix $\mathbf{M}(\mathbf{w^N})$ is such that*
1) the set $S_2(\mathbf{M}(\mathbf{w^N}))$ is finite,
2) the set S_0 defined by

$$S_0 = \{i \in S : (\mathbf{M}(\mathbf{w^N}))_{i,i} > 0\} \tag{105}$$

is nonempty,
3)

$$(\mathbf{M}(\mathbf{w^N}))_{i,j} > 0 \ \ if \ \ i \in S_0, \ \ and \ \ j \in S_0$$

4)

$$(\mathbf{M}(\mathbf{w^N}))_{i,j} = 0 \ \ if \ \ i \in S_2(\mathbf{M}(\mathbf{w^N})) \setminus S_0,$$

then we say that **Condition P** *is satisfied.*

Proposition 9.2 *If Condition P is satisfied then Condition B1 is also satisfied.*

Proof. Our arguments will be similar to those given by Kochman and Reeds in [19], section 5, in their proof of Theorem 2 in [19].

It is not difficult to convince oneself that it is no loss of generality to assume that S is a finite or infinite set consisting of consecutive positive integers starting with the number 1. It is also clear that we can assume that the set S_0 defined by (105) is such that

$$S_0 = \{1, 2, ..., d\}$$

where d is a positive integer, and that the set $S_2(\mathbf{M}(\mathbf{w^N}))$ is such that

$$S_2(\mathbf{M}^N(\mathbf{w^N})) = \{1, 2, ..., L\}$$

where L is a positive integer $\geq d$.

Next let us write

$$G = \mathbf{M}(\mathbf{w^N}),$$

$$A = \{(\mathbf{M}(\mathbf{w^N}))_{i,j} : i \in S_0, \ \ j \in S_0 \ \},$$

$$B = \{(\mathbf{M}(\mathbf{w^N}))_{i,j} : i \in S_0, \ \ j \in S_2(\mathbf{M}(\mathbf{w^N})) \setminus S_0 \ \},$$

$$C = \{(\mathbf{M}(\mathbf{w^N}))_{i,j} : i \in S \setminus S_2(\mathbf{M}(\mathbf{w^N})), \ \ j \in S_0 \ \}$$

and

$$D = \{(\mathbf{M}(\mathbf{w^N}))_{i,j} : i \in S \setminus S_2(\mathbf{M}(\mathbf{w^N})), \ \ j \in S_2(\mathbf{M}(\mathbf{w^N})) \setminus S_0 \ \}.$$

We can then write

$$G = \begin{pmatrix} A \; B \; 0 \\ 0 \;\; 0 \;\; 0 \\ C \; D \; 0 \end{pmatrix}$$

where each 0 denotes a zero-matrix of appropriate format, and where the $0-$ *matrix* in the first column has the same number of rows as the number of columns in B. (In case $S_2(\mathbf{M}(\mathbf{w^N})) = S$ the third row and the third column are omitted.)

By induction it is straight forward to prove that

$$G^n = \begin{pmatrix} A \\ 0 \\ C \end{pmatrix} A^{n-2} \begin{pmatrix} \; A \; B \; 0 \; \end{pmatrix}$$

if $n \geq 2$,

Since Condition P is satisfied, it follows that A is a finite-dimensional square matrix with strictly positive elements. Therefore, by Perron's theorem (see e.g [13], vol II, Theorem 8.1), it follows that there exist a number $\lambda > 0$ and a rank-1 matrix $\overline{A}$ with strictly positive elements, such that

$$\lim_{n \to \infty} ||A^n/\lambda^n - \overline{A}|| = 0. \tag{106}$$

where the norm $|| \cdot ||$ is for example defined by (7).

Next let us define

$$W_0 = (1/\lambda^2) \begin{pmatrix} A \\ 0 \\ C \end{pmatrix} \overline{A} \begin{pmatrix} \; A \; B \; 0 \; \end{pmatrix}.$$

Since the elements of the matrices C and B are uniformly bounded and $S_2(W_0)$ is a finite set $||W_0||$ exists and clearly $||W_0|| > 0$. By using (106) it is elementary to prove that for all $i \in S$

$$\lim_{n\to\infty} ||e^i G^n/\lambda^n - e^i W_0|| = 0$$

and also that

$$\lim_{n\to\infty} ||G^n/\lambda^n|| - ||W_0|| = 0. \quad (107)$$

and therefore by defining

$$W = W_0/||W_0||$$

it follows that

$$\lim_{n\to\infty} ||e^i G^n/||G^n|| - e^i W|| = 0 \quad (108)$$

for all $i \in S$ and thereby we have verified Condition B1. $\square$

Next let us introduce the terminology *subrectangular matrix*, a notion introduced in [16] in a more special setting.

Definition 9.1 *Let* $M = \{(M)_i, j : i \in S, j \in S\}$ *be a matrix such that if*

$$(M)_{i_1,j_1} \neq 0 \text{ and also } (M)_{i_2,j_2} \neq 0$$

then also

$$(M)_{i_1,j_2} \neq 0 \text{ and } (M)_{i_2,j_1} \neq 0,$$

then we call M *a* **subrectangular** *matrix.*

Our next result is a generalization of Theorem A of [16] from the case when S is finite to the case when S is denumerable. Recall that $\mathcal{G}'(S)$ is the set of partitions for which each associated tr.pr.m P is irreducible, aperiodic and positively recurrent.

Proposition 9.3 *Let* S *be a denumerable set and let* $\mathcal{M} = \{M(w) : w \in \mathcal{W}\} \in \mathcal{G}'(S)$ *and suppose also that*
1) there exist an integer N_1 *and an element* $\mathbf{a^{N_1}}$ *in* $\mathcal{W}^{N_1}$*, such that* $\mathbf{M}(\mathbf{a^{N_1}})$ *is a non-zero subrectangular matrix,*
2) there exist an integer N_2 *and an element* $\mathbf{b^{N_2}}$ *in* $\mathcal{W}^{N_2}$*, such that* $S_2(\mathbf{M}(\mathbf{b^{N_2}}))$ *is a finite set.*

It then follows that $\mathbf{P}_{\mathcal{M}}$ *is asymptotically stable.*

Proof. From Proposition 9.2, Proposition 9.1 and Theorem 1.1 it follows that it suffices to prove that Condition P is satisfied.

In order to do this we have to find an integer N and an element $\mathbf{w^N} = (w_1, w_2, ..., w_N) \in \mathcal{W}^N$ such that the matrix $\mathbf{M}^N(\mathbf{w^N})$ satisfies the hypotheses of Condition P.

First let $j_0 \in S_2(M(b_1)M(b_2)...M(b_{N_2}))$ and $m \in S_1(\mathbf{M}(\mathbf{b^{N_2}}))$ be such that $(\mathbf{M}(\mathbf{b^{N_2}}))_{m,j_0} > 0$.

Next let $k \in S_2(\mathbf{M}(\mathbf{a^{N_1}}))$. Since P is irreducible there exist an integer N_3 and an element $\mathbf{c^{N_3}} \in \mathcal{W}^{N_3}$ such that $(\mathbf{M}(\mathbf{c^{N_3}}))_{k,m} > 0$.

Next let $i \in S_1(\mathbf{M}(\mathbf{a^{N_1}}))$. Since $\mathbf{M}(\mathbf{a^{N_1}})$ is subrectangular $(\mathbf{M}(\mathbf{a^{N_1}}))_{i,k} > 0$.

Again using the fact that P is irreducible it follows that we can find an integer N_4 and an element $\mathbf{d^{N_4}} \in \mathcal{W}^{N_4}$ such that $(\mathbf{M}(\mathbf{d^{N_4}}))_{j_0,i} > 0$.

Now set $N = N_1 + N_2 + N_3 + N_4$, define the element $\mathbf{w^N} \in \mathcal{W}^N$ by

$$\mathbf{w^N} = (\mathbf{d^{N_4}}, \mathbf{a^{N_1}}, \mathbf{c^{N_3}}, \mathbf{b^{N_2}})$$

and set

$$G = \mathbf{M}(\mathbf{w^N}) = \mathbf{M}(\mathbf{d^{N_4}})\mathbf{M}(\mathbf{a^{N_1}})\mathbf{M}(\mathbf{c^{N_3}})\mathbf{M}(\mathbf{b^{N_2}}).$$

That $S_2(G)$ is a finite set follows from the fact that $S_2(\mathbf{M}(\mathbf{b^{N_2}}))$ is a finite set, and hence hypothesis 1) of Condition P is fulfilled. By the construction of G it is clear that $(G)_{j_0,j_0} > 0$ since $(\mathbf{M}(\mathbf{d^{N_4}}))_{j_0,i} > 0$, $(\mathbf{M}(\mathbf{a^{N_1}}))_{i,k} > 0$, $(\mathbf{M}(\mathbf{c^{N_3}}))_{k,m} > 0$ and $(\mathbf{M}(\mathbf{b^{N_2}}))_{m,j_0} > 0$, and hence hypothesis 2) is fulfilled.

Furthermore if we have three nonnegative square matrices A_1, A_2 and A_3 of the same format such that

$$A_1 = A_2 A_3$$

and either A_2 or A_3 - or both - are subrectangular, it is easily proved that A_1 is also subrectangular. Therefore G is a subrectangular matrix since $\mathbf{M}(\mathbf{a^{N_1}})$ is subrectangular. Defining $S_0 = \{i : (G)_{i,i} > 0\}$ and using the fact that G is subrectangular it follows that

$$(G)_{i,j} > 0 \;\; if\; i \in S_0 \;\; and\; j \in S_0$$

and hence hypothesis 3) is fulfilled.

It remains to show that $(G)_{i,j} = 0$ if $i \in S_2(G) \setminus S_0$, and $j \in S_2(G)$. Let us assume that $(G)_{i,j} > 0$ for $i \in S_2(G) \setminus S_0$ and $j \in S_2(G)$. Since $i \in S_2(G) \setminus S_0$ there exists an element i' such that $G(i', i) > 0$. Since $(G)_{j_0,j_0} > 0$, $(G)_{i,j} > 0$, $(G)_{i',i} > 0$ and G is a subrectangular matrix, it follows that both $(G)_{(i,j_0)} > 0$ and $(G)_{j_0,i} > 0$. Again using the fact that the matrix G is subrectangular it follows that $(G)_{i,i} > 0$. This however contradicts the fact that $i \notin S_0$. Hence hypothesis 4) of Condition P is fulfilled and hence Condition P is satisfied. $\square$

10 A random walk example.

One drawback with Condition P introduced in the previous section is hypothesis 1), which requires that the set $S_2(\mathbf{M}(\mathbf{w^N}))$ is a finite set.

In this section we shall look at an example in which we verify that Condition B holds although the set $S_2(\mathbf{M}(\mathbf{w^N}))$ is infinite for every positive integer N and every $\mathbf{w^N} \in \mathcal{W}^N$.

Let $S = \{..., -2, -1, 0, 1, 2, ...\}$, let $S_{odd} = \{i \in S,\ i \;\; odd\}$ and $S_{even} = \{i \in S,\ i \;\; even\}$. For $i \in S$, let $a_i,\ b_i$ and c_i be positive numbers satisfying

$$a_i + b_i + c_i = 1.$$

Let $P \in PM(S \times S)$ be defined such that

$$(P)_{i,i} = b_0, \forall i \in S,$$

$$(P)_{i,i+1} = (P)_{-i,-(i+1)} = c_i,\ \forall i \geq 0$$

and

$$(P)_{i,i-1} = (P)_{-i,-(i-1)} = a_i,\ \forall i \geq 1.$$

Let $\mathcal{M} = \{M(1), M(2)\}$ be a partition of P such that if i is odd then the $i : th$ column of P is equal to the $i : th$ column of $M(1)$ and if i is even then the $i : th$ column of P is equal to the $i : th$ column of $M(2)$.

Theorem 10.1 *Let the tr.p.m P and the partition $\mathcal{M}$ be defined as above, and suppose also that*

$$\sum_{n=1}^{\infty}\prod_{i=1}^{n} c_{i-1}/a_i < \infty. \tag{109}$$

A) If $b_i = b_0,\ \forall i \in S$ then $\mathbf{P}_{\mathcal{M}}$ is asymptotically stable.
B) If there exists $i_0 \in S$ such that $b_{i_0} > \sup\{b_i : i \in S,\ i \neq i_0\}$ then $\mathbf{P}_{\mathcal{M}}$ is asymptotically stable.

Proof. From the definition of P it is clear that P is aperiodic and irreducible. That P is positively recurrent follows from (109), (see e.g [23], Problem 18, chapter 2). We let π denote the unique probability vector such that $\pi P = \pi$.

We first assume that $b_i = b_0,\ \forall i \in S$ and shall prove the assertion of the theorem by verifying that Condition B holds.

Set $1 - b_0 = \alpha$. Clearly $a_i + c_i = \alpha,\ \ and\ \ b_i = 1-\alpha,\ \ \forall i \in S$. In this case the set K of probability vectors on S is defined by

$$K = \{x = ((x)_i,\ -\infty < i < \infty) : (x)_i \geq 0, \sum_{-\infty}^{\infty}(x)_i = 1\}.$$

To each $x \in K$ we associate two other vectors which we denote by x^1 and x^2, and which we define by

$$(x^1)_i = (x)_i,\ i \in S_{odd},\ \ (x^1)_i = 0,\ i \in S_{even}$$
$$(x^2)_i = (x)_i,\ i \in S_{even},\ \ (x^2)_i = 0,\ i \in S_{odd}. \tag{110}$$

Clearly

$$x = x^1 + x^2.$$

It is easily seen that if $i \in S_{odd}$ then $||e^iM(1)|| = b_0 = 1-\alpha,\ \ ||e^iM(2)|| = \alpha$ and that if $i \in S_{even}$ then $||e^iM(1)|| = \alpha,\ \ ||e^iM(2)|| = b_0 = 1-\alpha$ from which follows that

$$||xM(1)|| = ||x^1||(1-\alpha) + ||x^2||\alpha,\ \ x \in K \tag{111}$$

and

$$||xM(2)|| = ||x^1||\alpha\ +\ ||x^2||(1-\alpha),\ \ x \in K. \tag{112}$$

Next, let $M = M(1)M(2)$. Let us first note that if $x \in K$ is such that $(x)_i = 0,\ i \in S_{odd}$ which thus implies that $x = x^2$, then from (111) and (112) it follows that

$$||xM(1)M(2)|| = ||xM(1)||[xM(1)]M(2)|| = \alpha \cdot \alpha = \alpha^2 \tag{113}$$

since $S_2(M(1)) = S_{odd}$.

We shall next compute $(M)_{i,j}$ when $i \in S_{even}$. For $i \in S_{even}$ and $i \geq 2$ we find that

$$(M)_{i,i} = (M)_{-i,-i} = (M(1))_{i,i+1}(M(2))_{i+1,i} + (M(1))_{i,i-1}(M(2))_{i-1,i} =$$

$$c_i \cdot a_{i+1} + a_i \cdot c_{i-1}.$$

Furthermore, we find that

$$(M)_{i,i+2} = (M)_{-i,-i-2} = (M(1))_{i,i+1}(M(2))_{i+1,i+2} + (M(1))_{i,i-1}(M(2))_{i-1,i} =$$

$$c_i \cdot c_{i+1},$$

and that

$$(M)_{i,i-2} = (M)_{-i,-i+2} = (M(1))_{i,i-1}(M(2))_{i-1,i-2} =$$

$$a_i \cdot a_{i-1}.$$

For $i \in S_{even}, i \neq 0$ and $k \neq i,\ k \neq i+2,\ k \neq i-2$ it is clear that we have $(M)_{i,k} = 0$. For $i = 0$ we find

$$(M)_{0,0} = (\alpha/2)a_1 + (\alpha/2)a_1 = \alpha a_1; \quad (M)_{0,2} = (M)_{0,-2} = (\alpha/2)c_1;$$

and

$$(M)_{0,j} = 0 \ \ otherwise.$$

Since

$$\sum_{k \in S_{even}} (M)_{i,k} = ||e_i M|| = \alpha^2, \ \ \forall i \in S_{even}$$

because of (113), we can conclude that the matrix $(1/\alpha^2)M$ can be considered as a tr.pr.m on S_{even}.

We now define the matrices $A = \{(A)_{i,j}, i \in S, j \in S\}$ and $B = \{(B)_{i,j}, i \in S, j \in S\}$ by

$$(A)_{i,j} = (M)_{i,j},\ i \in S_{even},\ j \in S,$$

$$(A)_{i,j} = 0, \ \ otherwise,$$

and

$$(B)_{i,j} = (M)_{i,j} - (A)_{i,j}, \ \ i \in S,\ j \in S.$$

Evidently $M = A + B$. Since $S_2(A) \subset S_{even}$, $S_1(B) \subset S_{odd}$ and $S_2(B) \subset S_{even}$ it is evident that

$$(A+B)\cdot(A+B) = A^2 + BA = MA$$

and more generally that for $n = 2, 3, ...$

$$M^n = (A+B)^n = A^n + BA^{n-1} = MA^{n-1}. \qquad (114)$$

Since A/α^2 can be considered as a tr.p.m on S_{even}, the Markov chain generated by A/α^2 is an irreducible, aperiodic random walk on S_{even} and since

$$\sum_{n=0}^{\infty}\prod_{i=0}^{n}(M)_{2i,2(i+1)}/(M)_{2(i+1),2i} =$$

$$\sum_{n=0}^{\infty}\prod_{i=0}^{n} c_{2i}c_{2i+1}/(a_{2i+2}a_{2i+1}) = \sum_{n=0}^{\infty}\prod_{i=0}^{2n+1} c_i/(a_{i+1}) < \infty$$

because of (109), it follows that this random walk is positively recurrent and therefore there exists a probability vector q on S such that

$$(q)_i = 0, \ \ if \ \ i \in S_{odd}$$

and

$$qA/(\alpha^2) = q.$$

Next, let $\rho > 0$ be given. In order to verify that Condition B is satisfied we have to show that there exists an element $i_0 \in S$, such that if the compact set C is such that

$$\mu(C \cap E_{i_0}((\pi)_{i_0}/2)) \geq (\pi)_{i_0}/3, \quad \forall\, \mu \in \mathcal{P}(K|\pi), \tag{115}$$

then we can find an integer N and an element $\mathbf{w^N} \in \mathcal{W}^N$ such that

$$i_0 \in S_1(\mathbf{M}(\mathbf{w^N})) \tag{116}$$

and

$$||[x\mathbf{M}(\mathbf{w^N})] - [e^{i_0}\mathbf{M}(\mathbf{w^N}]|| < \rho, \qquad \forall x \in C \cap E_{i_0}((\pi)_{i_0}/2)).$$

We choose $i_0 = 0$, and we let $C_0 \subset K$ be a compact set such that (115) is satisfied with C replaced by C_0. Set

$$C' = \{xM(1)/||xM(1)|| : x \in C_0\}.$$

Since $||xM(1)|| \geq \min\{\alpha, 1-\alpha\}, \quad \forall x \in K$, it follows easily that C' is also a compact set. From the general theory on Markov chains (see e.g [21], chapter 2), it therefore also follows that we can find an integer N_1 such that if $n \geq N_1$ then

$$||y(A/\alpha^2)^n - q|| < \rho/2, \;\; if \;\; y \in C'. \tag{117}$$

We now choose the integer $N = 2(N_1+1)$ and choose the sequence $w_i, i = 1, 2, ..., N$ such that $w_i = 1$, if i is odd and $w_i = 2$, if i is even. Then

$$\mathbf{M}(\mathbf{w}^N) = M(1)M(2)M(1)M(2)...M(1)M(2) = M^{N_1+1},$$

where as before $M = M(1)M(2)$. Clearly $i_0 \in S_1(M^{N_1+1})$ if $i_0 = 0$.

Next, let $x \in C_0$ be chosen arbitrarily, and set $z_1 = [xM]$ and $z_0 = [e^0 M]$. Using the fact that $||yM|| = \alpha^2, \; \forall y \in K'$ such that $y = y^2$, where y^2 is defined by (110), we now find, by using (113), (114), (117) and the scaling property (23)(see Lemma 2.1) that

$$||[xM^{N_1+1}] - [e^0 M^{N_1+1}]|| =$$

$$||[z_1 A^{N_1}] - [z_0 A^{N_1}]|| =$$

$$||(z_1 A^{N_1}/(\alpha^{2(N_1)}) - z_0 A^{N_1}/(\alpha^{2(N_1)}))|| \leq$$

$$||z_1 A^{N_1}/(\alpha^{2(N_1)}) \; - \; q|| + ||z_0 A^{N_1}/(\alpha^{2(N_1)}) \; - \; q|| < \rho$$

and since $x \in C_0$ was chosen arbitrarily, we can conclude that Condition B is satisfied. Since the tr.pr.m P is irreducible, aperiodic, and positively recurrent the conclusion of the theorem follows from Theorem 1.1.

It remains to consider the case when there exists $i_0 \in S$ such that $b_{i_0} > \sup\{b_i : i \in S, \; i \neq i_0\}$. Let us first assume that $i_0 \in S_{odd}$. Define the matrix $D = \{(D)_{i,j} : i \in S, j \in S\}$ by

$$(D)_{i,i} = (M(1))_{i,i}$$

$$(D)_{i,j} = 0, \; if \;\; i \neq j.$$

Clearly $||D|| = b_{i_0}$, and

$$\lim_{n\to\infty} D^n/||D^n|| = e^{i_0 c} e^{i_0}$$

where $e^{i_0{}^c}$ denotes the column vector obtained by taking the transpose of e^{i_0}.

Next let us note that

$$M(1)^n = M(1)D^{n-1} \tag{118}$$

for $n \geq 2$. Set

$$W_1 = e^{i_0{}^c} e^{i_0}$$

and

$$W = M(1)W_1/||M(1)W_1||.$$

Using (118) it is now easy to verify that for all $i \in S$

$$\lim_{n\to\infty} ||e^i M(1)^n/||M(1)^n|| - e^i W|| = 0$$

and thereby we have verified that Condition B1 is satisfied and therefore Condition B is also satisfied, and since the tr.pr.m P is irreducible, aperiodic and positively recurrent, the conclusion again follows from Theorem 1.1.

If instead $b_{i_0} > \sup\{b_i : i \in S,\ i \neq i_0\}$ and $i_0 \in S_{even}$ we can argue in a similar way using the matrix $M(2)$ instead of $M(1)$. □

11 Convex functions and barycenters.

As usual, let S be a denumerable set and let K denote the set of probability vectors on S. As was mentioned in the introduction (see subsection 1.3) the set K is a convex set. For the definitions of the set $C_{convex}[K]$ and the measure ψ_x see subsection 1.8.

Following Choquet (see e.g [6]) we now make the following definition.

Definition 11.1 *(Compare e.g. [6], Definition 26.6.) Let $q \in K$ and let $\mu \in \mathcal{P}(K|q)$ and also $\nu \in \mathcal{P}(K|q)$. We say that μ is* **more diffuse** *than ν and write $\mu \succeq \nu$ and $\nu \preceq \mu$ if*

$$\langle u, \mu \rangle \geq \langle u, \nu \rangle, \ \ \forall u \in C_{convex}[K].$$

Next, for each $q \in K$ we define the subset $\mathcal{P}_d(K|q)$ of $\mathcal{P}(K|q)$ as follows:

Definition 11.2 *The subset $\mathcal{P}_d(K|q)$ of $\mathcal{P}(K|q)$ consists of all measures $\nu \in \mathcal{P}(K|q)$ such that $\nu = \sum_{k=1}^{\infty} \alpha_k \delta_{x_k}$ where $\{x_k,\ k = 1, 2, ...\}$ is a sequence of elements in K and $\{\alpha_k, k = 1, 2, ...\}$ is a sequence of non-negative numbers satisfying $\sum_{k=1}^{\infty} \alpha_k = 1$.*

Proposition 11.1 *Let $q \in K$ and let $\mu \in \mathcal{P}_d(K|q)$. Then μ is always more diffuse than δ_q and less diffuse than ψ_q which symbolically can be expressed as*

$$\delta_q \preceq \mu \preceq \psi_q.$$

Proof. Let $q \in K$ and let $\mu \in \mathcal{P}_d(K|q)$ be such that

$$\mu = \sum_{k=1}^{\infty} \alpha_k \delta_{x_k}$$

where $\{x_k, k = 1, 2, ...\}$ is a set of elements in K and $\{\alpha_k, k = 1, 2, ...\}$ is a sequence of non-negative numbers satisfying $\sum_{k=1}^{\infty} \alpha_k = 1$. Since $\mu \in \mathcal{P}(K|q)$ it follows that $\overline{b}(\mu) = q$ and hence $\sum_{k=1}^{\infty} \alpha_k x_k = q$. Therefore, if $u \in C_{convex}[K]$, it follows that

$$\langle u, \mu \rangle = \sum_{k=1}^{\infty} \alpha_k u(x_k) \geq u(\sum_{k=1}^{\infty} \alpha_k x_k) = u(q) = \langle u, \delta_q \rangle$$

which implies that $\mu \succeq \delta_q$, and it also follows that

$$\langle u, \mu \rangle = \sum_{k=1} \alpha_k u(x_k) \leq \sum_{k=1}^{\infty} \alpha_k (\sum_{i=1}^{\infty} (x_k)_i u(e^i)) =$$

$$\sum_{i=1}^{\infty} u(e^i) \sum_{k=1}^{\infty} \alpha_k (x_k)_i = \sum_{i=1}^{\infty} u(e^i) q_i = \langle u, \psi_q \rangle$$

which implies that $\mu \preceq \psi_q$. □

Next recall that a continuous bounded convex function u belongs to $C'_{convex}[K]$ if the function u can be obtained as

$$u = \sup\{v_n : n \in \mathcal{N}\}$$

where $\mathcal{N}$ is an arbitrary index set, and each v_n is an affine function on K such that

$$v_n(x) = x a_n^c + b_n$$

where thus $a_n \in l^\infty(S)$ and each b_n is a real number.

Proposition 11.2 *Let S be a denumerable set and let $\mathcal{M} \in \mathcal{G}(S)$. Then $u \in C'_{convex}[K] \Rightarrow T_{\mathcal{M}} u \in C_{convex}[K]$.*

Proof. Let $u \in C'_{convex}[K]$. By definition this also implies that $u \in C[K]$. Since the tr.pr.f $\mathbf{P}_{\mathcal{M}}$ induced by $\mathcal{M}$ is Lipschitz continuous, the space $(K, \mathcal{E})$ is a complete, separable, metric space and the set $Lip[K]$ is measure determining, it is easily proved, that $T_{\mathcal{M}} u \in C[K]$ (the Feller property).

Next let $x, y \in K$ be two arbitrary points, let $0 < \lambda < 1$ and set $z = \lambda x + (1-\lambda) y$. To prove that $T_{\mathcal{M}} u \in C_{convex}[K]$ we have to prove that

$$T_{\mathcal{M}} u(z) \leq \lambda T_{\mathcal{M}} u(x) + (1-\lambda) T_{\mathcal{M}} u(y). \tag{119}$$

Recall that the set $\mathcal{W}_{\mathcal{M}}(\xi)$ for $\xi \in K$ is defined by $\mathcal{W}_{\mathcal{M}}(\xi) = \{w \in \mathcal{W} : ||\xi M(w)|| > 0\}$.

Next, let $l_1(S) = \{x = (x)_i, i \in S) : \sum_{i \in S} |(x)_i| < \infty\}$. For $\xi \in l_1(S)$ and $a \in l^\infty(S)$, we write $a(\xi) = \xi a^c$.

Now clearly, if $v : l_1(S) \to \mathtt{R}$ is defined by $v(\xi) = a(\xi) + b$ where $a \in l^\infty(S)$ and $b \in \mathtt{R}$, and $x_1 \in K$ and $x_2 \in K$, then if $\alpha_1 \geq 0$, $\alpha_2 \geq 0$ and $\alpha_1 + \alpha_2 = 1$, it follows that

$$v(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 v(x_1) + \alpha_2 v(x_2) \tag{120}$$

and

$$a(x_1 + x_2) = a(x_1) + a(x_2). \tag{121}$$

Furthermore, if $u \in C'_{convex}[K]$, then u can be written

$$u = \sup\{v_n : n \in \mathcal{N}\}$$

where $\mathcal{N}$ is an arbitrary index set and each v_n is an affine function on K such that

$$v_n(x) = xa_n^c + b_n$$

where thus $a_n \in l^\infty(S)$ and b_n is a real number.

We now find by using (120) and (121) that

$$T_{\mathcal{M}}u(z) = \sum_{w \in \mathcal{W}_{\mathcal{M}}(z)} ||zM(w)|| \sup_{n \in \mathcal{N}} v_n([zM(w)]) =$$

$$\sum_{w \in \mathcal{W}_{\mathcal{M}}(z)} ||zM(w)|| \sup_{n \in \mathcal{N}} ((zM(w)/||zM(w)||)a_n^c + b_n) =$$

$$\sum_{w \in \mathcal{W}_{\mathcal{M}}(z)} \sup_{n \in \mathcal{N}} (zM(w)a_n^c + b_n||zM(w)||) =$$

$$\sum_{w \in \mathcal{W}_{\mathcal{M}}(z)} \sup_{n \in \mathcal{N}} (\lambda xM(w)a_n^c + b_n\lambda||xM(w)|| + (1-\lambda)yM(w)a_n^c + (1-\lambda)b_n||yM(w)||) \leq$$

$$\sum_{w \in \mathcal{W}_{\mathcal{M}}(z)} \sup_{n \in \mathcal{N}} (\lambda xM(w)a_n^c + b_n\lambda||xM(w)||) +$$

$$\sum_{w \in \mathcal{W}_{\mathcal{M}}(z)} \sup_{n \in \mathcal{N}} ((1-\lambda)yM(w)a_n^c + b_n(1-\lambda)||yM(w)||) =$$

$$\sum_{w \in \mathcal{W}_{\mathcal{M}}(x)} \sup_{n \in \mathcal{N}} (\lambda xM(w)a_n^c + b_n\lambda||xM(w)||) +$$

$$\sum_{w \in \mathcal{W}_{\mathcal{M}}(y)} \sup_{n \in \mathcal{N}} ((1-\lambda)yM(w)a_n^c + b_n(1-\lambda)||yM(w)||) =$$

$$\lambda T_{\mathcal{M}}u(x) + (1-\lambda)T_{\mathcal{M}}u(y)$$

and hence (119) holds, which was what we wanted to prove. □

Proposition 11.3 *Let S be a denumerable set, let $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{G}(S)$ and let P_1, P_2 be the tr.pr.ms associated to $\mathcal{M}_1$ and $\mathcal{M}_2$ respectively. Let $q \in K$ and suppose that*

$$q = qP_1 = qP_2.$$

Then, if $u \in C'_{convex}[K]$

$$\langle u, \delta_q \rangle \leq \langle u, \breve{P}_{\mathcal{M}_2}\delta_q \rangle \leq \langle u, \breve{P}_{\mathcal{M}_2}\breve{P}_{\mathcal{M}_1}\delta_q \rangle \leq \langle u, \breve{P}_{\mathcal{M}_2}\breve{P}_{\mathcal{M}_1}\psi_q \rangle \leq \langle u, \breve{P}_{\mathcal{M}_2}\psi_q \rangle \leq \langle u, \psi_q \rangle.$$

Proof. From Theorem 5.2 follows that $\breve{P}_{\mathcal{M}_2}\delta_q \in \mathcal{P}(K|q)$. Since also $\breve{P}_{\mathcal{M}_2}\delta_q \in \mathcal{P}_d(K|q)$ the first inequality now follows from Proposition 11.1.

To prove the second inequality, we first use (11) to obtain

$$\langle u, \breve{P}_{\mathcal{M}_2}\breve{P}_{\mathcal{M}_1}\delta_q \rangle = \langle T_{\mathcal{M}_2}u, \breve{P}_{\mathcal{M}_1}\delta_q \rangle$$

and then, using the fact that $T_{\mathcal{M}_2}u \in C_{convex}[K]$ if $u \in C'_{convex}[K]$ and the fact that $\breve{P}_{\mathcal{M}_1}\delta_q \in \mathcal{P}_d(K|q)$, it follows from Proposition 11.1 that

$$\langle T_{\mathcal{M}_2}u, \breve{P}_{\mathcal{M}_1}\delta_q \rangle \geq \langle T_{\mathcal{M}_2}u, \delta_q \rangle = \langle u, \breve{P}_{\mathcal{M}_2}\delta_q \rangle.$$

Hence

$$\langle u, \breve{P}_{\mathcal{M}_2}\breve{P}_{\mathcal{M}_1}\delta_q\rangle \geq \langle u, \breve{P}_{\mathcal{M}_2}\delta_q\rangle$$

and thereby the second inequality is proved.

To prove the third inequality we first use a) of Theorem 2.1 to obtain

$$\langle u, \breve{P}_{\mathcal{M}_2}\breve{P}_{\mathcal{M}_1}\delta_q\rangle = \langle u, \breve{P}_{\mathcal{M}_1\mathcal{M}_2}\delta_q\rangle.$$

Since also $\breve{P}_{\mathcal{M}_1\mathcal{M}_2} \in \mathcal{P}_d(K|q)$, it follows from Proposition 11.2 that if $u \in C'_{convex}[K]$, then $T_{\mathcal{M}_1\mathcal{M}_2}u \in C_{convex}[K]$ and hence by Proposition 11.1 and (11), it follows that

$$\langle u, \breve{P}_{\mathcal{M}_2}\breve{P}_{\mathcal{M}_1}\delta_q\rangle = \langle T_{\mathcal{M}_1\mathcal{M}_2}u, \delta_q\rangle \leq \langle T_{\mathcal{M}_1\mathcal{M}_2}u, \psi_q\rangle = \langle u, \breve{P}_{\mathcal{M}_1\mathcal{M}_2}\psi_q\rangle.$$

From a) of Theorem 2.1 and the preceding inequality it now follows that

$$\langle u, \breve{P}_{\mathcal{M}_2}\breve{P}_{\mathcal{M}_1}\delta_q\rangle \leq \langle u, \breve{P}_{\mathcal{M}_1\mathcal{M}_2}\psi_q\rangle = \langle u, \breve{P}_{\mathcal{M}_2}\breve{P}_{\mathcal{M}_1}\psi_q\rangle$$

and thereby the third inequality is proved.

Next, using (11), Proposition 11.2 and Proposition 11.1 again, we find that

$$\langle u, \breve{P}_{\mathcal{M}_2}\breve{P}_{\mathcal{M}_1}\psi_q\rangle =$$

$$\langle T_{\mathcal{M}_2}u, \breve{P}_{\mathcal{M}_1}\psi_q\rangle \leq \langle T_{\mathcal{M}_2}u, \psi_q\rangle = \langle u, \breve{P}_{\mathcal{M}_2}\psi_q\rangle$$

which proves the fourth inequality, and the last inequality follows again from Proposition 11.1 and the fact that $\breve{P}_{\mathcal{M}_2}\psi_q \in \mathcal{P}_d(K|q)$. □

As an immediate corollary we obtain

Proposition 11.4 *Let S be a denumerable space, let $P \in PM(S\times S)$, let $q \in K$ satisfy $q = qP$, let $\mathcal{M}$ be a partition of P and let $u \in C'_{convex}[K]$. Then, for $n = 1, 2, ...,$*

$$\langle u, \breve{P}^n_{\mathcal{M}}\delta_q\rangle \leq \langle u, \breve{P}^{n+1}_{\mathcal{M}}\delta_q\rangle \leq \langle u, \breve{P}^{n+1}_{\mathcal{M}}\psi_q\rangle \leq \langle u, \breve{P}^n_{\mathcal{M}}\psi_q\rangle.$$

Proof. Follows from Proposition 11.3 and Corollary 2.1. □.

12 A martingale.

As usual, let S be a denumerable set and let K denote the set of probability vectors on S. Let $P \in PM(S\times S)$ and suppose that $\pi \in K$ is such that $\pi = \pi P$. Let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ be a partition of P, and let $\{Z_n(\pi), n = 0, 1, 2, ...\}$ denote the sequence of stochastic variables with values in $(K, \mathcal{E})$ generated by $\mathbf{P}_{\mathcal{M}}$ and the initial distribution δ_π.

Next let $\mathcal{A}$ denote the discrete $\sigma - algebra$ on $\mathcal{W}$. For $n = 1, 2, ...$ let as before $\mathcal{W}^n = \prod_{i=1}^n \mathcal{W}_i$ where $\mathcal{W}_i = \mathcal{W}, i = 1, 2, ..., n$ and let $\mathcal{A}^n$ denote the discrete $\sigma - algebra$ on $\mathcal{W}^n$.

Now define α_1 on $(\mathcal{W}, \mathcal{A})$ by

$$\alpha_1(B) = \sum_{w\in B} ||\pi M(w)||, \quad B \in \mathcal{A},$$

and, for $n = 2, 3, ...,$ define α_n on $(\mathcal{W}^n, \mathcal{A}^n)$ by

$$\alpha_n(B^n) = \sum_{\mathbf{w^n}\in B^n} ||\pi M(w_n)M(w_{n-1})...M(w_1)||, \quad B^n \in \mathcal{A}^n.$$

Lemma 12.1 *For $n = 1, 2, ...$ the set function α_n is a probability measure on $(\mathcal{W}^n, \mathcal{A}^n)$. Furthermore α_{n+1} is an extension of α_n.*

Proof. First note that

$$\sum_{(w_1,w_2,...,w_{n+1})\in\mathcal{W}^{n+1}} ||\pi M(w_{n+1})M(w_n)M(w_{n-1})...M(w_1)|| =$$

$$\sum_{(w_1,w_2,...w_n)\in\mathcal{W}^n} ||\pi P M(w_n)M(w_{n-1})...M(w_1)|| =$$

$$\sum_{(w_1,w_2,...,w_n)\in\mathcal{W}^{n-1}} ||\pi M(w_n)...M(w_1)||,$$

since $\sum_{w\in\mathcal{W}} M(w) = P$ and $\pi P = \pi$. By induction follows that

$$\sum_{(w_1,w_2,...,w_{n+1})\in\mathcal{W}^{n+1}} ||\pi M(w_{n+1})M(w_n)...M(w_1)|| =$$

$$\sum_{w_1\in\mathcal{W}} ||\pi M(w_1)|| = ||\pi P|| = ||\pi|| = 1$$

and since also $||\pi M(w_n)M(w_{n-1})...M(w_1)|| \geq 0$ it follows that each α_n is a probability measure on $(\mathcal{W}^n, \mathcal{A}^n)$ respectively. That α_{n+1} is an extension of α_n for each n follows from the fact that

$$\sum_{w_{n+1}\in\mathcal{W}} ||\pi M(w_{n+1})M(w_n)M(w_{n-1})...M(w_1)|| =$$

$$||\pi P M(w_n)M(w_{n-1})...M(w_1)|| =$$

$$||\pi M(w_n)...M(w_1)||. \ \square$$

Next, let $\mathcal{W}^\infty = \prod_{i=1}^\infty \mathcal{W}_i$ where as above $\mathcal{W}_i = \mathcal{W}$, and let $\mathcal{A}^\infty$ be the least $\sigma - algebra$ containing all sets of the form $B_n \times \prod_{i=n+1}^\infty \mathcal{W}_i$ where $n = 1, 2, ...$ and $B_n \in \mathcal{A}^n$. Let $\overline{\alpha}$ be the extension of the sequence $\{\alpha_n, n = 1, 2, ...\}$ of probability measures on $(\mathcal{W}^n, \mathcal{A}^n)$, $n = 1, 2, ...$, to the space $(\mathcal{W}^\infty, \mathcal{A}^\infty)$. We denote an element in $\mathcal{W}^\infty$ by $\overline{w}$ and write

$$\overline{w} = (w_n,\ n = 1, 2, ...)$$

For $n = 1, 2, ...$ define $Y_n : \mathcal{W}^\infty \to \mathcal{W}$ by $Y_n'(\overline{w}) = w_n$, and define $Z_n' : \mathcal{W}^\infty \to K$ by

$$Z_n'(\overline{w}) = [\pi M(w_n)M(w_{n-1})...M(w_1)]$$

if $||\pi M(w_n)M(w_{n-1})...M(w_1)|| > 0$ and $Z_n'(\overline{w}) = \pi$ *otherwise.*

That Y_n', $n = 1, 2, ...$ are stochastic variables on the probability space $(\overline{\alpha}, \mathcal{W}^\infty, \mathcal{A}^\infty)$ with values in $\mathcal{W}$ is obvious, and that also Z_n', $n = 1, 2, ...$ are stochastic variables on $(\overline{\alpha}, \mathcal{W}^\infty, \mathcal{A}^\infty)$ with values in K is not too difficult to prove if one uses arguments similar to those that are needed in order to prove that the tr.pr.f $\mathbf{P}_\mathcal{M}(\cdot, \cdot)$, induced by a partition, is measurable, in the first variable. (See Proposition 14.1.)

From the definition of the measures α_n, $n = 1, 2...$ and the definition of Y_n', $n = 1, 2, ..$ it is obvious that the following equality is true, which we state without further proof.

Lemma 12.2 *For n=1,2,...*

$$Pr[Y'_1 = w_n, Y'_2 = w_{n-1}, ..., Y'_n = w_1] = ||\pi M(w_1)M(w_2)...M(w_n)||.$$

From Lemma 12.2 the following equality follows.

Lemma 12.3 *For* $n = 1, 2, ...,$

$$Pr[Z'_n \in B] = Pr[Z_n(\pi) \in B], \ \ \forall B \in \mathcal{E}.$$

Proof. Set

$$\mathcal{W}^{*n}(B) = \{(w_1, w_2, ..., w_n) \in \mathcal{W}^n : [\pi M(w_n)M(w_{n-1})...M(w_1)] \in B\}.$$

Then, we first note that

$$Pr[Z'_n \in B] = \sum_{(w_1,w_2,...,w_n)\in\mathcal{W}^{*n}(B)} \alpha_n(w_1, w_2, ..., w_n) =$$

$$\sum_{(w_1,w_2,...,w_n)\in\mathcal{W}^{*n}(B)} ||\pi M(w_n)M(w_{n-1})...M(w_1)||.$$

Next considering $Z_n(\pi)$ it is clear from the definition of $Z_n(\pi)$ that

$$Pr[Z_n(\pi) \in B] = \sum_{(w_1,w_2...,w_n)\in\mathcal{W}_{\mathcal{M}^n}(\pi,B)} ||\pi M(w_1)M(w_2)...M(w_n)||$$

where $\mathcal{W}_{\mathcal{M}^n}(\pi, B)$ is defined by

$$\mathcal{W}_{\mathcal{M}^n}(\pi, B) = \{(w_1, w_2, ..., w_n) \in \mathcal{W}^n : ||\pi M(w_1)M(w_2)...M(w_n)|| > 0$$

$$and \ \ [\pi M(w_1)M(w_2)...M(w_n)] \in B \ \ \}, \quad B \in \mathcal{E}.$$

But if $(w_1, w_2, ...w_n) \in \mathcal{W}_{\mathcal{M}^n}(\pi, B)$ then clearly $(w_n, w_{n-1}, ..., w_1) \in \mathcal{W}^{*n}(B)$ from which follows that

$$Pr[Z_n(\pi) \in B] = \sum_{(w_n,w_{n-1},...,w_1)\in\mathcal{W}^{*n}(B)} ||\pi M(w_n)M(w_{n-1})...M(w_1)|| = Pr[Z'_n \in B]. \ \square$$

The following martingale relation also follows easily.

Proposition 12.1 *For* $n = 1, 2, 3, ...$ *and* $i \in S$

$$E[(Z'_{n+1}(\overline{w}))_i \mid \mathcal{A}^n] = (Z'_n(\overline{w}))_i$$

Proof. First note that the set

$$\mathcal{W}^{*n,+} = \{(w_1, w_2, ...w_n) : ||\pi M(w_n)M(w_{n-1})...M(w_1)|| > 0\}$$

is such that $\alpha_n(\mathcal{W}^{*n,+}) = 1$. Next note that if

$$(w_1, w_2, ..., w_n) \in \mathcal{W}^{*n,+}$$

then

$$Pr[Y'_{n+1} = w \mid Y'_1 = w_1, Y'_2 = w_2, ...Y'_n = w_n] =$$

$$||\pi M(w)M(w_n)M(w_{n-1})...M(w_1)||/||\pi M(w_n)M(w_{n-1})...M(w_1)||.$$

Hence if $(w_1, w_2, ...w_n) \in \mathcal{W}^{*n,+}$

$$E[(Z'_{n+1}(\overline{w}))_i \mid Y'_1 = w_1, Y'_2 = w_2, ..., Y'_n = w_n] =$$

$$\sum_{w_{n+1}\in\mathcal{W}} (\pi M(w_{n+1})M(w_n)M(w_{n-1})...M(w_1))_i/||\pi M(w_{n+1})qM(w_n)M(w_{n-1})...M(w_1)||\cdot$$

$$||\pi M(w_{n+1})qM(w_n)M(w_{n-1})...M(w_1)||/||\pi M(w_n)M(w_{n-1})...M(w_1)|| =$$

$$\sum_{w_{n+1}\in\mathcal{W}} (\pi M(w_{n+1})M(w_n)M(w_{n-1})...M(w_1))_i/||\pi M(w_n)M(w_{n-1})...M(w_1)|| =$$

$$(\pi \cdot P \cdot M(w_n)M(w_{n-1})...M(w_1))_i/||\pi M(w_n)M(w_{n-1})...M(w_1)||) =$$

$$(\pi M(w_n)M(w_{n-1})...M(w_1))_i/||\pi M(w_n)M(w_{n-1})...M(w_1)||) = (Z'_n(\overline{w}))_i. \ \Box$$

Corollary 12.1 *The sequence $\{Z'_n,\ n = 1, 2, ...\}$ converges almost surely.*

Proof. Since, for each $i \in S$, $\{(Z'_n)_i, n = 1, 2, ...\}$ is a sequence of uniformly bounded, stochastic variables, it follows from Proposition 12.1 and the martingale convergence theorem (see e.g. [12]), that $\{(Z'_n)_i,\ n = 1, 2, ...\}$ converges almost surely for each $i \in S$. Since S is a denumerable set, it then follows that also $\{Z'_n,\ n = 1, 2, ...\}$ converges almost surely. □

13 Blackwell's entropy formula.

Let $h : [0, 1] \to [0, 1/(e \cdot \ln(2))]$ be defined by

$$h(t) = -t\ln(t)/\ln(2), \ \ if \ \ 0 < t \leq 1 \qquad and \ \ h(0) = 0.$$

In this section we shall define the *entropy* and the *entropy rate* for a partition in $\mathcal{G}^{',\mathcal{Z}}(S)$, (see section 2 for the definitions of $\mathcal{G}^{',\mathcal{Z}}(S)$), and shall prove that the formula obtained by Blackwell for the process $\{Y_n, ;\ \infty < n < \infty\}$ described in the introduction (see subsection 1.7) also holds under a more general situation.

Definition 13.1 *Let S be a denumerable set, let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\} \in \mathcal{G}^{',\mathcal{Z}}(S)$, let $P \in PM(S \times S)$ be the associated tr.pr.m and let $\pi \in K^+$ be such that $\pi P = \pi$.*

We define the **entropy**, *$H(\mathcal{M})$, of the partition $\mathcal{M} \in \mathcal{G}^{',\mathcal{Z}}(S)$, by*

$$H(\mathcal{M}) = \sum_{w\in\mathcal{W}} h(||\pi M(w)||),$$

and we define the **entropy rate** *$H_R(\mathcal{M})$ of the partition $\mathcal{M} \in \mathcal{G}^{',\mathcal{Z}}(S))$ by*

$$H_R(\mathcal{M}) = \lim_{n\to\infty} (H(\mathcal{M}^{n+1}) - H(\mathcal{M}^n)).$$

That $H(\mathcal{M})$ is well-defined is clear since we assume that $\mathcal{M} \in \mathcal{G}^{',\mathcal{Z}}(S)$. That also $H_R(\mathcal{M})$ is well-defined follows from the following theorem. Recall that δ_q is defined by $\delta_q(\{q\}) = 1$ and that ψ_q is defined by $\psi_q(\{e^i\}) = (q)_i,\ i \in S$. The following theorem is similar to Theorem 4.4.1 of [7].

Theorem 13.1 *Let S be a denumerable set, let $\pi \in K^+$, and let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\} \in \mathcal{G}_\pi(S) \cap \mathcal{G}^{',\mathcal{Z}}(S)$.*

Then

a) for $n = 1, 2, ...$

$$\sum_{w\in\mathcal{W}} \int_K h(||yM(w)||)\breve{P}^n_{\mathcal{M}}\psi_\pi(dy) \le \sum_{w\in\mathcal{W}} \int_K h(||yM(w)||)\breve{P}^{n+1}_{\mathcal{W}}\psi_\pi(dy) \le$$

$$\sum_{w\in\mathcal{W}} \int_K h(||yM(w)||)\breve{P}^{n+1}_{\mathcal{M}}\delta_\pi(dy) \le \sum_{w\in\mathcal{M}} \int_K h(||yM(w)||)\breve{P}^n_{\mathcal{M}}\delta_\pi(dy).$$

b)

$$H_R(\mathcal{M}) = \lim_{n\to\infty} \sum_{w\in\mathcal{W}} \int_K h(||yM(w)||)\breve{P}_{\mathcal{M}^n}\delta_\pi(dy) \qquad (122)$$

Proof. For each $w \in \mathcal{W}$ define $g_w : K \to \mathtt{R}$ by

$$g_w(x) = h(||xM(w)||)$$

It is easily seen that if we define $f : [0,1] \to \mathtt{R}$ by

$$f(t) = \inf\{h(\kappa) + h'(\kappa)(t-\kappa) : 0 < \kappa \le 1\}$$

then

$$f(t) = h(t).$$

Hence

$$g_w(x) = f(||xM(w)||) = \inf\{h(\kappa) + h'(\kappa)(||xM(w)|| - \kappa) : 0 < \kappa \le 1\}.$$

Since the mapping $\rho : K \to [0,1]$ defined by $\rho(x) = ||xM(w)||$ is a linear function on K it clearly follows that for κ such that $0 < \kappa \le 1$ the mapping from K to $\mathtt{R}$ defined by

$$h(\kappa) + h'(\kappa)(||xM(w)|| - \kappa)$$

is an affine mapping. Therefore the function $f_w(x)$ defined by $f_w(x) = -g_w(x)$ belongs to $C'_{convex}[K]$ for each $w \in \mathcal{W}$. We can thus apply Corollary 11.1 to f_w, and find that for each $w \in \mathcal{W}$ and $n = 1, 2, ...$

$$\langle f_w, \breve{P}^n_{\mathcal{M}}\delta_\pi\rangle \le \langle f_w, \breve{P}^{n+1}_{\mathcal{M}}\delta_\pi\rangle \le \langle f_w, \breve{P}^{n+1}_{\mathcal{M}}\psi_\pi\rangle \le \langle f_w, \breve{P}^n_{\mathcal{M}}\psi_\pi\rangle,$$

and by multiplying by (-1) and adding over w we obtain the conclusion of part a) of the theorem.

Next by using the inequalities of part a) we conclude that

$$\{\sum_{w\in\mathcal{M}} \int_K h(||yM(w)||)\breve{P}_{\mathcal{M}^n}\delta_\pi(dy), \quad n = 1, 2, ...\}$$

is a decreasing sequence bounded from below and hence the right hand side of (122) exists. From part b) of Theorem 2.1 follows that $H(\mathcal{M}^n)$ exists for $n = 1, 2, ...$ and that

$$H(\mathcal{M}^{n+1}) - H(\mathcal{M}^n) = \sum_{w\in\mathcal{M}} \int_K h(||yM(w)||)\breve{P}_{\mathcal{M}^n}\delta_\pi(dy).$$

Hence

$$\lim_{n\to\infty}(H(\mathcal{M}^{n+1})-H(\mathcal{M}^n)) = H_R(\mathcal{M}) = \lim_{n\to\infty}\sum_{w\in\mathcal{W}}\int_K h(||yM(w)||)\breve{P}_{\mathcal{M}^n}\delta_\pi(dy),$$

and thereby part b) of the theorem is proved. □

The entropy formula in the next theorem originates from the paper [4] by Blackwell from 1957.

Theorem 13.2 *Let S be a denumerable set, let $\pi \in K^+$, and let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\} \in \mathcal{G}_\pi(S) \cap \mathcal{G}^{',\mathcal{Z}}(S)$. Assume also that the tr.pr.f $\mathbf{P}_\mathcal{M}$ is asymptotically stable, and let μ denote the unique stationary measure of $\mathbf{P}_\mathcal{M}$.*

Then, for each $x \in K$

$$H_R(\mathcal{M}) = \lim_{n\to\infty}\sum_{w\in\mathcal{W}} h(||yM(w)||)\breve{P}_{\mathcal{M}^n}\delta_x(dy) =$$

$$\sum_{w\in\mathcal{M}}\int_K h(||yM(w)||)\mu(dy).$$

Proof. Since 1) $||yM(w)||$ is a continuous function of y for each $w \in \mathcal{W}$, 2) the function h is a continuous function and 3) $\mathbf{P}_\mathcal{M}$ is asymptotically stable, it follows for each $x \in K$ - and in particular for $x = \pi$ - that

$$\lim_{n\to\infty}\int_K h(||yM(w)||)\mathbf{P}^n_\mathcal{M}(x,dy) = \lim_{n\to\infty}\int_K h(||yM(w)||)\breve{P}_{\mathcal{M}^n}\delta_x(dy) =$$

$$\int_K h(||yM(w)||)\mu(dy) =$$

for each $w \in \mathcal{W}$. Therefore, since all integrands are positive, we can conclude by using b) of Theorem 13.1 that for all $x \in K$

$$H_R(\mathcal{M}) = \lim_{n\to\infty}\sum_{w\in\mathcal{W}}\int_K h(||yM(w)||)\breve{P}_{\mathcal{M}^n}\delta_\pi(dy) =$$

$$\lim_{n\to\infty}\sum_{w\in\mathcal{W}}\int_K h(||yM(w)||)\breve{P}_{\mathcal{M}^n}\delta_x(dy) = \sum_{w\in\mathcal{M}}\int_K h(||yM(w)||)\mu(dy). \quad \square$$

14 $\mathbf{P}_\mathcal{M}$ is a transition probability function.

Theorem 14.1 *Let S be a denumerable set, and let $\mathcal{M} = \{M(w) : w \in \mathcal{W}\}$ be a partition of $P \in PM(S \times S)$. Define the function $\mathbf{P}_\mathcal{M} : K \times \mathcal{E} \to [0,1]$ by*

$$\mathbf{P}_\mathcal{M}(x,B) = \sum_{w\in\mathcal{W}_\mathcal{M}(x,B)} ||xM(w)||, \;\; x \in K, \;\; B \in \mathcal{E}$$

where

$$\mathcal{W}_\mathcal{M}(x,B) = \{w \in \mathcal{W} : ||xM(w)|| > 0, xM(w)/||xM(w)|| \in B\}.$$

Then, for every $x \in K$, $\mathbf{P}_{\mathcal{M}}(x, \cdot)$ is a probability measure on $(K, \mathcal{E})$, and also, for every $B \in \mathcal{E}$,

$$\mathbf{P}_{\mathcal{M}}(\cdot, B)$$

is a measurable function. Hence $\mathbf{P}_{\mathcal{M}} : K \times \mathcal{E} \to [0,1]$ is a transition probability function according to the usual definition in probability theory. (See e.g. [25], Definition 1.8.)

Proof. To prove that $\mathbf{P}_{\mathcal{M}}(x, \cdot)$ is a probability measure for each $x \in K$ let us first note that $\mathbf{P}_{\mathcal{M}}(x, K) = \sum_w ||xM(w)|| = ||xP|| = 1$. Furthermore if $B_1, B_2 \in \mathcal{E}$ are such that $B_1 \cap B_2 = \emptyset$ then clearly also $\mathcal{W}_{\mathcal{M}}(x, B_1) \cap \mathcal{W}_{\mathcal{M}}(x, B_1) = \emptyset$ and more generally if $\{B_n,\ n = 1, 2, ...\}$ is a sequence of disjoint sets then clearly $\{\mathcal{W}_{\mathcal{M}}(x, B_n),\ n = 1, 2, ...\}$ is also a sequence of disjoint sets, from which follows that if $\{B_n,\ n = 1, 2, ...\}$ is a sequence of disjoint sets then

$$\mathbf{P}_{\mathcal{M}}(x, \cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mathbf{P}_{\mathcal{M}}(x, B_n).$$

Hence $\mathbf{P}_{\mathcal{M}}(x, \cdot)$ is a probability measure for each $x \in K$.

It thus remains to prove that if $B \in \mathcal{E}$ then $\mathbf{P}_{\mathcal{M}}(\cdot, B)$ is $\mathcal{E} - measurable$. We first prove the following proposition.

Proposition 14.1 *Let S be a denumerable set and let M denote a square, non-negative $S \times S$ matrix, such that $0 < \sup\{||xM|| : x \in K\} \leq 1$. Let $K^1 = \{x \in K : ||xM|| > 0\}$. For an arbitrary $F \subset K$ we define $K^1_F = \{x \in K^1 : xM/||xM|| \in F\}$, and $u_F : K \to [0,1]$ by*

$$u_F(x) = ||xM|| \;\; if \;\; x \in K^1_F$$

and

$$u_F(x) = 0 \;\; if \;\; x \notin K^1_F.$$

Further let

$$\mathcal{F} = \{F \subset K : u_F \;\; is \;\; \mathcal{E} - measurable\}.$$

Then:
a) If F is an open set then $F \in \mathcal{F}$.
b) If $F_0 = F_1 \cap F_2$ where $F_1 \in \mathcal{F}$ and $F_2 \in \mathcal{F}$ then $F_0 \in \mathcal{F}$.
c) If $F_0 = \cup_{\{i=1,2,...\}} F_i$ and $F_i \in \mathcal{F}, i = 1, 2, ...$ then $F_0 \in \mathcal{F}$.
d) If $F_0 = K \setminus F_1$ and $F_1 \in \mathcal{F}$ then $F_0 \in \mathcal{F}$.
e) If $B \in \mathcal{E}$ then u_B is $\mathcal{E} - measurable$.

Proof of Proposition 14.1. For an arbitrary $F \subset K$ define $I_{K^1_F} : K \to \{0,1\}$, by

$$I_{K^1_F}(x) = 1 \;\; if \;\; x \in K^1 \;\; and \;\; xM/||xM|| \in F$$

$$I_{K^1_F}(x) = 0 \;\; otherwise.$$

From the definition of u_F it is clear that we can express u_F by

$$u_F(x) = ||xM|| I_{K^1_F}(x).$$

Next suppose $F_0 = F_1 \cap F_2$. Then $I_{K^1_{F_0}}(x) = 1$ if and only if $x \in K^1$ *and* $xM/||xM|| \in F_1 \cap F_2$ from which it follows that

$$I_{K^1_{F_0}}(x) = \min\{I_{K^1_{F_1}}(x), I_{K^1_{F_2}}(x)\}.$$

Hence if $F_i \in \mathcal{F}, i = 1, 2$ and $F_0 = F_1 \cap F_2$ then we can rewrite $u_{F_0}(x)$ as follows:

$$u_{F_0}(x) = ||xM|| I_{K^1_{F_0}}(x) = ||xM|| \min\{I_{K^1_{F_1}}(x), I_{K^1_{F_2}}(x)\} =$$

$$\min\{||xM|| I_{K^1_{F_1}}(x), ||xM|| I_{K^1_{F_2}}(x)\} =$$

$$\min\{u_{F_1}(x), u_{F_2}(x)\}.$$

Since $F_i \in \mathcal{F}, i = 1, 2$ it follows from the definition of $\mathcal{F}$ that u_{F_1} and u_{F_2} are $\mathcal{E} - measurable$, and since the minimum of two measurable functions is also measurable it follows that u_{F_0} is also $\mathcal{E} - measurable$ and hence $F_0 \in \mathcal{F}$. Thereby assertion b) of the proposition is proved.

To prove assertion c) we assume that $F_0 = \cup_{\{i=1,2,...\}} F_i$ and that $F_i \in \mathcal{F}, i = 1, 2, ...$. Then we can write

$$u_{F_0}(x) = ||xM|| I_{K^1_{F_0}}(x) = ||xM|| I_{K^1_{\cup_{\{i=1,2,...\}} F_i}}(x) =$$

$$||xM|| \sup_{\{i=1,2,...\}} I_{K^1_{F_i}}(x)\} =$$

$$\sup\{||xM|| I_{K^1_{F_i}}(x) : i = 1, 2, ...\} =$$

$$\sup\{u_{F_i}(x) : i = 1, 2, ...\}$$

which implies that u_{F_0} is $\mathcal{E} - measurable$ since if $\{f_1, f_2, ...\}$ is a denumerable set of measurable functions then $\sup\{f_1, f_2, ...\}$ is also measurable.

In order to prove assertion d) we can write

$$u_{F_0}(x) = ||xM|| I_{K^1_{F_0}}(x) = ||xM|| I_{K^1_{K \setminus F_1}}(x) =$$

$$||xM||(I_{K^1}(x) - I_{K^1_{F_1}}(x)) = ||xM|| - u_{F_1}(x)$$

and since both $||xM||$ and $u_{F_1}(x)$ are $\mathcal{E} - measurable$ as functions of x it follows that $u_{F_0}(x)$ is $\mathcal{E} - measurable$, and hence $F_0 \in \mathcal{F}$.

From assertions b), c) and d) it follows that $\mathcal{F}$ is a $\sigma - algebra$ and since $\mathcal{E}$ is the least $\sigma - algebra$ containing the open sets, we can conclude that part e) follows when we also have proved assertion a) of the proposition.

The proof of proposition a) is fairly standard but somewhat tedious. Thus let us assume that $F \subset K$ is an open set. For $-\infty < b < \infty$ let

$$K_b = \{x \in K : u_F(x) < b\}.$$

Furthermore define a_0 by

$$a_0 = \sup\{||xM|| : x \in K\}.$$

To prove that u_F is measurable it suffices to prove that $K_b \in \mathcal{E}$ for all $b \in \mathtt{R}$. Since $0 \leq u_F(x) \leq a_0$, it is clear that $K_b = \emptyset$ if $b \leq 0$ and $K_b = K$ if $b > a_0$, and since both $\emptyset \in \mathcal{E}$ and $K \in \mathcal{E}$, it remains to show that $K_b \in \mathcal{E}$ for $0 < b \leq a_0$.

We first consider the set $K'_0 = \{x \in K : u_F(x) = 0\}$. Now from the definition of u_F we find that the set K'_0 can be written

$$K'_0 = (K \setminus K^1) \cup (K^1 \setminus K^1_F).$$

Since the function $x \to ||xM||$ is a continuous function it is clear that the set $K^1 = \{x \in K : ||xM|| > 0\}$ belongs to $\mathcal{E}$. Next consider the set $K_F^1 = \{x \in K^1 : xM/||xM|| \in F\}$. Now if $x_0 \in K$ is such that $||x_0M|| > 0$ then, since the set $B \in \mathcal{E}$ is assumed to be open, there exists a number $\rho > 0$ such that if $||y - x_0|| < \rho$ then $y \in K_B^1$, from which follows that K_B^1 is an open set. (That such a number $\rho > 0$ exists follows from the inequality (38) since it implies that

$$||(xM/||xM|| - yM/||yM||)|| \leq 2(||xM - yM||)/||xM||$$

if both $||xM|| > 0$ and $||yM|| > 0$.) Therefore K, K^1 and K_F^1 all belong to $\mathcal{E}$ and hence $K_0' \in \mathcal{E}$.

Next let $a, b \in \mathtt{R}$ be two numbers satisfying $0 < a < b \leq a_0$ but otherwise arbitrary. Define

$$K_{a,b} = \{x \in K : a < u_F(x) < b\}.$$

Then again utilizing the assumption that the set F is an open set, we find that if $x \in K_{a,b}$ then $a < u_F(y) < b$ if $y \in B(x, \rho)$ and $\rho > 0$ is sufficiently small; hence $K_{a,b}$ is an open set and therefore belongs to $\mathcal{E}$.

Since K_b, when $0 < b \leq a_0$, can be written

$$K_b = K_0 \cup K_0' \cup_{n=1}^{\infty} K_{1/n,b}$$

and we have showed that each of these sets belongs to $\mathcal{E}$, it follows that $K_b \in \mathcal{E}$ when $0 < b \leq a_0$ and thereby we have proved that $K_b \in \mathcal{E}$ for all $b \in \mathtt{R}$. Hence u_F is $\mathcal{E} - measurable$, and the proof of Proposition 14.1 is completed. $\Box$

We now return to the proof of Theorem 14.1. Let $B \in \mathcal{E}$ be fixed but arbitrary. For $w \in \mathcal{W}$, define $u_{w,B} : K \to [0, 1]$ by

$$u_{w,B}(x) = ||xM(w)|| \;\; if \;\; x \in K_{w,B}^1$$

and

$$u_{w,B}(x) = 0 \;\; if \;\; x \notin K_{w,B}^+.$$

where

$$K_{w,B}^1 = \{x \in K_w^1 : xM(w)/||xM(w)|| \in B\}$$

and where

$$K_w^1 = \{x \in K : ||xM(w)|| > 0\}.$$

With these functions at our disposal we can write

$$P_{\mathcal{M}}(x, B) = \sum_{w \in \mathcal{W}_{\mathcal{M}}(x,B)} ||xM(w)|| = \sum_{w \in \mathcal{W}} u_{w,B}(x).$$

But from e) of Proposition 2.5 follows that $u_{w,B}$ is $\mathcal{E} - measurable$ for each $w \in \mathcal{W}$. Since $\mathcal{W}$ is denumerable and $u_{w,B} \geq 0$ it follows that $P_{\mathcal{M}}(\cdot, B)$ is $\mathcal{E} - measurable$, which was what we wanted to prove. $\Box$

15 Acknowledgements.

I am deeply grateful to my brother Sten Kaijser for always having the possibility to discuss my work with him. I also want to thank Arne Enqvist, Torkel Erhardsson and Robert Forchheimer for valuable and constructive comments.

References

[1] K.J. Åström, " Optimal Control of Markov Processes with Incomplete State Information", *J Math Anal Appl*, vol 10, (1965).

[2] L.E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains" *Ann of Math Stat*, vol 37, (1966).

[3] P. Billingsley, *Convergence of Probability Measures*, Wiley , New York , NY, 1968.

[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, T Ryden, *Inference in Hidden Markov Models*, Springer Series in Statistics, Springer, (2005).

[6] G.Choquet, *Lectures on Analysis, Volume II, Representation Theory*, Benjamin, Inc, New York, 1969.

[7] T.Cover, J. Thomas, *Elements of Information Theory*, Wiley and Sons, Inc, New York, 1991.

[8] G.B DiMasi, L. Stettner, "Ergodicity of hidden Markov models" *Math Control Signals systems*, vol 17, (2005).

[9] C. Dorea and A. Pereira, "A note on a variation of Doeblin's condition for uniform ergodicity of Markov chains", *Acta Math Hungar*, vol 110, no 4, pp 287-92, (2006).

[10] R.M. Dudley, *Real analysis and probability*, Cambridge University Press, 2002.

[11] Y. Ephraim and N. Merhav, "Hidden Markov Processes", *IEEE Trans Information Theory*, vol 48, no 6, pp 1518-69, (2002).

[12] W. Feller, *An introduction to Probability Theory and Its Applications, volume II, second edition*, Wiley, 1971.

(1973)

[13] F. Gantmacher, *Matrix theory,vol II*, Chelsea Publishing, NY , 1965.

[14] R. van Händel, "Observability and non-linear filtering", *Prob Theory and Related Fields*, vol 145, (2009).

[15] T. Holliday, P. Glynn and A. Goldsmith, "Capacity of Finite State Channels based on Lyaponov Exponents of Random Matrices" , *IEEE Transactions on Information Theory*, vol 52, (2006).

[16] T. Kaijser, "Some limit theorems for partially observed Markov chains" *Ann of Prob*, vol 3, (1975).

[17] L.V. Kantorovich, "On the transfer of mass (in Russian)", *Dokl Akad Nauk USSR* vol 37, (1942). Translated in Managements Science, vol 5, (1959).

[18] H. Kesten , (Private communication.) 1974.

[19] F. Kochman and J. Reeds, "A simple proof of Kaijser's unique ergodicity result for hidden Markov α-chains", *Ann Appl Prob*, vol 16, No 4, (2006).

[20] H. Kunita, "Asymptotic behaviour of the nonlinear filtering errors of Markov Processes", J Multivariate Analysis , vol 1, (1971)

[21] T. Lindwall, *Lectures on the coupling method*, Wiley and Sons, 1992.

[22] K.R. Parthasarathy, *Probability measures on metric spaces*, Academic Press, New York 1967

[23] N.U. Prabhu, *Stochastic Processes*, McMillan, New York, 1965.

[24] L.R. Rabiner, and B.H. Juang , "An introduction to Hidden Markov Models", *IEEE, ASSP Magazine* (1986).

[25] D. Revuz , *Markov chains*, North-Holland/American Elsevier, 1975.

[26] M. Rudemo, "State estimation for partially observed Markov chains", J Math Anal Appl, vol 44, (1973)

[27] A. Vershik, " Kantorovich metric: Initial history and little-known applications", *J Math Sciences*, vol 133, (2006).