Classification of Markov Chains on $\mathbb{R}^k$

Niels Richard Hansen
Department of Statistics and Operations Research
University of Copenhagen
Universitetsparken 5
2100 Copenhagen
Denmark
email: richard@math.ku.dk

November 28, 2000

Abstract

In this paper it is shown that the behavior of a Markov chain on $\mathbb{R}^k$ to a large extent is determined by the conditional mean values and the conditional variances. First it is shown that geometric drift (or drift) towards a compact set using the simple and well known drift function $V(x) = 1 + |x|^2$ is completely characterized by these two conditional moments, and also uniform ergodicity can be derived on the basis of these moments. Secondly a special class of Markov chains, called affine Markov chains, are considered and a new kind of drift function is introduced. Using this drift function we derive another criteria for geometric drift towards a compact set again based on the two conditional moments.

Keywords: Markov chains; affine Markov chains; conditional moments; drift functions; geometric drift; geometric ergodicity; uniform ergodicity.

1 Introduction

The study of a Markov chain on a general state space is often divided into two different parts, where we first investigate some basic regularity properties like ϕ-irreducibility and periodicity and secondly we try to classify the chain as being either transient or recurrent, and in the recurrent case we furthermore try to establish ergodicity, geometric ergodicity or uniform ergodicity. Under the assumption of ϕ-irreducibility and aperiodicity one can often verify some kind of ergodicity directly by using a *drift criteria*. This is treated briefly in appendix A.

To apply the drift criteria, one should also know the so-called *small sets*, and if the state space is $\mathbb{R}^k$, these are in many cases the same as the bounded sets, or at least all the compact sets are small. The approach to classification of Markov chains on $\mathbb{R}^k$ being ergodic or geometrically ergodic can therefore in many cases be divided as follows:

1) Establish ϕ-irreducibility, aperiodicity and that the compact sets are all small.

2) Establish drift or geometric drift towards some compact set.

This paper will almost exclusively deal with the second question of establishing drift or geometric drift.

In some cases the task of showing 1) is trivial, like if the transition probabilities have a strictly positive and continuous density w.r.t. the Lebesgue measure. In other cases it can be extremely difficult to show 1), for instance if the one-step transitions are singular in some sense and we have to iterate the transitions to expose the behavior of the chain and hence to show ϕ-irreducibility for some ϕ. Such situations occur for instance whenever we bring a (non-Markov) scalar process to state space form. Results about ϕ-irreducibility for this kind of Markov chains can be found in for instance [6], [7] or [2] and [1] also treats state space modeling techniques.

The verification of 2) is independent of 1), and we will show that it is very closely related to the knowledge of the conditional mean values and the conditional variances – at least if one uses the drift function $1 + |x|^2$. This is more or less well known, but will be made completely explicit in section three. We will also show how uniform ergodicity can be verified with the knowledge of these conditional moments.

A special class of Markov chains called affine Markov chains is also introduced, and we derive a new criterion for geometric drift towards a compact set for this class of Markov chains. Again the criterion is stated in terms of the conditional means and conditional variances. To show this criterion we will introduce a new class of drift functions, which in some sense are multidimensional versions of cosh.

Application of the results in section 3 to ARCH(1)-like models from time series analysis are also briefly discussed.

It should be mentioned that most of the results are stated in terms of Markov kernels instead of Markov chains.

2 Markov Chains on $\mathbb{R}^k$

We will consider Markov kernels on the state space $(\mathbb{R}^k, \mathbb{B}_k)$ with $\mathbb{B}_k$ denoting the Borel-σ-algebra on $\mathbb{R}^k$. We let $\mathbb{R}^k$ be equipped with the usual inner product denoted by $\langle\cdot,\cdot\rangle$ and the related 2-norm given by $|x|^2 = \langle x, x\rangle$ for $x \in \mathbb{R}^k$. The set of linear operators on $\mathbb{R}^k$ can be identified with the set of $k \times k$ matrices $M(k)$ and the set of invertible operators – which is often called the general linear group and denoted $GL(k)$ – is precisely the invertible matrices. The set $M(k)$ is itself a finite dimensional vector space of dimension k^2 and can be equipped with several different (but of course equivalent) norms. One of these norms is the operator norm defined as

$$||A|| = \sup_{|x|\leq 1} |Ax|$$

for $A \in M(k)$.

We are going to give a number of criteria for a Markov kernel P on $\mathbb{R}^k$ to have drift or geometric drift (see appendix A) towards a compact set in terms of conditional moments. Under usual regularity assumptions on P – that is ϕ-irreducibility, aperiodicity and the compact sets being small – these criteria guaranty P to be ergodic or geometrically ergodic – see theorem A.2 in appendix A. For a treatment of Markov chains and Markov kernels on a general state space we refer to [6] or [8].

If P is a Markov kernel we define recursivly $P^1 = P$ and

$$P^m(x, A) = \int P^{m-1}(y, A)P(x, dy)$$

for $m \geq 2$, hence P^m is the m-step transition probabilities for a Markov chain with transition probabilities P.

In the following we will use the convention that a Markov kernel P has second order moments if

$$\int |y|^2 P(x, dy) < \infty$$

for all $x \in \mathbb{R}^k$.

Definition 2.1 *Let P be a Markov kernel and suppose that P^m has second order moments. The m-step mean value map is then*

$$\xi^m(x) = \int y P^m(x, dy)$$

and the m-step variance map is

$$\Sigma^m(x) = \int yy^T P^m(x, dy) - \xi^m(x)\xi^m(x)^T.$$

Note that $\xi^m : \mathbb{R}^k \to \mathbb{R}^k$ and $\Sigma^m : \mathbb{R}^k \to M(k)$ are both measurable maps. The m-step variance map can always be written as

$$\Sigma^m(x) = \Phi^m(x)\Phi^m(x)^T \tag{1}$$

for some matrix $\Phi^m(x)$ since $\Sigma^m(x)$ is positive semidefinite. A canonical choice would of course be to use the unique positive semidefinite square root $\Sigma^m(x)^{\frac{1}{2}}$, but we will not restrict ourself to that. Hence we will call any measurable map $\Phi^m : \mathbb{R}^k \to M(k)$, for which $\Phi^m(x)$ satisfies (1) for all $x \in \mathbb{R}^k$, an m-step scale map. In the case $m = 1$ we will always drop the "1-step", that is we will just say mean value map, variance map etc., and we will write $\xi(x)$, $\Sigma(x)$ and $\Phi(x)$. With these conventions we can observe that the m-step mean value map is the mean value map for P^m.

If $(X_n)_{n\in\mathbb{N}_0}$ is a Markov chain with transition probability P we have that

$$\xi^m(x) = E(X_{n+m}|X_n = x)$$

and

$$\Sigma^m(x) = V(X_{n+m}|X_n = x)$$

for all $m \in \mathbb{N}$. This should be interpreted either as if the right-hand sides *by definition* are equal to the left-hand sides, or as an almost sure equality if $E(|X_{n+m}|^2) < \infty$ and the right-hand sides are defined using ordinary conditional mean values. Note that time-homogeneity of the Markov chain makes the expressions above independent of n.

Since the m-step mean value map and variance map only make sense if P^m has second order moments, we will make the convention that whenever we discuss m-step mean value and variance maps, we will implicitly assume that P^m has second order moments.

It can often be quite difficult or impossible to find P^m from P and it is therefore just as impossible to calculate ξ^m and Σ^m for $m \geq 2$ directly from the definition. However, it follows from the Chapman-Kolmogorov equations that if P^{m-1} has second order moments and if $\mathrm{tr}(\Sigma^{m-1} + \xi^{m-1}(\xi^{m-1})^T)$ is integrable with respect to $P(x,\cdot)$ for all $x \in E$ then P^m has second order moments and

$$\xi^m(x) = \iint z\, P^{m-1}(y,dz)P(x,dy) = \int \xi^{m-1}(y)P(x,dy) \tag{2}$$

and

$$\Sigma^m(x) = \int \Sigma^{m-1}(y) + \xi^{m-1}(y)\xi^{m-1}(y)^T P(x,dy) - \xi^m(x)\xi^m(x)^T. \tag{3}$$

These recursion formulas can be used to calculate ξ^m as well as Σ^m. Observe also that if $\xi(x) = 0$ for all $x \in \mathbb{R}^k$ it follows from (2) that $\xi^m(x) = 0$ for all $m \in \mathbb{N}$. In this case (3) reduces to

$$\Sigma^m(x) = \int \Sigma^{m-1}(y)P(x,dy). \tag{4}$$

2.1 Affine Markov Chains

The two conditional moments considered above reflect some of the dependencies in the Markov chain but in general not all kinds of dependencies. We will now consider the situation where the conditional distribution of X_{n+1} given X_n is only affected by X_n on the mean value and the scale. The definition is most easily understood in terms of Markov chains instead of kernels.

Definition 2.2 *A Markov chain $(X_n)_{n\in\mathbb{N}_0}$ is affine if there exist measurable maps $\tilde{\xi}:\mathbb{R}^k\to\mathbb{R}^k$ and $\tilde{\Phi}:\mathbb{R}^k\to GL(k)$ and a sequence of iid stochastic variables $(W_n)_{n\in\mathbb{N}}$ independent of X_0 such that*

$$X_{n+1}=\tilde{\xi}(X_n)+\tilde{\Phi}(X_n)W_{n+1} \tag{5}$$

for $n\in\mathbb{N}_0$.

The sequence $(W_n)_{n\in\mathbb{N}}$ is called the innovation sequence and the distribution of W_1 (and hence of all the W_n's) the innovation distribution. If ν is the innovation distribution we can identify the transition probability as the Markov kernel

$$P(x,\cdot)=A(x)(\nu) \tag{6}$$

where $A(x):\mathbb{R}^k\to\mathbb{R}^k$ is the affine map $A(x)(y)=\tilde{\xi}(x)+\tilde{\Phi}(x)y$. Thus if we define an *affine Markov kernel* to be a kernel of the form (6) for some $\tilde{\xi}$, $\tilde{\Phi}$ and a probability measure ν, we have that the transition probability for an affine Markov chain is an affine Markov kernel. Furthermore, if $(X_n)_{n\in\mathbb{N}_0}$ is a Markov chain with a transition probability P being affine, the sequence $(W_n)_{n\in\mathbb{N}}$ defined by

$$W_n=\tilde{\Phi}(X_{n-1})^{-1}(X_n-\tilde{\xi}(X_{n-1}))$$

is iid with W_n having distribution ν, independent of X_0 and

$$X_n=\tilde{\xi}(X_{n-1})+\tilde{\Phi}(X_{n-1})W_n$$

for all $n\in\mathbb{N}$.

Definition 2.3 *An affine Markov chain (or kernel) is called regular if the innovation distribution has second order moment and if the variance matrix is non-singular.*

In the regular case it is clear that we can always assume ν to be *normalized*, i.e. to have $\int x\mu(dx)=0$ and $\int xx^T\nu(dx)=I$ by changing $\tilde{\xi}$ and $\tilde{\Phi}$. If ν is normalized it is clear that $\xi=\tilde{\xi}$ and $\Phi=\tilde{\Phi}$, that is $\tilde{\xi}$ is the mean value map and $\tilde{\Phi}$ is some scale map. We will from now on always assume that the innovation distribution is normalized in the regular case.

Though we are not going to discuss detailed criteria for ϕ-irreducibility etc. it should be mentioned that if P is an affine Markov chain for which the innovation distribution has a strictly positive density with respect to the Lebesgue measure λ, then P is λ-irreducible and aperiodic. If the density as well as ξ and Φ are continuous then the compact sets are small.

It should also be mentioned that Tong in [10] introduces a similar class of Markov chains, dividing the Markov chain into a deterministic term and a noise term. The affine class dicussed here is not as general but a lot easier to handle, and in section 4 we will give a general and rather easy to use criteria for geometric drift. Furthermore, many models used in practice fall into this class of affine Markov chains.

3 Drift, Geometric Drift and Uniform Ergodicity

The following trivial observation will turn out to be quite useful, if we want to use a drift function like $V(x) = 1 + |x|^2$. If ν is any distribution on $\mathbb{R}^k$ with second order moments we have that

$$\int |x|^2 \nu(dx) = |\xi|^2 + \mathrm{tr}(\Sigma), \tag{7}$$

where $\xi = \int x \ \nu(dx)$ and $\Sigma = \int x x^T \nu(dx) - \xi\xi^T$.

Theorem 3.1 *If P be a Markov kernel with m-step mean value map ξ^m and m-step variance map Σ^m being bounded on the compact sets, then P has m-step drift towards a compact set with drift function*

$$V(x) = |x|^2$$

if and only if

$$\limsup_{|x|\to\infty} |\xi^m(x)|^2 + \mathrm{tr}(\Sigma^m(x)) - |x|^2 < 0. \tag{8}$$

Proof: If (8) holds then by (7)

$$\begin{aligned} P^m V(x) &= |\xi^m(x)| + \mathrm{tr}(\Sigma^m(x)) \\ &= |x|^2 - \epsilon + (|\xi^m(x)|^2 + \mathrm{tr}(\Sigma^m(x)) - |x|^2 + \epsilon) \\ &\leq V(x) - \epsilon + b1_C(x) \end{aligned}$$

where $\epsilon > 0$ and C is a compact set such that

$$|\xi^m(x)|^2 + \mathrm{tr}(\Sigma^m(x)) - |x|^2 + \epsilon < 0$$

on C^c, and $b = \sup_{x\in C} |\xi^m(x)|^2 + \mathrm{tr}(\Sigma^m(x)) + \epsilon < \infty$ by assumption. On the other hand, if there exists a compact set C such that

$$P^m V(x) = |\xi^m(x)| + \mathrm{tr}(\Sigma^m(x)) \leq |x|^2 - \epsilon$$

for all $x \in C^c$, then (8) clearly holds. $\square$

Theorem 3.2 *A Markov kernel P with m-step mean value map ξ^m and m-step variance map Σ^m being bounded on compact sets has m-step V-geometric drift towards a compact set with drift function*

$$V(x) = 1 + |x|^2 \tag{9}$$

if and only if

$$\limsup_{|x|\to\infty} \frac{|\xi^m(x)|^2 + \mathrm{tr}(\Sigma^m(x))}{|x|^2} < 1. \tag{10}$$

Proof: From (7) we get the conditionally expected drift to be

$$\int V(y)P^m(x,dy) \quad = \quad 1 + |\xi^m(x)|^2 + \mathrm{tr}(\Sigma^m(x)).$$

Assume that (10) holds and put

$$\alpha = \limsup_{|x|\to\infty} \frac{|\xi^m(x)|^2 + \mathrm{tr}(\Sigma^m(x))}{|x|^2}.$$

Then choose $\beta < 1$ with $\alpha < \beta$, and we have that for some suitable compact set C

$$\begin{aligned} \int V(y)P^m(x,dy) &= 1 - \frac{|\xi^m(x)|^2 + \mathrm{tr}(\Sigma^m(x))}{|x|^2} + \frac{|\xi^m(x)|^2 + \mathrm{tr}(\Sigma^m(x))}{|x|^2} V(x) \\ &\leq \beta V(x) + b 1_C(x) \end{aligned}$$

with $b = \sup_{x\in C} 1 + |\xi^m(x)|^2 + \mathrm{tr}(\Sigma^m(x)) < \infty$ by assumption. If P has m-step V-geometric drift towards a compact set C, it follows immediately that for $x \in C^c$,

$$1 + |\xi^m(x)|^2 + \mathrm{tr}(\Sigma^m(x)) \leq \beta(1 + |x|^2)$$

for some $\beta < 1$ and (10) is clearly satisfied. $\square$

On $M(k)$ there is an inner product given by the trace, i.e. $(A,B)_{\mathrm{tr}} = \mathrm{tr}(AB^T)$. The related 2-norm is then $||A||_2^2 = (A,A)_{\mathrm{tr}} = \mathrm{tr}(AA^T)$. If Φ^m is an m-step scale map, that is $\Sigma^m(x) = \Phi^m(x)\Phi^m(x)^T$, we get that $\mathrm{tr}(\Sigma^m(x)) = ||\Phi^m(x)||_2^2$. Thus the criterion for geometric drift in theorem 3.2 is a restriction upon the growth of $\xi^m(x)$ in the norm $|\cdot|$ and $\Phi^m(x)$ in the norm $||\cdot||_2$ for $|x| \to \infty$.

Example 3.3 Consider a Markov kernel P with second order moments, with

$$\xi(x) = Ax \tag{11}$$

for some matrix $A \in M(k)$ and with

$$\Sigma(x) = \Omega + Bxx^T B^T \tag{12}$$

for a matrix $B \in M(k)$ and a positive definite matrix Ω. We have this kind of conditional means and variances if we for instance consider an ARCH(1)-model. By induction using (3) we find that

$$\xi^m(x)\xi^m(x)^T + \Sigma^m(x) = (A \otimes A + B \otimes B)^m(xx^T) + \sum_{i=0}^{m-1}(A \otimes A + B \otimes B)^i(\Omega)$$

where $A \otimes A$ is the linear operator on $M(k)$ given by

$$A \otimes A(C) = ACA^T$$

and similarly for $B \otimes B$. This gives that

$$\begin{aligned}
\limsup_{|x|\to\infty} \frac{|\xi^m(x)|^2 + \mathrm{tr}(\Sigma^m(x))}{|x|^2} &= \limsup_{|x|\to\infty} \frac{\mathrm{tr}(\xi^m(x)\xi^m(x)^T + \Sigma^m(x))}{|x|^2} \\
&= \limsup_{|x|\to\infty} \frac{\mathrm{tr}((A \otimes A + B \otimes B)^m(xx^T))}{|x|^2} \\
&= \sup_{|x|=1} \mathrm{tr}((A \otimes A + B \otimes B)^m(xx^T)) \\
&\leq k||(A \otimes A + B \otimes B)^m||.
\end{aligned}$$

If the spectral radius of $A \otimes A + B \otimes B$ is less than 1, it follows from the spectral radius formula that $||(A \otimes A + B \otimes B)^m|| \to 0$ for $m \to \infty$, and (10) is fulfilled for m large enough. This provides us with a well known criteria for m-step geometric drift towards a compact set, i.e. that the numerically largest eigenvalue of $(A \otimes A + B \otimes B)$ should be less than 1, for the ARCH(1)-model using the drift function V_2, but the setup here is considerably more general. We do not have any restrictions on the Markov kernel P except that it should have conditional moments given by (11) and (12). If, furthermore, P is ϕ-irreducible, aperiodic and the compact sets are small, then P is geometrically ergodic with invariant distribution having second order moments – see theorem A.2 in appendix A.

The use of the drift function $1 + |x|^2$ for a simple ARCH-model has been considered in for instance [4] and for more general ARCH-models in [5]. One can find earlier results in [3] on the weak stationarity of GARCH-models and other references to the ARCH-literature can be found there. $\diamond$

Theorem 3.4 *Let P be a ϕ-irreducible and aperiodic Markov kernel with the compact sets being small. If P^m has second order moments and if both the m-step mean value map ξ^m and the m-step variance map Σ^m are bounded, then P is uniformly ergodic.*

Proof: It follows from Chebychevs inequality that

$$P^m(x, B(0,R)^c) \leq \frac{\mathrm{tr}(\Sigma^m(x)) + |\xi^m(x)|^2}{R^2} \leq \frac{\gamma}{R^2}$$

for some $\gamma < \infty$ since both ξ^m and Σ^m are bounded. For $\epsilon \in]0,1[$ choose R such that $\frac{\gamma}{R^2} \leq \epsilon$. Let $(X_n)_{n\in\mathbb{N}_0}$ be a Markov kernel with transition probability P started in $x \in E$, then by induction

$$\mathbb{P}_x(\tau_{B(0,R)} > nm) \leq \mathbb{P}_x(|X_m| \geq R, |X_{2m}| \geq R, \dots, |X_{nm}| \geq R) \leq \epsilon^n$$

with $\tau_{B(0,R)}$ being the first return time to $B(0,R)$. This gives us the following bound on the expected return time to $B(0,R)$

$$\begin{aligned} E_x(\tau_{B(0,R)}) &= \sum_{n=1}^{\infty} n\mathbb{P}_x(\tau_{B(0,R)} = n) \\ &= \sum_{k=0}^{\infty} \mathbb{P}_x(\tau_{B(0,R)} > k) \\ &\leq m\sum_{k=0}^{\infty} \mathbb{P}_x(\tau_{B(0,R)} > mk) \\ &\leq m\sum_{k=0}^{\infty} \epsilon^k, \end{aligned}$$

and therefore

$$\sup_x E_x(\tau_{B(0,R)}) \leq m\sum_{k=0}^{\infty} \epsilon^k < \infty.$$

From theorem 16.0.2 in [6] we get that P is uniformly ergodic, since $B(0,R)$ is small for all R by assumption. □

4 A New Class of Drift Functions

If the variance map is bounded, the criterion for geometric drift given by (10) reduces to

$$\limsup_{|x|\to\infty} \frac{|\xi(x)|}{|x|} < 1$$

In this section we will show that it is possible to give a less restrictive criteria for geometric ergodicity in the affine case under the assumption of a bounded variance map and that the innovation distribution has sufficiently rapid decaying tails

Define for $s \in \mathbb{R}$ the function $V_s : \mathbb{R}^k \to \mathbb{R}$ by

$$V_s(x) = \int_{S^{k-1}} \exp(s\langle y, x\rangle)\omega(dy) \tag{13}$$

with ω being the normalized surface measure on the sphere S^{k-1}. We will in this section show that V_s can be used as a drift function, and that it is especially useful when we consider affine Markov kernels. Note that in the one dimensional case

$V_s(x) = \cosh(sx)$. Since the surface measure ω on S^{k-1} is invariant under rotations, we can fix a unit vector e and for each $x \in \mathbb{R}$ find a rotation O_x, such that $O_x x = |x|e$, which gives us that

$$\begin{aligned}\int_{S^{k-1}} \exp(s\langle y, x\rangle)\omega(dy) &= \int_{S^{k-1}} \exp(s|x|\langle y, e\rangle)\omega(dy) \\ &= \int_{-1}^{1} \exp(s|x|t)\mu(dt)\end{aligned}$$

with μ being the transformation of ω under the map

$$y \mapsto \langle y, e\rangle.$$

Before we can apply this drift function we make a useful observation. If μ is a probability measure on $[-1, 1]$, we define

$$\phi(h) = \int_{-1}^{1} \exp(ht)\mu(dt) \tag{14}$$

and we then have an exponential family of probability measures given by

$$\nu_h(B) = \frac{1}{\phi(h)} \int_B \exp(ht)\mu(dt) \tag{15}$$

for $h \in \mathbb{R}$.

Lemma 4.1 *If $\mu([q, 1]) > 0$ for all $q < 1$, the probability measures ν_h converge weakly to δ_1 for $h \to \infty$.*

Proof: By the definition we have for $q < 1$

$$\begin{aligned}\phi(h) &= \int_{-1}^{1} \exp(ht)\mu(dt) \\ &\geq \int_{q}^{1} \exp(ht)\mu(dt) \\ &\geq \exp(hq)\mu([q, 1])\end{aligned}$$

Then if $p < q$ it follows that

$$\begin{aligned}\nu_h([-1, p]) &= \frac{1}{\phi(h)} \int_{-1}^{p} \exp(ht)\mu(dt) \\ &\leq \frac{1}{\mu([q, 1])} \exp(-hq) \exp(hp)\mu([-1, p]) \\ &= \frac{\mu([-1, p])}{\mu([q, 1])} \exp(h(p - q)) \to 0\end{aligned}$$

for $h \to \infty$. Since all the measures ν_h are concentrated on $[-1,1]$, it follows that

$$\nu_h \xrightarrow{w} \delta_1$$

for $h \to \infty$. $\square$

Consider a regular affine Markov kernel P with (normalized) innovation distribution ν. From now on we will assume that ν has rapidly decaying tails in the sense that

$$\int \exp(s_0|x|)\nu(dx) < \infty \tag{16}$$

for some $s_0 > 0$, and we can define the map

$$\psi(s,y) = \int \exp(s\langle y,x\rangle)\nu(dx)$$

for $(s,y) \in [0,s_0] \times B(0,1)$, with $B(0,1) = \{x \in \mathbb{R}^k \mid |x| \le 1\}$ being the closed unit ball. By decreasing s_0 we can increase $B(0,1)$ to be any closed ball, and hence we can assume that ψ is defined on $[0,s_0] \times B(0,d)$ for some suitable large d and sufficiently small s_0 if necessary.

Theorem 4.2 *If P is a regular affine Markov kernel with ν satisfying (16), with Φ being bounded, with ξ being bounded on compact sets and with*

$$\limsup_{|x|\to\infty} |\xi(x)| - |x| < 0 \tag{17}$$

then P has geometric drift towards a compact set with drift function $V_{\tilde{s}}$ for some suitable $\tilde{s} > 0$.

Proof: By Tonelli and the integral transformation theorem we get

$$\begin{aligned}
PV_s(x) &= \int V_s(z)P(x,dz) \\
&= \int\int_{S^{k-1}} \exp(s\langle y,z\rangle)\omega(dy)P(x,dz) \\
&= \int_{S^{k-1}}\int \exp(s\langle y,\xi(x)+\Phi(x)z\rangle)\nu(dz)\omega(dy) \\
&= \int_{S^{k-1}} \exp(s\langle y,\xi(x)\rangle)\int \exp(s\langle \Phi(x)^T y,z\rangle)\nu(dz)\omega(dy) \\
&= \int_{S^{k-1}} \exp(s\langle y,\xi(x)\rangle)\psi(s,\Phi(x)^T y)\omega(dy)
\end{aligned}$$

for $s > 0$ sufficiently small.

Let $K = \{\Phi(x)y \in \mathbb{R}^k \mid x \in \mathbb{R}^k \text{ and } y \in S^{k-1}\}$ be the picture of the map $(x,y) \mapsto \Phi(x)y$, which is bounded since Φ is bounded. Therefore K is contained in a compact, convex box $B_r = \{x \in \mathbb{R}^k \mid |x|_\infty \leq r\}$ for some $r > 0$, with $|x|_\infty = \max_{i=1,\dots,k} |x_i|$ being the max-norm. Assume now that s_0 is chosen sufficiently small such that that ψ is defined on $[0, s_0] \times B(0, d)$ with $B_r \subseteq B(0, d)$.

The box B_r has exactly 2^k extreme points, which we will denote by $z_1, \dots, z_{2^k}$, and the map

$$z \mapsto \psi(s, z)$$

for fixed $s \leq s_0$ is convex, and hence it attains its maximal value over a compact, convex set in one of the extreme points. With

$$\widetilde{\psi}(s) = \max_{j=1,\dots,2^k} \psi(s, z_j)$$

we therefore get that

$$\psi(s, z) \leq \widetilde{\psi}(s)$$

for all $z \in B_r \supseteq K$. This gives us

$$PV_s(x) \leq \widetilde{\psi}(s) \int_{S^{k-1}} \exp(s\langle y, \xi(x)\rangle)\omega(dy) = \widetilde{\psi}(s) \int_{-1}^{1} \exp(s|\xi(x)|t)\mu(dt).$$

Now assume that R is given such that for $|x| \geq R$ we have

$$|\xi(x)| \leq -\alpha + |x|$$

for some $\alpha > 0$. Put

$$\phi_j(s) = \psi(s, z_j) \exp(-s\alpha) = \int \exp(s(\langle z_j, x\rangle - \alpha))\nu(dx)$$

for $j = 1, \dots, 2^k$, then ϕ_j is convex and differentiable with

$$\phi_j'(0) = \int \langle z_j, x\rangle \nu(dx) - \alpha = -\alpha$$

and since

$$\phi_j(0) = 1$$

it follows that we can find s_j such that

$$\phi_j(s) < 1$$

for all $s \leq s_j$. With $\tilde{s} = \min_{j=1,\dots,2^k}\{s_j\} > 0$, we get that

$$c = \widetilde{\psi}(\tilde{s}) \exp(-\tilde{s}\alpha) < 1.$$

The map $u \mapsto \int_{-1}^{1} \exp(ut)\mu(dt)$ is increasing for $u \geq 0$, and therefore

$$\int_{-1}^{1} \exp(\tilde{s}|\xi(x)|t)\mu(dt) \leq \int_{-1}^{1} \exp(-\tilde{s}\alpha t) \exp(\tilde{s}|x|t)\mu(dt)$$

for $|x| \geq R$ and since $t \mapsto \exp(-\tilde{s}\alpha t)$ is continuous on $[-1, 1]$ it follows from lemma 4.1 that

$$\frac{\int_{-1}^{1} \exp(-\tilde{s}\alpha t) \exp(\tilde{s}|x|t)\mu(dt)}{\int_{-1}^{1} \exp(\tilde{s}|x|t)\mu(dt)} \to \exp(-\tilde{s}\alpha)$$

for $|x| \to \infty$. This shows that

$$\begin{aligned} \frac{PV_{\tilde{s}}(x)}{V_{\tilde{s}}(x)} &\leq \widetilde{\psi}(\tilde{s}) \frac{\int_{-1}^{1} \exp(-\tilde{s}\alpha t) \exp(\tilde{s}|x|t)\mu(dt)}{\int_{-1}^{1} \exp(\tilde{s}|x|t)\mu(dt)} \\ &\to \widetilde{\psi}(\tilde{s}) \exp(-\tilde{s}\alpha) = c < 1 \end{aligned}$$

for $|x| \to \infty$ and it follows that for fixed $\beta \in]c, 1[$ we can choose an $R_0 \geq R$, such that for $|x| \geq R_0$ we have

$$\frac{PV_{\tilde{s}}(x)}{V_{\tilde{s}}(x)} \leq \beta.$$

Since ξ is bounded on compact sets, we have

$$PV_{\tilde{s}}(x) \leq \beta V_{\tilde{s}}(x) + b 1_{B(0,R_0)}(x)$$

with $b = \sup_{|x| \leq R_0} PV_{\tilde{s}}(x) < \infty$. □

Note that if (10) is satisfied, then

$$\limsup_{|x| \to \infty} |\xi(x)| - |x| = \limsup_{|x| \to \infty} |x| (\frac{|\xi(x)|}{|x|} - 1) = -\infty < 0$$

and therefore (17) holds too.

5 Concluding Remarks

The theorems presented in section 2 completely characterize to what extend the drift function $V(x) = 1 + |x|^2$ ($V(x) = |x|^2$) can be used to show geometric ergodicity (ergodicity) for a Markov chain on $\mathbb{R}^k$. Furthermore, we showed that $V(x) = 1 + |x|^2$ can be used with a minimum of knowledge about the Markov kernel considered.

If we, however, make further assumptions, i.e. affinity, on the structure of the Markov kernel considered, we are able to use other drift functions. This gives us theorem 4.2, which is expected to be close to the optimal condition for geometric ergodicity as long as the scale map is bounded.

In the one-dimensional case, Tjøstheim derives in [9] for affine Markov chains with constant variance a result about geometric ergodicity similar to the one presented in section 3. We have shown in section 4 that one can obtain a stronger result in this setup if the innovation distribution has light tails.

5.1 Ackowledgement

I am very grateful to Ernst Hansen for his many critical comments on earlier versions of this paper, and to him and Anders Rahbek for showing me unpublished work related to ergodicity of ARCH-processes. Furthermore, I would like to thank MaPhySto and CAF for the financial support.

A Appendix: Drift and Ergodicity

It is well known that ergodicity and geometric ergodicity can be shown using the so-called drift inequalities like it is described in Meyn and Tweedies monograph [6]. In this appendix we will briefly treat some useful generalizations of these drift inequalities.

Definition A.1 *A Markov kernel P on $(E, \mathbb{E})$ is said to have m-step drift towards a set C with drift function $V : E \to [0, \infty[$, if*

$$P^m V(x) = \int V(y) P^m(x, dy) \leq V(x) - \epsilon + b 1_C(x) \tag{18}$$

for some $\epsilon > 0$ and some $b < \infty$.

A Markov kernel P is said to have m-step geometric drift towards a set C with drift function $V : E \to [1, \infty]$ if $V(x)$ is finite for some $x \in E$, and if

$$P^m V(x) = \int V(y) P^m(x, dy) \leq \beta V(x) + b 1_C(x) \tag{19}$$

for some $\beta < 1$ and some $b < \infty$.

If the Markov kernel P has an invariant probability measure π, then P is said to be ergodic if

$$||P^n(x, \cdot) - \pi||_{tv} \to 0$$

for $n \to \infty$ and for all $x \in E$, and V-geometrically ergodic on $A \in \mathbb{E}$ if

$$||P^n(x, \cdot) - \pi||_V \leq R V(x) \rho^n$$

for some function $V : E \to [1, \infty]$, finite on A, some $R < \infty$, $\rho < 1$ and all $x \in A$. For the definition of the total variation norm and the V-norm see [6].

With these definitions, the following theorem holds:

Theorem A.2 *Suppose that P is ϕ-irreducible and aperiodic.*

If P has m-step drift towards a small set with drift function V, then P has an invariant probability measure and P is ergodic.

If P has m-step geometric drift towards a small set with drift function V, then P has an invariant probability measure π and there exists a function $\tilde{V} \geq V$, such that $\pi(\tilde{V}) < \infty$ and such that P is $\tilde{V}$-geometrically ergodic on $\{x \in E \mid \tilde{V}(x) < \infty\}$.

A complete proof is not given, but a few details should be mentioned. Part one of the theorem is proved by using that P^m has ordinary drift towards a small set with an everywhere finite drift function, and then using theorem 11.3.5 and 13.0.1 in [6] plus the fact that the total variation norm $||P^n(x,\cdot) - \pi||$ is decreasing (proposition 13.3.2 in [6]). For the second part one introduces

$$\tilde{V}(x) = \sum_{i=0}^{n-1} \gamma^{\frac{i}{n}} P^i V(x) \geq V(x)$$

for some sufficiently large n and a constant $\gamma > 1$. Then one shows that P has $\tilde{V}$-geometric drift towards a small set, and the theorem follows. For details on the second part one can consult lemma 7 in [4]. If $P^i V(x) < \infty$ for all $x \in E$ and all $i \in \mathbb{N}_0$ it therefore also follows that P is $\tilde{V}$-geometrically ergodic on E.

References

[1] Aoki, Masanao. *State Space Modeling of Time Series.* Springer-Verlag, 1987.

[2] Doukhan, Paul. *Mixing, Properties and Examples.* Springer-Verlag, 1994.

[3] Engle, R. F. and Kroner, K. F. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11:122–150, 1995.

[4] Hansen, E. and Rahbek, A. Stationarity and assymptotics of multivariate ARCH time series with an application to robustness of cointegration analysis. Preprint 12, Department of Statistics and Operations Research, University of Copenhagen, 1998.

[5] Kristensen, Dennis. Ergodicity and inference in multivariate ARCH-models: A markovian approach. Master thesis, Department of Statistics and Operations Research, University of Copenhagen, 2000.

[6] Meyn, S. P. and Tweedie, R. L. *Markov Chains and Stochastic Stability.* Springer-Verlag, 1996.

[7] Mokkadem, A. Propriétés de mélange des processus autorégressifs polynomiaux. *Ann. Inst. Henri Poincarre*, 26(4):219–260, 1990.

[8] Nummelin, E. *General Irreducible Markov Chains and Non-negative Operators.* Cambridge University Press, 1984.

[9] Tjøstheim, D. Non-linear time series and Markov chains. *Adv. Appl. Prob.*, 22(2):587–611, 1990.

[10] Tong, H. *Non-linear Time Series. A Dynamical System Approach.* Oxford University Press, 1990.