STATISTICAL THEORY OF NON-LINEAR STOCHASTIC SYSTEMS I.

László Gerencsér      Zalán Mátyás      Gábor Molnár-Sáska
Zsanett Orlovits

(Lecture Notes)

Computer and Automation Research Institute of
the Hungarian Academy of Sciences, SZTAKI
Stochastic Systems Research Group

2004.

© All rights reserved.
Contents

1 A basic estimation problem 2
2 Parametric Markov processes 8
3 The basic algorithm 9
4 The ODE-method 13
5 $L^q$ estimates 23
6 Appendix 28
1 A basic estimation problem

The LMS-method. Let \((x_n, y_n)\) be a second order stationary process, \(x_n \in \mathbb{R}^d, y_n \in \mathbb{R}\). Find \(\theta \in \mathbb{R}^d\) such that

\[
E(y_n - \theta^T x_n)^2
\]

is minimized. The problem is quadratic in \(\theta\). The solution is

\[
\theta^* = (E x_n x_n^T)^{-1}(E x_n y_n).
\]

A low-complexity procedure to find \(\theta^*\) is the LMS-method (least mean squares):

\[
\theta_{n+1} = \theta_n + \gamma x_n (y_n - x_n^T \theta_n).
\]

This is a widely studied procedure, for an excellent reference see [6]. The signed LMS method (more precisely: the signed error LMS method) is

\[
\theta_{n+1} = \theta_n + \gamma x_n \text{sgn}(y_n - x_n^T \theta_n)
\]

see Bucklew, Kurtz, Sethares [2]. The reformulation of the LMS problem: solve

\[
E x_n (y_n - x_n^T \theta) = 0.
\]

Now we move to a Markovian framework. The classical definition of a Markov-process: given a measurable space \((\mathcal{X}, \mathcal{A})\), where \(\mathcal{X}\) is a separable, complete metric space (i.e. a Polish space).

**Definition 1.1** A transition kernel \(P(x; A)\), \(x \in \mathcal{X}, A \in \mathcal{A}\) is a function of two variables such that

1. \(P(x; \cdot)\) is a measure for all \(x \in \mathcal{X}\)
2. \(P(\cdot; A)\) is measurable for all \(A \in \mathcal{A}\).

**Definition 1.2** An \(\mathcal{X}\)-valued process \(x = (x_n)\), \(n = 0, 1\ldots\) over some probability space \((\Omega, \mathcal{F}, Q)\) is a homogenous Markov process with transition kernel \(P(x; A)\) if \(Q\)-almost surely

\[
P(X_{n+1} \in A|X_n, \ldots, X_0) = P(X_{n+1} \in A|X_n)
\]

and

\[
P(X_{n+1} \in A|X_n = x) = P(x; A).
\]

**Remark 1.1** In BMP we have \(\mathcal{X} = \mathbb{R}^k\), but the Euclidean structure is not really needed.
A modern view of Markov processes: let \((U, \mathcal{B})\) be a measurable space, and let
\[ f : \mathcal{X} \times U \to \mathcal{X} \]
be a measurable mapping. Let \(U = (U_i), i = 1, 2, \ldots\) be an \(U\)-valued i.i.d. sequence over some probability space \((\Omega, \mathcal{F}, Q)\). Then the process defined by
\[ X_{n+1} = f(X_n, U_{n+1}), \quad x_0 = \xi \]
with \(\xi\) being independent of \(U\) is a homogeneous Markov-process. (Prove this!)

The mapping
\[ T_n : x \to f(x, U_n) \]
is a random mapping of \(\mathcal{X}\) into itself. We thus can also write
\[ X_{n+1} = T_{n+1}X_n \]
The corresponding transition kernel is
\[ P(x; A) = Q(Tx \in A). \tag{1} \]
Conversely if a given transition kernel can be written in this form, then we say that \(T\) realizes \(P(x; A)\). This is not just an example for a Markov process, as the following theorem shows.

**Theorem 1.1** (see [7]) Let \(\mathcal{X}\) be a Polish space, and let \(A = \mathcal{B}(\mathcal{X})\) be its Borel \(\sigma\)-algebra. Let \(P(x; A)\) be a Markov kernel with \(x \in \mathcal{X}, A \in A\). Then \(P\) can be realized by a random mapping \(T\) of the form
\[ T : x \to f(x, U), \]
where \(U\) can be taken to be a uniformly distributed in \([0, 1]\).

Discussion. The realization of Markov-process is described in Kifer [7], see also [9]. An open problem(?) : assume that \(P_\theta(x, A)\) is Hölder-continuous in \(\theta\) in the Prohorov-metric. Does it follow that there exists a realization \(f\) such that \(f\) is Hölder-continuous in \(\theta\) for all fixed \((x, u)\)?

**Remark 1.2** This is at first sight an unexpected result. Assume e.g. that
\[ x_{n+1} = Ax_n + Be_{n+1}, \]
where \(x_n\) is vector-valued, \((e_n)\) is Gaussian i.i.d., and \(A\) is stable, i.e. its eigenvalues are all in the open unit disc of the complex plane \(\{z : |z| < 1\}\).
Then, it is easy to see that there exists a stationary solution. And now suddenly we claim that the very same process can be defined by a recursion of the form

\[ x_{n+1} = f(x_n, U_{n+1}), \]

where \((U_n)\) is i.i.d. in \([0, 1]\).

**Morale:** The same transition kernel may have a number of different realization!

**Local averaging.** Let \(f : \mathcal{X} \to \mathbb{R}\) be a bounded, measurable function. Define a function \(Pf\) by

\[ Pf(x) = \int_{\mathcal{X}} f(y) P(x, dy), \]

or in BMP-s notation:

\[ \pi f(x) = \int_{\mathcal{X}} f(y) \pi(x, dy). \]

We can also write

\[ Pf(x) = E(f(X_{n+1})|X_n = x). \]

\(Pf\) is thus the local average of \(f\) over positions reached in one step starting from \(x\).

**Proposition 1.1** \(Pf\) is a bounded measurable function.

(Hint for proof: Use the fact that \(P(x, A)\) is measurable in \(x\).) A difference operator acting on \(f\) is defined as

\[ g = Pf - f. \]

**Definition 1.3** A probability measure \(\mu\) on \((\mathcal{X}, \mathcal{A})\) is invariant for the transition kernel \(P(x; A)\) if for all \(A \in \mathcal{A}\)

\[ \mu(A) = \int_{\mathcal{X}} \mu(dx) P(x; A). \]

An alternative definition: if the distribution of \(X_0\) is \(\mu\), then the distribution of \(X_1\) will be also \(\mu\).

**Proposition 1.2** If the distribution of \(X_0\) is invariant, then \((x_n)\) is stationary. (Prove it!)
Proposition 1.3 Assume that $\mu$ is an invariant measure for the transition kernel $P(x;A)$. Let $f$ be an arbitrary bounded, measurable function, $f : \mathcal{X} \to \mathbb{R}$. Then

$$\int_{\mathcal{X}} (Pf - f)(x)\mu(dx) = 0.$$ 

Proof.

$$\int_{\mathcal{X}} Pf(x)\mu(dx) = \int_{\mathcal{X}} \mu(dx) \int_{\mathcal{X}} P(x,dy)f(y).$$

Interchanging the order of integration we get

$$\int_{\mathcal{X}} f(y) \int_{\mathcal{X}} \mu(dx)P(x,dy).$$

Now

$$\int_{\mathcal{X}} \mu(dx)P(x,dy) = \mu(dy)$$

due to the invariance of $\mu$, thus we get

$$\int_{\mathcal{X}} f(y)\mu(dy)$$

This step is a bit sloppy, but it can be made precise which proves the proposition. 

Remark 1.3 The definition of $Pf$ can be extended to unbounded measurable functions, assuming that for all $x \in \mathcal{X}$

$$\int_{\mathcal{X}} |f(y)|P(x,dy) < \infty,$$

or equivalently all $X_0 = x$

$$E_x|f(x_1)| < \infty.$$

Problem: Extend the Proposition 1.3 to unbounded functions.

The Poisson-equation. Assume that $\mu$ is an invariant measure on $(\mathcal{X},A)$ for the transition kernel $P(x;A)$. Let $k : \mathcal{X} \to \mathbb{R}$ be a bounded measurable function such that

$$\int_{\mathcal{X}} k(x)\mu(dx) = 0.$$
The Poisson equation: find a bounded measurable function \( f : \mathcal{X} \to \mathbb{R} \) such that
\[
Pf - f = k.
\]
The solution of a Poisson-equation is a kind of integration of \( k \) along the Markov-process.

**Remark 1.4** If \( f \) is a solution of the Poisson-equation, then for any constant \( c \) on \( \mathcal{X} \) \( f + c \) is also a solution.

**Remark 1.5** If the Poisson-equation does have a bounded, measurable solution \( f \) for all \( k \) satisfying
\[
\int_{\mathcal{X}} k(x) \mu(dx) = 0,
\]
then for any bounded measurable function \( k \) the modified Poisson-equation
\[
Pf - f = k - k_0,
\]
with
\[
k_0 = \int_{\mathcal{X}} k(x) \mu(dx)
\]
has a bounded measurable solution.

The usefulness of Poisson-equations is demonstrated by the following theorem.

**Proposition 1.4** Assume that for \( k : \mathcal{X} \to \mathbb{R} \) we have
\[
Pf - f = k,
\]
where \( k, f \) are both bounded and measurable functions. Then
\[
\frac{1}{N} \sum_{n=1}^{N} k(x_n)
\]
converges to 0 almost surely.

**Remark 1.6** Note that the existence of an invariant measure is not assumed!
Proof. Write
\[ \sum_{n=1}^{N} k(x_n) = \sum_{n=1}^{N} (Pf(x_n) - f(x_n)). \]
Couple the terms \( Pf(x_{n+1}) - f(x_n) \), and write the above sum as
\[ \sum_{n=1}^{N-1} (Pf(x_{n+1}) - f(x_n)) + Pf(x_1) - f(x_N). \]
Since the terms in the sum are bounded martingale differences, and \( f, Pf \) are bounded, the proposition follows from the law of large numbers for martingales, see [5]. ■

Extension of the definition of \( Pf \) to unbounded \( f \):

**Proposition 1.5** Assume that the initial distribution of \( X \) is \( \mu \) and
\[ E_{\mu}|f(X_1)| < +\infty. \]
Then
\[ Pf = E[f(X_1)|X_0 = x] \]
is well defined \( \mu \)-a.s.

Proof. Let \( f = f^+ - f^- \), where \( f^+ , f^- \) are the positive and negative parts of \( f \), respectively. Then
\[
\int_{\mathcal{X}} Pf^+ \mu(dx) = \int_{\mathcal{X}} \mu(dx) \int_{\mathcal{X}} f^+(x_1) P(x, dx_1) =
\]
\[ = \int_{\mathcal{X}} f^+(x_1) \int_{\mathcal{X}} \mu(dx) P(x, dx_1) < +\infty \]
by the assumption. Thus \( Pf^+ \) is finite \( \mu \) a.s. Similarly for \( f^- \), thus the claim follows.

We can rephrase the above as follows: if \( f \in L_1(\mu) \) then also \( Pf \in L_1(\mu) \).

**Proposition 1.6** Let \( p \geq 1 \), and assume that \( f \in L_p(\mu) \). Then also \( Pf \in L_p(\mu) \), and
\[ ||Pf||_p \leq ||f||_p, \]
where \( || \cdot ||_p \) denotes the \( L_p(\mu) \)-norm.
Proof. Let $f \geq 0$. Then by convexity
$$\left( \int_{\mathcal{X}} P(x, dy) f(y) \right)^p \leq \int_{\mathcal{X}} P(x, dy) f^p(y).$$
Integrating with respect to $\mu(dx)$ we get the claim as above. ■

Extension of Proposition 1.4:

Proposition 1.7 Let $p > 1$ and let $f \in L_p(\mu)$. Then
$$\frac{1}{N} \sum_{n=1}^{N} k(x_n)$$
converges to $0$ $\mu \times P$ almost surely. Prove it!

2 Parametric Markov processes

Let $\theta \in D \subset \mathbb{R}^d$, where $D$ is an open domain. A parametric family of Markov-transition kernels is a function
$$P_\theta(x; A),$$
where for each $\theta, x \ P(\cdot; A)$ is a probability measure in $(\theta, x)$, and for each $A \in \mathcal{A} \ P(\cdot; A)$ is measurable in $(\theta, x)$.

Definition 2.1 $(X_n)$ is a Markov-process with time-varying dynamics $(\theta_n)$, if for all $x \in \mathcal{X}$, $A \in \mathcal{A}$
$$P(X_{n+1} \in A | X_n = x) = P_{\theta_n}(x, A),$$
and here $\theta_n$ is $\mathcal{F}_n$-measurable, where
$$\mathcal{F}_n = \sigma\{X_0, \ldots, X_n\}.$$

Alternative definition: A parametric family of Markov processes is given by a random mapping $T : D \times \mathcal{X} \to \mathcal{X}$ defined by
$$(\theta, x) \to f_\theta(x, U)$$
where $U$ is an $\mathcal{U}$-valued random variable and $f_\theta(x, u)$ is measurable in $(\theta, x, u)$. If $\theta$ is fixed then the mapping $x \to f_\theta(x, U)$ will be denoted by $T_\theta$.  

8
Remark 2.1 By Theorem 1.1 given above if $\mathcal{X}$ is a Polish space, then $P_\theta(x, A)$ can be realized by a random mapping.

Consider local averaging for Markov processes with time-varying dynamics. Let $f(\theta, x)$, $\theta \in D$, $x \in \mathcal{X}$ be a non-negative measurable function.

Proposition 2.1 We have

$$E[f(\theta_n, X_{n+1})|\mathcal{F}_n] = \int_{\mathcal{X}} f(\theta_n, y) P_{\theta_n}(X_n, dy).$$

Hint: The conditional distribution of $(\theta_n, X_{n+1})$ given $X_n$ is $P_\theta(X_n, dy)$.

3 The basic algorithm

Let $H = H(\theta, x)$ be an $\mathbb{R}^d$-valued function defined on $D \times \mathcal{X}$. We will use a dual interpretation of $H$: for each fixed $x$ it will be considered a vector field on $D$, and for each fixed $\theta$ it will be considered as a function over $\mathcal{X}$.

Assume for a moment that for any $\theta \in D$ the Markov chain with transition kernel $P_\theta(x, A)$ has a unique invariant probability measure $\mu_\theta$ and define

$$h(\theta) = \int_{\mathcal{X}} H(\theta, x) \mu_\theta(dx).$$

Then our objective is to solve the equation

$$h(\theta) = 0.$$

The basic algorithm: Let $\theta_0 = a$, $X_0 = x$ be the initial conditions and define recursively

$$\begin{align*}
\theta_{n+1} &= \theta_n + \gamma_{n+1} H(\theta_n, X_{n+1}), \\
X_{n+1} &= f(\theta_n, X_n, U_{n+1})
\end{align*}$$

with step sizes $\gamma_n$, satisfying the following condition.

Condition A.1.

$$\gamma_n > 0, \quad \gamma_n \searrow 0 \quad \text{and} \quad \sum \gamma_n = \infty.$$

We assume the existence of a norm $|x|$ in $\mathcal{X}$, with the only property $|x| \geq 0$. The analysis of this algorithm is the objective of Part II. of BMP.
Condition A.5.-0. We assume that the parametric Markov-process \((X_n)\) with transition kernel \(P_\theta(x, A)\) is \(L_q\)-stable for any \(q \geq 1\) uniformly in \(\theta\) in the following sense: for any \(q \geq 1\), any compact set \(Q \subset D\), for all \(\theta \in Q\), \(x \in \mathcal{X}\) we have for all \(n \geq 1\)

\[
E_\theta[|X_n|^q |X_0 = x] \leq C_q(1 + |x|^q)
\]

where \(C_q\) also depends on \(Q\).

This assumption is implicit in Condition A.5. of BMP. The advantage of this condition is that it is defined purely in terms of the Markov process and not in terms of the basic algorithm.

Condition A.3. There exists a \(q' \geq 1\) such that in any compact set \(Q \subset D\) we have for all \(\theta \in Q\)

\[
|H(\theta, x)| \leq C(1 + |x|^q),
\]

where \(C\) depends on \(Q\).

Combining this with Condition A.5.-0, it follows that if the initial distribution of \(X_0\) is \(\mu\) and for some \(q > 1\)

\[
\int_\mathcal{X} |x|^{qq'} \mu(dx) < +\infty
\]

then

\[
\sup_n E|H(\theta, X_n)|^{qq'} < +\infty,
\]

uniformly in \(\theta\) for \(\theta \in Q\).

Remark 3.1 It follows that if \(X_0 = x\) a constant then, replacing \(qq'\) by \(q\), for any \(q \geq 1\) we have

\[
\sup_n E|H(\theta, X_n)|^q < +\infty,
\]

uniformly in \(\theta\) for \(\theta \in Q\).

In the literature we also find conditions for the existence of exponential moments requiring that there exists an \(\varepsilon > 0\) such that

\[
\sup_n E \exp\left[\varepsilon \frac{\partial}{\partial \theta} H(\theta, X_n)\right] < +\infty
\]

uniformly in \(\theta\) for \(\theta \in Q\). The dual role of these two kinds of conditions will be explained later.
Condition A.4. (i). It will be assumed that for all $\theta \in D$ there exists a constant $h(\theta)$ such that the Poisson-equation

$$(I - \Pi_\theta) \nu(x) = H(\theta, x) - h(\theta) \tag{P1}$$

has a solution. This solution will be denoted by $\nu_\theta = \nu_\theta(x)$.

Condition A.4. (iii)/a. There exists a $q' \geq 1$ such that for all compact domains $Q \subset D$ we have for $\theta \in Q$

$$|\nu_\theta(x)| \leq C(1 + |x|^{q'})$$

where $C$ also depends on $Q$.

The latter condition, combined with the condition that

$$E_\theta \left(|X_1|^{q'} \mid X_0 = x\right) < +\infty$$

for any $x \in \mathcal{X}$ implies that $\Pi_\theta \nu_\theta$ is well defined.

In view of the above conditions and Proposition 1.4 it follows that for any fixed $\theta \in D$ and $X_0 = x$ we have

$$\lim_{n} \frac{1}{N} \sum_{n=1}^{N} H(\theta, X_n) = h(\theta)$$

with probability 1 or $Q - a.s.$

Remark 3.2 The same conclusion holds $Q \times d\mu - a.s.$ if $X_0$ has an initial distribution $\mu$ such that

$$\int_{\mathcal{X}} |x|^{q} \mu(dx) < +\infty.$$

An important feature of the basic algorithm is that $H$ may be discontinuous in $\theta$. But we require that "averaging $H$" over $\mathcal{X}$, yielding $h$, will smooth out eventual discontinuities.

Condition A.4. (ii). $h$ is locally Lipschitz continuous in $D$.

A similar, but less direct condition is imposed on the local average of $H$. Apply $\Pi_\theta$ to the Poisson equation and write

$$w_\theta(x) = \Pi_\theta \nu_\theta(x)$$

to get

$$(I - \Pi_\theta) w_\theta(x) = \Pi_\theta H_\theta(x) - h(\theta) \tag{P2}$$
where $H_\theta$ denotes the partial function $H(\theta, x)$ with $\theta$ fixed.

**Condition A.4. (iii)/b.** There exists a $q' \geq 1$ and a $\lambda$ with $1/2 \leq \lambda \leq 1$ such that for all compact $Q \subset D$, $\theta, \theta' \in Q$ and any $x \in X$ we have

$$|w_\theta(x) - w_\theta'(x)| \leq C|\theta - \theta'|^\lambda(1 + |x|^{-q'})$$

where $C$ depends on $Q$.

**Remark 3.3** Connection between $\nu_\theta$ and $w_\theta$. If $\nu_\theta(x)$ is a solution of the Poisson equation

$$(I - \Pi_\theta) \nu(x) = H(\theta, x) - h(\theta)$$

then rearranging this we get

$$\nu_\theta(x) = \Pi_\theta \nu_\theta(x) + H_\theta(x) - h(\theta) = w_\theta(x) + H_\theta(x) - h(\theta).$$

Conversely, if $w_\theta(x)$ is a solution of the Poisson-equation (P2), then

$$\nu_\theta(x) = w_\theta(x) + H_\theta(x) - h(\theta) \quad (P1')$$

is a solution of the Poisson equation (P1). Indeed, it is easy to see that $\Pi_\theta \nu_\theta$ is well defined. Now

$$(I - \Pi_\theta) \nu(x) = \left(\Pi_\theta H_\theta(x) - h(\theta)\right) + (H_\theta(x) - \Pi_\theta H_\theta(x)) = H_\theta(x) - h(\theta).$$

Condition A.4. (iii)/a and Condition A.5.-0. imply

$$w_\theta(x) \leq C(1 + |x|^{-q'})$$

for $\theta \in Q, x \in X$. Conversely, (2) implies A.4. (iii)/a.

**Example 3.1** We give an example for the smoothing effect of $\Pi_\theta$: let $d = 1$, $X = \mathbb{R}$ and

$$H(\theta, x) = \text{sgn}(\theta - x).$$

Let $\Pi$ be independent of $\theta$, then

$$\Pi H_\theta(x) = \int \text{sgn}(\theta - x) \, dP(x) = \int \text{sgn} \, dP(x - \theta).$$

If $dP(x - \theta) = \varphi(x - \theta)dx$ with a smooth (infinitely differentiable) kernel $\varphi$ then $\Pi H_\theta(x)$ will be smooth with respect to $\theta$.  

12
Remark 3.4 The verification of Condition A.4. is the subject of Chapter 2 of Part II. of BMP.

The stopped process. We enforce the boundedness of the process \((\theta_n)\) by introducing

\[
\tau(Q) = \min\{n : \theta_n \not\in Q\}
\]

and restricting our analysis to the time-domain \(0 \leq n \leq \tau(Q)\). We impose the following “boundedness condition”, which is a purely technical condition, the verification of which will require substantial extra work.

Condition A.5. For all \(q > 0\) and \(Q \subset D\) compact, \(\theta \in Q\) and all \(n\) we have

\[
E_{x,a} \{I(n < \tau)(1 + |X_{n+1}|^q)\} \leq C_q(1 + |x|^q)
\]

where \(C_q\) depends also on \(Q\).

What is this condition about? It extends A.5.-0 significantly by requiring that \((X_n)\) is \(L_q\)-stable in a sense even if it is driven by a time-varying dynamics determined by the sequence \((\theta_n)\)!

Example 3.2 Let \(\mathcal{X} = \mathbb{R}^k\), \(d = 1\) and consider the Markov-process

\[
X_{n+1} = A(\theta)X_n + U_n
\]

where \(U_n\) is i.i.d., and for all \(q \geq 1\)

\[
E|U_n|^q < +\infty.
\]

Assume that \(A(\theta)\) is stable for all \(\theta \in D\), i.e. all its eigenvalues are strictly less than 1 in absolute value. Then Condition A.5.-0 is satisfied. However, it does not follow that for any sequence \((\theta_n) \in Q \subset D\) the process

\[
X_{n+1} = A(\theta_n)X_n + U_n
\]

will be \(L_q\)-bounded for any \(X_0 = x\). (Hint: Find two matrices \(A_1, A_2\) such that they are both stable but \(A_1A_2\) is not.)

4 The ODE-method

Define the associated ODE as

\[
\dot{y} = h(y) \quad y_0 = a.
\]

(ODE)

We expect that under suitable conditions \(\theta_n\) will “track” the solution trajectory of (ODE). The definition and quantification of the tracking error
is a basic step in classical ODE-method. In BMP a different route is taken: the
starting point is a Lyapunov-function for the ODE and this is then used to
analyse the convergence of our basic algorithm.

Let $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$ be a $C^2$ function such that
$$M_2 := \sup_{\theta \in \mathbb{R}^d} ||\Phi''(\theta)|| < +\infty.$$ Write our algorithm as
$$\theta_{n+1} = \theta_n + \gamma_{n+1} h(\theta_n) + \gamma_{n+1}(H(\theta_n, X_{n+1}) - h(\theta_n))$$
and consider the evolution of $\Phi(\theta_n)$. Define
$$\varepsilon_n = \varepsilon_n(\Phi) = \Phi(\theta_{n+1}) - \Phi(\theta_n) - \gamma_{n+1}\Phi'(\theta_n)h(\theta_n).$$ Let
$$R(\theta, \theta') = \Phi(\theta') - \Phi(\theta) - (\theta' - \theta)\Phi'(\theta)$$
where $\Phi'(\theta)$ is the gradient of $\Phi$. Then
$$|R(\theta, \theta')| \leq \frac{1}{2}M_2|\theta - \theta'|^2.$$ \textbf{Remark 4.1} We need this global condition to control cases when $\theta$ exits $Q$. Rewriting $\varepsilon_k$: subtract and add $(\theta_{k+1} - \theta_k)\Phi'(\theta_k)$. Thus we get
$$\varepsilon_k = R(\theta_{k+1}, \theta_k) + \Phi'(\theta_k)(\theta_{k+1} - \theta_k - \gamma_{k+1} h(\theta_k)).$$ \textbf{Remark 4.2} Note that the ”standard” local tracking error $\theta_{k+1} - \theta_k - \gamma_k h(\theta_k)$ shows up here. Thus any result along the lines of classical ODE method can also be used for a Lyapunov-function analysis!

Taking into account the recursion defining $\theta_{k+1}$ and taking out $\gamma_{k+1}$ we get
$$\varepsilon_k = R(\theta_{k+1}, \theta_k) + \gamma_{k+1}\Phi'(\theta_k)(H(\theta_k, X_{k+1}) - h(\theta_k)).$$ We would like to estimate the partial sum of $\varepsilon_k$ for reasons that will become clear later in the Lyapunov-analysis. The critical objects are the second terms. For $\theta_k = \theta = \text{const.}$, and $\gamma_{k+1} = \gamma = \text{const.}$ we could simply use the Poisson-equation to deal with partial sums of these terms. The variation of $\theta_k$ and $\gamma_{k+1}$ requires additional devices. First, using Condition A.3. (ii), write, assuming $\theta \in Q$
$$H(\theta_k, X_{k+1}) - h(\theta_k) = \nu_{\theta_k}(X_{k+1}) - \Pi_{\theta_k}\nu_{\theta_k}(X_{k+1}).$$
Then generate a martingale difference on the right hand side in the usual way by writing it as
\[
(\nu_{\theta_k}(X_{k+1}) - (\Pi_{\theta_k} \nu_{\theta_k})(X_k)) + ((\Pi_{\theta_k} \nu_{\theta_k})(X_k) - (\Pi_{\theta_k} \nu_{\theta_k})(X_{k+1})).
\]
Taking into account the condition \(|\nu_{\theta_k}(x)| \leq C(1 + |x|^{d'})\) and Condition A.5, we get that for any \(q \geq 1\)
\[
I(k < \tau) \nu_{\theta_k}(X_{k+1})
\]
is \(L_q\)-bounded uniformly in \(k\), and thus
\[
\Phi'(\theta_k) I(k < \tau) (\nu_{\theta_k}(X_{k+1}) - (\Pi_{\theta_k} \nu_{\theta_k})(X_k))
\]
is uniformly \(L_q\)-bounded martingale difference. Define
\[
\epsilon^k = \gamma_{k+1} \Phi'(\theta_k) I(k < \tau) [\nu_{\theta_k}(X_{k+1}) - (\Pi_{\theta_k} \nu_{\theta_k})(X_k)].
\]

**Time-varying telescopic sums.** Let
\[
\Psi_{\theta}(x) = \Phi'(\theta) \Pi_{\theta} \nu_{\theta}(x)
\]
and consider
\[
\sum_{k=0}^{n-1} \gamma_{k+1} (\Psi_{\theta_k}(X_k) - \Psi_{\theta_k}(X_{k+1})).
\]
Using a variant of Abel-summation (which is a discrete version of partial integration) we get
\[
\gamma_1 \Psi_{\theta_0}(X_0) - \gamma_n \Psi_{\theta_{n-1}}(X_n) + \sum_{k=1}^{n-1} \gamma_{k+1} (\Psi_{\theta_k}(X_k) - \Psi_{\theta_{k-1}}(X_k)) + \sum_{k=1}^{n-1} (\gamma_{k+1} - \gamma_k) \Psi_{\theta_{k-1}}(X_k).
\]
Note that in the first sum the \(\Psi\)-s are evaluated of the same \(X_k\).

**Remark 4.3** For \(\theta_k = \theta = \text{const.} \) we would get the usual Abel summation.

A comparison lemma. The lemma below has been used for ODE analysis of stochastic approximation processes in [3]. Consider the ordinary differential equation
\[
\dot{y}_t = F(t,y_t), \quad y_s = \xi, \ s \geq 1.
\]
The solution of the above ODE will be denoted by \(y(t,s,\xi)\) in the time interval where it exists and is unique.

**Condition C.** \(F = (F(t,y))\) is defined for \(t \geq 1, y \in D\) where \(D \subset \mathbb{R}^p\) is an open domain, and \(F\) is continuously differentiable in \((t,y)\). It is assumed that there exists a compact domain \(D_0' \subset D\) such that for all \(\xi \in D_0'\) the solution \(y(t,s,\xi) \in D\) is defined for all \(1 \leq s \leq t < \infty\).
Lemma 4.1 Assume that Condition C is satisfied. Let \((x_t), 1 \leq s \leq t < \infty\) be a continuous, piecewise continuously differentiable curve such that \(x_t \in D'_0\) for \(t \geq s\) and \(x_s = y_s = \xi \in D'_0\). Then for \(t \geq s\)

\[
x_t - y_t = \int_s^t \frac{\partial}{\partial \xi} y(t, r, x_r) \left( \dot{x}_r - F(r, x_r) \right) \, dr.
\]  

\[\text{(4)}\]

Proof. We shall use subscripts to indicate partial derivatives. Write \(z_r = y(t, r, x_r)\). Obviously the left hand side of (4) can be written as \(z_t - z_s\) and we have

\[
z_t - z_s = \int_s^t z'_r \, dr = \int_s^t \left( y_r(t, r, x_r) + y_\xi(t, r, x_r) \dot{x}_r \right) \, dr.
\]  

\[\text{(5)}\]

Taking into account the equality \(y_r(t, r, x_r) = -y_\xi(t, r, x_r) \cdot F(t, x_r)\) we get the lemma.

The associated ODE. We have written our basic algorithm as

\[
\theta_{n+1} = \theta_n + \gamma_{n+1} h(\theta_n) + \epsilon_n.
\]  

\[\text{(6)}\]

A natural way to approximate this stochastic algorithm is to use the deterministic recursion

\[
z_{n+1} = z_n + \gamma_{n+1} h(z_n) \quad z_0 = a.
\]  

\[\text{(7)}\]

This recursion is the Euler-approximation (with step sizes \(\gamma_n\)) of the ordinary differential equation

\[
\dot{\bar{\theta}}(t) = h(\bar{\theta}(t)) \quad \bar{\theta}(0) = a.
\]  

\[\text{(8)}\]

The general solution with \(\bar{\theta}(s) = a\) is \(\bar{\theta}(t; s, a)\). Let \(T > 0\) be fixed. Define

\[
m(T) = \inf \{k \geq 0 : \gamma_1 + \ldots + \gamma_{k+1} \geq T \}
\]

and for \(n = 1, \ldots, m(T)\) define

\[
t_n = \sum_{k=1}^n \gamma_k.
\]

Let \(Q_0, Q'_0\) be compact sets such that \(Q_0 \subset \text{int} \ Q'_0\) and let

\[
\delta_0 = \inf \{|\theta - \eta| : \theta \in Q_0, \eta \in Q'_0\}.
\]
**Condition A.6.** There exist compact sets $Q_0 \subseteq Q'_0 \subseteq D$ such that 

$$\overline{\theta}(t; 0, a) \in Q'_0 \quad \text{for all} \quad a \in Q_0, \ 0 \leq t \leq T.$$ 

We compare $\theta_n$ and $\overline{\theta}(t_n)$ for $n \leq \nu$ where 

$$\nu = m(T) \wedge \tau(Q'_0).$$

For $\overline{\theta}(t_n)$ we have the recursion 

$$\overline{\theta}(t_{n+1}) - \overline{\theta}(t_n) = \gamma_{n+1} h(\overline{\theta}(t_n)) + \alpha_n,$$

where, using the fact that $h$ is Lipschitz continuous, we get 

$$|\alpha_n| \leq L \gamma_{n+1}^2.$$

(Prove it!) Alternatively, using the notation $\overline{\theta}_n = \overline{\theta}(t_n)$ we get 

$$\overline{\theta}_{n+1} = \overline{\theta}_n + \gamma_{n+1} h(\overline{\theta}_n) + \alpha_n. \quad (9)$$

Subtracting (9) from (6) and iterating the result we are led to 

$$\theta_n - \overline{\theta}_n = \sum_{k=0}^{n-1} \gamma_{k+1}(h(\theta_k) - h(\overline{\theta}_k)) + \sum_{k=0}^{n-1} (\varepsilon_k - \alpha_k).$$

Using again the Lipschitz condition on $h$ and multiplying both sides by $I(n \leq \nu)$ we get 

$$I(n \leq \nu)|\theta_n - \overline{\theta}_n| \leq L \sum_{k=0}^{n-1} \gamma_{k+1} I(n \leq \nu)|\theta_k - \overline{\theta}_k| +$$

$$+ |\sum_{k=0}^{n-1} I(n \leq \nu)\varepsilon_k| + L \sum_{k=0}^{n-1} \gamma_{k+1}^2.$$ 

Now we can use the discrete version of the Bellman-Gronwall lemma.

**Theorem 4.1** (Discrete Bellman-Gronwall lemma.) Let $(f_i, g_i)_{i=1}^n$ and $c$ be non-negative real numbers, $f_0 = 0$ and $c \geq 0$ such that for $k = 1 \ldots n$

$$f_k \leq \sum_{i=1}^{k} g_i f_{i-1} + c.$$ 

Then $f_n \leq f^*_n$ where $f^*_k$ is defined by 

$$f^*_k = \sum_{i=1}^{k} g_i f^*_{i-1} + c. \quad (10)$$

It follows that $f_n \leq c \exp(\sum_{i=2}^{n} g_i).$
Proof. Prove by induction that $f_n \leq f_n^*$. The proof of $f_n^* \leq c \exp\left(\sum_{i=2}^{n} g_i\right)$: the dynamic form of equality (10) is

$$f_k^* = f_{k-1}^* + g_k f_{k-1}^* \quad \text{with} \quad f_1^* = c,$$

from which

$$f_n^* = c \prod_{i=2}^{n} (1 + g_i).$$

The proof is completed using the inequality $1 + x \leq e^x$. ■

Applying the Bellman-Gronwall lemma with $f_k = I(n \leq \nu) |\theta_k - \bar{\theta}_k|$ we get

$$I(n \leq \nu) |\theta_n - \bar{\theta}_n| \leq (U_{1n} + U_{2n}) \exp(L \sum_{k=1}^{m(T)} \gamma_k)$$

where

$$U_{1n} = \sup_{1 \leq m \leq n} |\sum_{k=0}^{m-1} I(\nu \geq k) \varepsilon_k|, \quad U_{2n} = L \sum_{k=0}^{m(T)-1} \gamma_{k+1}^2.$$ 

Take supremum over $n$ for $n \leq m(T)$ of both sides. From here

$$E \left\{ \sup_{n \leq \nu} I(n \leq \nu) |\theta_n - \bar{\theta}_n| \leq (2E(U_{1n}^2) + 2U_{2n}^2) \exp(2L \sum_{k=1}^{n} \gamma_k) \right\}.$$ 

Remark 4.4 Since at time $n = \tau(Q_0)$ we have $|\theta_n - \bar{\theta}(t_n)| \geq \delta_0$, we get that for any $0 < \delta < \delta_0$

$$\{ \sup_{n \leq \nu} |\theta_n - \bar{\theta}(t_n)| \geq \delta \} = \{ \sup_{n \leq m(T)} |\theta_n - \bar{\theta}(t_n)| \geq \delta \}.$$ 

Now using the upper bound obtained for $E(U_{1n}^2)$ (see BMP [1], Part II., Proposition 7), the Cauchy-Schwartz inequality for $U_{2n}^2$ and applying the Markov-inequality, we get the following theorem.

Theorem 4.2 Assume Conditions A.1.-A.5. are satisfied and $Q_0, Q, T$ and $\delta_0$ are as above. Then there exist constants $C, L, s$ such that for $0 < \delta < \delta_0$ and all $a \in Q, \ x \in X$ we have

$$P_{x,a} \left( \sup_{n \leq m(T)} |\theta_n - \bar{\theta}_n| > \delta \right) \leq \frac{C}{\delta^2} (1 + |x|^s)(1 + T) \exp(2LT) \sum_{k=1}^{m(T)} \gamma_k^2.$$

18
Shift in time. Let \( P_{n;x,a} \) denote the conditioned distribution of the process \((X_{n+k}, \theta_{n+k})\) with initial values \(X_n = x, \theta_n = a\). Let us now start the process at time \(n\). Let \( \bar{\theta}(t; s, a) \) denote the solution of ODE (8) with initial value \( \bar{\theta}(s) = a \) and define

\[
m(n, T) = \inf\{k \geq n : \gamma_{n+1} + \ldots + \gamma_{k+1} \geq T\}.
\]

Then

\[
P_{n;x,a} \left( \sup_{n \leq k \leq m(n, T)} |\theta_k - \bar{\theta}(t_k; t_n, a)| > \delta \right) \leq C \delta^2 (1 + |x|^s) (1 + T) \exp(2LT) \sum_{k=n}^{\infty} \gamma_k^2.
\]

I.e. if \( \sum \gamma_k^2 < \infty \) then the trajectories of the stochastic algorithm initialized at time \(n\) follow the deterministic trajectories of the ODE with probability increasing to 1.

Asymptotic analysis. Assume that there exists a unique \( \theta^* \in D \) such that \( h(\theta^*) = 0 \), and that this \( \theta^* \) is a globally asymptotically stable equilibrium point. Then Krasovskii [8] implies the following.

**Condition A.7.** There exist a twice continuously differentiable positive function \( U : D \to \mathbb{R}_+ \) and a constant \( C \leq +\infty \) such that

1. \( U(\theta) \to C \) if \( \theta \to \partial D \) or \( |\theta| \to \infty \);
2. \( U(\theta) < C \) for \( \theta \in D \);
3. \( U'(\theta) \cdot h(\theta) \leq 0 \) for all \( \theta \in D \).

Thus \( D \) is a level set of \( U \). Generally, for \( c \in \mathbb{R} \) define the level sets

\[
K(c) = \{ \theta ; U(\theta) \leq c \}.
\]

Note that due to Condition A.7. (i) the level set \( K(c) \) is a compact set whenever \( c < C \). To analyze the algorithm we will use a level-crossing argument. Define the stopping times

\[
\tau(c) = \inf\{n \geq 1 ; \theta_n \notin K(c)\}
\]
\[
\nu(c) = \inf\{n \geq 1 ; \theta_n \in K(c)\}.
\]

**Lemma 4.2** Assume Conditions A.1. - A.7. hold and let \( c_0 \) be such that \( \theta^* \in K(c_0) \). Then for any \( c_0 < c_1 < c_2 < C, \forall a \in K(c_2), \forall x \in \mathcal{X} \)

\[
\nu(c_1) < \infty \quad P_{x,a} \text{ a.s. on } \{\tau(c_2) = +\infty\},
\]

i.e. if \( \theta_n \) does not exit \( K(c_2) \) at all then it enters \( K(c_1) \) almost surely.
Proof. Let $\Phi$ be a twice continuously differentiable function on $\mathbb{R}^d$ with bounded second derivatives which coincides with $U$ on $K(c_2)$ (prove that there exists such a function!). Define the transformed error-process as

$$
\varepsilon_k(\Phi) = \Phi(\theta_{k+1}) - \Phi(\theta_k) - \gamma_{k+1} \Phi'(\theta_k) h(\theta_k).
$$

We have to show that the set

$$
\Omega_0 = \{\nu(c_1) = \tau(c_2) = +\infty\}
$$

has measure 0. Notice that $\tau(c_2) = +\infty$ implies $\tau = +\infty$. Choose $\alpha > 0$ such that $U'(\theta) h(\theta) < -\alpha$ for all $\theta$ satisfying $c_1 \leq U(\theta) \leq c_2$. Then on $\Omega_0$ we have

$$
c_1 - c_2 \leq \Phi(\theta_{m(n,T)}) - \Phi(\theta_n) = \sum_{k=n}^{m(n,T)-1} \gamma_{k+1} U'(\theta_k) h(\theta_k) + \sum_{k=n}^{m(n,T)-1} \varepsilon_k(\Phi).
$$

Since the first term is majorated by $-\alpha T$ we conclude that the series $\sum_k \varepsilon_k(\Phi)$ diverges. This, however, contradicts our earlier result stating that on $\{\tau = +\infty\}$ the former series converges almost surely. ■

Theorem 4.3 Assume Conditions A.1. - A.7. hold and let $c_0$ be such that $\theta^* \in K(c_0)$. Then for any $0 < c < c_2 < C$, $\forall a \in K(c), \forall x \in X$

$$
\theta_n \rightarrow K(c_0) \text{ P}_{x,a} \text{ a.s. on } \{\tau(c_2) = +\infty\}.
$$

Proof. Suppose indirectly that $\limsup \ U(\theta_n) > c$ for some $c > c_0$. Choose $c_0 < c_1 < c$ and define

$$
\nu_1 = \inf\{n; \theta_n \in K(c_1)\} \quad \tau_1 = \inf\{n; \theta_n \notin K(c)\}
$$

$$
\nu_k = \inf\{n > \tau_{k-1}; \theta_n \in K(c_1)\} \quad \tau_k = \inf\{n > \nu_k; \theta_n \notin K(c)\}.
$$

Due to the previous lemma $\nu_k < \infty$ for any $k \geq 1$ while $\tau_k < \infty$ by our indirect assumption. Now choose $\Phi \in C^2(\mathbb{R}^d)$ in such a way that it coincides with $U$ on $K(c_2)$ and $\inf\{\Phi(\theta); \theta \notin K(c_2)\} = c$. Noting that

$$
\Phi(\theta_\nu_n) - \Phi(\theta_\nu_n) \geq c - c_1 > 0
$$

and using the same argument as in Lemma 4.2 we yield a contradiction. ■

To see that our theorem is indeed useful, we need to know something about the measure of the set $\{\tau(c) = +\infty\}$. Using once again the train of thought of Lemma 4.2 a lower bound follows easily from the bound given for the $L_2$ norms of the transformed error process.
Lemma 4.3 Assume Conditions A.1 - A.7 hold and let $c_1 < c_2 < C$. Then there exist constants $B$ and $s$ such that for all $\forall a \in K(c_1)$, $\forall x \in X$

$$P_{x,a} \{ \tau(c_2) < +\infty \} \leq B(1 + |x|^s) \sum_{k=1}^{\infty} \gamma_k^{1+\lambda}.$$  

Putting together Lemma 4.3 and Theorem 4.3 and using a shift in time, we are led to the following result.

Theorem 4.4 Assume Conditions A.1 - A.7 hold and $c_0 \in \mathbb{R}$ is such that $\{ \theta \mid U'(\theta) h(\theta) = 0 \} \subseteq K(c_0)$. Then for any compact set $Q \subseteq D$ there exist constants $B$ and $s$ such that for all $n \geq 0$, $\forall a \in Q$, $\forall x \in X$

$$P_{n,x,a} \{ \theta_k \rightarrow K(c_0) \} \geq 1 - B(1 + |x|^s) \sum_{k=n+1}^{\infty} \gamma_k^{1+\lambda}.$$  

Relaxing Condition A.5.

Condition A’5. For all $q \geq 1$, and $Q \subset D$ compact we have

(i)’ $\sup_{\theta \in Q} \int \Pi_{\theta}(x, dy)|y|^q \leq C|x|^q + \beta$;

In addition there exists a positive integer $r$ such that

(i) $\sup_{\theta \in Q} \int \Pi_{\theta}^r(x, dy)|y|^q \leq \bar{\alpha}|x|^q + \beta$;

with $\bar{\alpha} < 1$.

Example 4.1 Consider a linear dynamics

$$X_{n+1} = A(\theta_n)X_n + U_{n+1}$$

such that $A(\theta)$ is stable for $\theta \in D$ and continuous in $\theta$, furthermore for all $q \geq 1$

$$E|U_n|^q < +\infty.$$  

Then Condition A’5. is satisfied. (Prove it!)

Lemma 4.4 For all positive integers $l$ there exists a constant $C_l$ such that for all $n \geq 1$

$$E(|X_{n+l}|^q | F_n) \leq C_l(1 + |x|^q)$$
Proof. For \( l = 1 \) the proposition follows directly from Condition A’5. For general \( l \)

\[
E(|X_{n+l}|^q | \mathcal{F}_{n+l-1}) \leq C_1(1 + |X_{n+l-1}|^q).
\]

Taking conditional expectation with respect to \( \mathcal{F}_n \) and using an inductive hypothesis for \( l - 1 \) we get the claim. ■

Let now \( \nu \) be a stopping time such that for \( \nu \geq n+1 \) we have \( \theta_n \in Q \). Then we have

**Lemma 4.5** For all positive integer \( l \) there exists a constant \( C_l \) such that for all \( n \geq 1 \)

\[
E[I\{\nu \geq n+l\}|X_{n+l}|^q | \mathcal{F}_n] \leq C_l I\{\nu \geq n+1\}(1 + |X_n|^q).
\]

Proof. The key difference is at \( l = 1 \). Now we have

\[
E[I\{\nu \geq n+1\}|X_{n+1}|^q | \mathcal{F}_n] = I\{\nu \geq n+1\}E[|X_{n+1}|^q | \mathcal{F}_n]
\]

since \( I\{\nu \geq n+1\} \) is \( \mathcal{F}_n \)-measurable. Now

\[
E[|X_{n+1}|^q | \mathcal{F}_n] = \int_X |y|^q P_{\theta_n}(X_n, dy).
\]

Multiplying by \( I\{\nu \geq n+l\} \) and taking supremum for \( \theta \in Q \) we get

\[
I\{\nu \geq n+1\}E[|X_{n+1}|^q | \mathcal{F}_n] \leq I\{\nu \geq n+1\}\sup_{\theta \in Q} \int_X |y|^q P_{\theta}(X_n, dy)
\]

which is bounded by

\[
I\{\nu \geq n+1\}C_1(1 + |X_n|^q),
\]

as stated. For \( l > 1 \) we proceed by induction. We have

\[
E[I\{\nu \geq n+l\}|X_{n+l}|^q | \mathcal{F}_{n+l-1}] \leq I\{\nu \geq n+l\}C_1(1 + |X_{n+l-1}|^q)
\]

by the already proven inequality for \( l = 1 \). Increase the right hand side by

\[
E[I\{\nu \geq n+l\}|X_{n+l}|^q | \mathcal{F}_{n+l-1}] \leq I\{\nu \geq n+l\}C_1(1 + |X_{n+l-1}|^q)
\]

Replacing \( I\{\nu \geq n+l\} \) by \( I\{\nu \geq n+l-1\} \), take conditional expectation with respect to \( \mathcal{F}_n \) and apply the inductive hypothesis to get an upper bound of the required form. ■
5 \(L^q\) estimates

In this section we investigate the problem discussed in Section 4 for the case \(\sum_n \gamma_n^\alpha < \infty\), where \(\alpha\) can be different from 2.

Our first goal is to investigate

\[
E\{ \sup_{n<k<m(n,T)} I(k \leq \nu(\epsilon, Q)|\sum_{i=n}^{k-1} \epsilon_i(\phi)|^q)\},
\]

where we set for \(\epsilon > 0\)

\[
\tau(Q) = \inf\{k > n : \theta_k \notin Q\}
\]

\[
\sigma(\epsilon) = \inf\{k \geq n : |\theta_k - \theta_{k-1}| > \epsilon\}
\]

\[
\nu(\epsilon, Q) = \inf\{\tau(Q), \sigma(\epsilon)\}.
\]

**Theorem 5.1** We assume conditions (A1)-(A4) and (A5'). For any regular function \(\phi\) with second derivatives, any compact subset \(Q \subset D\) and for all \(q \geq 2\) there exit constants \(B, s, \epsilon_0 > 0\) (\(\epsilon_0\) is independent of \(\phi\)) such that for all \(\epsilon \leq \epsilon_0, T > 0, x, a\) we have

\[
E\{ \sup_{n<k<m(n,T)} I(k \leq \nu(\epsilon, Q)|\sum_{i=n}^{k-1} \epsilon_i(\phi)|^q)\} \leq B(1 + T)^{q-1}(1 + |x|^s) \sum_{i=n+1}^{m(n,T)} \gamma_i^{1+q/2}
\]

(12)

The proof is similar to the proof of Proposition 7 in Part II. of BMP [1]. Note that on the right hand side we have the sum of the \((1 + q/2)\)-th power of the \(\gamma_i\) weights instead of the sum of the second powers. For the proof we will use the Burkholder inequality (see Hall and Heyde [5]) and a Hölder type inequality:

**Lemma 5.1** Let \((U_n : n \geq 0)\) be a sequence of real-valued random variables defined on a filtered probability space \((\Omega, \mathcal{F}, P)\), satisfying for some \(q \geq 2, E|U_n|^q < \infty, \) and \(E(U_{n+1}|\mathcal{F}_n) = 0.\) Then

\[
E(\sup_{n \leq k \leq m} |\sum_{i=n}^{k} U_i|^q) \leq C_q E(\sum_{i=n}^{m} U_i^2)^{q/2}.
\]

(13)

**Lemma 5.2** Let \(a_i \geq 0, b_i \in \mathbb{R}, u > 1, 0 < \delta < 1,\) then

\[
|\sum_{i=n}^{m} a_i b_i|^u \leq (\sum_{i=n}^{m} a_i^{\delta u/(u-1)})^{u-1}(\sum_{i=n}^{m} a_i^{(1-\delta)u}) |b_i|^u
\]

(14)
The following result gives an approximation for the distance from the mean trajectory.

Consider the compact sets \(Q_1, Q'_1, Q_2\) such that \(Q_1 \subset Q'_1 \subseteq \text{int } Q_2 \subset D\) and \(\bar{\theta}(t, 0, a) \in Q'_1\) for all \(a \in Q_1\).

**Lemma 5.3**

\[
\{ \sup_{n \leq m(T)} |\theta_n - \bar{\theta}(t_n)| \geq \delta \} \subset \{ \sup_{n \leq \nu \wedge m(T)} |\theta_n - \bar{\theta}(t_n)| \geq \delta \} \bigcup \{ \sigma \leq \tau, \sigma < m(T) \}
\]

**Proof.** We have

\[
\{ \sup_{n \leq m(T)} |\theta_n - \bar{\theta}(t_n)| \geq \delta \} \subset \{ \sup_{n \leq m(T)} |\theta_n - \bar{\theta}(t_n)| \geq \delta, m(T) \leq \nu \} \bigcup \{ \nu < m(T) \}.
\]

Using the following relationships

\[
\{ \nu < m(T) \} \subset \{ \sigma \leq \tau, \sigma < m(T) \} \bigcup \{ \tau \leq \sigma, \tau < m(T) \},
\]

we have the statement of the proposition.  

**Theorem 5.2** We assume the conditions (A1)-(A4) and (A5'). Suppose that \(Q_1 \subset Q'_1 \subseteq \text{int } Q_2\) are compact subsets of \(D\), and \(q \geq 2\). Then there exits constants \(B, s, L\), such that for all \(T > 0\), all \(\delta < \delta_0\), and all \(a \in Q_1\), all \(x\)

\[
P_{x,a}\{ \sup_{n \leq m(T)} |\theta_n - \bar{\theta}(t_n)| \geq \delta \} \leq \frac{B}{\delta^q} (1 + |x|^s)(1 + T)^{q-1} \exp(qLT) \sum_{k=1}^{m(T)} \gamma_k^{1+q/2}
\]

(15)

The proof of the theorem is based on Lemma 5.3 and the following lemmas.

**Lemma 5.4** For any compact subset \(Q\) of \(D\) and \(q \geq 2\), there exist constants \(A, s\) such that for all \(T > 0\), all \(\epsilon \leq \epsilon_0(q, Q)\), all \(a \in Q\), all \(x\), we have

\[
P_{x,a}(\sigma(\epsilon) \leq \tau(Q), \sigma(\epsilon) \leq m(T)) \leq \frac{A}{\epsilon^q} (1 + |x|^s) \sum_{k=1}^{m(T)} \gamma_k^q
\]

(16)
Lemma 5.5  Given compact subsets $Q_1 \subset Q_2$ of $D$ and $q \geq 2$, there exist constants $A, s$ such that for all $T$ and all $\epsilon \leq \epsilon_0(q, Q_2)$, all $a \in Q_1$, all $x$

$$E_{x,a}\{ \sup_{n \leq \nu \wedge m(T)} \left| \theta_n - \overline{\theta}(t_n)|^q \right| \} \leq A(1 + |x|^s)(1 + T)^q - 1 \exp(qLT) \sum_{k=1}^{m(T)} \gamma_k^{1+q/2}$$  \hspace{1cm} (17)

In the following we analyze the asymptotic properties of the algorithm under the following assumption:

(A6') There exists $\alpha > 1$ such that $\sum_n \gamma_n^\alpha < \infty$.

We retain assumption (A7) and the notation $K(c) = \{ \theta : U(\theta) \leq c \}$, 
$\tau(c) = \inf\{n : \theta \notin K(c)\}$, $\nu(c) = \inf\{n : \theta_n \in K(c)\}$. Let

$$q_0(\alpha) = \max\{2, 2(1 - \alpha)\},$$

i.e. if $q \geq q_0(\alpha)$, then we have both $q \geq 2$ and $1 + q/2 \geq \alpha$.

We shall suppose further that there exists a compact subset $F$ of $D$ satisfying

$$F = \{ \theta : U(\theta) \leq c_0 \} \supset \{ \theta : U'(\theta)h(\theta) = 0 \}$$  \hspace{1cm} (18)

Theorem 5.3 We assume (A1)-(A4), (A5'), (A6') and (A7), and suppose that $F$ is a compact set satisfying (18). Then for any compact subset $Q$ of $D$, and $q \geq q_0(\alpha)$, there exist constants $B, s$ such that for all $a \in Q$ and all $x$

$$P_{x,a}(\sigma_{n \leq \nu \wedge m(T)} |x|^s) \sum_{k=1}^{m(T)} \gamma_k^{1+q/2} \geq 1 - B(1 + |x|^s) \sum_{k=1}^{m(T)} \gamma_k^{1+q/2}$$  \hspace{1cm} (19)

The proof consists two steps. The first step is to prove that the set $\{ \sigma(\epsilon_0) = \infty, \tau(c_2) = \infty \}$ is large, the second step is to prove that on this set the convergence is satisfied. These steps are expressed in the following statements.

Proposition 5.1 Given $c_1, c_2, q$ satisfying $c_0 < c_1 < c_2 < C, q \geq q_0(\alpha)$, there exist $\epsilon_0, B, s$ such that for all $a \in K(c_1)$, all $x$

$$P_{x,a}(\sigma(\epsilon_0) = \infty, \tau(c_2) = \infty) \geq 1 - B(1 + |x|^s) \sum_{k=1}^{m(T)} \gamma_k^{1+q/2}$$  \hspace{1cm} (20)
Consider the complement of \( \{ \sigma(\epsilon_0) = \infty, \tau(c_2) = \infty \} \)
\[
\{ \tau(c_2) < \infty, \sigma(\epsilon_0) \geq \tau(c_2) \} \cup \{ \sigma(\epsilon_0) < \infty, \tau(c_2) \geq \sigma(\epsilon_0) \}.
\]

Using Theorem 5.1 with \( \epsilon_0 = \epsilon_0(q, K(c_2)) \) and the following lemma we get the first statement

**Lemma 5.6** Given \( c_1, c_2, q \) such that \( c_0 < c_1 < c_2 < C, q \geq q_0(\alpha) \), there exist \( \epsilon_0, B, s \) such that for \( \epsilon \leq \epsilon_0 \), \( a \in K(c_1) \),
\[
P_{x,a} \{ \tau(c_2) < \infty, \sigma(\epsilon) \geq \tau(c_2) \} \leq B(1 + |x|^s) \sum_{k \geq 1} \gamma_k^{1+q/2} \tag{21}
\]

The convergence of the algorithm towards the compact set \( F \) is expressed in the second step.

**Proposition 5.2** Suppose \( c \) and \( \epsilon \) satisfy \( c_0 < c < C \) and \( \epsilon \leq \epsilon_0(q, K(c)) \) for some \( q \geq q_0(\alpha) \). Then for all \( x \) and all \( a \) in the interior of \( K(c) \), \( \theta_n \) converges \( P_{x,a} \) a.s. towards \( F \) on \( \{ \tau(c) = \infty, \sigma(\epsilon) = \infty \} \).

At the end of the section we shall combine Theorem 5.2 and 5.3 to give a detailed description of the convergence of \( \theta_n \) when \( D \) is the domain of attraction of a point \( \theta^* \). We replace \((A7)\) by \((A7')\):

\((A7')\) There exists a positive function \( U \in C^2 \) on \( D \), such that \( U(\theta) \to C \leq \infty \) if \( \theta \to \partial D \) or \( |\theta| \to \infty \) and \( U(\theta) < C \) if \( \theta \in D \) which satisfies

1. \( U(\theta^*) = 0, U(\theta) > 0, \theta \in D, \theta \neq \theta^* \)
2. \( U'(\theta)h(\theta) < 0 \) for all \( \theta \in D, \theta \neq \theta^* \)

Clearly \( F = \{ \theta^* \} \) satisfy (18) for \( c_0 = 0 \).

**Theorem 5.4** We assume \((A1)-(A4)\) and \((A5')-(A7')\). Then for any compact subset \( Q \) of \( D \), for all \( q \geq q_0(\alpha) \), all \( \delta > 0 \), there exist constants \( B, s \), such that for all \( n \geq 0 \), all \( a \in Q \), all \( x \)
\[
P_{n,x,a} (\sup_{k \geq n} |\theta_k - \bar{\theta}(t_k, t_n, a)| > \delta) \leq B(1 + |x|^s) \sum_{k \geq n} \gamma_k^{1+q/2} \tag{22}
\]
References


6 Appendix

A summary of assumptions

We made the following assumptions:

(A.1.) $(\gamma_n)_{n \geq 1}$ is a decreasing sequence of positive real numbers such that $\sum_n \gamma_n = +\infty$.

(A.2.) There exists a family $\{\Pi_\theta : \theta \in \mathbb{R}^d\}$ of transition probabilities $\Pi_\theta(x, A)$ on $\mathcal{X}$ such that for any measurable subset $A$ of $\mathcal{X}$ we have

$$P[X_{n+1} \in A \mid \mathcal{F}_n] = \Pi_{\theta_n}(X_n, A).$$

(A.3.) There exists a $q' \geq 1$ such that in any compact set $Q \subset D$ we have for all $\theta \in Q$

$$|H(\theta, x)| \leq C(1 + |x|^q),$$

where $C$ depends on $Q$.

(A.4.) (i) For all $\theta \in D$ there exists a constant $h(\theta)$ such that the Poisson-equation

$$(I - \Pi_\theta) \nu(x) = H(\theta, x) - h(\theta)$$

has a solution. Denote the solution by $\nu_\theta = \nu_\theta(x)$.

(ii) $h$ is locally Lipschitz-continuous on $D$.

(iii)/a There exists a $q' \geq 1$ such that for all compact domains $Q \subset D$ we have for $\theta \in Q$

$$|\nu_\theta(x)| \leq C(1 + |x|^{q'})$$

where $C$ also depends on $Q$.

(iii)/b There exists a $q' \geq 1$ and a $\lambda$ with $1/2 \leq \lambda \leq 1$ such that for all compact $Q \subset D$, $\theta$, $\theta' \in Q$ and any $x \in \mathcal{X}$ we have for $w_\theta = \Pi_\theta \nu_\theta$

$$|w_\theta(x) - w_{\theta'}(x)| \leq C|\theta - \theta'|^\lambda(1 + |x|^{q'})$$

where $C$ depends on $Q$. 

28
(A.5.-0.) The Markov-process \((X_n)\) is \(L_q\)-stable for any \(q \geq 1\) uniformly in \(\theta\): for any \(q \geq 1\), any compact set \(Q \subset D\), for all \(\theta \in Q\), \(x \in \mathcal{X}\) we have for all \(n \geq 1\)

\[
E_\theta[|X_n|^q | X_0 = x] \leq C_q (1 + |x|^q)
\]

where \(C_q\) also depends on \(Q\).

(A.5.) For all \(q > 0\) and \(Q \subset D\) compact, \(\theta \in Q\) and all \(n\) we have

\[
E_{x,a} \{I(n < \tau(Q))(1 + |X_{n+1}|^q)\} \leq C_q (1 + |x|^q)
\]

where \(C_q\) depends also on \(Q\).

(A.6.) There exist compact sets \(Q_0 \subseteq \text{int } Q'_0 \subseteq D\) such that

\[
\bar{\theta}(t; 0, a) \in Q'_0 \quad \text{for all } a \in Q_0, \ 0 \leq t \leq T
\]

where \(\bar{\theta}(t; 0, a)\) denotes the general solution of the associated ODE

\[
\dot{\bar{\theta}}(t) = h(\bar{\theta}(t)) \quad \bar{\theta}(0) = a.
\]

(A.7.) There exist a twice continuously differentiable positive function \(U : D \mapsto \mathbb{R}_+\) and a constant \(C \leq +\infty\) such that

(i) \(U(\theta) \to C\) if \(\theta \to \partial D\) or \(|\theta| \to \infty\);

(ii) \(U(\theta) < C\) for \(\theta \in D\);

(iii) \(U'(\theta) \cdot h(\theta) \leq 0\) for all \(\theta \in D\).