Truncated Stochastic Approximation with Moving Bounds: Convergence

Teo Sharia

Department of Mathematics
Royal Holloway, University of London
Egham, Surrey TW20 0EX
e-mail: t.sharia@rhul.ac.uk

Abstract

In this paper we propose a wide class of truncated stochastic approximation procedures with moving random bounds. While we believe that the proposed class of procedures will find its way to a wider range of applications, the main motivation is to accommodate applications to parametric statistical estimation theory. Our class of stochastic approximation procedures has three main characteristics: truncations with random moving bounds, a matrix valued random step-size sequence, and dynamically changing random regression function. We establish convergence and consider several examples to illustrate the results.

Keywords: Stochastic approximation, Recursive estimation, Parameter estimation

1 Introduction

Stochastic approximation (SA) introduced by Robbins and Monro in 1951(see [16]) was created to locate a root of an unknown function when only noisy measurements of the the function can be observed. SA quickly became very popular resulting in interesting new developments and numerous applications across a wide range of disciplines. Comprehensive surveys of the SA technique including some recent developments can be found in [2], [3],[12], [13], [14].

In this paper we propose a wide class of truncated SA procedures with moving random bounds. While we believe that the proposed class of procedures will find its way to a wider range of applications, the main motivation is to accommodate applications to parametric statistical estimation theory. Our class of SA procedures has three main characteristics: truncations with random moving bounds, a matrix valued random step-size sequence, and dynamically changing random regression function. To introduce the main ideas, let us first consider the classical problem of finding a unique zero, say z^0, of a real valued function $R(z) : \mathbb{R} \to \mathbb{R}$ when only noisy measurements of R are available. Consider a recursive procedure defined as

$$Z_t = \Big[\ Z_{t-1} + \gamma_t \left(R(Z_{t-1}) + \varepsilon_t \right) \Big]_{\alpha_t}^{\beta_t}, \qquad t = 1, 2, \dots \tag{1.1}$$

where ε_t is a sequence of zero mean random variables and γ_t is a deterministic sequence of positive numbers. Here α_t and β_t are random variables with $-\infty \leq \alpha_t \leq \beta_t \leq \infty$ and $[v]_{\alpha_t}^{\beta_t}$ is the truncation operator, that is,

$$[v]_{\alpha_t}^{\beta_t} = \begin{cases} \alpha_t & \text{if } \ v < \alpha_t, \\ v & \text{if } \ \alpha_t \leq v \leq \beta_t, \\ \beta_t & \text{if } \ v > \beta_t. \end{cases}$$

We assume that the truncation sequence $[\alpha_t, \beta_t]$ contains z^0 for large t-s. For example, if it is known that z^0 belongs to (α, β), with $-\infty \leq \alpha \leq \beta \leq \infty$, one can consider truncations with expanding bounds to avoid possible singularities at the endpoints of the interval. That is, we can take $[\alpha_t, \beta_t]$ with some sequences $\alpha_t \downarrow \alpha$ and $\beta_t \uparrow \beta$. Truncations with expanding bounds may also be useful to overcome standard restrictions on growth of the corresponding functions. The most interesting case arises when the truncation interval $[\alpha_t, \beta_t]$ represents our auxiliary knowledge about z^0 at step t, which is incorporated into the procedure through the truncation operator. Consider for example a parametric statistical model. Suppose that $X_1, \dots, X_t$ are the i.i.d. r.v.'s. and $f(x, \theta)$ is the common probability density (pdf) function (w.r.t. some σ-finite measure) depending on an unknown parameter $\theta \in \mathbb{R}^m$. Consider the recursive estimation procedure for θ defined by

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{t} i(\hat{\theta}_{t-1})^{-1} \, \frac{f'^T(X_t, \hat{\theta}_{t-1})}{f(X_t, \hat{\theta}_{t-1})}, \qquad t \geq 1. \tag{1.2}$$

where f' is the row-vector of partial derivatives of f w.r.t. the components of θ, $i(\theta)$ is the one-step Fisher information matrix, and $\hat{\theta}_0 \in \mathbb{R}^m$ is some initial value. This estimator was introduced in [18] and studied in [9] and [11]. In particular, it has been shown that under certain conditions the recursive estimator $\hat{\theta}_t$ is asymptotically equivalent to the maximum likelihood estimator, i.e., it is consistent and asymptotically efficient. The analysis of (1.2) can be conducted by rewriting it in the form of stochastic approximation. Indeed, in the case of (1.2), let us fix θ and denote

$$R(z) = i(z)^{-1} E^{\theta} \left\{ \frac{f'^T(X_t, z)}{f(X_t, z)} \right\} \quad \text{and} \quad \varepsilon_t = i(\hat{\theta}_{t-1})^{-1} \left(\frac{f'^T(X_t, \hat{\theta}_{t-1})}{f(X_t, \hat{\theta}_{t-1})} - R(\hat{\theta}_{t-1}) \right).$$

Then, under the usual regularity assumptions, $R(\theta) = 0$, and ε_t is a martingale difference (w.r.t. the filtration $\mathcal{F}_t$ generated by the observations). So, (1.2) is a standard SA of type (1.1) without truncations (i.e., in the one dimensional case, $-\alpha_t = \beta_t = \infty$). However, the need of truncations may naturally arise from various reasons. One obvious consideration is that the functions in the procedure may only be defined for certain values of the parameter. In this case one would want the procedure to produce points only from this set. Truncations may also be useful when the standard assumptions such as restrictions on the growth rate of the relevant functions are not satisfied. More importantly, truncations may provide a simple tool to achieve an efficient use of information available in the estimation process. This information can be auxiliary information about the parameters, e.g. a set, possibly time dependent, that is known to contain the value of the unknown parameter. Suppose for instance that a consistent (i.e., convergent), but not necessarily efficient auxiliary estimator $\tilde{\theta}_t$ is available having a rate d_t. Then one can consider truncated procedure with shrinking bounds. The idea is to truncate the recursive procedure in a neighbourhood of θ by taking $[\alpha_t, \beta_t] = [\tilde{\theta}_t - \delta_t, \tilde{\theta}_t + \delta_t]$ with $\delta_t \to 0$. Such a procedure is obviously consistent since $\hat{\theta}_t \in [\tilde{\theta}_t - \delta_t, \tilde{\theta}_t + \delta_t]$ and $\tilde{\theta}_t \pm \delta_t \to \theta$. However, to construct an efficient estimator, care should be taken to ensure that the truncation intervals do not shrink to θ too rapidly, for otherwise $\hat{\theta}_t$ will have the same asymptotic properties as $\tilde{\theta}_t$ (see [24] for details in the case of AR processes).

Note that the idea of truncations with moving bounds is not new. For example, idea of truncations with shrinking bounds goes back to [9] and [11]. Truncations with expanding bounds in the context of recursive parametric estimation were considered in [19] (see also [24]). Truncations with adaptive truncation sets of the Robbins-Monro SA were introduced in [4], [5]. However, motivation as well as the actual truncation procedures for the latter method are very different from the ones considered in our paper. Truncations with adaptive truncation sets were further explored and extended in [1], [26], [27].

Let us now consider a general time series model given by a sequence $X_1, \dots, X_t$ of r.v.'s with the joint distribution depending on an unknown parameter $\theta \in \mathbb{R}^m$. Then one can consider the recursive estimator of θ defined by

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \gamma_t(\hat{\theta}_{t-1})\psi_t(\hat{\theta}_{t-1}), \qquad t \geq 1, \tag{1.3}$$

where $\psi_t(v) = \psi_t(X_1, \dots, X_t; v)$, $t = 1, 2, \dots$, are suitably chosen functions which may, in general, depend on the vector of all past and present observations and have the property that the process $\psi_t(\theta)$ is P^θ- martingale difference, i.e., $E^\theta\left\{\psi_t(\theta) \mid \mathcal{F}_{t-1}\right\} = 0$ for each t. For example, if $f_t(x, \theta) = f_t(x, \theta | X_1, \dots, X_{t-1})$ is the conditional probability density function (pdf) of the observation X_t given $X_1, \dots, X_{t-1}$, then one can obtain a likelihood type estimation prcesedure by choosing $\psi_t(v) = l_t(v) = f'_t(X_t, v)/f_t(X_t, v)$. Asymptotic behaviour of this type of procedures for non i.i.d. models was studied by a number of authors, see e.g., [8], [6], [15], [20] – [23].

In particular, the results in [23] show that to obtain an estimator with asymptotically optimal properties, one has to consider a state dependent matrix valued random step-size sequence. One possible choice is $\gamma_t(u)$ with the property

$$\triangle\gamma_t^{-1}(v) = \gamma_t^{-1}(v) - \gamma_{t-1}^{-1}(v) = E_\theta\{\psi_t(v) l_t^T(v) \mid \mathcal{F}_{t-1}\}$$

In particular, to obtain a recursive procedure which is asymptotically equivalent to the the maximum likelihood estimator, one has to take $l_t(v) = f_t'(X_t, v)/f_t(X_t, v)$ and $\gamma_t(v) = I_t(v)$, where $I_t(v)$ is the conditional Fisher information matrix (see [23] for details). To rewrite (1.3) in a SA form, let us assume that θ is an arbitrary but fixed value of the parameter and define

$$R_t(z) = E^\theta \left\{\psi_t(X_t, z) \mid \mathcal{F}_{t-1}\right\} \quad \text{and} \quad \varepsilon_t = \left(\psi_t(X_t, \hat\theta_{t-1}) - R_t(\hat\theta_{t-1})\right).$$

Obviously, $R_t(\theta) = 0$ for each t, and ε_t is a martingale difference.

Therefore, to be able to study these procedures in a unified manner, one needs to consider SA of the following form

$$Z_t = \Big[\ Z_{t-1} + \gamma_t(Z_{t-1})\big\{R_t(Z_{t-1}) + \varepsilon_t(Z_{t-1})\big\}\Big]_{U_t}, \quad t = 1, 2, \ldots$$

where $R_t(z)$ is predictable with the property that $R_t(z^0) = 0$ for all t's, $\gamma_t(z)$ is a matrix valued predictable step-size sequence, $U_t \subset \mathbb{R}^m$ is random sequence of truncation sets, and $Z_0 \in \mathbb{R}^m$ is some starting value (see Section 2 for more details).

The paper is organised as follows. In sections 2.2 we prove two lemmas on the convergence of the proposed SA procedures under very general conditions. The analysis is based on the well-known method of using convergence sets of nonnegative semimartingales. The decomposition into negative and positive parts in this lemmas turns out to be very useful in applications (see Example 3 in Section 2.4). In section 2.3 we give several corollaries in the case of state independent scalar random step-size sequence. In section 2.4 we consider examples. Proofs of some technical results are postponed to Section 4.

2 Convergence

2.1 Main objects and notation

Let $(\Omega, \mathcal{F}, F = (\mathcal{F}_t)_{t\geq 0}, P)$ be a stochastic basis satisfying the usual conditions. Suppose that

$$\{R_t(z) = R_t(z, \omega),\ z \in \mathbb{R}^m\} \quad \text{and} \quad \{\gamma_t(z) = \gamma_t(z, \omega),\ z \in \mathbb{R}^m\},$$

$t = 1, 2, \ldots,$ are vactor and matrix valued measurable random fields respectively such that for each $z \in \mathbb{R}^m$ the corresponding processes are predictable, i.e., $R_t(z)$ and $\gamma_t(z)$ are $\mathcal{F}_{t-1}$ measurable for each t. Suppose also that

$$\{\varepsilon_t(z) = \varepsilon_t(z, \omega),\ z \in \mathbb{R}^m\},$$

$t = 1, 2, \dots$, is a vector valued measurable random field such that for each $z \in \mathbb{R}^m$ the process $\varepsilon_t(z)$ is a martingale difference, i.e., $E\{\varepsilon_t(z) \mid \mathcal{F}_{t-1}\} = 0$. We also assume that

$$R_t(z^0) = 0$$

for each $t = 1, \dots$, where $z^o \in \mathbb{R}^m$ is a non-random vector. Suppose that $h = h(z)$ is a real valued function of $z \in \mathbb{R}^m$. We denote by $h'(z)$ the row-vector of partial derivatives of h with respect to the components of z, that is,

$$h'(z) = \left(\frac{\partial}{\partial z_1} h(z), \dots, \frac{\partial}{\partial z_m} h(z) \right).$$

Also, we denote by $h''(z)$ the matrix of second partial derivatives. The $m \times m$ identity matrix is denoted by $\mathbf{1}$.

Let $U \subset \mathbb{R}^m$ and define a truncation operator as a function $\big[z\big]_U : \mathbb{R}^m \longrightarrow \mathbb{R}^m$, such that

$$\big[z\big]_U = \begin{cases} z & \text{if } z \in U \\ z^* & \text{if } z \notin U, \end{cases}$$

where z^* is a point in the closure of U, that minimizes the distance to z (z^* is unique if U is convex).

Suppose that $z^o \in \mathbb{R}^m$. We say that a random sequence of sets $U_t = U_t(\omega)$ ($t = 1, 2, \dots$) from $\mathbb{R}^m$ is *admissible* for z^o if

- for each t and almost all ω, $U_t(\omega)$ is a closed convex subset of $\mathbb{R}^m$;
- for each t and $z \in \mathbb{R}^m$, the truncation $\big[z\big]_{U_t}$ is $\mathcal{F}_t$ measurable;
- $z^o \in U_t$ eventually, i.e., for almost all ω there exist $t_0(\omega) < \infty$ such that $z^o \in U_t(\omega)$ whenever $t > t_0(\omega)$.

Assume that $Z_0 \in \mathbb{R}^m$ is some starting value and consider the procedure

$$Z_t = \big[\, Z_{t-1} + \gamma_t(Z_{t-1}) \Psi_t(Z_{t-1}) \big]_{U_t}, \quad t = 1, 2, \dots \tag{2.1}$$

where

$$\Psi_t(Z_{t-1}) = R_t(Z_{t-1}) + \varepsilon_t(Z_{t-1}) \tag{2.2}$$

and $U_t \subset \mathbb{R}^m$ is admissible for z^o.

Convention.

• *Everywhere in the present work Convergence and all relations between random variables are meant with probability one w.r.t. the measure P unless specified otherwise. A sequence of random variables $(\zeta_t)_{t\geq 1}$ has some property **eventually** if for every ω in a set Ω_0 of P probability 1, the realisation $\zeta_t(\omega)$ has this property for all t greater than some $t_0(\omega) < \infty$.*

• *We will also assume that the* $\inf_{z\in U} h(z)$ *of a real valued function $h(z)$ is* 1 *whenever $U = \emptyset$.*

2.2 Convergence Lemmas

Lemma 2.1 *Let Z_t be a process defined by (2.1) with an admissible for $z^0 \in \mathbb{R}^m$ truncation sequence U_t. Let*

$$\{V_t(u) = V_t(u, \omega),\ u \in \mathbb{R}^m,\ t = 1, 2, \dots\}$$

is a random field, such that for each $u \in \mathbb{R}^m$, the process $V_t(u)$ is predictable, and for each fixed ω and t, $V_t(u) : \mathbb{R}^m \longrightarrow \mathbb{R}$ is a real valued nonnegative function having continuous and bounded partial second derivatives. Denote

$$\Delta_t = Z_t - z^0$$

and suppose that the following conditions are satisfied.

(L)

$$V_t\Big(\Delta_t\Big) \le V_t\Big(\Delta_{t-1} + \gamma_t(Z_{t-1})\Psi_t(Z_{t-1})\Big)$$

eventually.

(S)

$$\sum_{t=1}^{\infty}(1 + V_t(\Delta_{t-1}))^{-1}\left[\mathcal{N}_t(\Delta_{t-1})\right]^+ < \infty, \qquad P\text{-a.s.} \tag{2.3}$$

where

$$\begin{aligned}\mathcal{N}_t(u) &= V_t'(u)\gamma_t(z^o + u)R_t(z^o + u)\\ &\quad + \frac{1}{2}\sup_v \|V_t''(v)\| E\left\{\|\gamma_t(z^o + u)\Psi_t(z^o + u)\|^2 \mid \mathcal{F}_{t-1}\right\}.\end{aligned}$$

Then $V_t(Z_t - z^o)$ converges (P-a.s.) to a finite limit for any initial value Z_0. Furthermore,

$$\sum_{t=1}^{\infty}\left[\mathcal{N}_t(\Delta_{t-1})\right]^- < \infty, \qquad P\text{-a.s.} \tag{2.4}$$

Proof. As always (see the convention in 2.1), convergence and all relations between random variables are meant with probability one w.r.t. the measure P unless specified otherwise.

From condition (L), using the Taylor expansion,

$$\begin{aligned}V_t(\Delta_t) &\le V_t(\Delta_{t-1}) + V_t'(\Delta_{t-1})\gamma_t(z^o + \Delta_{t-1})\Psi_t(z^o + \Delta_{t-1})\\ &+\frac{1}{2}\left[\gamma_t(z^o + \Delta_{t-1})\Psi_t(z^o + \Delta_{t-1})\right]^T V_t''(\tilde{\Delta}_{t-1})\gamma_t(z^o + \Delta_{t-1})\Psi_t(\theta + \Delta_{t-1}),\end{aligned}$$

where $\tilde{\Delta}_{t-1} \in \mathbb{R}^m$. Taking the conditional expectation w.r.t. $\mathcal{F}_{t-1}$ yields

$$E\left\{V_t(\Delta_t) \mid \mathcal{F}_{t-1}\right\} \le V_t(\Delta_{t-1}) + \mathcal{N}_t(\Delta_{t-1}).$$

Using the obvious decomposition $\mathcal{N}_t(\Delta_{t-1}) = [\mathcal{N}_t(\Delta_{t-1})]^+ - [\mathcal{N}_t(\Delta_{t-1})]^-$, the previous inequality can be rewritten as

$$E\{V_t(\Delta_t) \mid \mathcal{F}_{t-1}\} \leq V_t(\Delta_{t-1})(1+B_t) + B_t - [\mathcal{N}_t(\Delta_{t-1})]^-, \tag{2.5}$$

eventually, where

$$B_t = (1+V_t(\Delta_{t-1}))^{-1}\,[\mathcal{N}_t(\Delta_{t-1})]^+.$$

By (2.3),

$$\sum_{t=1}^{\infty} B_t < \infty. \tag{2.6}$$

According to the Robbins-Siegmund Lemma (see e.g., [17]) inequalities (2.5) and (2.6) imply that (2.4) holds and $V_t(\Delta_t)$ converges to some finite limit. $\diamondsuit$

Everywhere below, we assume that the $\inf_{u\in U} v(u)$ of a function $v(u)$ is 1 whenever $U = \emptyset$.

Lemma 2.2 *Suppose that $V_t(Z_t - z^o)$ converges (P-a.s.) to a finite limit for any initial value Z_0, where V_t is defined in Lemma 2.1, and* (2.4) *holds. Suppose also that for each $\varepsilon \in (0,1)$,*

$$\inf_{\substack{\|u\|\geq\varepsilon \\ z^0+u\in U_t}} V_t(u) > \delta > 0 \tag{2.7}$$

eventually, for some δ. Suppose also that

(C) *For each $\varepsilon \in (0,1)$,*

$$\sum_{t=1}^{\infty} \inf_u\,[\mathcal{N}_t(u)]^- = \infty, \qquad P\text{-a.s.}$$

where the infimum is taken over the set $\{u : \varepsilon \leq V_t(u) \leq 1/\varepsilon;\ z^0 + u \in U_{t-1}\}$.

Then $Z_t \to \theta$ (P-a.s.), for any initial value Z_0.

Proof. As always (see the convention in 2.1), convergence and all relations between random variables are meant with probability one w.r.t. the measure P unless specified otherwise. Suppose that $V_t(\Delta_t) \to r \geq 0$ and there exists a set A with $P(A) > 0$, such that $r > 0$ on A. Then there exists $\varepsilon > 0$ and (possibly random) t_0, such that if $t \geq t_0$, $\varepsilon \leq V_t(\Delta_{t-1}) \leq 1/\varepsilon$ on A. Note also that $z^o + \Delta_{t-1} = Z_{t-1} \in U_{t-1}$. By **(C)**, these would imply that

$$\sum_{s=t_0}^{\infty} [\mathcal{N}_s(\Delta_{s-1})]^- \geq \sum_{s=t_0}^{\infty} \inf_u\,[\mathcal{N}_s(u)]^- = \infty$$

on the set A with $P^\theta(A) > 0$, where the infimums are taken over the sets specified in condition (C). This contradicts (2.4). Hence, $r = 0$ and so, $V_t(\Delta_t) \to 0$. Now,

$\Delta_t \to 0$ follows from (2.7) by contradiction. Indeed, suppose that $\Delta_t \not\to 0$ on a set, say B of positive probability. Then, for any fixed ω from this set, there would exist a sequence $t_k \to \infty$ such that $\|\Delta_{t_k}\| \geq \varepsilon$ for some $\varepsilon > 0$, and (2.7) would imply that $V_{t_k}(\Delta_{t_k}) > \delta > 0$ for large k-s, which contradicts the P-a.s. convergence $V_t(\Delta_t) \to 0$. $\diamondsuit$

2.3 Sufficient conditions

Everywhere in this subsection we assume that γ_t is state independent (i.e., constant w.r.t. z) non-negative scalar predictable process.

Corollary 2.3 *Let Z_t be a process defined by* (2.1) *with an admissible for $z^0 \in \mathbb{R}^m$ truncation sequence U_t. Suppose also that γ_t is a non-negative predictable scalar process and*

(C1)

$$\sup_{z \in U_{t-1}} \frac{\left[2(z-z^o)^T R_t(z) + \gamma_t E\left\{\|\Psi_t(z)\|^2 \mid \mathcal{F}_{t-1}\right\}\right]^+}{1+\|z-z^o\|^2} \leq q_t \qquad (2.8)$$

eventually, where

$$\sum_{t=1}^{\infty} q_t \gamma_t < \infty, \qquad P\text{-a.s.}$$

Then $\|Z_t - z^0\|$ converges (P-a.s.) to a finite limit.

Proof. Let us show that the conditions of Lemma 2.1 are satisfied with $V_t(u) = u^T u = \|u\|^2$ and the step-size sequence $\gamma_t(z) = \gamma_t \mathbf{I}$. Since $z^0 \in U_t$ for large t-s, definition of the truncation (see 2.1) implies that

$$\|Z_t - z^0\| \leq \left\|Z_{t-1} + \gamma_t \Psi_t(Z_{t-1}) - z^0\right\|,$$

eventually. Therefore (L) holds. Then, $V_t'(u) = 2u^T$ and $V_t''(u) = 2\mathbf{I}$, and so, for the process $\mathcal{N}_t(u)$ in (2.3) we have

$$\mathcal{N}_t(u) = 2u^T \gamma_t R_t(z^o + u) + \gamma_t^2 E\left\{\|\Psi_t(z^o+u)\|^2 \mid \mathcal{F}_{t-1}\right\} \qquad (2.9)$$

and

$$\frac{[\mathcal{N}_t(\Delta_{t-1})]^+}{1+V_t(\Delta_{t-1})} = \gamma_t \, \frac{\left[2\Delta_{t-1}^T R_t(z^o+\Delta_{t-1}) + \gamma_t E\left\{\|\Psi_t(z^o+\Delta_{t-1})\|^2 \mid \mathcal{F}_{t-1}\right\}\right]^+}{1+\|\Delta_{t-1}\|^2}$$

Since $z^o + \Delta_{t-1} = Z_{t-1} \in U_{t-1}$ and Z_{t-1} is $\mathcal{F}_{t-1}$ measurable, (2.3) follows from conditions (C1). $\diamondsuit$

Corollary 2.4 *Suppose that the conditions of Corollary 2.3 hold and*

(C2) *For each* $\varepsilon \in (0,1)$,

$$\sum_{t=1}^{\infty} \inf_u \left[\mathcal{N}_t(u)\right]^- = \infty, \qquad P\text{-}a.s.$$

where

$$\mathcal{N}_t(u) = 2u^T\gamma_t R_t(z^o+u) + \gamma_t^2 E\left\{\|\Psi_t(z^o+u)\|^2 \mid \mathcal{F}_{t-1}\right\}$$

and the infimum is taken over the set $\{u : \varepsilon \le \|u\| \le 1/\varepsilon;\ z^0+u \in U_{t-1}\}$.

Then $Z_t \to \theta$ *(*P*-a.s.), for any initial value* Z_0.

Proof. Let us show that the conditions of Lemma 2.2 are satisfied with $V_t(u) = u^Tu = \|u\|^2$ and $\gamma_t(z) = \gamma_t\mathbf{I}$. It follows from the proof of Corollary 2.3 that all the conditions of Lemma 2.1 hold with $V_t(u) = u^Tu$ and so, $\|Z_t - z^0\|$ converges and (2.4) holds. Conditions (2.7) also trivially holds and (C) is a consequence of (C2). $\diamondsuit$

Corollary 2.5 *Suppose that*

(1)

$$(z-z^0)^T R_t(z) < 0$$

for any $z \in U_t$, *eventually;*

(2)

$$\sup_{z\in U_{t-1}} \frac{E\left\{\|\Psi_t(z)\|^2 \mid \mathcal{F}_{t-1}\right\}}{1+\|z-z^o\|^2} \le r_t$$

eventually, where

$$\sum_{s=1}^{\infty} r_t\gamma_t^2 < \infty, \qquad P\text{-}a.s., \qquad P\text{-}a.s.$$

Then $\|Z_t - z^0\|$ *converges (*P*-a.s.) to a finite limit.*

Proof. The proof is immediate from Corollary 2.3 with $q_t = r_t\gamma_t$. $\diamondsuit$

Corollary 2.6 *Suppose that the conditions of Corollary 2.5 are satisfied and*

(CC) *for each* $\varepsilon \in (0,1)$,

$$\inf_{\substack{\varepsilon\le\|z-z^o\|\le 1/\varepsilon\\ z\in U_{t-1}}} -(z-z^0)^T R_t(z) > \nu_t \tag{2.10}$$

eventually, where

$$\sum_{s=1}^{\infty} \nu_t\gamma_t = \infty, \qquad P\text{-}a.s.$$

Then Z_t converges (P-a.s.) to z^0.

Proof. Let us show that the conditions of Corollary 2.4 hold. Indeed, by (1) and (2) of Corollary 2.5, conditions of Corollary 2.3 trivially hold with $q_t = r_t \gamma_t$. Then, using the obvious inequality $[a]^- \geq -a$, we have

$$[\mathcal{N}_t(u)]^- \geq -2u^T \gamma_t R(z^o + u) - \gamma_t^2 E\left\{\|\Psi_t(z^o + u)\|^2 \mid \mathcal{F}_{t-1}\right\}.$$

It follows from condition (2) of Corollary 2.5 that if we take the supremum of the conditional expectation above over the set $\{u : \varepsilon \leq \|u\| \leq 1/\varepsilon;\ z^0 + u \in U_{t-1}\}$ we obtain

$$\sup \frac{E\left\{\|\Psi_t(z^o + u)\|^2 \mid \mathcal{F}_{t-1}\right\}}{1 + \|u\|^2}\,(1 + \|u\|^2) \leq r_t(1 + \|1/\varepsilon\|^2).$$

Now, using (2.10) we take infimum over the set $\{u : \varepsilon \leq \|u\| \leq 1/\varepsilon;\ z^0 + u \in U_{t-1}\}$ and obtain

$$\inf\,[\mathcal{N}_t(u)]^- \geq 2\gamma_t \nu_t - \gamma_t^2 r_t(1 + \|1/\varepsilon\|^2).$$

Condition (C) of Lemma 2.2 is now immediate from (CC) and conditions (2) of Corollary 2.5. $\diamondsuit$

The following simple result is concerned with the classical stochastic approximation procedure with $R_t(z) = R(z)$. For simplicity we assume that the errors are state independent.

Corollary 2.7 *Suppose that*

$$\Psi_t(z) = R(z) + \varepsilon_t,$$

where $R(z)$ is a deterministic function and ε_t is a martingale difference error term which does not depend on z. Suppose also that

(i)

$$(z - z^0)^T R(z) < 0$$

for any $z \in U_t$ eventually;

(ii)

$$\sup_{z \in U_{t-1}} \frac{\|R(z)\|^2}{1 + \|z - z^o\|^2} \leq r_t$$

eventually, where

$$\sum_{s=t}^{\infty} r_t \gamma_t^2 < \infty, \qquad P\text{-a.s.};$$

(iii)

$$\sum_{s=t}^{\infty} E\left\{\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right\}\,\gamma_t^2 < \infty, \qquad P\text{-a.s.}$$

Then $\|Z_t - z^0\|$ converges (P-a.s.) to a finite limit.

Furthermore, if condition (CC) holds with $R_t(z) = R(z)$, then Z_t converges (P-a.s.) to z^0.

Proof. Since $E\{\varepsilon_t \mid \mathcal{F}_{t-1}\} = 0$, we have

$$E\left\{\|\Psi_t(z)\|^2 \mid \mathcal{F}_{t-1}\right\} = \|R(z)\|^2 + E\left\{\|\varepsilon_t\|^2 \mid \mathcal{F}_{t-1}\right\}$$

and

$$\frac{E\left\{\|\Psi_t(z)\|^2 \mid \mathcal{F}_{t-1}\right\}}{1 + \|z - z^o\|^2} \le \frac{\|R(z)\|^2}{1 + \|z - z^o\|^2} + E\left\{\|\varepsilon_t\|^2 \mid \mathcal{F}_{t-1}\right\},$$

condition (ii) and (iii) imply that condition (2) in Corollary 2.5 holds. The result now follows from Corollaries 2.5 and 2.6. ♢

2.4 Examples

Example 1 Suppose that l is any odd integer and consider the function

$$R(z) = -(z - z^0)^l,$$

and truncation sequence $[-\alpha_t, \alpha_t]$, where $\alpha_t \to \infty$ is a sequence of positive numbers. Then, condition (i) of Corollary 2.7 trivially holds. Suppose that

$$\sum_{t=1}^{\infty} \alpha_{t-1}^{2l}\ \gamma_t^2 < \infty.$$

Then, provided that the measurement errors satisfy condition (iii) of Corollary 2.7 and $\sum_{t=1}^{\infty} \alpha_{t-1} = \infty$, the truncated procedure

$$Z_t = \Big[\ Z_{t-1} + \gamma_t \left(R(Z_{t-1}) + \varepsilon_t\right)\Big]_{-\alpha_t}^{\alpha_t}, \quad t = 1, 2, \ldots$$

converges a.s. to z^0.

Indeed, for large t's,

$$\sup_{z\in[-\alpha_{t-1},\alpha_{t-1}]} \frac{\|R(z)\|^2}{1 + \|z - z^o\|^2} \le \sup_{z\in[-\alpha_{t-1},\alpha_{t-1}]} (z - z^0)^{2l} \le 4^l \alpha_{t-1}^{2l}$$

which implies condition (iii) of Corollary 2.7. For example, if the degree of the polynomial is known to be l (or at most l), and $\gamma_t = 1/t$, then one can take $\alpha_t = Ct^{\frac{1}{2l}-\delta}$, where C and δ are some positive constants and $\delta < 1/2l$. One can also take a truncation sequence which is independent of l, e.g., $\alpha_t = C \log t$, where C is a positive constant.

Example 2 Let $X_1, X_2, \ldots,$ be i.i.d. Gamma$(\theta, 1)$, $\theta > 0$. The the common pdf is

$$f(x, \theta) = \frac{1}{\mathbf{\Gamma}(\theta)} x^{\theta-1} e^{-x}, \quad \theta > 0, \ \ x > 0,$$

where $\mathbf{\Gamma}(\theta)$ is the Gamma function. Then

$$\frac{f'(x,\theta)}{f(x,\theta)} = \log x - \underbrace{\frac{d}{d\theta}\log\mathbf{\Gamma}(\theta)}_{\log'\mathbf{\Gamma}(\theta)}, \qquad i(\theta) = \underbrace{\frac{d^2}{d\theta^2\theta}\log\mathbf{\Gamma}(\theta)}_{\log''\mathbf{\Gamma}(\theta)},$$

where $i(\theta)$ is the one-step Fisher information. Then a likelihood type recursive estimation procedure (see also (1.2)) can be defined as

$$\hat{\theta}_t = \left[\hat{\theta}_{t-1} + \frac{1}{t\ \log''\mathbf{\Gamma}(\hat{\theta}_{t-1})}\Big(\log X_t - \log'\mathbf{\Gamma}(\hat{\theta}_{t-1})\Big)\right]_{\alpha_t}^{\beta_t}, \qquad t = 1, 2, \ldots \tag{2.11}$$

where $\alpha_t \downarrow 0$ and $\beta_t \uparrow \infty$ are sequences of positive numbers.

Everywhere in this example, $\mathcal{F}_t$ is the sigma algebra generated by $X_1, \ldots, X_t$, P^θ is the family of corresponding measures and $\theta > 0$ is an arbitrary but fixed value of the parameter.

Let us rewrite (2.11) in the form of the stochastic approximation, i.e.,

$$\hat{\theta}_t = \left[\hat{\theta}_{t-1} + \frac{1}{t}\Big(R(\hat{\theta}_{t-1}) + \varepsilon_t\Big)\right]_{\alpha_t}^{\beta_t}, \qquad t = 1, 2, \ldots \tag{2.12}$$

where (see Section 3 for details)

$$R(u) = R^\theta(u) = \frac{1}{\log''\mathbf{\Gamma}(u)} E^\theta\{\ln X_t - \log'\mathbf{\Gamma}(u)\} = \frac{1}{\log''\mathbf{\Gamma}(u)}\left(\log'\mathbf{\Gamma}(\theta) - \log'\mathbf{\Gamma}(u)\right)$$

and

$$\varepsilon_t = \frac{1}{\log''\mathbf{\Gamma}(\hat{\theta}_{t-1})}\Big(\log X_t - \log'\mathbf{\Gamma}(\hat{\theta}_{t-1})\Big) - R(\hat{\theta}_{t-1}).$$

Since $\hat{\theta}_{t-1}$ is $\mathcal{F}_{t-1}$ - measurable, we have $E^\theta\{\varepsilon_t \mid \mathcal{F}_{t-1}\} = 0$. Obviously, $R(\theta) = 0$, and since $\log'\mathbf{\Gamma}$ is increasing (see, e.g., [28], 12.16), condition **(1)** of Corollary 2.5 holds with $z^0 = \theta$. Based on the well known properties of the logarithmic derivatives of the gamma function, it is not difficult to show that (see, Section 3) that if

$$\sum_{t=1}^{\infty}\frac{\alpha_{t-1}^2}{t} = \infty \qquad \text{and} \qquad \sum_{t=1}^{\infty}\frac{\log^2\alpha_{t-1} + \log^2\beta_{t-1}}{t^2} < \infty \tag{2.13}$$

then all the conditions of Corollary 2.5 and 2.6 hold and therefore, $\hat{\theta}_t$ is consistent, i.e.,

$$\hat{\theta}_t \to \theta \qquad \text{as} \qquad t \to \infty \quad (P^\theta\text{-a.s.})$$

For instance, the sequences

$$\alpha_t = C_1(\log\,(t+2))^{-\frac{1}{2}} \quad \text{and} \quad \beta_t = C_2(t+2)$$

with some positive constants C_1 and C_2, obviously satisfy (2.13).

Note also, that since $\theta \in (0, \infty)$, it may seem unnecessary to use the upper truncations $\beta_t < \infty$. However, without upper truncations (i.e. if $\beta_t = \infty$), the standard restriction on the growth does not hold. Also, with $\beta_t = \infty$ the procedure fails condition (2) of Corollary 2.5 (see (3.7)).

Example 3 Consider an AR(1) process

$$X_t = \theta X_{t-1} + \xi_t, \tag{2.14}$$

where ξ_t is a sequence of random variables (r.v.'s) with mean zero. Taking $\psi_t(z) = X_{t-1}\left(X_t - zX_{t-1}\right)$ and $\gamma_t = \hat{I}_t = \hat{I}_0 + \sum_{s=1}^{t} X_{t-1}^2$, procedure (1.3) reduces to the recursive least squares (LS) estimator of θ, i.e.,

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \hat{I}_t^{-1} X_{t-1}\left(X_t - \hat{\theta}_{t-1}X_{t-1}\right), \tag{2.15}$$
$$\hat{I}_t = \hat{I}_{t-1} + X_{t-1}^2, \qquad t = 1, 2, \ldots$$

where $\hat{\theta}_0$ and $\hat{I}_0 > 0$ are any starting points.

For simplicity let us assume that ξ_t is a sequence of i.i.d. r.v.'s with mean zero and variance 1. Consistency of (2.15) can be derived from our results for any $\theta \in \mathbb{R}$ and without any further moment assumptions on the innovation process ξ_t. Indeed, assume that θ is an arbitrary but fixed value of the parameter. Then, using (2.14), we obtain

$$X_t - \hat{\theta}_{t-1}X_{t-1} = \xi_t + X_{t-1}(\theta - \hat{\theta}_{t-1}).$$

and (2.15) can be rewritten as

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \hat{I}_t^{-1}\left(X_{t-1}^2(\theta - \hat{\theta}_{t-1}) + X_{t-1}\xi_t\right). \tag{2.16}$$

So, (2.16) is a SA procedure with

$$R_t(z) = X_{t-1}^2(\theta - z), \tag{2.17}$$

$\varepsilon_t = X_{t-1}\xi_t, \quad \gamma_t = \hat{I}_t^{-1}$ and $U_t = \mathbb{R}$. Let us check condition (C1) of Corrolary 2.3 with $z^0 = \theta$ and $U_t = \mathbb{R}$. Since

$$E\left\{\|\Psi_t(z)\|^2 \mid \mathcal{F}_{t-1}\right\} = \|R_t(z)\|^2 + E\left\{\|\varepsilon_t\|^2 \mid \mathcal{F}_{t-1}\right\} = X_{t-1}^4(\theta - z)^2 + X_{t-1}^2, \tag{2.18}$$

denoting the expression in the square brackets in (2.8) by $w_t(z)$ we obtain

$$w_t(z) = -2X_{t-1}^2(z-\theta)^2 + \hat{I}_t^{-1}X_{t-1}^4(\theta - z)^2 + \hat{I}_t^{-1}X_{t-1}^2 \tag{2.19}$$

$$= -\delta X_{t-1}^2(z-\theta)^2 - X_{t-1}^2(z-\theta)^2\left((2-\delta) - \hat{I}_t^{-1}X_{t-1}^2\right) + \hat{I}_t^{-1}X_{t-1}^2 \tag{2.20}$$

for some $0 < \delta < 1$. Since $\hat{I}_t^{-1}X_{t-1}^2 \leq 1$, the positive part of the above expression does not exceed $\hat{I}_t^{-1}X_{t-1}^2$. This implies that (2.8) holds with $q_t = \hat{I}_t^{-1}X_{t-1}^2$. Now, note that if d_n is a nondecreasing sequence of positive numbers such that $d_t \to +\infty$ and $\triangle d_t = d_t - d_{t-1}$, then $\sum_{t=1}^{\infty} \triangle d_t / d_t = +\infty$ and $\sum_{t=1}^{\infty} \triangle d_t / d_t^2 < +\infty$. So, for $X_{t-1}^2 = \triangle \hat{I}_t$ and $\hat{I}_t \to \infty$ for any $\theta \in \mathbb{R}$ (see, e.g, Shiryayev [25], Ch.VII, §5) , we have

$$\sum_{t=1}^{\infty} \hat{I}_t^{-2} X_{t-1}^2 < \infty \quad \text{and} \quad \sum_{t=1}^{\infty} \hat{I}_t^{-1} X_{t-1}^2 = \infty \tag{2.21}$$

and since $q_t \gamma_t = \hat{I}_t^{-2} X_{t-1}^2$, (C1) follows. Therefore, $(\hat{\theta}_t - \theta)^2$ converges to a finite limit. To show convergence to θ, let us check condition (C2) of of Corrolary 2.4 with $z^0 = \theta$ and $U_t = \mathbb{R}$. Using (2.17) and (2.18), we have

$$\mathcal{N}_t(u) = -2\hat{I}_t^{-1} X_{t-1}^2 u^2 + \hat{I}_t^{-2} X_{t-1}^4 u^2 + \hat{I}_t^{-2} X_{t-1}^2 = \hat{I}_t^{-1} w_t(\theta + u),$$

where w_t is defined in (2.19). Since the middle term in (2.20) is non-positive, using the obvious inequality $[a]^- \geq -a$, we can write

$$[\mathcal{N}_t(u)]^- \geq \delta \hat{I}_t^{-1} X_{t-1}^2 u^2 - \hat{I}_t^{-2} X_{t-1}^2$$

and

$$\sum_{t=1}^{\infty} \inf_{\varepsilon \leq |u| \leq 1/\varepsilon} [\mathcal{N}_t(u)]^- = \infty$$

now follows from (2.21). So, by Corollary 2.3, $\hat{\theta}_t \to \theta$ ($P^\theta-$ a.s.).

3 Appendix

We will need the following properties of the Gamma function (see, e.g., [28], 12.16). $\log' \mathbf{\Gamma}$ is increasing, $\log'' \mathbf{\Gamma}$ is decreasing and continuous, and

$$\log'' \mathbf{\Gamma}(x) = \frac{1}{x^2} + \sum_{n=1}^{\infty} \frac{1}{(x+n)^2}.$$

The latter implies that

$$\log'' \mathbf{\Gamma}(x) \leq \frac{1}{x^2} + \sum_{n=1}^{\infty} \int_{n-1}^{n} \frac{dz}{(x+z)^2} = \frac{1}{x^2} + \frac{1}{x} = \frac{1+x}{x^2} \tag{3.1}$$

and

$$\log'' \mathbf{\Gamma}(x) \geq \sum_{n=0}^{\infty} \int_{n}^{n+1} \frac{dz}{(x+z)^2} = \frac{1}{x}. \tag{3.2}$$

Also (see [7], 12.5.4),

$$\log' \mathbf{\Gamma}(x) \leq \ln(x). \tag{3.3}$$

Then,

$$E^{\theta}\left\{\log X_1\right\} = \log'\mathbf{\Gamma}(\theta) \quad \text{ and } \quad E^{\theta}\left\{(\log X_1)^2\right\} = \log''\mathbf{\Gamma}(\theta) + \left(\log'\mathbf{\Gamma}(\theta)\right)^2 \tag{3.4}$$

and

$$E^{\theta}\left\{\left(\log X_1 - \log'\mathbf{\Gamma}(\theta)\right)^2\right\} = \log''\mathbf{\Gamma}(\theta).$$

Let us show that the conditions of Corollary 2.5 hold. Since

$$\Psi_t(u) = \frac{1}{\log''\mathbf{\Gamma}(u)}\left(\log X_t - \log'\mathbf{\Gamma}(u)\right),$$

using (3.4) and (3.2) we obtain

$$\begin{aligned}\frac{E\left\{\|\Psi_t(u)\|^2 \mid \mathcal{F}_{t-1}\right\}}{1+\|u-\theta\|^2} &= \frac{\log''\mathbf{\Gamma}(\theta) + \left(\log'\mathbf{\Gamma}(\theta) - \log'\mathbf{\Gamma}(u)\right)^2}{(\log''\mathbf{\Gamma}(u))^2(1+\|u-\theta\|^2)} \\ &\le \frac{u^2}{1+(u-\theta)^2}\left(\log''\mathbf{\Gamma}(\theta) + \left(\log'\mathbf{\Gamma}(\theta) - \log'\mathbf{\Gamma}(u)\right)^2\right).\end{aligned} \tag{3.5}$$

Now, $u^2/(1+(u-\theta)^2) \le C$. Here and further on in this subsection, C denotes various constants which may depend on θ. So, using (3.3) we obtain

$$\frac{E\left\{\|\Psi_t(u)\|^2 \mid \mathcal{F}_{t-1}\right\}}{1+\|u-\theta\|^2} \le C\left(\log''\mathbf{\Gamma}(\theta) + \log'\mathbf{\Gamma}(\theta)^2 + \log'\mathbf{\Gamma}(u)^2\right) \le C(1+\log^2(u)).$$

For large t's, since $\alpha_t < 1 < \beta_t$, we have

$$\sup_{u\in[\alpha_t,\beta_t]} \log^2(u) \le \left\{\sup_{\alpha_t\le u<1}\log^2(u) + \sup_{1<u\le\beta_t}\log^2(u)\right\} \le \log^2\alpha_t + \log^2\beta_t.$$

Condition (2) of Corollary 2.5 is now immediate from the second part of (2.13). It remains to check that (CC) of Corollary 2.6 holds. Indeed,

$$-(u-\theta)R(u) = \frac{(u-\theta)\left(\log'\mathbf{\Gamma}(u) - \log'\mathbf{\Gamma}(\theta)\right)}{\log''\mathbf{\Gamma}(u)}.$$

Since $\log'\mathbf{\Gamma}$ is increasing and $\log''\mathbf{\Gamma}$ is decreasing and continuous, we have that for each $\varepsilon \in (0,1)$,

$$\inf_{\substack{\varepsilon\le\|u-\theta\|\le 1/\varepsilon \\ u\in U_{t-1}}} -(u-\theta)R(u) \ge \frac{\inf_{\varepsilon\le\|u-\theta\|\le 1/\varepsilon}\left(\log'\mathbf{\Gamma}(u) - \log'\mathbf{\Gamma}(\theta)\right)(u-\theta)}{\sup_{u\in U_{t-1}}\log''\mathbf{\Gamma}(u)} \ge \frac{C}{\log''\mathbf{\Gamma}(\alpha_{t-1})} \tag{3.6}$$

where C is a constant that my depend on ε and θ. Since $\alpha_{t-1} < 1$ for large t's, it follows (3.1) that $1/\log''\mathbf{\Gamma}(\alpha_{t-1}) \ge \alpha_{t-1}^2/2$. Condition (CC) of Corollary 2.6 is now immediate from the first part of (2.13).

Note that with $\beta_t = \infty$ the procedure fails condition (2) of Corollary 2.5. Indeed, (3.5) and (3.1) implies that

$$\sup_{\alpha_t\le u}\frac{E\left\{\Psi_t^2(u) \mid \mathcal{F}_{t-1}\right\}}{1+(u-\theta)^2} \ge \sup_{\alpha_t\le u}\frac{\left\{\log''\mathbf{\Gamma}(\theta) + \left(\log'\mathbf{\Gamma}(\theta) - \log'\mathbf{\Gamma}(u)\right)^2\right\}u^4}{(1+u)^2(1+(u-\theta)^2)} = \infty \tag{3.7}$$

References

[1] Andrieu, C., Moulines, E. and Priouret, P. (2005). Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.* **44**, 283–312.

[2] Benveniste, A, Metivier, M. and Priouret , P. (1990). *Adaptive Algorithms and Stochastic Approximation.* Berlin and New York: Springer-Verlag.

[3] Borkar, V. S. (2008). *Stochastic approximation: A Dynamical Systems Viewpoint.* Cambridge University Press.

[4] Chen, H., Guo, L. and Gao, A. (1987). Convergence and robustness of the Robbins- Monro algorithm truncated at randomly varying bounds. *Stochastic Processes Appl.* **27**, 217231.

[5] Chen, H. and Zhu, Y.-M. (1986). Stochastic approximation procedures with randomly varying truncations. *Scientia Sinica 1* **29**, 914926.

[6] Campbell, K. (1982). Recursive computation of M-estimates for the parameters of a finite autoregressive process. *Ann. Statist.* **10**, 442-453.

[7] Cramer, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, Princeton.

[8] Englund, J.-E., Holst, U., and Ruppert, D. (1989). Recursive estimators for stationary, strong mixing processes – a representation theorem and asymptotic distributions. *Stochastic Processes Appl.* **31**, 203–222.

[9] Fabian, V. (1978). On asymptotically efficient recursive estimation. *Ann. Statist.* **6**, 854-867.

[10] Gu, M.G. and Li, S. (1998). A stochastic approximation algorithm for maximum-likelihoood estimation with incomplete data. *The Canadian Journal of Statistics* **26**, 567-582.

[11] Khas'minskii, R.Z., Nevelson, M.B. (1972). *Stochastic Approximation and Recursive Estimation.* Nauka, Moscow.

[12] Kushner, H. (2010). Stochastic approximation: a survey. *Wiley Interdisciplinary Reviews: Computational Statistics* **2**, 6, 87–96.

[13] Kushner, H. and Yin, G. (1997). *Stochastic Approximation Algorithms and Applications. Applications of Mathematics.* Springer-Verlag, New-York.

[14] Lai, T.L. (2003). Stochastic approximation, *Ann. Statist.* **31**, 391-406.

[15] Ljung, L. and Soderstrom, T. (1987). *Theory and Practice of Recursive Identification,* MIT Press.

[16] Robbins, H. and Monro, S. A stochastic approximation method, *Ann. Statist.* **22** (1951), 400–407.

[17] Robbins, H. and Siegmund, D. (1971). A convergence theorem for non-negative almost supermartingales and some applications. *Optimizing Methods in Statistics*. ed. J.S. Rustagi Academic Press, New York, 233–257.

[18] Sakrison, D.J. (1965). Efficient recursive estimation; application to estimating the parameters of a covariance function. *Internat. J. Engrg. Sci.* **3**, 461–483.

[19] Sharia, T. (1997). Truncated recursive estimation procedures, *Proc. A. Razmadze Math. Inst.* **115**, 149–159.

[20] Sharia, T. (1998). On the recursive parameter estimation for the general discrete time statistical model. *Stochastic Processes Appl.* **73**, **2**, 151–172.

[21] Sharia, T. (2008). Recursive parameter estimation: Convergence. *Statistical Inference for Stochastic Processes*. **11**, 2, pp. 157 – 175.

[22] Sharia, T. (2007). Rate of convergence in recursive parameter estimation procedures. *Georgian Mathematical Journal*. **14**, 4, pp. 721–736.

[23] Sharia, T. (2010). Recursive parameter estimation: Asymptotic expansion. *The Annals of The Institute of Statistical Mathematics* **62** 2, 343-362.

[24] Sharia, T. (2010). Efficient On-Line Estimation of Autoregressive Parameters. *Mathematical Methods of Statistics*. **19**, 2, 163-186.

[25] Shiryayev, A.N. (1984). *Probability,* Springer-Verlag, New York.

[26] Tadic, V. (1997) Stochastic gradient with random truncations, *European J. of Operational Research*, **101**, pp. 261–284.

[27] Tadic, V. (1998) Stochastic approximations with random truncations, state dependent noise and discontinuous dynamics, *Stochastics and Stochastics reports*. **64**, pp. 283–326.

[28] Whittaker, E. Watson, G. (1927). *A Course of Modern Analysis*. Cambridge University Press, Cambridge.