Proceedings of the 40th IEEE
Conference on Decision and Control
Orlando, Florida USA, December 2001

FrM08-4

THE MATHEMATICS OF NOISE-FREE SPSA

László Gerencsér
Computer and Automation Institute of
the Hungarian Academy of Sciences,
13-17 Kende u., Budapest, 1111, Hungary
gerencser@sztaki.hu

Zsuzsanna Vágó
Computer and Automation Institute of
the Hungarian Academy of Sciences,
13-17 Kende u., Budapest, 1111, Hungary
vago@oplab.sztaki.hu

Abstract

We consider discrete-time fixed gain stochastic approximation processes that are defined in terms of a random field that is identically zero at some point θ^*. The boundedness of the estimator process is enforced by a resetting mechanism. Under appropriate technical conditions the estimator sequence is shown to converge to θ^* with geometric rate almost surely. This result is in striking contrast to classical stochastic approximation theory where the typical convergence rate is $n^{-1/2}$. For the proof a discrete-time version of the ODE-method is developed and used, and the techniques of [8] are extended. The paper is motivated by the study of simultaneous perturbation stochastic approximation (SPSA) methods applied to noise-free problems and to direct adaptive control, see [13].

1 Introduction

Let $H(n,\theta,\omega)$ be a random field defined over some probability space $(\Omega,\mathcal{F},P)$ for $n\geq 1$ and $\theta\epsilon D\subset \mathbb{R}^p$, where D is a bounded open domain. Assume that for some $\theta^*\epsilon\mathbb{R}^p$ the random field identically vanishes, i.e. we have

$$H(n,\theta^*,\omega)=0. \quad (1)$$

The problem that we study is to determine θ^* via a stochastic approximation procedure based on observed values of $H(n,\theta,\omega)$.

Noise free SPSA: For a motivation consider the following problem: minimize a function $L(\theta)$ defined for $\theta\epsilon\mathbb{R}^p$, such that it is three-times continuously differentiable with respect to θ, and $L(.)$ has a unique global minimizing value θ^*. Assume that the computation of $L(.)$ is expensive and the gradient of $L(.)$ is not computable at all. To minimize $L(.)$ we estimate the gradient of $L(.)$ denoted by $G(\theta)=L_\theta(\theta)$. Following [19] consider random simultaneous perturbations of the components of θ as follows: first take a sequence of independent, identically distributed (i.i.d.) random variables, with time index n, $\Delta_{ni}(\omega),\ i=1,...,p$ defined over some probability space $(\Omega,\mathcal{F},\mathcal{P})$ satisfying certain weak technical conditions given in [19]. E.g. we may take a Bernoulli-sequence with

$$P(\Delta_{ni}(\omega)=+1)=1/2 \qquad P(\Delta_{ni}(\omega)=-1)=1/2.$$

Let $c>0$ be a fixed small positive number. For any $\theta\epsilon\mathbb{R}^p$ we evaluate $L(.)$ at two randomly, but symmetrically chosen points, $\theta+c\Delta_n(\omega)$ and $\theta-c\Delta_n(\omega)$, respectively. Then the i-th component of the gradient is estimated as

$$H_i(n,\theta,\omega)=\Delta_{ni}^{-1}(\omega)\frac{L(\theta+c\Delta_n(\omega))-L(\theta-c\Delta_n(\omega))}{2c}.$$

Set $H(n,\theta,\omega)=(H_1(n,\theta,\omega),...,H_p(n,\theta,\omega))$ and

$$\Delta_n^{-1}(\omega)=\left[\Delta_{n1}^{-1}(\omega),\ldots,\Delta_{np}^{-1}(\omega)\right]^T.$$

Using this gradient estimator with a decreasing $c=c_n$, where c_n tends to zero at a rate $1/n^\gamma$ with some $\gamma>0$, a stochastic approximation procedure with decreasing gain $1/n^a$ with some $0<a\leq 1$, called the simultaneous perturbations stochastic approximation or SPSA method, has been developed in [19]. SPSA methods have been analyzed under various conditions in [3], [9], [17] and [19], and simplified and improved versions have been developed in [20, 21].

In the special case, when L is a positive definite quadratic function we have

$$H(n,\theta,\omega)=\Delta_n^{-1}(\omega)\Delta_n^T(\omega)G(\theta)$$

for any size of the perturbation c. Since $G(\theta^*)=0$, we have identically

$$H(n,\theta^*,\omega)=0.$$

For general, non-quadratic cost functions we have to decrease the size of the perturbation, thus we chose $c=c_n$, where c_n tends to zero. Then the condition $H(n,\theta^*,\omega)=0$ will be satisfied asymptotically. An alternative class of problems where the condition $H(n,\theta^*,\omega)=0$ is satisfied exactly is described in [13] in connection with multivariable direct adaptive controller design.

A survey of previous results: The standard stochastic approximation procedure for (1) would be

$$\theta_{n+1}=\theta_n+\frac{1}{n+1}H(n+1,\theta_n,\omega) \qquad \theta_0=\xi. \quad (2)$$

Such general stochastic approximation procedures have been considered e.g. in [1, 2, 5, 18]. The asymptotic covariance matrix of the estimator process, denoted by S^*, has been determined under various conditions in [1, 2, 6, 18]. It is easy to see that $H(n, \theta^*, \omega) = 0$ implies

$$S^* = \lim_{n\to\infty} n\mathrm{E}(\theta_n - \theta^*)(\theta_n - \theta^*)^T = 0, \qquad (3)$$

hence the convergence rate is better than the standard rate $n^{-1/2}$. But how much better can it be? A straightforward, but tiresome calculation induces us to consider fixed gain stochastic approximation processes of the form

$$\theta_{n+1} = \theta_n + \lambda H(n+1, \theta_n, \omega) \qquad \theta_0 = \xi. \qquad (4)$$

Fixed gain recursive estimation processes of this general form have been widely used in the engineering literature. An important example is the well-known LMS-algorithm of adaptive filtering, the stability properties of which had been studied in [7, 12, 16], assuming some form of stationarity. The most complete characterization of LMS processes has been given recently in [14]. General classes of algorithms given above has been considered in [1] and [18]. Assuming that

$$G(\theta) = \mathrm{E}H(n, \theta, \omega). \qquad (5)$$

is independent of n, and that θ^* is an asymptotic stable equilibrium point for G has the estimator sequence θ_n is related to the solution of the associated ordinary differential equation

$$\dot{y}_t = \lambda G(y_t) \qquad y_0 = \xi.$$

This method of proof is often called the ODE-method.

In [8] $H(n, \theta, \omega)$ is assumed to be an L-mixing process. A key condition in the analysis presented in [8] is that the essential supremum of the random variable $|H(n, \theta, \omega) - G(\theta)|$ is sufficiently small for all $\theta \epsilon D$, to ensure a priori that θ_n will stay in a prescribed compact domain $D_0 \subset D$. The conclusion is that the L_q-norms of the tracking error is of the order of magnitude $\lambda^{1/2}$.

A *new feature* of the present paper is that the condition $H(n, \theta^*, \omega) = 0$ is imposed, and that assumption ensuring a priori that θ_n will stay in a prescribed compact domain $D_0 \subset D$, is removed. Instead, the condition $\theta_n \epsilon D_0$ is enforced by using a resetting mechanism. The relaxation of the conditions on the upper bound of $|H(n, \theta, \omega) - G(\theta)|$ is essential for noise free SPSA.

We have the surprising result that, in spite of the fact that, predictably, we have frequent resetting, θ_n does converge to θ^* almost surely, and the rate of convergence is geometric. A heuristic argument in favour this result is that when θ_n gets close to θ^* then the effect of the noise is negligible, and also the occurrence of a resetting becomes less likely. The analysis is based on a discrete-time ODE-method.

In the special case of SPSA methods applied to noise-free optimization we get fixed gain SPSA methods. These have been first considered in [10] for general random fields, i.e. without the extra condition $H(n, \theta^*, \omega) = 0$. The present paper is a follow-up of [11].

2 The basic result

The p-dimensional Euclidean-space will be denoted by $\mathbb{R}^p$. The Euclidean-norm of a vector x will be denoted by $|x|$, the operator norm of a matrix A will be denoted be $\|A\|$. Assume $\theta^* = 0$ and that $H(n, \theta, \omega)$ can be written in the form

$$H(n, \theta, \omega) = A(n, \theta, \omega)\theta$$

where the $p \times p$ matrix-valued random-field $A(n, \theta, \omega)$ satisfies the conditions below. Thus we come to consider a quasi-linear random iterative process of the form

$$\theta_{n+1} = \theta_n + \lambda A(n+1, \theta_n, \omega)\theta_n. \qquad (6)$$

If $A(n+1, \theta_n, \omega)$ is in fact independent of θ, then (6) is a random linear difference equation with state-independent transition-matrices, the stability of which has been studied extensively already in [15].

In the conditions below we use the notations given in the Appendix. Define the $p \times p$ matrix-valued random-field $\Delta A/\Delta\theta$, for $\theta, \theta + h$ $h \neq 0$, by

$$(\Delta A/\Delta\theta)(n, \theta, \theta+h, \omega) = |A(n, \theta+h, \omega) - A(n, \theta, \omega)|/|h|.$$

Condition 2.1 *The matrix-valued random-fields A and $\Delta A/\Delta\theta$ are defined and bounded for $n \geq 1$, θ, $\theta + h \epsilon D$, $h \neq 0$, where D is a bounded domain:*

$$||A(n, \theta, \omega)|| \leq K', \quad ||(\Delta A/\Delta\theta)(n, \theta, \theta + h, \omega)|| \leq L'.$$

A key technical condition that ensures a stochastic averaging effect is the following:

Condition 2.2 *A and $\Delta A/\Delta\theta$ are L-mixing uniformly in θ for $\theta \epsilon D$ and in $\theta, \theta+h$ for $\theta, \theta+h \epsilon D$, $h \neq 0$, respectively, with respect to a pair of families of σ-algebras $(\mathcal{F}_n, \mathcal{F}_n^+), n \geq 1$.*

For the sake of convenience we assume:

Condition 2.3 *The mean field $\mathrm{E}H(n, \theta, \omega)$ is independent of n, i.e. we can write*

$$G(\theta) = \mathrm{E}H(n, \theta, \omega). \qquad (7)$$

It follows, that $B(\theta) = \mathrm{E}A(n,\theta,\omega)$ is also independent of n. Write

$$\overline{A}(n,\theta,\omega) = A(n,\theta,\omega) - \mathrm{E}A(n,\theta,\omega).$$

Condition 2.4 *The function G defined on D is continuous and bounded in y together with its first and second partial derivatives, say*

$$|G(y)| \le K, \quad \|\partial G(y)/\partial y\| \le L \quad \|\partial^2 G(y)/\partial y^2\| \le L. \tag{8}$$

Domains: We develop a discrete-time version of the ODE method which can be applied with ease for the problem considered in this paper. The process θ_n will be compared with the discrete-time deterministic process (z_n) defined by

$$z_{n+1} = z_n + \lambda G(z_n), \quad z_0 = \xi = \theta \epsilon D_\theta, \tag{9}$$

where D_θ is a compact domain to which the estimator sequence θ_n will be confined by a resetting technique. Let $z(n,m,\xi)$ denote the solution of (9) with initial condition $z_m = \xi$. Define the time-homogeneous mapping associated with (9) $\xi \to z_n(\xi) = z(n,0,\xi)$. Let $D_\theta \subset D$ be a subset of D such that for $\theta \epsilon D_\theta$ we have $z_n(\theta) \epsilon D$ for any $n \ge 0$. For any fixed n the image of D_θ under z_n will be denoted as $z_n(D_\theta)$ i.e. $z_n(D_\theta) = \{z : z = z(n,0,\theta), \ \theta \epsilon D_\theta\}$. The union of these sets will be denoted by $z(D_\theta)$, i.e.

$$z(D_\theta) = \{z : z = z(n,0,\theta) \quad \text{for some} \quad n \ge 0, \ \theta \epsilon D_\theta\}.$$

It can be proved that, under suitable technical conditions, $z(D_\theta) \subset D_z \subset D$ where D_z is some compact domain.

The associated continuous-time ODE is defined as

$$\dot{y}_t = \lambda G(y_t), \qquad y_s = \xi = z \epsilon D_z, \qquad s \ge 0. \tag{10}$$

The solution of (10) will be denoted by $y(t,s,\xi)$. The time-homogeneous flow associated with (10) is defined as the mapping $\xi \to y_t(\xi) = y(t,0,\xi)$. Let D_z be such that for $z \epsilon D_z$ we have $y_t(z) \epsilon D$ for any $t \ge 0$. For any fixed t the image of D_z under y_t will be denoted as $y_t(D_z)$ i.e. $y_t(D_z) = \{y : y = y(t,0,z), \ z \epsilon D_z\}$. The union of these sets will be denoted by $y(D_z)$, i.e.

$$y(D_z) = \{y : y = y(t,0,z) \quad \text{for some} \quad t \ge 0, \ z \epsilon D_z\}.$$

For any set D_0 write

$$S(D_0, \varepsilon) = \{\theta : |\theta - z| < \varepsilon \text{ for some } z \epsilon D_0\}.$$

Finally the interior of a compact domain D_0 is denoted by int D_0. We will require the following stability condition:

Condition 2.5 *There exist compact domains $D_\theta \subset D_z \subset D_y \subset D$ and $d > 0$ such that $0 \ \epsilon \ \mathrm{int} D_\theta$ and*

$$S(y(D_\theta), d) \subset D_z \quad \text{and} \quad y(D_z) \subset D_y \subset D. \tag{11}$$

Condition 2.6 *The ordinary differential equation (10) satisfies the following stability conditions: for any $\xi = z \epsilon D_z$ and for any $\varepsilon > 0$ there exists a $T > 0$ such that $t - s > T$ implies $|y(t,s,\xi)| \le \varepsilon$. Moreover the Jacobian-matrix $G_y(0)$ has all its eigenvalues in the open left half plane.*

It follows, that for some $C_0 > 0$ and $\alpha > 0$ we have for all $0 \le s \le t, \ z \epsilon D_z$

$$\|\frac{\partial}{\partial z} y(t,s,z)\| \le C_0 e^{-\alpha(t-s)}. \tag{12}$$

We can assume that $C_0 \ge 1$.

Resetting: Let D_ξ be a compact domain such that

$$0 \epsilon \ \mathrm{int} D_\xi \quad \text{and} \quad S(D_\xi, d') \subset \mathrm{int D}_\theta \tag{13}$$

for some $d' > 0$. Assume that $\xi = \widehat{\theta}_0 \epsilon D_\xi$. At time n we first define a tentative value θ_{n+1-} following (4) as

$$\theta_{n+1-} = \theta_n + \lambda H(n+1, \theta_n, \omega) \tag{14}$$

and then we set

$$\begin{array}{lcll} \theta_{n+1} & = & \theta_{n+1-} & \text{if} \quad \theta_{n+1-} \epsilon D_\theta \\ \theta_{n+1} & = & \theta_0 & \text{if} \quad \theta_{n+1-} \not\epsilon D_\theta. \end{array} \tag{15}$$

Condition 2.7 *It is assumed that for some $0 < r < R$ we have*

$$D_\xi \subset S(0,r) \subset S(0,R) \subset D_\theta. \tag{16}$$

Theorem 2.1 *Assume that Conditions 2.1-2.7 are satisfied. Let $\xi = \theta_0 \epsilon D_\xi$ and assume that $C_0^3 r/R < 1$. Then there exists a γ with $0 < \gamma < 1$ and a positive random variable $C(\omega)$ such that for sufficiently small λ we have*

$$|\theta_N| \le C(\omega) \gamma^{\lambda N}.$$

3 Outline of the proof

First we need two simple lemmas on the discrete flow defined by (9).

Lemma 3.1 *Assume that Conditions 2.4, 2.5 and 2.6 are satisfied. Let y_t be the solution of (10) and let z_n be the solution of (9) with $y_0 = z_0 \epsilon D_\theta$. Then if $d > C_0 \alpha^{-1} \cdot \lambda L K$ then z_n will stay in D_z for all n, and $|z_n - y_n| \le C_0 \alpha^{-1} \cdot \lambda L K$ for all $n \ge 0$. In addition for any $\alpha' < \alpha$ we have $|z_n| \le C_0 e^{-n\lambda\alpha'}$ whenever λ is sufficiently small.*

An interesting property of the discrete flow defined by (9) is that it inherits exponential stability with respect to initial perturbations if λ is sufficiently small.

Lemma 3.2 *Assume that Conditions 2.4, 2.5 and 2.6 are satisfied. Then* $d > C_0\alpha^{-1} \cdot \lambda LK$ *implies that* z_n *will stay in* D_z *for all* n*, moreover for any* $0 < \alpha' < \alpha$ *and* $n \geq m$ *we have*

$$||\frac{\partial}{\partial \xi} z(n,m,\xi)|| \leq C_0 e^{-\lambda\alpha'(n-m)}, \qquad (17)$$

whenever λ *is sufficiently small.*

A local approximation: In what follows the discrete-time parameter n will be replaced by t and n will stand for a rescaled discrete time-index. Let T be a fixed positive integer. Let us subdivide the set of integers into intervals of length T. Let n be a non-negative integer and let $\tau(nT)$ denote the first integer $t > nT$ for which $\theta_t \notin D_\theta$. In the interval $[nT, (n+1)T-1]$ we consider the solution of (9) starting from θ_{nT} at time nT. This will be denoted by $\overline{z}_t$, i.e. $\overline{z}_t$ is defined by

$$\overline{z}_{t+1} = \overline{z}_t + \lambda G(\overline{z}_t), \qquad \overline{z}_{nT} = \theta_{nT}.$$

We can also write $\overline{z}_t = z(t, nT, \theta_{nT})$ for $nT \leq t \leq (n+1)T$. The definition of $\overline{z}_t$ is non-unique for $t = (n+1)T$, therefore we use the notation $\overline{z}_{(n+1)T-} = z((n+1)T, nT, \theta_{nT})$ and $\overline{z}_{(n+1)T} = \theta_{(n+1)T}$. A key step in the derivation is to get an upper bound for $|\theta_t - \overline{z}_t|$.

We need the following simple observation: for $s \geq nT$ we have

$$\overline{z}_s = z(s, nT, \theta_{nT}) = z(s, nT, \theta_{nT}) - z(s, nT, 0) =$$

$$\int_0^1 \frac{\partial z}{\partial \xi}(s, nT, \lambda\theta_{nT}) d\lambda \cdot \theta_{nT}.$$

The presence of the multiplicative term $|\theta_{nT}|$ on the right hand side is a key feature that ensures convergence with exponential rate. In the lemma below the definition of $\theta_{\tau(nT)}$ will be temporarily changed for the sake of convenience to denote the value of θ_t at time $\tau(nT)$ prior to resetting.

Lemma 3.3 *For any* T *we have*

$$\sup_{nT \leq t \leq (n+1)T \wedge \tau(nT)} |\theta_t - \overline{z}_t| \leq c^* \eta_n^* |\theta_{nT}| \qquad (18)$$

where η_n^* *is defined in terms of* $\overline{H}$ *as follows:*

$$\eta_n^* = \sup_{\substack{nT \leq t \leq (n+1)T \\ \theta \varepsilon D_\theta}} || \sum_{s=nT}^{t-1} \lambda \overline{A}(s+1, z(s, nT, \theta), \omega)|| \quad (19)$$

and $c^* = C_0(1 + \lambda L)^T$.

Remark: Lemma 2.2 of [8] implies that the process (η_n^*) is L-mixing with respect to $(\mathcal{F}_{nT}, \mathcal{F}_{nT}^+)$. In particular, choosing $T = [(\lambda\alpha)^{-1}] + 1$ we get for any $2 < q < \infty$ and $r > p$

$$M_q(\eta^*) \leq C_q \lambda^{1/2}, \qquad (20)$$

where C_q is independent of λ.

In the lemma below we use the following notation: if a resetting takes place at time nT then θ_{nT} denotes the value of θ_t at time nT prior to resetting.

Now if no resetting takes place over several periods then the repeated application of (18) leads to the following result:

Lemma 3.4 *Let* $T = [(\lambda\alpha)^{-1}] + 1$ *and assume that no resetting takes place for* $0 \leq t \leq nT$ *with some positive integer* n*. Then we have*

$$|\theta_{nT}| \leq C_0 |\theta_0| \cdot \Pi_{k=1}^n (\beta + C_0 c^* \eta_{k-1}^*)$$

with $0 < \beta < 1$*, where the latter is independent of* λ*.*

The starting points and the endpoints of the above interval can be replaced by arbitrary, possible random time-moments.

For small λ the averaging effect is ensured by taking large T. Note that β is independent of λ, but $\eta_{k-1}^* = O_M(\lambda^{1/2})$ can be made very small.

If a resetting does take place in $(nT, (n+1)T]$, then the following multiplicative inequality can be derived, that will replace (18) in the analysis:

$$|\theta_{(n+1)T}| \leq \frac{r}{R}(C_0 + 2c^* \eta_n^*)^2 |\theta_{nT}|. \qquad (21)$$

The proof the Theorem 2.1 can now be completed by applying fairly straightforward estimations for random products.

4 Simulation results

Consider a quadratic function $L(\theta) = \frac{1}{2}\theta^T A \theta$, with some symmetric positive definite A. Then $H(k, \theta) = \Delta_k^{-1} \Delta_k^T G(\theta)$. But $G(\theta) = A\theta$, hence we get the following recursion for θ_k:

$$\theta_{k+1} = (I - \lambda \Delta_k^{-1} \Delta_k^T A)\theta_k. \qquad (22)$$

Applying Oseledec's multiplicative ergodic theorem we can easily get the claim of Theorem 2.1. Consider now non-quadratic problems of the form

$$L(\theta) = \frac{1}{2}\theta^T A(\theta)\theta$$

with

$$A(\theta) = A + u(\theta)u(\theta)^T$$

where $u(\theta) = D\theta$ with some fixed matrix D.

We have tested the fixed gain SPSA method (4) for randomly generated problems of dimension 20, where the elements of D were chosen uniformly in the range of $[0, c_p]$. In each experiment we had $N = 500$ iterations.

The top Lyapunov-exponent is approximated by

$$\frac{1}{N} \log |\theta_N - \theta^*|.$$

In Figures 1. and 2. the approximate top Lyapunov exponent is plotted against the step-size λ for three, increasingly non-quadratic problems for plain SPSA and second order SPSA proposed in [21].

5 Conclusion

Stochastic approximation processes based on vanishing random fields with a resetting mechanism converge with geometric rate almost surely. The result is in striking contrast to classical stochastic approximation theory where the typical convergence rate is $n^{-1/2}$. The paper is motivated by the study of simultaneous perturbation stochastic approximation (SPSA) methods applied to noise-free problems.

An open problem is the following: The condition $C_0^3 r/R < 1$ becomes $r/R < 1$ if $C_0 = 1$. This is equivalent to saying of (12) that for some $\alpha > 0$ we have for all $0 \leq s \leq t,\ z \epsilon D_z$

$$\|\frac{\partial}{\partial z} y(t, s, z)\| \leq e^{-\alpha(t-s)}. \qquad (23)$$

An interesting problem is to find useful sufficient conditions under which (23) will be satisfied for non-linear systems with some Euclidean norm.

6 Appendix

In this section the basic concepts of the theory of L-mixing processes developed in [4] will be presented. Let a probability space $(\Omega, \mathcal{F}, P)$ be given, let $D \subset \mathbb{R}^p$ be an open domain and let $(x_n(\theta))$: $\Omega \times \mathbb{Z} \times D \to \mathbb{R}^n$ be parameter-dependent stochastic process. We say that $(x_n(\theta))$ is M-bounded if for all $1 \leq q < \infty$

$$M_q(x) = \sup_{\substack{n \geq 0 \\ \theta \epsilon D}} \mathrm{E}^{1/q} |x_n(\theta)|^q < \infty.$$

We shall use the same terminology if θ or t degenerate into a single point.

Let $(\mathcal{F}_n), n \geq 0$ be a family of monotone increasing σ-algebras, and $(\mathcal{F}_n^+), n \geq 0$ be a monotone decreasing family of σ-algebras. We assume that for all $n \geq 0, \mathcal{F}_n$ and $\mathcal{F}_n^+$ are independent. For $n \leq 0$ we set $\mathcal{F}_n^+ = \mathcal{F}_0^+$. A stochastic process $(x_n(\theta)), n \geq 0$ is L-mixing with respect to $(\mathcal{F}_n, \mathcal{F}_n^+)$ uniformly in θ if it is $(\mathcal{F}_n)$-measurable, M-bounded and if we set for $1 \leq q < \infty$

$$\gamma_q(\tau, x) = \gamma_q(\tau) = \sup_{\substack{n \geq \tau \\ \theta \epsilon D}} \mathrm{E}^{1/q} |x_n(\theta) - \mathrm{E}(x_n(\theta)|\mathcal{F}_{n-\tau}^+)|^q$$

where τ is a positive integer then

$$\Gamma_q = \Gamma_q(x) = \sum_{\tau=1}^{\infty} \gamma_q(\tau) < \infty.$$

Acknowledgement

This work was supported by the National Research Foundation of Hungary (OTKA) under Grant no. T 032932 and by Bolyai János Research Fellowship.
The first author expresses his thanks to James C. Spall and Stacy D. Hill for cooperating in this research and providing a number of useful comments, and to Håkan Hjalmarsson for calling his attention to reference [13].

References

[1] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations.* Springer-Verlag, Berlin, 1990.

[2] A.N. Borodin. A stochastic approximation procedure in the case of weakly dependent observations. *Theory of Probability and Appl.*, 24:34–52, 1979.

[3] H.F. Chen, T.E. Duncan, and B. Pasik-Duncan. A Kiefer-Wolfowitz algorithm with randomized differences. *IEEE Trans. Automat. Contr.*, 44:442–453, 1999.

[4] L. Gerencsér. On a class of mixing processes. *Stochastics*, 26:165–191, 1989.

[5] L. Gerencsér. Rate of convergence of recursive estimators. *SIAM J. Control and Optimization*, 30(5):1200–1227, 1992.

[6] L. Gerencsér. A representation theorem for the error of recursive estimators. In *Proc. of the 31st IEEE Conference on Decision and Control, Tucson*, pages 2251–2256, 1992.

[7] L. Gerencsér. Rate of convergence of the LMS algorithm. *Systems and Control Letters*, 24:385–388, 1995.

[8] L. Gerencsér. On fixed gain recursive estimation processes. *J. of Mathematical Systems, Estimation and Control*, 6:355–358, 1996. Retrieval code for full electronic manuscript: 56854.

[9] L. Gerencsér. Rate of convergence of moments for a simultaneuous perturbation stochastic approximation method for function minimization. *IEEE Trans. Automat. Contr.*, 44:894–906, 1999.

[10] L. Gerencsér, S. D. Hill, and Zs. Vágó. Optimization over discrete sets via SPSA. In *Proceedings of the 38-th Conference on Decision and Control, CDC'99*, pages 1791–1794. IEEE, 1999.

[11] L. Gerencsér and Zs. Vágó. SPSA in noise free optimization. In *Proceedings of the American Control Conference*, pages 3284–3288. IEEE, 2000.

[12] L. Györfi and H. Walk. On the average stochastic approximation for linear regression. *SIAM J. Control and Optimization*, 34(1):31–61, 1996.

[13] H. Hjalmarsson. Efficient tuning of linear multivariable controllers using iterative feedback tuning. *Int. J. Adapt.Control Signal Process.*, 13:553–572, 1999.

[14] J. A. Joslin and A. J. Heunis. Law of the iterated logarithm for a constant-gain linear stochastic gradient algorithm. *SIAM J. on Control and Optimization*, 39:533–570, 2000.

[15] R.Z. Khasminskii. Necessary and sufficient conditions for the asymptotic stability of linear stochastic systems. *Th. Prob. Appl.*, 12:144–147, 1967.

[16] H.J. Kushner and A. Shwartz. Weak convergence and asymptotic properties of adaptive filters with constant gains. *IEEE Trans. Informat. Theory*, 30(2):177–182, 1984.

[17] H.J. Kushner and G. Yin. *Stochastic Approximation Algorithms and Applications.* Springer Verlag. New York, 1997.

[18] L. Ljung and T. Söderström. *Theory and practice of recursive identification.* The MIT Press, 1983.

[19] J.C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Automat. Contr.*, 37:332–341, 1992.

[20] J.C. Spall. A one-measurement form of simultaneous perturbation stochastic approximation. *Automatica*, 33:109–112, 1997.

[21] J.C. Spall. Adaptive stochastic approximation by the simultaneous perturbation method. *IEEE Trans. Automat. Contr.*, 45:1839–1853, 2000.

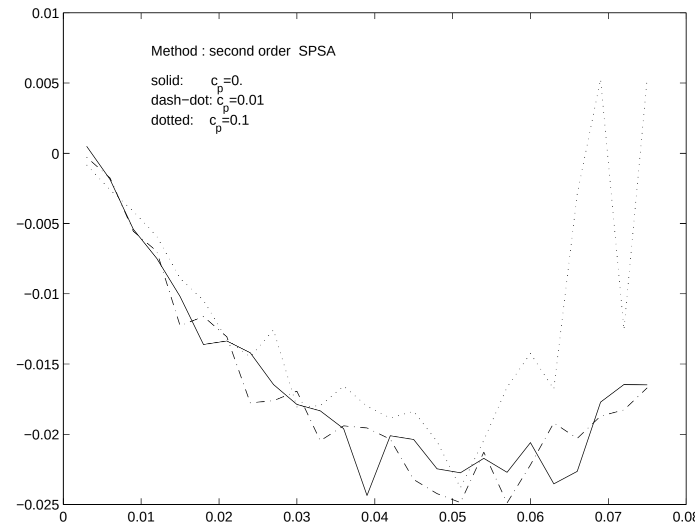

Figure 1:

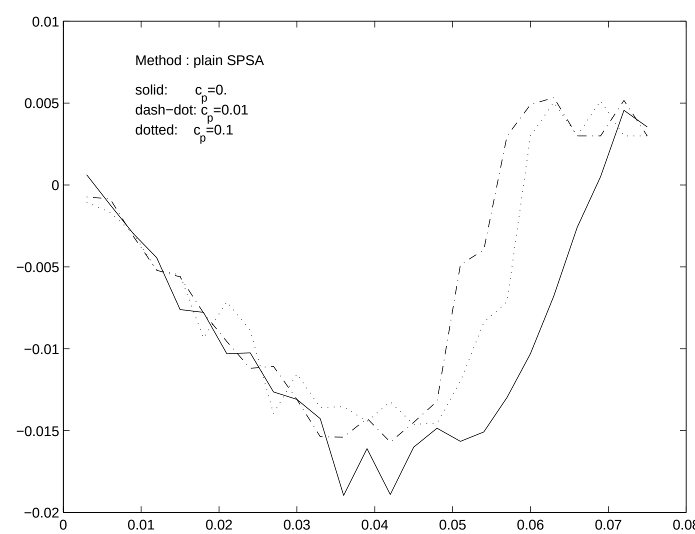

Figure 2: