Copyright © 2002 IFAC
15th Triennial World Congress, Barcelona, Spain

IDENTIFICATION OF LINEAR PLANTS WITH LARGE SIGNAL-TO-NOISE RATIO

A.L. Bunich and V.A. Lototsky

Institute of Control Sciences, Russian Academy of Sciences
65, Profsoyuznaya, Moscow, 117997, Russia.
Tel.: +7(095)3349201, fax: +7(095)3348759
E-mail: lotfone@ipu.rssi.ru

Abstract. A principally new approach to the design of identification algorithms is suggested. Under this approach, the consistency of an estimate is reached by the special nonlinear transformation of discrepancy rather than by diminishing the algorithm step value. *Copyright © 2001 IFAC*

Keywords: system identification, model adjustment, dead zone algorithms, test signals.

1. INTRODUCTION

Both conditions of consistent parameter estimation and the available convergence rate of identification algorithms (in the least mean squares (LMS) sense) depend on the uncertainty level in the description of the plant and disturbances. Without noises, there exist computationally simple enough algorithms of recurrent estimation (like Kaczmarz algorithm and its modification, Shi and Kozin, 1986), which guarantee an exponential convergence rate. When the unmeasured disturbances can not be neglected the asymptotically reachable rate of transient processes in the identifier depends on the sample volume N as $O(N^{\lambda})$ with $\lambda \leq 1$ determined by the noise distribution class capacity. In particular, for regular parameter estimation problems, $\lambda = 1$ and the asymptotic convergence rate is determined by the lower limit of Cramer – Rao information inequality or its various modifications for dynamic plant identification (Kashiap and Rao, 1976). In the non-parametric identification problems, the limiting algorithm speed is less in the order and is determined by the smoothness of the noise distribution for the estimated plant characteristic (Krzyak, 1986; Nazin and Yuditsky, 1999; Ibragimov and Khasminsky, 1979). The slow convergence is indeed an obstacle for practical use of stochastic approximation algorithms in the real time control systems. On the other hand, for less uncertainty there are estimation algorithms with a higher order convergence rate, determined by the noise distribution specifics (Ibragimov and Khasminsky, 1979).

The identifier adjusts the model using the feedback on the discrepancy between outputs of the plant and its adjustable model like in a classic control system structure with the error feedback. However, there is an essential difference due to the fact that in the identification problems the discrepancy gain factor is diminishing during the adjustment to meet the estimate consistency conditions. Such a step value decrease may be stated *a priori* (as in an ordinary stochastic

approximation algorithm) or chosen based on current observations as in the least mean squares algorithm and its modifications with recurrent matrix inversion (Poznyak and Tikhonov, 1990; Evans *et al.,* 1994; Ljung, 1987), and also in the scheme of Goodwin *et al.* (1981). Due to the quasi-stationary nature of real-life control plants, the problem naturally arises to finding principally different recursive estimation schemes where the convergence condition is met by special nonlinear discrepancy transformation rather than by diminishing the step value. Such an algorithm with a discrepancy transform of the "dead zone" type was suggested by Bunich (1983); it was shown that for certain classes of distributions with border peculiarities a higher
order convergence rate is really achieved compared to the asymptotically reachable rate in regular estimation problems. Nevertheless, the applicability of the algorithm was limited by strict constraints on noise distributions.

In what follows, the noise distribution class constraints are essentially relaxed and the identification algorithm convergence is investigated under limited noises with *a priori* upper border.

2. THE PROBLEM STATEMENT AND DESCRIPTION OF RECURSIVE ESTIMATION PROCEDURE

Consider an identification problem of a linear plant with discrete time $t = 1, 2, ...,$ described by "input – output" equation

$$y_t = \phi_t^T \theta^* + v_t \tag{1}$$

where $\phi_t \in R^N$ stands for regressor, $\theta^* \in R^N$ is unknown parameter, $v_t \in R^1$ is unmeasured plant output disturbance (noise). The observations are a random vectors sequence $(y_t, \phi_t)^{\infty}_{t=0} \in R^{N+1}$ on the probability space $(\tilde{U}, \Im, P)$. The noise is limited with *a priori* given upper bound, i.e., positive constant δ:

$$|v_t| < \delta \quad P\text{-}a.s. \quad \text{(almost surely)} \tag{2}$$

Given an observation sample, one is to estimate parameter θ^*.

In the absence of information about the useful statistical plant properties such as ergodicity one can not guarantee consistent plant parameter estimation, and is to limit himself to *functional identification.* For example, in a periodic observation sequence, the functional identification may be treated as a solution of a finite set of linear inequalities with respect to (w.r.t.) θ^* which are determined by plant equation (1) and the noise constraint (2). To solve this problem, the relaxation procedure by Motzkin and Schoenberg (1954) can be used with discrepancy transformation of the "dead zone" type:

$$\begin{gathered}\theta_t = \theta_{t-1} + \phi_t f(\xi_t) \, \| \phi_t \|^{-2}, \\ \xi_t = y_t - \phi_t^T \theta_{t-1}, f(\xi_t) = \\ =(|\xi_t| - \delta) + sign\, \xi_t.\end{gathered} \tag{3}$$

An essentially wider class of functional identification and adaptive control problems based on algorithm (3) and its modifications (so called "widened strip") was considered by Fomin *et al.* (1981) in the context of the recurrent target inequalities theory.

For a time-invariant observation scheme with defined on $(\Omega, \Im, P)$ measure P saving shift transform $T: \Omega \rightarrow \Omega$ the statistic properties of this transform will be the worse, the higher is the speed of approximation by T-periodic automorphisms. The limit case of no new data corresponds to the Motzkin and Schoenberg (1954) scheme with a periodic observation sequence. When the automorphism T possesses useful statistical properties under certain constraints one can obtain consistent plant parameter estimation. Then (3) becomes the Motzkin and Schoenberg (1954) scheme generalization for stochastic identification problems and is in fact, a stochastic relaxation procedure (Ermolyev, 1976).

As algorithm (3) characteristic contains a dead zone, the global convergence of estimates will demand imposing additional conditions on the observed variables. For example, consistent estimation is impossible in a passive identification problem when the regressor has a finite probability density (w.r.t. Lebesgue measure). In identification problems with randomized strategy there are methods of signal-to-noise ratio improvement by choosing a test signal in the control channel. The effect of large signal-to-noise ratio and the means of obtaining it are crucial in the design and convergence investigation of identification algorithms in this paper.

3. STATIC PLANT IDENTIFICATION

The first industrial applications of control systems with identifier (Danilov and Rajbman, 1973) concerned the adaptive compensator design for a system with linear static plant $Y_t = \phi_t^T \theta^* + K u_t + v_t$ where Y_t stands for output to stabilize, ϕ_t is the observed disturbance vector, K is a given gain factor w.r.t. scalar control u_t, v_t is white noise, and θ^* is unknown parameter vector. Defining $y_t = Y_t - K u_t$ and using estimate $\theta_{t-1}(y_0^{t-1}, \phi_0^{t-1})$ of parameter θ^* to adjust the compensator $u_t = - K^{-1} \phi_t^T \theta_{t-1}$ one finds that control quality is determined by the choice of a plant identification algorithm (1) in the context of the estimation problem stated above.

Consider the distance function $d(\theta, P_t)$ from the point $\theta \in R^N$ to a rectangular strip $P_t = \{a \in R^N$: $|y_t - \phi_t^T a| \le \delta\}$. To investigate the algorithm convergence (3), let us use the following statement (no proof is supplied here) which is a direct consequence of geometric interpretation of algorithm (3) and distance function definition.

Lemma 1. Function $d(\theta, P_t)$ is convex, Lip $d(\theta, P_t)=$ $=1$, and

$$\sum_{t>0} d^2(\theta_{t-1}, P_t) \le ||\theta^* - \theta_0||^2,$$

$$|d^2(\theta, P_t) - |\phi_t^T(\theta^* - \theta)|^2 \, ||\phi_t||^{-2}| \le$$

$$\le 4\,\delta^2 ||\phi_t||^{-2} + 4\delta \, ||\theta^* - \theta|| \, ||\phi_t||^{-1}.$$

The statement that follows develops the idea of consistent estimator design without diminishing the identification algorithm step value.

Lemma 2. Let $\{\phi_t=\phi_0(T^t\omega)\}^{\infty}_{t=0}$ be a stationary ergodic sequence with values in R^N defined on $(\Omega, \Im, P)$, and the probability measure generated by the distribution function of this sequence is equivalent to Lebesgue measure on R^N. Then $\limsup_{t\to\infty} d(\theta, P_t)=$ $= ||\theta^* - \theta||$ *a.s. (almost surely)* $\forall\, \theta \in R^N$.

The restrictions assumed in Lemma 2 of the observed disturbances distribution class can be relaxed. However, the simple example of an estimation problem following the scheme (3) with a uniformly bounded observation sequence shows that without such restrictions the identifiability conditions are not met.

The following statement provides consistent estimation conditions for a stationary observed input variables sequence satisfying the weak dependence condition.

Theorem 1. Let under the conditions of Lemma 2 $\{y_t, \phi_t\}^{\infty}_{t=0}$ be an m–dependent stationary process in R^{N+1} and the following condition of limiting non-singularity holds

$$\limsup_{A\to\infty} tr\,\{\Gamma^{-1}(A)\} < \infty, \qquad (4)$$

$$\Gamma(A) = cov\,[\phi_t ||\phi_t||^{-1} \;/\; ||\phi_t|| > A\,].$$

Then the estimator (3) is strongly consistent, i.e., $\lim \theta_t = \theta^*$ *a.s. for* $t\to\infty$ $\forall \theta_0 \in R^N$.

Let us explain the limiting non-singularity condition (4). It is easy to give an example of a stationary observation sequence $\{\phi_t\}_{t=0}^{\infty}$ with the distribution concentrated on some proper sub-space of R^N (for large enough $||\phi_t||$). In this case, the condition (4) is not met and the global convergence of algorithm (3) cannot be guaranteed.

4. DYNAMIC PLANT IDENTIFICATION USING TEST SIGNALS

In the previous section the possibility of plant parameter consistent estimation was related to a large ratio of amplitudes of observed and unobserved disturbances. However, the global convergence of estimates (3) is principally impossible for limited observed variables. Such a situation occurs in the linear system identification with limited unmeasurable disturbance v_t. The system consists of discrete-time control plant

$$a(\nabla, \theta)Y_t = b(\nabla, \theta)U_t + v_t\,, \;\; a(0, \theta) = 1,$$
$$b(0, \theta) = 0\,, \;\; deg\, a = deg\, b = s \qquad (5)$$

and stabilizing controller

$$\alpha(\nabla, \theta)U_t=\beta(\nabla, \theta)Y_t,\; deg\,\alpha=deg\,\beta=s,\; \alpha(0)=1 \qquad (6)$$

where ∇ denotes one step delay operator. The data (Y_t, U_t) for $t<0$ are fixed. The class $\Re$ of stabilizing regulators (6) for plant (5) is large enough. In particular, one can obtain any stable polynomial $g(z)$, $deg\, g=2s$, $g(0)=1$ as characteristic polynomial of a closed-loop system, solving pole adjustment problem in class $\Re$ (Fomin *et al.,* 1981).

The problem of identification of plant (5) is to estimate vector $\theta = col\,(-a_1,\dots, -a_s;\, b_1,\dots, b_s)$ from observations "input – output". To have consistent estimates of the parameter, different schemes with control randomization by a measurable test signal are widely used (Saridis, 1977). Consider the randomized control system with plant equation

$$a(\nabla, \theta)y_t = b(\nabla, \theta)u_t + v_t, \qquad (7)$$

and randomized by test signal w_t feedback

$$\alpha(\nabla)u_t = \beta(\nabla)y_t + w_t\,,\; w_t = A_t e_t. \qquad (8)$$

In the right-hand side of (8) $e_t = \pm 1$ is the white noise Bernoulli sequence independent of plant disturbance. The amplitudes A_t are to meet the control admissibility conditions and to provide large enough discrepancy between plant and model outputs (overlapping the dead zone). It results in following conditions

$$\limsup_{t\to\infty} A_t = \infty, \qquad \sum_{t=1}^{T} A_t^2 = o(T).$$

Varying the test signal amplitude within these limits one can control the identification speed. The condition of randomized control admissibility limits the unit test signal power

$$\limsup_{T\to\infty} T^{-1} \sum_{t=1}^{T} E_P(y_t^2 + u_t^2) < \infty.$$

In the class of admissible controls the limit control quality functional can be defined:

$$J = \limsup_{T\to\infty} T^{-1} \sum_{t=1}^{T} E_P\, y_t^2. \qquad (9)$$

Both, the disturbances and the test signal are assumed to belong to a probability space $(\Omega, \Im, P)$, and control availability conditions and quality functional (9) ensemble averaging operation denotes the expectation w.r.t. measure P (in what follows, the subscript in notation E_p of this averaging will be omitted). The control admissibility condition is met also for unlimited but rare enough test signals.

Consider an auxiliary system S with output $u^*_t = u_t - U_t$ and input $y_t = y_t - Y_t$:

$$\begin{aligned} \alpha^*(\nabla, \theta)u^*_t &= \beta^*(\nabla, \theta)y^*_t + w_t, \\ a(\nabla, \theta)y^*_t &= b(\nabla, \theta)u^*_t. \qquad (S) \\ \alpha^*(\nabla, \theta) &= \alpha(\nabla) - \beta_0 b(\nabla, \theta), \\ \beta^*(\nabla, \theta) &= \beta(\nabla) - \beta_0\, a(\nabla, \theta) \end{aligned}$$

The characteristic polynomial of this system coincides with that of system (5, 6). “The controller” (second equation of (S)) has a measurement delay, and polynomials $a(.,.)$, $b(.,.)$ are co-prime due to the controllability of plant (7). The difference order of the “controller” due to (6) is not less than the plant order, which implies the controllability of matrix pair (A, B), $B=col(1, 0, ..., 0)$ in a standard closed loop system representation

$$\Delta x_{t+1} = A\,\Delta x_t + B\, w_{t+1}, \qquad (10)$$

$$\Delta x_t = col(u^*_t, \ldots, u^*_{t-s+1}, y^*_t, \ldots y^*_{t-s+1}),\ \Delta x_0 = 0.$$

Let a test signal be switched on during the steps $t(k)-2s+1, \ldots, t(k)$, $k=1, 2, \ldots;\ t(1) > 2s$. Iterating equation (10), taking into account the white noise w_t, and the conditional averaging, one obtains a matrix inequality (Fomin *et al.*, 1981)

$$E(\Delta x_t \Delta x^T_t / v_0^{\,t-2s}\, w_0^{\,t-2s} \geq \lambda \sigma_{t-2s+1}^2 I_{2s} \text{ a.s., } t=t(k), \quad (11)$$

where constant λ is positive due to the controllability of pair (A, B), and σ^2_{t-2s+1} is the variance of a test signal at the beginning of a new randomization cycle on step $t(k)-2s+1$. To derive (11), one is to take into account the assumed independence of the test signal and plant disturbance, and the monotonic growth of the test signal variance during a randomization cycle.

Let us consider the routine of the test signal amplitude and switching time control. Denote as $n(t)$ the total number of instances when the test signal is switched on during steps $0, 1, \ldots, t$. Define a set of natural numbers $M = \bigcup_{k>1} [t(k) - 2s+1, t(k)]$,

$t(k+1)>t(k)+2s$ with characteristic function $\chi_M(t)$, $t=1, 2, \ldots$ and define amplitudes as $A_t = A(t)\chi_M(t)$, where the monotonic number sequence $\{A(t), t=1, 2, \ldots\}$ meets the following conditions

$$A^2(t)=o(t/n(t)), \qquad \lim A(t)=\lim n(t)=\infty \quad (12)$$

which guarantees unbounded test signal amplitudes with zero relative power per step:

$$\sum_{t=1}^{T} Ew_t^2 \leq n(t)A^2(t) = o(t). \qquad (13)$$

The zero power condition is preserved when a test signal passes through a stable system (7, 8) providing availability (in the sense of (9)) of control for randomization scheme (8) and the same limit value of quadratic quality functional as in the system (5, 6) without randomization.

In relation to adopted randomization scheme note the following. It is known that in dynamic plant identification problems one can get convergence rate (in the LMS sense) of the order $o(t^{-1})$ even for smooth disturbances distribution, for example, using the LMS techniques for auto-regression parameter estimation for “heavy tails” noise distribution (see Hannan and Kanter (1977)). So to control the convergence rate, it is natural to use randomization scheme (8) where variances of test signals and time intervals between their switching on are increasing unlimitedly.

To estimate parameters of the plant (7) let us use the dead zone algorithm (3) with regressor $\phi_t = col\,(y_{t-1} \ldots y_{t-s}, u_{t-1} \ldots u_{t-s})$:

$$\phi^T_t\theta = [1 - a(\nabla, \theta)]y_t + b(\nabla, \theta)u_t.$$

The following statement not only validates the use of an identification algorithm for rather general assumptions of disturbances but also allows us to evaluate a convergence rate.

Theorem 2. Let the feedback (6) be stabilizing for control plant (5), and for control system (7, 8) the following conditions hold:
1. Bernoulli sequence of random variables $\{e_t\}$ and disturbance $\{v_t\}$, $t=0, 1, \ldots, \infty$ are independent.
2. Random variables sequence $\{v_t\}$ meets the condition $|v_t| < \delta$ a.s. with a priori given bound δ.
3. Test signal amplitudes A_t and the number $n(t)$ of its switching on during the steps $[t(k)-2s+1, t(k)]$, $0<t(k)<t+1$, $k=1, 2, \ldots$ meet the conditions:
$A_t = A(t)^\tau \chi_M(t)$, $A^2(t) = n(t)\ln n(t)$, $\lim n(t)=\infty$, $\lim A(t+1)/A(t) = 1$, $M = \bigcup_k [t(k) - 2s+1, t(k)]$.

Then:

1. For any bounded set $Q \subset R^{2s}$ *the estimate (3) uniformly w.r.t. initial condition* $\theta_0 \in Q$ *is strongly consistent, i.e.,*

$$\lim_{t\to\infty} \sup_{\theta_0 \in Q} \varphi_t(\theta_0)=0 \text{ a.s.}, \quad \varphi_t(\theta_0)=||\theta-\theta_t(\theta_0)||.$$

2. The following hold:

$$\liminf_{t\to\infty} [A^2(t)\, E\Delta_{t-2s}]<\infty, \quad \Delta_{t-2s} = ||\theta-\theta_{t-2s}||^2.$$

Corollary.

1. If under Theorem 2 conditions test signal meets condition $n(t)=(t/\ln t \ln\ln t)^{1/2}$ *for large enough t, then control strategy generated by feedback (7) is admissible.*

2. If under Theorem 2 conditions $n(t)=t/\ln\ln t$*, then*

$$\liminf_{t\to\infty}(t E\Delta_t)=0, \quad \Delta_t = ||\theta-\theta_t||^2.$$

Note that in the second statement of the Corollary, the control admissibility condition does not hold. The achievement of convergence rate of order $o(t^{-1})$ does not contradict to Cramer – Rao lower bound, as in the identification problem with randomized control strategy (8) the observations sample is not homogeneous.

5. CONCLUSION

The convergence analysis of stochastic relaxation procedure "dead zone" was based on the effect of algorithm linearization for large signal-to-noise ratio which offers an opportunity for consistent plant parameter estimation without diminishing the step value. The consistency is preserved also when the identifier is switched on only in randomization steps, which simplifies the adaptive identification strategy. Moreover, instead of test signal amplitude control program one can use feedback w.r.t. discrepancy. The suggested recurrent parameter estimation procedure is computationally simple and can be used for real-time control in systems with identifier in closed loop.

REFERENCES

Bunich, A.L. (1983). Fast convergent identification algorithm for linear plant with limited noise. *Avtomatika i Telemekhanika,* 8, 101 – 107 (in Russian; English translation: *Automation and Remote Control).*

Danilov, F.A. and N.S.Rajbman (eds.) (1973). *Adaptive Control of Tube Rolling Accuracy.* Metallurgiya, Moscow (in Russian).

Ermolyev, Yu.M. (1976). *Stochastic Programming Methods.* Nauka, Moscow.

Evans, R.J., C. Zhang and Y.G.Soh (1994). Bounded error estimation using dead zone and bounded ellipsoid. *Int. J. of Adaptive Contr. and Signal Processing,* **8,** 31 – 42.

Goodwin, G., P. Ramadage and P. Caines (1981). Discrete time stochastic adaptive control, *SIAM Jornal on Control and Optimization,* **19**, 829 - 853.

Fomin, V.N., A.L.Fradkov and V.A.Yakubovich (1981). *Adaptive Control of Dynamic Plants.* Nauka, Moscow (in Russian).

Hannan, E.J. and M. Kanter (1977). Autoregressive processes with infinite variance. *J. Appl . Probab*. **14**, 411 – 415.

Ibragimov, I.A. and R.Z.Khasminsky (1979). *Asymptotic Estimation Theory.* Nauka, Moscow (in Russian).

Kashiap, R.L. and A.R.Rao (1976). *Dynamic Stochastic Models from Empirical Data*. Academic Press, New York.

Kornfeld, N.P., Ya.G.Sinay and S.V.Fomin (1980). *Ergodic Theory.* Nauka, Moscow (in Russian).

Krzyak A. (1986). The rates convergence of kernel regression estimates and classification rules. *IEEE Trans. on Inf. Theory,* **IT – 32**, 5, 668 – 679.

Ljung L. (1987). *System Identidication*: *Theory for User.* Prentice – Hall, Englewood Cliffs, N.J.

Motzkin Th. and I.J. Schoenberg (1954). The relaxation method for linear inequalities. *Canadian J. of Math.*, **6,** 393 – 404.

Nazin, A.V. and A.B.Yuditsky (1999). Reachable information bounds in adaptive control problems for nonlinear stochastic plants. *Avtomatika i Telemekhanika, 3*, 180 – 196 (in Russian; English translation: *Automation and Remote Control).*

Poznyak, A.S. and S.N.Tikhonov (1990). Strong consistency of extended least mean squares with nonlinear discrepancy transform. *Avtomatika i Telemekhanika, 8*, 119 – 128 (in Russian; English translation: *Automation and Remote Control).*

Saridis, G. (1977). *Self-Organizing Control of Stochastic Systems.* Marcel Dekker, Inc., New York.

Shi, D.H. and F.Kozin (1986). An almost sure convergence of adaptive algorithms*. IEEE Trans. on Automatic Control,* **AC-31,** 471 – 474.

APPENDIX

Proof of Lemma 2. Let us use the inequality $(\theta, P_t) \geq$ $\geq [\phi_t^T(\theta - \theta^*) - 2\delta]||\phi_t||^{-1}$. Let $\varepsilon \in (0, 1)$ and $\chi_U(x)$ be a characteristic function of open set $U = \{x \in R^N: \; |x^T(\theta-\theta^*)|>2\delta+(1-\varepsilon)\, ||\theta-\theta^*||\, ||x||\}$, $\theta \neq \theta^*$. This set contains a point $x=\lambda(\theta-\theta^*)$, where $\lambda\varepsilon\|\theta-\theta^*\|^2>2\delta$ and so is not empty. Due to Poincaré theorem (Kornfeld *et al.*, 1980) $\limsup \chi_U(\phi_t)=1$ a.s., so $\limsup d(\theta, P_t)\geq (1-\varepsilon)||\theta - \theta^*\|$ a.s. As ε is arbitrary from (0, 1), $\limsup d(\theta, P_t) \geq|| \theta - \theta^*||$ and inverse inequality follows from definition of function $d(.,.)$.

Proof of Theorem 1. Due to Lemma 1, $Lip\ d(\theta, P_t)=1$ so for $t>m$ one has

$$\sum_{i=0}^{m}|d(\theta_{t-i-1}, P_t)-d(\theta_{t-i}, P_t)| \le \sum_{i=0}^{m}||\theta_{t-i-1}-\theta_{t-i}|| \le$$

$$\le m^{1/2}\{\sum_{i=0}^{m}||\theta_{t-i-1}-\theta_{t-i}||^2\}^{1/2},\ \sum_{t=m}^{\infty}d^2(\theta_{t-m}, P_t)\le m^2||\theta^*-\theta_0||^2,$$

$$\sum_{t=0}^{\infty}E[d^2(\theta_t,P_{t+m})/||\phi_{t+m}||>A]P(||\phi_{t+m}||>A)<\sum_{t=0}^{\infty}Ed^2(\theta_t,P_{t=m})<$$

$<\infty$. Together with the second inequality from Lemma 1 we obtain:

$$|E\{[d^2(\theta_t, P_{t+m})-|\phi_{t+m}^T(\theta^*-\theta_t)|^2||\phi_{t+m}||^{-2}]/||\phi_{t+m}||>A\}|\le$$

$4\delta^2A^{-2}+4\delta||\theta^*-\theta_0||A^{-1}$, from which it follows that

$$\limsup_{t\to\infty}|E\{|\phi_{t+m}^T(\theta^*-\theta_t)|^2||\phi_{t+m}||^{-2}]/||\phi_{t+m}||>A\}|=O(A^{-1})$$

Due to the condition of m-dependence of observations $\{\phi_t\}_{t=0}^{\infty}$ one gets $\limsup_{t\to\infty} tr\{\Gamma(A)cov(\theta^*-\theta_t)\} = O(A^{-1})$ consistency of estimator (3) follows from limit non-singularity of $\Gamma(A)$, and strong consistency follows from the monotonic character of $\|\theta^*-\theta_t\|$ sequence.

Proof of Theorem 2. Let us define states x_t, X_t of randomized system (7, 8) and non-randomized one (5, 6) respectively as:

$$x_t=\phi_{t+1} = col\ (y_t,\dots,y_{t-s+1}, u_t,\dots, u_{t-s+1}),$$

$$X_t = col\ (Y_t,\dots,Y_{t-s+1}, U_t,\ \dots,\ U_{t-s+1}),\ \delta_t=x_t-X_t. \tag{A.1}$$

Let us use a matrix inequality

$$E(\delta_t\delta_t^T / v_0^{t-2s}, w_0^{t-2s}) \ge \lambda\sigma_{t-2s+1}^2 I_{2s} \text{ a.s.}, t=t(k), \tag{A.2}$$

where $\sigma_{t-2s+1}^2 = A_{t-2s+1}^2$ stands for test signal variance at the beginning of a randomization cycle, constant $\lambda > 0$ is minimal eigenvalue of matrix

$\sum_{k=0}^{2s-1} A^k BB^T(A^k)^T$, which is positive definite due to controllability of pair (A, B), and $t = t(k)$. As the disturbance v_t is limited, random variables $\|X_t\|$ are also bounded by non-random constant, depending on initial conditions, so from (A.2) one gets

$$E(x_t /v_0^{t-2s}, w_0^{t-2s})\ge(C_1A_t^2-C_2A_t-C_3)I \text{ a.s.}, t=t(k) \tag{A.3}$$

with certain positive constants C_i, $i=1, 2, 3$. As the estimate θ_{t-2s} is F_t–measurable with F_t being a σ-algebra generated by random variable (v_0^{t-2s}, w_0^{t-2s}), from (A.3) denoting $z_t = x_t^T(\theta - \theta_{t-2s})$, one gets for $t=t(k)$:

$$Ez_t^2 \ge (C_1A_t^2 - C_2A_t - C_3)\Delta_{t-2s} \tag{A.4}$$

Introduce random variables $\eta_t=|f(z_t + v_{t+1})|\ ||x_t||^{-1}$, $Z_t = x^T_t(\theta - \theta_t)$. So $E\eta_t^2 = o(1)$, as $E\{f^2(v_{t+1}+Z_t)||x_t||^{-2}\} \to 0$ for $t \to \infty$ and

$$|f(z_t+v_{t+1})-f(Z_t+v_{t+1})|\ ||x_t||^{-1}\le||\theta_t-\theta_{t-2s}||\le\sum_{i=0}^{2s-1}||\theta_{t-i}-\theta_{t-i-1}||, \tag{A.5}$$

where random variables $||\theta_{t-i} - \theta_{t-i-1}||$ have zero limit a.s. for $t \to \infty$ and are majored by the constant $2\|\theta-\theta_0\|$. As $|z_t|\le 2\delta+\eta_t||x_t||$ and $||x_t|| \le C_4A_t + C_5$, $C_{4,5} > 0$ (easily derived from (12)), accounting for (A.4) and the monotonic property of the sequence $\|\theta-\theta_t\|$, the strong consistency of estimator θ_t follows. Moreover, as this estimator due to (3) is a.s. continuous function of initial value θ_0, the monotonic sequence of continuous functions $\varphi_t(\theta_0)= \|\theta-\theta_t(\theta_0)\|$ converge to zero uniformly over any compact set Q from R^{2s}, for $t\to\infty$.

Let us now evaluate convergence rate.

Lemma 3. For control system (7, 8) the following inequality hold

$$C_4A_t+C_5\ge\|\underline{x_t}\|\quad C_4,\ C_5>0,\ t=t(k),\ k>1.$$

(The proof is rather straightforward, but tiresome and is not given due to lack of space).

From Lemma 3 and (A.5), denoting

$$\zeta_t=|f(Z_t+v_{t+1})|\ \|x_t\|^{-1}+ \sum_{i=0}^{2s-1}\|\theta_{t-i}\theta_{t-i-1}\|$$

one gets

$$\zeta_t[C_4A(t)+C_5]+2\delta\ge|z_t|,\quad C_6\ge \sum_{t=0}^{\infty}\zeta_t^2,$$

$C_7+C_8A^2(t)E\zeta_t^2 \ge Ez_t^2$ with some nonrandom constants C_i. As $\Sigma\ (k\ln k)^{-1}$ diverges, one has $\liminf\ (n\ln n\ E\zeta_{t(n)}^2) = 0$ for $n\to\infty$, so $C_7 \ge \liminf Ez_{t(n)}^2$ for $A^2(t)= n(t)\ln n(t)$ and from (A.4) it follows $\liminf\ [A^2(t)Ez_{t(n)}^2]<\infty$ for $t\to\infty$. Moreover, for $A^2(t) = n(t)\ln n(t)$, $n(t)= (t/\ln t\ \ln\ln t)^{1/2}$ the control admissibility condition (13) holds. When $n(t) = t/\ln\ln t$, $A^2(t)=n(t)\ln n(t)$ control is not admissible but $A^2(t)t^{-1}\to\infty$ for $t\to\infty$ what implies given in the Corollary evaluation: $\liminf\ (tE\Delta_t)=0$ for $t\to\infty$.

Note that randomization together with (13) does not diminish control quality in the sense of (9).