Linear Regression and Filtering Under Nonstandard Assumptions (Arbitrary Noise)

Oleg Granichin

***Abstract*—This note is devoted to parameter estimation in linear regression and filtering, where the observation noise does not possess any "nice" probabilistic properties. In particular, the noise might have an "unknown-but-bounded" deterministic nature. The basic assumption is that the model regressors (inputs) are random. Optimal rates of convergence for the modified stochastic approximation and least-squares algorithms are established under some weak assumptions. Typical behavior of the algorithms in the presence of such deterministic noise is illustrated by numerical examples.**

***Index Terms*—Filtering, linear regression, parameter estimation, prediction, randomized algorithm.**

I. Introduction

This note is concerned with filtering and parameter estimation of a class of linear models described later. Let θ_n, $n = 0, 1, \ldots$, denote a signal process not directly observable, and let φ_n be a known process. Let y_n denote an observation process that is given by

$$y_n = \varphi_n^{\mathrm{T}} \theta_n + \text{observation noise}.$$

The objective of the problems considered in this note is to estimate θ_{n+1} based on the observation up to time n.

Traditionally, the observation noise is assumed to be a mutually independent and zero-mean. These assumptions are often hard to justify in practice and without them, the validity of many algorithms is questionable in engineering applications. For example, it is known that the standard "least-squares method" or the "maximum likelihood method" give wrong estimates if the observation noise has an "unknown-but-bounded" deterministic nature or it is a probabilistic "dependent" sequence (the enemy jams the signal). Therefore, it is important to investigate the capability of filtering or linear-regression (LR) parameter estimation under minimal assumptions on the statistical characteristics of the observation noise.

The main contribution of this note is the try to avoid the standard requirements on the observation noise. In the case of noncentered correlated noise and even nonrandom noise, the LR parameters can be efficiently estimated although it seems surprising at first glance. This can be done under certain conditions when the inputs (regressors) φ_n are random and available. Moreover, the optimal LR parameter estimation algorithms have the same rates of convergence as in the "standard" case. The idea of using random inputs in order to eliminate the bias effect was suggested by Fisher [1] as the randomization principle in experiment design. Apart from the experiment design problem where the regressors can be randomized by an experimenter, random inputs occur in many problems of identification, filtration, recognition, maneuvered target tracking, telecommunications, and manufacturing.

Recursive algorithms for LR parameter estimation were considered for the case of random inputs in many works; see [2]–[7]. Polyak and Tsypkin [2], [6] studied the rate of convergence of such algorithms. They obtained the optimal algorithms with the best possible rates of convergence. All these papers made standard assumptions on the noise process, namely, the noise was assumed to be a sequence of mutually independent or weakly dependent random variables with zero-mean.

In the case of *unknown but bounded* nonrandom noises, the minimax problem or H_∞ approach were usually considered; see, for example, [8]–[10]. The advantage of these approaches is that they do not require any specific assumptions on statistic properties of the noise. The disadvantage is, however, that the accuracy of the estimation depends on the noise level directly. The quality of estimates is not good when the noise level is high.

In [11]–[14], the problem of LR parameter estimation was considered under nonstandard assumptions on the observation noise. Goldenshluger and Polyak [12] studied this problem with an arbitrary noise for case of centered random input signals and time-invariant LR parameters. But the algorithms proposed in [12] do not achieve the optimal rate of convergence in the general case. For the case of almost arbitrary noise sequences, Granichin [11], [14] considered the problem with a time-varying vector of unknown parameters. The author presented estimation algorithms for the mean vector of unknown parameters. The possibility of getting strongly consistent parameter estimates was also discussed by Ljung and Guo [13] when the noises were bounded and deterministic and the input sequence was suitably chosen.

The filtering problem in the case of random inputs in an observation channel was considered by Zhang [15], [16] with non-Gaussian disturbances and by Granichin [17] with almost arbitrary noise.

This note is organized as follows. In Section II, we formulate the LR problem under consideration and state the main assumptions on the inputs (regressors) and the noises (disturbances). In Section III, we present and study the randomized stochastic approximation and least-squares algorithms for the estimation of a mean vector of the unknown LR parameters. Conditions for almost sure and mean-square convergence of estimates are given in Theorems 1 and 3, respectively. An upper bound on the mean-square convergence rate is given in Theorem 2 for the estimates of the randomized stochastic approximation algorithm. In Section IV, we study the problem of linear filtering in the case of random inputs in an observation channel and mixed type uncertainties. We propose to use the randomized least mean squares algorithm for a special case of the prediction problem. An upper bound on the squared mean value of the prediction error is given in Theorem 4. It can be made small by an appropriate choice of the probabilistic distribution of inputs. In Section V, we give numerical simulation examples comparing the performance of our schemes with the standard LMS or KF estimates. When the observation noise is bounded but does not satisfy the standard statistical properties, the numerical results indicate the two important facts: the schemes suggested in this note often outperform standard algorithms; the averaged prediction errors are significantly less than the squared observation noise level. In the Appendix, we go present the proofs of main results.

II. LR Problem Formulation and Main Assumptions

Consider the linear regression model

$$y_n = \varphi_n^{\mathrm{T}} \theta_n + v_n \quad \theta_n = \theta + w_n, \qquad n = 0, 1, \ldots \tag{1}$$

where $y_n \in \mathrm{R}^1$ is an observation output made at time n, the input vector $\varphi_n \in \mathrm{R}^d$ is available at time n, $v_n \in \mathrm{R}^1$ and $w_n \in \mathrm{R}^d$ represent noises (disturbances). The objective is to find an estimates sequence $\{\hat{\theta}_n\}$ which converges to the vector θ of unknown parameters. Each estimate $\hat{\theta}_n$ should be based on the observations $y_i, \varphi_i, i \leq n$.

We adopt the following notations. The Euclidean norm of a vector x from $\mathbb{R}^d$ is denoted by $\|x\|$. $\mathrm{E}\{\cdot\}$ is the mathematical expectation

Manuscript received March 23, 2002; revised May 2, 2003, December 13, 2003, February 2, 2004, and April 23, 2004. Recommended by Associate Editor R. Srikant.

The author is with the Department of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg 198904, Russia (e-mail: Oleg_granichin@mail.ru).

Digital Object Identifier 10.1109/TAC.2004.835585

symbol. The trace of a matrix A is denoted by Tr[A]. $A > 0$ means that A is a positive definite matrix. The maximum (minimum) eigenvalue of A is denoted by $\lambda_{\max}(A)(\lambda_{\min}(A))$, and the Euclidean norm of A is defined as its maximum singular value, i.e.,

$$\|A\| = \sqrt{\lambda_{\max}(AA^{\mathrm{T}})}.$$

Let $\mathcal{F}_n$ be the σ-algebra of probabilistic events generated by $\{\varphi_0, \ldots, \varphi_n, w_0, \ldots, w_n, v_1, \ldots, v_n\}$, $\hat{\mathcal{F}}_n$ be the σ-algebra generated by $\{\varphi_0, \ldots, \varphi_n, w_0, \ldots, w_n, v_0, \ldots, v_{n+1}\}$ and $\tilde{\mathcal{F}}_n$ be the σ-algebra generated by $\{\varphi_0, \ldots, \varphi_n, w_0, \ldots, w_{n+1}, v_0, \ldots, v_{n+1}\}$, $\mathcal{F}_{n-1} \subset \hat{\mathcal{F}}_{n-1} \subset \tilde{\mathcal{F}}_{n-1} \subset \mathcal{F}_n$.

We make the following assumptions.

A) The inputs $\{\varphi_n\}$ form a sequence of independent identically distributed random vectors with known bounded expectation $\|\mathrm{E}\{\varphi_n\}\| = M_\varphi < \infty$. For each n the vector φ_n is independent of $\tilde{\mathcal{F}}_{n-1}$, and the centered random vector $\Delta_n = \varphi_n - \mathrm{E}\{\varphi_n\}$ have a symmetric distribution function $\mathrm{P}(\cdot)$ (i.e. $\mathrm{P}(\Omega) = \mathrm{P}(-\Omega)$ for any Borel set $\Omega \subset \mathbb{R}^d$), and $\mathrm{E}\{\Delta_n \Delta_n^{\mathrm{T}}\} = \mathrm{B} > 0$, $\mathrm{E}\{\|\Delta_n\|^4\} \le M_4 < \infty$.

B) $\forall n$ $\mathrm{E}\{w_n\} = 0$, w_n is independent of $\hat{\mathcal{F}}_{n-1}$, $\{v_n\}$ and $\{w_n\}$ satisfy either

i)

$$\mathrm{E}\left\{v_n^2 | \mathcal{F}_{n-1}\right\} \le \sigma_v^2 < \infty, \text{ a.s., } \mathrm{E}\left\{\|w_n\|^2\right\} \le \sigma_w^2 < \infty$$

or

ii)

$$\mathrm{E}\left\{v_n^2\right\} \le \sigma_v^2 < \infty, \ \mathrm{E}\left\{w_n w_n^{\mathrm{T}}\right\} \le \mathrm{Q}_w < \infty$$

conditions, where σ_v, σ_w are some constants, Q_w is a symmetric matrix.

Note, that in the LR parameters estimation problem with random inputs $\{\varphi_n\}$ the standard requirements on the observation noise $\{v_n\}$ are somewhat different (see, e.g., [2]). In particular, it is usually assumed that $\{v_n\}$ is a sequence of identically distributed random variables with zero-mean and v_n is independent of each other and φ_n.

III. LR Parameters Estimation

We first examine a randomized algorithm of the stochastic approximation type for the observation model (1)

$$\hat{\theta}_n = \hat{\theta}_{n-1} - \alpha_n \Gamma \Delta_n \left(\varphi_n^{\mathrm{T}} \hat{\theta}_{n-1} - y_n\right), \qquad n = 1, 2, \ldots \tag{2}$$

where $\alpha_n \ge 0$ is a nonrandom step-size and Γ is a positive definite symmetric matrix. We assume that the initial estimate $\hat{\theta}_0$ is an arbitrary nonrandom vector from $\mathbb{R}^d$.

Theorem 1: Let Assumption A) be satisfied for the model (1) inputs and

$$\sum_{n=1}^{\infty} \alpha_n = \infty, \qquad \alpha_n \to 0 \quad \text{as} \quad n \to \infty. \tag{3}$$

If Assumption Bi) holds for the model (1) noises and $\sum_{n=1}^{\infty} \alpha_n^2 < \infty$, *then* for estimates generated by the algorithm (2) we have $\hat{\theta}_n \to \theta$ a.s. as $n \to \infty$.

If Assumption Bii) holds, *then* $\mathrm{E}\{(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)^{\mathrm{T}}\} \to 0$ as $n \to \infty$.

The following theorem establishes the rate of convergence of the sequence of estimates generated by (2).

Theorem 2: Let Assumptions A) and Bii) be fulfilled, $\alpha_n = n^{-1}$, and $-\Gamma\mathrm{B} + (1/2)\mathrm{I}$ be a Hurwitz matrix, i.e., all its eigenvalues lie in the left half-plane.

Then, for the estimates of (2), we have

$$\mathrm{E}\left\{(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)^{\mathrm{T}}\right\} \le n^{-1}\mathrm{S} + o(n^{-1}) \tag{4}$$

where the matrix S is the solution of the matrix equation

$$\begin{aligned} \Gamma\mathrm{BS} + \mathrm{SB}\Gamma - \mathrm{S} &= \Gamma\mathrm{R}\Gamma \\ \mathrm{R} &= \left(\sigma_v^2\left(1 + M_\varphi^2 \rho\right) + M_\varphi^2 \mathrm{Tr}[\mathrm{Q}_w]\right)\mathrm{B} \\ &\quad + \mathrm{E}\left\{\Delta_n \Delta_n^{\mathrm{T}} \mathrm{Q}_w \Delta_n \Delta_n^{\mathrm{T}}\right\} \end{aligned} \tag{5}$$

with any $\rho > 0$.

Proofs of Theorems 1 and 2 are given in the Appendix.

If $\Gamma = \mathrm{B}^{-1}$, $\mathrm{Tr}[\mathrm{Q}_w] = 0$ and $M_\varphi = 0$, then the (5) for the matrix S can be explicitly solved: $\mathrm{S} = \mathrm{B}^{-1}\mathrm{RB}^{-1}$. For (2), which assumes the form

$$\hat{\theta}_n = \hat{\theta}_{n-1} - (n\mathrm{B})^{-1}\varphi_n \left(\varphi_n^T \hat{\theta}_{n-1} - y_n\right) \tag{6}$$

we get here

$$\mathrm{E}\left\{(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)^{\mathrm{T}}\right\} \le n^{-1}\sigma_v^2 \mathrm{B}^{-1} + o(n^{-1}).$$

For the last algorithm (6), we have obtained almost the same convergence rate as the best possible rate in the case where the noise v_n is an independent random variable with zero-mean; see [2]. Moreover, this choice of α_n and Γ has been shown in [2] is an optimal for the similar kind algorithms.

Remark: If Assumption Bii) in Theorem 2 holds as an equality, the inequality in the bound on the convergence rate can also be replaced by an equality.

For the same regression model of observations (1), we consider now the estimates generated by the following randomized least squares method:

$$\begin{cases} \hat{\theta}_n = \hat{\theta}_{n-1} - \Gamma_n \Delta_n \left(\varphi_n^{\mathrm{T}} \hat{\theta}_{n-1} - y_n\right) \\ \Gamma_n = \Gamma_{n-1} - \frac{\Gamma_{n-1}\Delta_n \Delta_n^T \Gamma_{n-1}}{\left(1 + \Delta_n^T \Gamma_{n-1} \Delta_n\right)}, \ \Gamma_0 = \gamma_0^{-1}\mathrm{I} \end{cases} \tag{7}$$

where $\gamma_0 > 0$ is some regularization parameter; see [3] and [4]. We assume again that the initial estimate $\hat{\theta}_0$ is an arbitrary nonrandom vector from $\mathbb{R}^d$.

Theorem 3: Let Assumption A) be fulfilled.

If Assumption Bi) is satisfied, *then* for the estimates generated by the algorithm (7) we have $\hat{\theta}_n \to \theta$ a.s. as $n \to \infty$.

If $|v_n| \le C_v$, $\|w_n\| \le C_w$, $\|\Delta_n\| \le C_\Delta$ a.s., then for the algorithm (7), $\mathrm{E}\{(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)^{\mathrm{T}}\} \to 0$ as $n \to \infty$. Here, $C_v, C_w, C_\Delta < \infty$ are constants.

The proof of Theorem 3 is also given in Appendix.

IV. Prediction of Signal (Filtering)

We confine our consideration to the following special case of the filtering problem: the one-step prediction when the observations are governed by equations

$$y_n = \varphi_n^{\mathrm{T}} \theta_n + v_n, \qquad n = 0, 1, \ldots. \tag{8}$$

Here, $y_n \in \mathbb{R}^1$ is an observation made at time n and φ_n is a d-dimensional vector that is known at time n, $v_n \in \mathbb{R}^1$ represents an and a vector process $\{\theta_n\}$, $\theta_n \in \mathbb{R}^d$ is generated by a zero-mean white wide-sense noise sequence $\{w_n\}$ through a stable linear filter

$$\theta_{n+1} = \mathrm{A}\theta_n + w_{n+1} \quad \theta_0 \in \mathbb{R}^d \tag{9}$$

where A is a known stable matrix (i.e., $\|\mathrm{A}\| < 1$).

The objective is to find an estimate of θ_{n+1} based on the observations y_i, φ_i up to time n. Let $\hat{\theta}_{n+1}$ be the current estimate of the vector θ_{n+1}. Quality of prediction (performance of filtering) is defined by the mean squared mismatch

$$\mathrm{E}\left\{\|\hat{\theta}_{n+1} - \theta_{n+1}\|^2\right\}.$$

It is usually assumed that the vectors $\{\varphi_n\}$ in the observation model (8) are defined by a deterministic sequence. Here, we will suppose that they are random and satisfy Assumption A).

We study the behavior of the estimates of the randomized least mean squares (RLMS) algorithm

$$\hat{\theta}_{n+1} = \mathrm{A}\hat{\theta}_n - \alpha\mathrm{A}\Gamma\Delta_n\left(\varphi_n^{\mathrm{T}}\hat{\theta}_n - y_n\right), \qquad \Delta_n = \varphi_n - M_\varphi \tag{10}$$

where $n = 0, 1, \ldots, \alpha > 0$ is a step-size, and Γ is a positive–definite symmetric matrix. We suppose that the initial value $\hat{\theta}_0$ is a vector from $\mathbb{R}^d$.

By substituting (8) and (9) into (10), we can write the prediction error as follows:

$$\begin{aligned}\hat{\theta}_{n+1} - \theta_{n+1} = \mathrm{A}\left(\mathrm{I} - \alpha\Gamma\Delta_n\Delta_n^{\mathrm{T}}\right)(\hat{\theta}_n - \theta_n) \\ -\alpha\mathrm{A}\Gamma\Delta_n\left(\mathrm{E}\{\varphi_n\}^{\mathrm{T}}(\hat{\theta}_n - \theta_n) - v_n\right) - w_{n+1}.\end{aligned}$$

Suppose that Assumptions A) and Bii) are satisfied. We carry out conditional averaging in the σ-algebra $\mathcal{F}_n$ and $\tilde{\mathcal{F}}_{n-1}$ successively. By virtue of the independence of the random vectors Δ_n and w_{n+1} and the assumption regarding the symmetric nature of the distribution $\mathrm{P}(\cdot)$, we conclude that

$$\begin{aligned}&\mathrm{E}\left\{\|\hat{\theta}_{n+1} - \theta_{n+1}\|^2|\tilde{\mathcal{F}}_{n-1}\right\} \\ &\quad\le \left(1 - 2\alpha\lambda_{\min}(\mathrm{B}\Gamma) + \alpha^2\|\Gamma\|^2 M_4^4\right)\|\mathrm{A}\|^2\|\hat{\theta}_n - \theta_n\|^2 \\ &\quad+ \alpha^2\left(\mathrm{E}\{\varphi_n\}^{\mathrm{T}}(\hat{\theta}_n - \theta_n) - v_n\right)^2\|\Gamma\|^2\mathrm{Tr}[\mathrm{B}] + \mathrm{Tr}[\mathrm{Q}_w].\end{aligned}$$

By taking the unconditional expectation on both sides of the last formula, for any $\rho > 0$, we obtain the following bound of the mean value of the prediction error:

$$\begin{aligned}\mathrm{E}\left\{\|\hat{\theta}_{n+1} - \theta_{n+1}\|^2\right\} \le b(\alpha,\rho)\mathrm{E}\left\{\|\hat{\theta}_n - \theta_n\|^2\right\} \\ +\alpha^2(1 + M_\varphi\rho)\|\Gamma\|^2\mathrm{Tr}[\mathrm{B}]\sigma_v^2 + \mathrm{Tr}[\mathrm{Q}_w]\end{aligned}$$

where

$$\begin{aligned}b(\alpha,\rho) = \left(1 - 2\alpha\lambda_{\min}(\mathrm{B}\Gamma) + \alpha^2\|\Gamma\|^2 M_4\right)\|\mathrm{A}\|^2 \\ +\alpha^2\left(M_\varphi + \frac{1}{\rho}\right)M_\varphi\|\Gamma\|^2\mathrm{Tr}[\mathrm{B}].\end{aligned} \tag{11}$$

The last inequality gives rise to the following theorem.

Theorem 4: Let the sequences $\{y_n\}, \{\varphi_n\}, \{v_n\}, \{\theta_n\}$, and $\{w_n\}$ be related by (8), (9), $\alpha > 0$, and $\Gamma > 0$.

If Assumptions A) and Bii) are satisfied, *then* for the prediction errors of the estimates $\{\hat{\theta}_n\}$ generated by the algorithm (10), the inequalities

$$\begin{aligned}\mathrm{E}\left\{\|\hat{\theta}_{n+1} - \theta_{n+1}\|^2\right\} \le \frac{\mathrm{Tr}[\mathrm{Q}_w] + \alpha^2(1 + M_\varphi\rho)\|\Gamma\|^2\mathrm{Tr}[\mathrm{B}]\sigma_v^2}{1 - b(\alpha,\rho)} \\ +b(\alpha,\rho)^n\mathrm{E}\left\{\|\hat{\theta}_0 - \theta_0\|^2\right\}\end{aligned}$$

are satisfied for any $\rho > 0$ and sufficiently small α such that $b(\alpha,\rho) < 1$. Here, the constant $b(\alpha,\rho)$ is determined by (11).

Theorem 4 allows us to study dependence of the filtering performance on the algorithm step-size α. Let us suppose that $\Gamma = \mathrm{B}^{-1}$, $\mathrm{E}\{\|\hat{\theta}_0 - \theta_0\|^2\} = 0$, $\|\mathrm{A}\|^{-2} = 1 + \mathcal{O}(\alpha^3)$ and α is a sufficiently small. We denote

$$r(\rho) = \frac{\left(M_4 + \left(M_\varphi + \frac{1}{\rho}\right)M_\varphi\mathrm{Tr}[\mathrm{B}]\right)\lambda_{\min}^{-2}(\mathrm{B})}{2}.$$

Then, from Theorem 4, we obtain

$$\mathrm{E}\left\{\|\hat{\theta}_{n+1} - \theta_{n+1}\|^2\right\} \le D(\alpha,\rho) + \mathcal{O}(\alpha^2)$$

where

$$\begin{aligned}D(\alpha,\rho) = \frac{1}{2}\mathrm{Tr}[\mathrm{Q}_w] \\ \times\left(\frac{1}{\alpha} + r(\rho) + \left(r(\rho)^2 + \frac{(1 + M_\varphi\rho)\mathrm{Tr}[\mathrm{B}]\sigma_v^2}{\mathrm{Tr}[\mathrm{Q}_w]\lambda_{\min}^2(\mathrm{B})}\right)\alpha\right).\end{aligned} \tag{12}$$

The last expression roughly indicates the tradeoff between the filtering ability and the noise sensitivity. In the case of $M_\varphi = 0$, a similar result can be obtained as a corollary to [8, Th. 4] for a tracking problem.

By minimizing $D(\alpha,\rho)$ in α and ρ, we establish

$$\alpha^\star = \left(r(\rho^\star)^2 + \frac{(1 + M_\varphi\rho^\star)\mathrm{Tr}[\mathrm{B}]\sigma_v^2}{\mathrm{Tr}[\mathrm{Q}_w]\lambda_{\min}^2(\mathrm{B})}\right)^{-\frac{1}{2}}$$

where $\rho^\star$ is the minimum point of the function

$$\begin{aligned}\bar{D}(\rho) = \frac{1}{2}\mathrm{Tr}[\mathrm{Q}_w] \\ \times\left(r(\rho) + 2\sqrt{r(\rho)^2 + \frac{(1 + M_\varphi\rho)\mathrm{Tr}[\mathrm{B}]\sigma_v^2\lambda_{\min}^{-2}(\mathrm{B})}{\mathrm{Tr}[\mathrm{Q}_w]}}\right).\end{aligned} \tag{13}$$

If $M_\varphi = 0$, then the function $D(\alpha,\rho)$ is independent of ρ. In this case, we have

$$\alpha^\star = 2\lambda_{\min}^2(\mathrm{B})\sqrt{\frac{\mathrm{Tr}[\mathrm{Q}_w]}{M_4^2\mathrm{Tr}[\mathrm{Q}_w] + 2\lambda_{\min}^2(\mathrm{B})\mathrm{Tr}[\mathrm{B}]\sigma_v^2}}$$

and

$$\bar{D}^\star = \frac{\mathrm{Tr}[\mathrm{Q}_w]}{4\lambda_{\min}^2(\mathrm{B})}\left(M_4 + 2\sqrt{M_4^2 + \frac{2\lambda_{\min}^2(\mathrm{B})\mathrm{Tr}[\mathrm{B}]\sigma_v^2}{\mathrm{Tr}[\mathrm{Q}_w]}}\right).$$

From this equation, we easily conclude that it is best to use the probabilistic Bernoulli distribution on (± 1) for simulating the random vectors φ_n when they can be chosen arbitrarily from the d-dimensional cube $[-1, 1]^d$.

Let $d = 1$, $\sigma_w^2 = \mathrm{Tr}[\mathrm{Q}_w] \ll \sigma_v^2$, let $\{\varphi_n\}$ be a scalar Bernoulli independent process (equal to $\pm\bar{\varphi}$ with the same probability). Then, $\alpha^\star\mathrm{A}\Gamma \approx \sigma_w/|\bar{\varphi}|\sigma_v$. Note that this value coincides approximately with the limiting value of the Kalman coefficient for the optimal Kalman

TABLE I
AVERAGED ERRORS OF VARIOUS ALGORITHMS

	$\tilde{D}((14))$	$\tilde{D}((15))$	$\tilde{D}((16))$
$v_n = 4.0 * (rand() - 0.5)$	0.5309	0.1803	0.1256
$v_n = 0.1 * sin(n) + 1.9 * sign(50 - n \bmod 100)$	0.5700	2.8254	2.2640
$v_n = 2.0$	0.5954	3.1387	2.5335
$v_n = -2.0$	0.7826	3.4989	3.9582

filter when the observation noises $\{v_n\}$ are independent and equal to $\pm\sigma_v$ with the identical probability.

V. NUMERICAL EXAMPLES

Let us discuss the example of applying the RLMS algorithm (10) to the prediction problem for the scalar ($d = 1$) signal $\{\theta_n\}$ generated through a stable linear filter (9) with $\mathrm{A} = 0.9999$ and $\theta_0 = 0$ by an independent process $\{w_n\}$ which is uniformly distributed over the interval $[-(1/3), 1/3]$: $\mathrm{E}\{w_n\} = 0$, $\mathrm{E}\{w_n^2\} = 2/81$. The quantities y_n and φ_n are available at each time n. They are related to the signal θ_n by the (8) with the bounded noise (disturbance): $|v_n| \leq 2$.

Random variables generated by the uniform distribution over the interval [0.5; 1.5] were used in simulating the sequence $\{\varphi_n\}$. The observations were made over the interval from $n = 1$ to 199. The filtering quality was defined as the averaged squared mismatch

$$\tilde{D}\left(\{\hat{\theta}_n\}\right) = \frac{1}{199}\sum_{n=1}^{199} \|\hat{\theta}_n - \theta_n\|^2.$$

Minimization of the function $\bar{D}(\rho)$ of (13) in ρ provides $\rho^\star = 0.269$. Hence, for the RLMS algorithm (10) we obtain the optimal step-size to be $\alpha^\star = 11.3808$, $\Gamma = 1/48$ and the bound on the filtering error is $\bar{D}^\star = 1.3699$. Note, this bound is less than $\sigma_v^2 = 4$.

In the presence of different kinds of observation noises, we compare the prediction errors of estimates generated by the following three algorithms:

- RLMS

$$\hat{\theta}_{n+1} = 0.9999\left(\hat{\theta}_n - 0.2371(\varphi_n - 1.0)(\varphi_n\hat{\theta}_n - y_n)\right) \tag{14}$$

- LMS

$$\hat{\theta}_{n+1} = 0.9999\left(\hat{\theta}_n - 0.2371\varphi_n(\varphi_n\hat{\theta}_n - y_n)\right) \tag{15}$$

- Kalman filter (KF)

$$\begin{aligned}\hat{\theta}_{n+1} &= 0.9999\hat{\theta}_n - k_n\varphi_n(\varphi_n\hat{\theta}_n - y_n)\\ k_n &= \frac{0.9999\gamma_{n-1}}{\frac{16}{3} + \gamma_{n-1}\varphi_n^2}\\ \gamma_n &= \gamma_{n-1}0.9999^2 - \frac{\varphi_n^2\gamma_{n-1}^2}{\frac{16}{3} + \gamma_{n-1}\varphi_n^2} + \frac{2}{81}\\ \gamma_0 &= 0.\end{aligned} \tag{16}$$

The numerical results are summarized in Table I. The Kalman filter (16) is known to provide the optimal estimates in the case of a Gaussian independent observation noise, LMS estimates (15) are sufficiently efficient for the independent zero-mean noises $\{v_n\}$ and $\{w_n\}$. Therefore, for the independent zero-mean noises the estimates generated by the algorithms (15) and (16) exhibit rather good behavior despite the high level of the observation noise. For the constant unknown observation noise or zero-mean, but insufficiently "diverse" noise, the prediction errors of the algorithms (15) and (16) are comparable with the squared observation noise level. At the same time, in all situations, the performance index of RLMS estimates is noticeably less than the squared observation noise level.

VI. CONCLUDING REMARKS

For random inputs, the requirements on the observation noise to ensure convergence are very moderate for the proposed algorithms. In particular, the noise is allowed to have "unknown-but-bounded" deterministic nature. For this reason, these algorithms can be useful in many applications. Numerical simulation has demonstrated their efficiency for various kinds of noises.

APPENDIX

In this appendix, we give the proofs of Theorems 1–3. For notational simplification, we denote $\eta_n = \hat{\theta}_{n-1} - \theta_n$, $\xi_n = v_n - \mathrm{E}\{\varphi_n\}^{\mathrm{T}}\eta_n$, $\mathrm{D}_n = (\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)^{\mathrm{T}}$.

Proof of Theorem 1

From (1) and (2), we can obtain

$$\begin{aligned}\|\hat{\theta}_n - \theta\|^2 = &\left((\hat{\theta}_{n-1} - \theta)^{\mathrm{T}} - \eta_n^{\mathrm{T}}\alpha_n\Delta_n\Delta_n^{\mathrm{T}}\Gamma\right)\\ &\times\left((\hat{\theta}_{n-1} - \theta) - \alpha_n\Gamma\Delta_n\Delta_n^{\mathrm{T}}\eta_n\right)\\ &+ \alpha_n^2\xi_n^2\Delta_n^{\mathrm{T}}\Gamma^2\Delta_n\\ &+ \left((\hat{\theta}_{n-1} - \theta)^{\mathrm{T}} - \eta_n^{\mathrm{T}}\alpha_n\Delta_n\Delta_n^{\mathrm{T}}\Gamma\right)\alpha_n\xi_n\Gamma\Delta_n\\ &+ \alpha_n\xi_n\Delta_n^{\mathrm{T}}\Gamma\left((\hat{\theta}_{n-1} - \theta) - \alpha_n\Gamma\Delta_n\Delta_n^{\mathrm{T}}\eta_n\right).\end{aligned}$$

First, we note that the relations

$$\begin{aligned}\mathrm{E}\left\{\eta_n^{\mathrm{T}}\alpha_n\xi_n\Gamma\Delta_n|\tilde{\mathcal{F}}_{n-1}\right\} &= \eta_n^{\mathrm{T}}\alpha_n\xi_n\Gamma\mathrm{E}\{\Delta_n|\tilde{\mathcal{F}}_{n-1}\}\\ &= \eta_n^{\mathrm{T}}\alpha_n\xi_n\Gamma\mathrm{E}\{\Delta_n\} = 0\end{aligned}$$

are satisfied by virtue of Assumption A) and that also

$$\begin{aligned}\mathrm{E}\left\{(\hat{\theta}_{n-1} - \theta)^{\mathrm{T}}\alpha_n\xi_n\Gamma\Delta_n|\tilde{\mathcal{F}}_{n-1}\right\} &= 0\\ \mathrm{E}\left\{\eta_n^{\mathrm{T}}\alpha_n^2\Delta_n\Delta_n^{\mathrm{T}}\Gamma^2\Delta_n\xi_n|\tilde{\mathcal{F}}_{n-1}\right\} &= 0\end{aligned}$$

by virtue of symmetric nature of the distribution $\mathrm{P}(\cdot)$. Therefore, we have

$$\begin{aligned}&\mathrm{E}\left\{\|\hat{\theta}_n - \theta\|^2|\hat{\mathcal{F}}_{n-1}\right\}\\ &\leq \|\hat{\theta}_{n-1} - \theta\|^2 - \eta_n^{\mathrm{T}}\alpha_n\mathrm{B}\Gamma(\hat{\theta}_{n-1} - \theta)\\ &- (\hat{\theta}_{n-1} - \theta)^{\mathrm{T}}\alpha_n\Gamma\mathrm{B}\eta_n\\ &+ \alpha_n^2\|\eta_n\|^2\mathrm{E}\left\{\mathrm{Tr}\left[\Delta_n\Delta_n^{\mathrm{T}}\Gamma^2\Delta_n\Delta_n^{\mathrm{T}}\right]\right\} + \alpha_n^2\xi_n^2\,\mathrm{Tr}[\Gamma\mathrm{B}\Gamma].\end{aligned}$$

By averaging over w_n and v_n sequentially, we derive

$$\begin{aligned}
&\mathrm{E}\left\{\|\hat{\theta}_n - \theta\|^2 | \mathcal{F}_{n-1}\right\} \\
&\quad\le \|\hat{\theta}_{n-1} - \theta\|^2 \\
&\quad\times \left(\mathrm{I} + \alpha_n^2 \left(\|\Gamma\|^2 \mathrm{E}\left\{\|\Delta_n\|^4\right\} + 2M_\varphi^2 \mathrm{Tr}[\Gamma \mathrm{B} \Gamma]\right)\right) \\
&\quad- (\hat{\theta}_{n-1} - \theta)^{\mathrm{T}} \alpha_n (\mathrm{B}\Gamma + \Gamma\mathrm{B})(\hat{\theta}_{n-1} - \theta) \\
&\quad+ \alpha_n^2 \left(\sigma_w^2 \|\Gamma\|^2 \mathrm{E}\left\{\|\Delta_n\|^4\right\} + \left(2\sigma_v^2 + M_\varphi^2 \sigma_w^2\right) \mathrm{Tr}[\Gamma\mathrm{B}\Gamma]\right).
\end{aligned}$$

Since $\sum_{n=1}^{\infty} \alpha_n^2 (1 + \mathrm{Tr}[\Gamma\mathrm{B}\Gamma]) < \infty$, we get by the Robbins–Siegmund lemma [18] that a finite limit $\lim_{n\to\infty} \|\hat{\theta}_n - \theta\|^2$ exists, and, additionally, the series $\sum_{n=1}^{\infty} (\hat{\theta}_{n-1}-\theta)^{\mathrm{T}} \alpha_n (\mathrm{B}\Gamma + \Gamma\mathrm{B})(\hat{\theta}_{n-1} - \theta) < \infty$ converges. From $\sum \alpha_n = \infty$, it follows that $\|\hat{\theta}_n - \theta\|^2 \to 0$ a.s. which proves the first assertion of Theorem 1.

To prove the second part, analogous to the previous proof, from (1) and (2) we obtain

$$\begin{aligned}
&\mathrm{E}\{\mathrm{D}_n | \hat{\mathcal{F}}_{n-1}\} \\
&\quad\le \mathrm{D}_{n-1} - \alpha_n \left(\Gamma\mathrm{B}\mathrm{D}_{n-1} + \mathrm{D}_{n-1}^{\mathrm{T}} \mathrm{B}\Gamma\right) \\
&\quad+\alpha_n^2 \left(\|\mathrm{D}_{n-1}\| \mathrm{E}\{\|\Delta_n\|^4\} \Gamma^2 \right.\\
&\qquad+\Gamma\left(\left(v_n^2\left(1 + M_\varphi^2 \rho\right) + \|\mathrm{D}_{n-1}\| \rho^{-1}\right.\right. \\
&\qquad\quad\left.\left.\left. + M_\varphi^2 \mathrm{Tr}[\mathrm{Q}_w]\right)\mathrm{B} + \mathrm{E}\left\{\Delta_n \Delta_n^{\mathrm{T}} \mathrm{Q}_w \Delta_n \Delta_n^{\mathrm{T}}\right\}\right)\Gamma\right).
\end{aligned}$$

By taking the unconditional expectation and using the first part of Assumption Bii) we get for the matrices $\mathrm{V}_n := \mathrm{E}\{\mathrm{D}_n\}$ that

$$\begin{aligned}
\mathrm{V}_n \le \mathrm{V}_{n-1} - \alpha_n (\Gamma\mathrm{B}\mathrm{V}_{n-1} + \mathrm{V}_{n-1}\mathrm{B}\Gamma) \\
+\alpha_n^2 \Gamma\mathrm{R}\Gamma + \alpha_n^2 \mathcal{O}\left(\|\mathrm{V}_{n-1}\|\right)
\end{aligned}$$

where R is defined by Theorem 2. By [19, Lemma 3], we conclude $\mathrm{V}_n \to 0$ as $n \to \infty$.

Proof of Theorem 2

We first demonstrate the matrix (5) is solvable. We rewrite it in the form

$$\left(\Gamma\mathrm{B} - \frac{1}{2}\mathrm{I}\right)\mathrm{S} + \mathrm{S}\left(\mathrm{B}\Gamma - \frac{1}{2}\mathrm{I}\right) = \Gamma\mathrm{R}\Gamma.$$

Since $-\Gamma\mathrm{B} + (1/2)\mathrm{I}$ is a Hurwitz matrix, according to the Lyapunov lemma, there is a positive–definite matrix S which is the solution of the corresponding matrix equation.

Let us return to the last inequality from the proof of Theorem 1. Denote $\mathrm{W}_n = n\mathrm{V}_n - \mathrm{S}$. Then, from the conditions of Theorem 2, we get

$$\begin{aligned}
\mathrm{W}_n \le \mathrm{W}_{n-1} - (n-1)^{-1}\left(\Gamma\mathrm{B} - \frac{1}{2}\mathrm{I}\right)\mathrm{W}_{n-1} - (n-1)^{-1} \\
\times \mathrm{W}_{n-1}\left(\mathrm{B}\Gamma - \frac{1}{2}\mathrm{I}\right) + n^{-2}\mathcal{O}\left(\|\mathrm{W}_{n-1}\|\right).
\end{aligned}$$

Consequently, by applying again [19, Lemma 3], we conclude $\mathrm{W}_n \to 0$ as $n \to \infty$ and, thus, Theorem 2 is proved.

Proof of Theorem 3

The following auxiliary results of [12] will be used in the proof.

Lemma 1: Under the conditions of Theorem 3, the following facts hold.

a) $\sum_{n=1}^{\infty} \Delta_n^{\mathrm{T}} \Gamma_n^2 \Delta_n < \infty$ a.s. and $\sum_{n=1}^{\infty} \|\Delta_n\|^4 \lambda_{max}^2(\Gamma_n) < \infty$ a.s.

b) $\sum_{n=1}^{\infty} \Delta_n^{\mathrm{T}} \Gamma_n \Delta_n = \infty$ a.s.

By substituting (7) into (1), we have

$$\begin{aligned}
\|\hat{\theta}_n - \theta\|^2 = &\left((\hat{\theta}_{n-1} - \theta)^{\mathrm{T}} - \eta_n^{\mathrm{T}} \Delta_n \Delta_n^{\mathrm{T}} \Gamma_n\right) \\
&\times \left((\hat{\theta}_{n-1} - \theta) - \Gamma_n \Delta_n \Delta_n^{\mathrm{T}} \eta_n\right) + \xi_n^2 \Delta_n^{\mathrm{T}} \Gamma_n^2 \Delta_n \\
&+ \left((\hat{\theta}_{n-1} - \theta)^{\mathrm{T}} - \eta_n^{\mathrm{T}} \Delta_n \Delta_n^{\mathrm{T}} \Gamma_n\right) \xi_n \Gamma_n \Delta_n \\
&+ \xi_n \Delta_n^{\mathrm{T}} \Gamma_n \left((\hat{\theta}_{n-1} - \theta) - \Gamma_n \Delta_n \Delta_n^{\mathrm{T}} \eta_n\right).
\end{aligned}$$

Taking the conditional expectation of both sides of the last relation with respect to σ-algebra $\tilde{\mathcal{F}}_{n-1}$ we obtain by Assumption A)

$$\begin{aligned}
&\mathrm{E}\left\{\|\hat{\theta}_n - \theta\|^2 | \tilde{\mathcal{F}}_{n-1}\right\} \\
&\quad\le \|\hat{\theta}_{n-1} - \theta\|^2 + \mathrm{E}\left\{\xi_n^2 | \tilde{\mathcal{F}}_{n-1}\right\} \mathrm{E}\left\{\Delta_n^{\mathrm{T}} \Gamma_n^2 \Delta_n | \tilde{\mathcal{F}}_{n-1}\right\} \\
&\quad- \eta_n^{\mathrm{T}} \mathrm{E}\left\{\Delta_n \Delta_n^{\mathrm{T}} \Gamma_n | \tilde{\mathcal{F}}_{n-1}\right\} \times (\hat{\theta}_{n-1} - \theta) - (\hat{\theta}_{n-1} - \theta)^{\mathrm{T}} \\
&\qquad\qquad \mathrm{E}\left\{\Gamma_n \Delta_n \Delta_n^{\mathrm{T}} | \tilde{\mathcal{F}}_{n-1}\right\} \eta_n \\
&\quad+ \eta_n^{\mathrm{T}} \mathrm{E}\left\{\Delta_n \Delta_n^{\mathrm{T}} \Gamma_n^2 \Delta_n \Delta_n^{\mathrm{T}} | \tilde{\mathcal{F}}_{n-1}\right\} \eta_n.
\end{aligned}$$

Hence, averaging over w_n and v_n sequentially, we derive

$$\begin{aligned}
&\mathrm{E}\left\{\|\hat{\theta}_n - \theta\|^2 | \mathcal{F}_{n-1}\right\} \\
&\quad\le \left(1 + \mathrm{E}\left\{2M_\varphi^2 \Delta_n^{\mathrm{T}} \Gamma_n^2 \Delta_n + \|\Delta_n\|^4 \lambda_{max}^2(\Gamma_n) | \mathcal{F}_{n-1}\right\}\right) \\
&\quad\times \|\hat{\theta}_{n-1} - \theta\|^2 - (\hat{\theta}_{n-1} - \theta)^{\mathrm{T}} \\
&\quad\times \mathrm{E}\left\{\left(\Gamma_n \Delta_n \Delta_n^{\mathrm{T}} + \Delta_n \Delta_n^{\mathrm{T}} \Gamma_n\right) | \mathcal{F}_{n-1}\right\} (\hat{\theta}_{n-1} - \theta) \\
&\quad+ \sigma_w^2 \mathrm{E}\left\{\|\Delta_n\|^4 \lambda_{max}^2(\Gamma_n) | \mathcal{F}_{n-1}\right\} \\
&\quad+ \left(2\sigma_v^2 + M_\varphi^2 \sigma_w^2\right) \mathrm{E}\left\{\Delta_n^{\mathrm{T}} \Gamma_n^2 \Delta_n | \mathcal{F}_{n-1}\right\}.
\end{aligned}$$

By applying the Robbins–Siegmund lemma [18] to the last relation, we conclude that by virtue of Lemma 1 assertion a) the sequence $\{\|\hat{\theta}_n - \theta\|^2\}$ has a finite limit a.s. and

$$\sum_{n=1}^{\infty} (\hat{\theta}_{n-1} - \theta)^{\mathrm{T}} \mathrm{E}\left\{\Gamma_n \Delta_n \Delta_n^{\mathrm{T}} + \Delta_n \Delta_n^{\mathrm{T}} \Gamma_n | \hat{\mathcal{F}}_{n-1}\right\} \times (\hat{\theta}_{n-1} - \theta) < \infty.$$

By virtue of Lemma 1 assertion b), $\sum_{n=1}^{\infty} \mathrm{E}\{\Gamma_n \Delta_n \Delta_n^{\mathrm{T}} | \hat{\mathcal{F}}_{n-1}\} = \infty$, and therefore $\|\hat{\theta}_n - \theta\|^2 \to 0$ a.s. as $n \to \infty$. The proof of the first part of Theorem 3 is completed.

From the constructing logic of the estimates (7), it follows $\hat{\theta}_n = \Gamma_n \sum_{k=1}^{n} \Delta_k (y_k - \mathrm{E}\{\varphi_k^{\mathrm{T}}\}\theta_{n-1})$, $\Gamma_n = (\sum_{k=1}^{n} \Delta_k \Delta_k^{\mathrm{T}} + \gamma_0 \mathrm{I})^{-1}$. By denoting $\epsilon_k = \xi_k + \Delta_k^{\mathrm{T}}(\theta_k - \theta)$, we obtain

$$\begin{aligned}
\mathrm{D}_n = \Gamma_n &\left(\gamma_0^2 \theta\theta^{\mathrm{T}} + \sum_{k=1}^{n} \Delta_k \Delta_k^{\mathrm{T}} \epsilon_k^2 - \gamma_0 \theta \sum_{i=1}^{n} \Delta_i^{\mathrm{T}} \epsilon_i \right.\\
&\left. - \gamma_0 \sum_{i=1}^{n} \Delta_i \epsilon_i \theta^{\mathrm{T}} + \sum_{\substack{i,j=1 \\ i\ne j}}^{n} \Delta_i \Delta_j^{\mathrm{T}} \epsilon_i \epsilon_j\right)\Gamma_n.
\end{aligned}$$

It is not so hard to show $\mathrm{E}\{\Gamma_n \Delta_i \Delta_j^{\mathrm{T}} \epsilon_i \epsilon_j \Gamma_n\} = 0$ and $\mathrm{E}\{\Gamma_n \gamma_0 \theta \Delta_i^{\mathrm{T}} \epsilon_i \Gamma_n\} = 0$ for any $i \ne j$, $i,j = 1,\ldots,n$ (see, e.g., [12]). On the basis of the conditions on the boundedness of v_n, w_n, Δ_n, we obtain for a sufficiently large n

$$\mathrm{E}\{\mathrm{D}_n\} = \mathrm{E}\left\{\Gamma_n \left(\gamma_0^2 \theta\theta^{\mathrm{T}} + \sum_{k=1}^{n} \Delta_k \Delta_k^{\mathrm{T}} \epsilon_k^2\right)\Gamma_n\right\} \le \hat{C}\mathrm{E}\{\Gamma_n\}$$

with some constant $\hat{C}$. Since $\sum_{k=1}^{n} \Delta_k \Delta_k^{\mathrm{T}} \to \infty$ a.s. as $n \to \infty$ and $\|\Gamma_n\| \leq \gamma_0^{-1}$ from the Lebesgue theorem of dominating sequence we conclude that $\mathrm{E}\{\mathrm{D}_n\} \to 0$ as $n \to \infty$. This completes the proof of Theorem 3.

ACKNOWLEDGMENT

The author would like to thank B. Polyak and the reviewers for their constructive comments.

REFERENCES

[1] R. A. Fisher, *The Design of Experiments*. Edinburgh, U.K.: Oliver and Boyd, 1935.
[2] B. T. Polyak and Y. Z. Tsypkin, “Adaptive estimation algorithms (convergence, optimality, stability),” *Automat. Remote Control*, vol. 40, pp. 378–389, 1979.
[3] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.
[4] P. C. Young, *Recursive Estimation and Time-Series Analysis. An Introduction*. Berlin, Germany: Springer-Verlag, 1984.
[5] A. Krieger and E. Masry, “Convergence analysis of adaptive linear estimation for dependent stationary processes,” *IEEE Trans. Inform. Theory*, vol. 34, pp. 177–182, Jan. 1988.
[6] B. T. Polyak, “New stochastic approximation type procedure,” *Automat. Remote Control*, vol. 51, pp. 937–946, 1990.
[7] A. Juditsky, “A stochastic estimation algorithm with observation averaging,” *IEEE Trans. Automat. Contr.*, vol. 38, pp. 794–798, May 1993.
[8] L. Guo, L. Ljung, and G.-J. Wang, “Necessary and sufficient conditions for stability of LMS,” *IEEE Trans. Automat. Contr.*, vol. 42, pp. 761–770, May 1997.
[9] W. H. Fleming and W. M. McEneaney, “Risk sensitive and robust nonlinear filtering,” in *Proc. 36th IEEE Conf. Decision Control*, 1997, pp. 1088–1093.
[10] W. M. McEneaney, “Robust/H_∞ filtering for nonlinear systems,” *Syst. Control Lett.*, vol. 33, pp. 315–325, 1998.
[11] O. N. Granichin, “A stochastic approximation algorithm with perturbations in the input for identification of static nonstationary plant,” *Vestnik Leningr. Univ., Math.*, vol. 21, pp. 92–93, 1988.
[12] A. V. Goldenshluger and B. T. Polyak, “Estimation of regression parameters with arbitrary noise,” *Math. Meth. Stat.*, vol. 2, pp. 18–29, 1993.
[13] L. Ljung and L. Guo, “The role of model validation for assessing the size of the unmodeled dynamics,” *IEEE Trans. Automat. Contr.*, vol. 42, pp. 1230–1239, Aug. 1997.
[14] O. N. Granichin, “Estimating the parameters of linear regression in an arbitrary noise,” *Automat. Remote Control*, vol. 63, pp. 25–35, 2002.
[15] Q. Zhang, “Optimal filtering of discrete-time hybrid systems,” *J. Optim. Theory Appl.*, vol. 100, pp. 123–144, 1999.
[16] ——, “Hybrid filtering for linear systems with non-Gaussian disturbances,” *IEEE Trans. Automat. Contr.*, vol. 45, pp. 50–61, Jan. 2000.
[17] O. N. Granichin, “Nonminimax filtering in unknown irregular constrained observation noise,” *Automat. Remote Control*, vol. 63, pp. 1482–1488, 2002.
[18] H. Robbins and D. Siegmund, “A convergence theorem for nonnegative almost super-martingales and some applications,” in *Optimizing Methods in Statistics*, J. S. Rustagi, Ed. New York: Academic, 1971, pp. 233–257.
[19] B. T. Polyak, “Convergence and rate of convergence of recursive stochastic algorithms. Linear case,” *Automat. Remote Control*, vol. 38, pp. 537–542, 1977.

Disturbance Propagation in Vehicle Strings

Pete Seiler, Aniruddha Pant, and Karl Hedrick

***Abstract*—This note focuses on disturbance propagation in vehicle strings. It is known that using only relative spacing information to follow a constant distance behind the preceding vehicle leads to string instability. Specifically, small disturbances acting on one vehicle can propagate and have a large effect on another vehicle. We show that this limitation is due to a complementary sensitivity integral constraint. We also examine how the disturbance to error gain for an entire platoon scales with the number of vehicles. This analysis is done for the predecessor following strategy as well as a control structure where each vehicle looks at both neighbors.**

***Index Terms*—Complementary sensitivity integral, interconnected systems, string stability.**

I. INTRODUCTION

The problem, in its most basic form, is to move a collection of vehicles from one point to another point. One application of this work is an automated highway system (AHS) [2] where the goal is to reduce traffic congestion by using closed loop control. To maximize the traffic throughput, the vehicles travel in closely spaced platoons (Fig. 1). Centralized control is impractical for medium to large sized platoons. Thus, a decentralized controller should be used. Furthermore, treating the vehicles independently is an unsafe approach because the intervehicle spacings are required to be small. A reasonable decentralized control strategy is for each vehicle to use a radar to keep a fixed distance behind the preceding vehicle (Fig. 2). The reference trajectory for the ith vehicle is a fixed distance, δ_i, behind the preceding vehicle: $r_i = x_{i-1} - \delta_i$. The feedback loops are coupled and it is possible for disturbances acting on one vehicle to propagate and affect other vehicles in the string. In fact, we show that for any linear control law $K(s)$ it is possible for a small disturbance acting on one vehicle to have an arbitrarily large effect on another vehicle.

The possibility of disturbance propagation in vehicle strings has been known for some time. Chu showed that an infinite string of vehicles could not be stabilized using the strategy depicted in Fig. 2 with a proportional control law [3]. A similar result was shown via a transfer function analysis [4]. In the early 1990s, renewed interest in AHS spurred further research on the control of vehicle strings [5]–[11]. Swaroop developed rigorous definitions of string stability and relations to error propagation transfer functions [9]. The research on vehicle strings can be generalized and studied as a spatially invariant system [12].

To summarize, we note that many researchers have shown that “string stability” cannot be obtained when vehicles use only relative spacing information to maintain a constant distance behind their predecessor. All of these results have been for specific control laws. In this note, we show that if vehicles use only relative spacing information, then we have “string instability” for *any* linear controller. It is well known [5], [9] that this string instability can be corrected by using a constant time headway policy (δ_i is proportional to the vehicle

Manuscript received June 18, 2003; revised November 10, 2003. Recommended by Associate Editor U. Jonsson. This work was supported in part by the Office of Naval Research under Grant N00014-99-10756. This work was presented at the 2002 American Control Conference [1].
P. Seiler is with the University of Illinois, Urbana-Champaign, Urbana, IL 61801 USA (e-mail: pseiler@uiuc.edu).
A. Pant is with Tata Research Development and Design Centre, Pune 411 013, India (e-mail: aniruddha_pant@pune.tcs.co.in).
K. Hedrick is with the University of California, Berkeley, CA 94720 USA (e-mail: khedrick@me.berkeley).
Digital Object Identifier 10.1109/TAC.2004.835586