Nonasymptotic Results for Finite-Memory WLS Filters

Maciej Niedźwiecki and Lei Guo, *Member, IEEE*

***Abstract*—The paper presents what we believe to be the first nonasymptotic analysis of properties of weighted least squares (WLS) adaptive filters used for identification of time-varying systems. We show that the problem of mean-square boundedness of WLS estimates is closely related to the problem of invertibility—in the mean sense—of the corresponding regression matrix. We discuss necessary and sufficient conditions for such invertibility to hold. Based on that, a number of results are derived paralleling those already obtained for least mean-square (LMS) filters and the problem of "statistical robustness" of the WLS estimator is briefly mentioned.**

I. Introduction

Consider the following time-varying stochastic system

$$y(t) = \alpha^T(t)\phi(t) + n(t) \tag{1}$$

where $\phi(t) = [u_1(t), \cdots, u_r(t)]^T$ is the measurable input vector, $\alpha(t) = [a_1(t), \cdots, a_r(t)]^T$ is the unknown (time-dependent) parameter vector and $\{n(t)\}$ denotes the unobservable (scalar) measurement noise. We will assume that

A1: The noise process $\{n(t)\}$ is a sequence of zero-mean independent and identically distributed random variables and $E[n^2(t)] = \rho_0$.

A2: The input process $\{\phi(t)\}$, independent of $\{n(t)\}$, is a sequence of identically distributed m-dependent random vectors (i.e., $\exists m$ such that $\forall t$ sequences $\{\phi(i), i \le t\}$ and $\{\phi(i), i \ge t + m\}$ are independent) and $E[\phi(t)\phi^T(t)] = R_0 > 0$.

A3: Time-varying parameters form a sequence $\{\alpha(t)\}$, independent of $\{\phi(t)\}$ and $\{n(t)\}$, which is bounded in the mean square sense, i.e.,

$$E\left[\|\alpha(t)\|^2\right] \le A < \infty \qquad \forall t.$$

Assumption about m-dependence of the input sequence is not critical and will be relaxed to include weaker mixing (asymptotic independence) and covariance conditions later on.

Note, that in the case where

$$u_i(t) = u(t-i), \qquad i = 1, \cdots, r \tag{2}$$

(1) specializes to the dynamic finite impulse response (FIR) model widely used in adaptive filtering, e.g., for the purpose of adaptive equalization of communication channels.

If parameters in (1) vary sufficiently slowly with time, the method of weighted least squares (WLS) can be used for the purpose of tracking of $\alpha(t)$. Let $\{w(t)\}$ denote the nonnegative and nonincreasing weighting sequence, such that

$$\sum_{t=0}^{\infty} w(t) = 1 \tag{3}$$

(the normalization constraint (3) is not essential for our analysis and was introduced for the sake of notational convenience). Assuming, for convenience, that the infinite observation history is available at the instant t, the WLS estimator can be defined in the following way

$$\begin{aligned}\alpha(t) &= \arg\min_{\alpha} \sum_{i=0}^{\infty} w(i)[y(t-i) - \alpha^T\phi(t-i)]^2 \\ &= \left(\sum_{i=0}^{\infty} w(i)\phi(t-i)\phi^T(t-i)\right)^{-1} \\ &\quad \cdot \left(\sum_{i=0}^{\infty} w(i)y(t-i)\phi(t-i)\right) = \tilde{R}^{-1}(t)\tilde{S}(t)\end{aligned} \tag{4}$$

with obvious definitions of $\tilde{R}(t)$ and $\tilde{S}(t)$.

In practice, the requirement that the WLS estimator should be recursively computable limits our choice of $w(t)$ to several standard windows. If, for example, the exponential window is used ($w(t) = (1-\lambda)\lambda^t$, $0 < \lambda < 1$) one can replace (4) by the following recursive algorithm [1]:

$$\begin{aligned}\hat{\alpha}(t) &= \hat{\alpha}(t-1) + D(t)\phi(t)\epsilon(t) \\ \epsilon(t) &= y(t) - \alpha^T(t-1)\phi(t)\end{aligned} \tag{5}$$

where the matrix $D(t)$ can be updated using the well-known formula

$$D(t) = \frac{1}{\lambda}\left[D(t-1) - \frac{D(t-1)\phi(t)\phi^T(t)D(t-1)}{\lambda + \phi^T(t)D(t-1)\phi(t)}\right]. \tag{6}$$

Similar (but basically two-step) algorithms can be derived for the sliding rectangular window ($w(t) = 1/N$ for $t < N$ and $= 0$ for $t \ge N$) (see e.g., [1]). Various fast versions of the WLS algorithm are also available but they usually require "safety jacketing" because of possible numerical ill-conditioning (see [2]).

If the data-dependent adaptation matrix $D(t)$ in (5) is replaced by a small adaptation gain μ one arrives at the so-called least mean-square (LMS) algorithm

$$\hat{\alpha}(t) = \hat{\alpha}(t-1) + \mu\phi(t)\epsilon(t). \tag{7}$$

Although computationally less demanding than the WLS algorithm, the LMS algorithm may suffer from a very slow initial convergence-a disadvantageous effect if rapid adaptation is required. Despite this difference, both algorithms have very similar parameter tracking properties (see e.g., [3]).

Manuscript received March 2, 1989; revised May 20, 1990. Paper recommended by Past Associate Editor, B. Sridhar.

M. Niedźwiecki is with the Institute of Computer Science, Technical University of Gdańsk, ul. Majakowskiego 11/12, Gdańsk, Poland.

L. Guo is with the Institute of Systems Science, Academia Sinica, Beijing 100080, People's Republic of China.

IEEE Log Number 9041415.

0018-9286/91/0200-0198$01.00 © 1991 IEEE

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 29,2025 at 00:49:48 UTC from IEEE Xplore. Restrictions apply.

While the statistical properties of LMS filters seem to be well-explored and documented, the situation is less clear for the WLS filters. So far, all the analyses were based on asymptotic arguments, i.e., strictly speaking, they dealt with the case where the effective length of the window tended to infinity.

Almost no precise results seem to exist for strictly finite-length windows—the recent paper of Macchi and Eweda [19] being the only noticeable exception.[1] However, even the results presented in [19] rely critically on the assumption about invertibility of a certain stochastic regression matrix [19, Assumption A2] which is postulated but is very difficult to verify.

In this paper, we present what we believe to be the first nonasymptotic analysis of properties of the WLS estimator based on realistic and verifiable assumptions. We show that the problem of the mean-square boundedness of $\hat{\alpha}(t)$ is closely related to the problem of invertibility—in the mean sense—of the regression matrix $\tilde{R}(t)$ in (4). We discuss conditions under which such invertibility is guaranteed if sufficiently strong mixing (asymptotic independence) and/or covariance conditions are imposed on $\{\phi(t)\}$. Based on that, a number of results can be derived paralleling those obtained for LMS estimators by Macchi and Eweda [4], [5] and the problem of "statistical robustness" of the WLS estimator can be properly addressed.

II. Review of Known Results

From among a large number of results on stability and tracking bounds for the LMS algorithm, we would like to point to a sequence of insightful papers by Macchi and Eweda [4]–[7] (see also [20] and [21]).

Assuming that the input sequence is stationary and m-dependent the authors were able to show that, for sufficiently small but nonzero gain μ

$$E\left[\|\hat{\alpha}(t) - \alpha(t)\|^2\right] \le C(\mu) < \infty. \tag{8}$$

In the constant parameter case ($\alpha(t) = \alpha_0$), we have [6]

$$C(\mu) = C_1\mu \tag{9}$$

that is, the bound on random fluctuations in steady state decreases with the stepsize in a linear way.

If system parameters vary with time, the choice of μ becomes a trade-off between the steady-state accuracy and tracking ability of the estimation algorithm. Assuming, for example, that $\{\alpha(t)\}$ evolves according to the random walk model, it is possible to show that [6]

$$C(\mu) = C_1\mu + \frac{C_2}{\mu} \tag{10}$$

which illustrates the need for the compromise mentioned previously.

Denote by ι the equivalent width of the window $\{w(t)\}$ (equivalent number of observations)

$$\iota = 1/\sum_{t=0}^{\infty} w^2(t) \tag{11}$$

deciding upon the "memory" of the WLS filter [8]. One can argue that the quantity $1/\iota$ determines the adaptation gain of the WLS algorithm, i.e., it plays exactly the same role as the stepsize μ in the LMS filter.

[1] This paper was brought to our attention during the revision process.

Basically, two different approaches were used to analyse properties of WLS estimators:

- the approach based on Taylor series approximations (for any weighting sequence) (see e.g., [8]–[11])
- the approach based on ODE approximations (for exponential weighting) (see e.g., [12]–[15])

In both cases, the derived results hold only asymptotically, that is for $\iota \to \infty$. Almost no results seem to exist if ι is finite and fixed. For example, for the constant parameter case only a considerably weaker version of (8), (9) is available. According to Eweda and Macchi [7], for arbitrarily small $\epsilon > 0$, the estimation error $\|\hat{\alpha}(t) - \alpha_0\|^2$ has an upper bound proportional to $1/\iota$ with *probability* $1 - \epsilon$.

The finite mean-square tracking bound established subsequently in [19] rests on an implicit assumption that there exist an integer N_0 and a constant $0 < c < \infty$ such that $\forall\, N \ge N_0$, $\forall\, t$

$$E\left\{\left[\lambda_{\min}\left(\sum_{i=0}^{N-1} \phi(t-i)\phi^T(t-i)\right)\right]^{-8}\right\} < c. \tag{12}$$

It turns out, however, that verification of assumptions similar to (12) is far from being obvious and constitutes the very core of the tracking assessment problem. We will consider this issue in more detail in Section IV.

III. "Idealized" WLS Estimator

Since under A2, we have

$$\tilde{R}(t) = \sum_{i=0}^{\infty} w(i)\phi(t-i)\phi^T(t-i) \underset{\iota\to\infty}{\mapsto} R_0 \tag{13}$$

where convergence takes place either in the mean-square sense or with probability one [9], for sufficiently large ι one can attempt to replace the regression matrix $\tilde{R}(t)$ in (4) by its expectation. The resulting "idealized" WLS estimator

$$\hat{\hat{\alpha}}(t) = R_0^{-1}\tilde{S}(t) \tag{14}$$

is analytically easy to handle. Moreover, provided that the difference

$$E\left[\|\hat{\alpha}(t) - \hat{\hat{\alpha}}(t)\|^2\right] \tag{15}$$

is sufficiently small one can infer about properties of the WLS estimator $\hat{\alpha}(t)$ by analyzing properties of its "idealized" counterpart—that was the line of thinking in [9], [10]. One of the points behind studying properties of (15) is that, via the inequality

$$E\left[\|\hat{\alpha}(t) - \alpha(t)\|^2\right] \le 2E[\|\hat{\alpha}(t) - \hat{\hat{\alpha}}(t)\|]^2 + 2E\left[\|\hat{\hat{\alpha}}(t) - \alpha(t)\|^2\right] \tag{16}$$

boundedness of (15) implies boundedness of the mean-square tracking error (under A1–A3 boundedness of the second term on the right-hand side of (16) can be shown quite easily).

Not surprisingly, the problem of boundedness of (15) can be related to the problem of invertibility, in the mean sense, of the matrix $\tilde{R}(t)$. In particular, we have the following:

Lemma 1: Under assumptions A1–A3 we have

$$E\left[\|\hat{\alpha}(t) - \hat{\hat{\alpha}}(t)\|^2\right] = 0(tr\{E[\Delta(t)]\}) \tag{17}$$

where $\Delta(t) = \tilde{R}^{-1}(t) - R_0^{-1}$.

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 29,2025 at 00:49:48 UTC from IEEE Xplore. Restrictions apply.

Proof:

$$\Delta\alpha(t) = \hat{\alpha}(t) - \hat{\hat{\alpha}}(t) = \Delta(t)\tilde{S}(t) = \sum_{i=0}^{\infty} x(i)v(i)$$

where

$$x(i) = \sqrt{w(i)}\,\Delta(t)\,\phi(t-i), \qquad v(i) = \sqrt{w(i)}\left[\phi^T(t-i)\alpha(t-i) + n(t-i)\right].$$

Using the Schwartz inequality, one gets

$$E\left[\|\Delta\alpha(t)\|^2\right] \le E\left[\sum_{i=0}^{\infty}\|x(i)\|^2\right]E\left[\sum_{i=0}^{\infty}v^2(i)\right].$$

Observe that

$$E\left[\sum_{i=0}^{\infty}\|x(i)\|^2\right] = E\left[\operatorname{tr}\left\{\sum_{i=0}^{\infty}x(i)x^T(i)\right\}\right] = \operatorname{tr}\left\{E\left[\Delta(t)\,\tilde{R}(t)\,\Delta(t)\right]\right\} = \operatorname{tr}\left\{E\left[\Delta(t)\right]\right\}.$$

Similarly

$$E\left[\sum_{i=0}^{\infty}v^2(i)\right] \le \sum_{i=0}^{\infty}E\left[\|\alpha(t-i)\|^2\right] \cdot E\left[\operatorname{tr}\left\{w(i)\phi(t-i)\phi^T(t-i)\right\}\right] + \rho_0 \le A\operatorname{tr}\{R_0\} + \rho_0 = 0(1).$$

IV. Invertibility of the Regression Matrix

A. Preliminary Considerations

According to Lemma 1, proving boundedness of the mean-square tracking error amounts to finding conditions under which

$$E\left[\tilde{R}^{-1}(t)\right] < \infty,$$

i.e., under which the stochastic regression matrix $\tilde{R}(t)$ is invertible in the mean sense. Let $\{w'(t)\}$:

$$w'(t) = \begin{cases} c & \text{for } t < N \\ 0 & \text{for } t \ge N. \end{cases} \tag{18}$$

denote a rectangular window "inscribed" in $\{w(t)\}$, see Fig. 1. Since $\tilde{R}(t) \ge cR(t)$ where

$$R(t) = \sum_{i=0}^{N-1}\phi(t-i)\phi^T(t-i)$$

the overbounding technique can be used, i.e., the boundedness results for general windows are implied by the corresponding results for the "inscribed" rectangular windows. However, even under uniform weighting the problem of invertibility of the stochastic regression matrix is conspicuously absent from the statistical literature.

Observe that

$$\left(R^{-1}(t)\right)_{ij} + \frac{\left(\operatorname{adj} R(t)\right)_{ij}}{\det R(t)}$$

and hence, using the Hölder inequality, one get ($k > 1$)

$$E\left[\left(R^{-1}(t)\right)_{ij}\right] \le \left[E\left[\left(\det R(t)\right)^{-k}\right]\right]^{\frac{1}{k}} \cdot \left[E\left[\left(\operatorname{adj} R(t)\right)_{ij}^{\frac{k}{k-1}}\right]\right]^{\frac{k-1}{k}}. \tag{19}$$

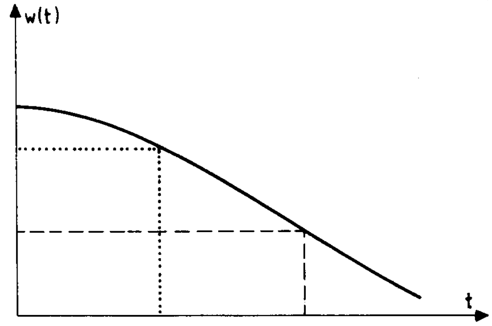


Fig. 1. Rectangular windows "inscribed" in $\{w(i)\}$

Imposing some moment conditions on $u_i(t)$, $i = 1,\cdots,r$, namely

A4: $\exists\epsilon > 0$ such that

$$E\left[\left(u_i(t)\right)^{2(r-1)+\epsilon}\right] < \infty, \qquad \forall t,\ i = 1,\cdots,r$$

and choosing k sufficiently large so that $2(r-1)/(k-1) \le \epsilon$, one can guarantee boundedness of the second factor on the right-hand side of (19). Therefore to prove invertibility of $R(t)$ in the mean it suffices to find conditions under which the determinant of $R(t)$ is invertible in kth-moment

$$E\left[\left(\det R(t)\right)^{-k}\right] < \infty \tag{20}$$

for sufficiently large k.

B. Need for Additional Constraints

Quite obviously, in order to satisfy (20) one needs

$$P\left(\det R(t) = 0\right) = 0 \tag{21}$$

which can not be guaranteed by imposing only moment conditions on $\{\phi(t)\}$, such as

$$E\left[\phi(t)\phi^T(t)\right] > 0 \qquad \forall t. \tag{22}$$

We will clarify this point by means of the following example:

Example: Consider the case where (2) holds and $\{u(t)\}$ is an i.i.d. sequence taking only two values: $+1$ and -1 with probabilities p $(0 < p < 1)$ and $1 - p$, respectively. Then it is straightforward to show that for any finite N (21) is not fulfilled even though (22) is. ■

Quite clearly, additional constraint is needed to rule out cases such as the one above. The following result, which can be thought of as a refinement of (21), will be very useful for our further purposes.

Lemma 2: The determinant of $R(t)$ is invertible in kth moment provided that $\exists x_0$, $\eta > 0$, $L \ge k + 1$ such that $\forall x$: $x_0 > x > 0$ and $\forall t$ it holds

$$P\left(\det R(t) \le x\right) \le \eta x^L. \tag{23}$$

Proof: Using (23) one gets

$$E\left[\left(\det R(t)\right)^{-k}\right] = \int_0^{\infty} x^{-k}\,dP\left(\det R(t) \le x\right) = k\int_0^{\infty} x^{-k-1}P\left(\det R(t) \le x\right)dx \le k\int_0^{x_0}\eta x^{L-k-1}\,dx + x_0^{-k} < \infty.$$

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 29,2025 at 00:49:48 UTC from IEEE Xplore. Restrictions apply.

C. Explicit Invertibility Condition

We will show that the implicit condition (23) can be met if N in (18) is sufficiently large and

A5: $\exists \gamma, \delta, x_0 > 0$ such that $\forall x$: $x_0 > x > 0$, and $\forall t$

$$\sup_{\|\beta\|=1} P\left(\left(\beta^T\phi(t)\right)^2 < x\right) \le \gamma x^{\delta}. \quad (24)$$

In order to simplify the analysis we will derive the main result under the assumption that $\{\phi(t)\}$ is an i.i.d. sequence and then we will extend it to more general cases.

Lemma 3 (Key Technical Lemma): Suppose that $\{\phi(t)\}$ is an i.i.d. sequence with finite second-order moments. Denote

$$R(t) = \sum_{j=0}^{N-1} \phi(t-i)\phi^T(t-i).$$

Then $E[R^{-1}(t)] < \infty$ for some deterministic integer $N < \infty$ if and only if the condition A5 holds.

Proof:

a) Sufficiency: For convenience, take $N = rM$. We have

$$R(t) = \sum_{j=0}^{M-1} H_j(t), \qquad H_j(t) = \sum_{k=0}^{r-1} \phi(t-jr-k)\phi^T(t-jr-k)$$

since for nonnegative definite matrices A and B we have $\det(A+B) \ge \det A + \det B$, it holds

$$\det R(t) \ge \sum_{j=0}^{M-1} \det H_j(t)$$

and consequently

$$P(\det R(t) \le x) \le P\left(\sum_{j=0}^{M-1} \det H_j(t) \le x\right) \le \left[P(\det H_0(t) \le x)\right]^M$$

since the matrices $H_j(t)$ are mutually independent. We shall prove, by induction, that under conditions implied by A5, the following proposition is true.

Proposition: $\exists \xi_r, \tau_r > 0$ such that $\forall x$: $x_0 > x > 0$ and $\forall t$

$$P(\det H_0(t) \le x) = P\left(\det\left[\sum_{j=0}^{r-1} \phi(t-j)\phi^T(t-j)\right] \le x\right) \le \xi_r x^{\tau_r}. \quad (25)$$

Observe that $\det H_0(t) = (\det W(t))^2$ where

$$W(t) = \begin{bmatrix} u_1(t) & \cdots & u_1(t-r+1) \\ \vdots & & \vdots \\ u_r(t) & \cdots & u_r(t-r+1) \end{bmatrix}$$

and

$$\det W(t) = \sum_{i=1}^{r} (-1)^{i-1} u_i(t) z_i(t) \triangleq \phi^T(t) z(t)$$

where $z(t)$ is a collection of corresponding minors of the matrix $W(t)$. Note that

$$z_i^2(t) = \det\left(\sum_{j=1}^{r-1} \phi_{[i]}(t-j)\phi_{[i]}^T(t-j)\right)$$

$$\phi_{[i]}(t) = \left[u_1(t), \cdots, u_{i-1}(t), u_{i+1}(t), \cdots, u_r(t)\right]^T.$$

Rewrite $\det H_0(t)$ in the form $\det H_0(t) = \|z(t)\|^2 (\beta^T(t)\phi(t))^2$ where $\beta(t) = z(t)/\|z(t)\|$ for all $z(t) \ne 0$. Using simple calculations based on conditional probabilities, one gets, for any $x < x_0$

$$\begin{aligned} P(\det H_0(t) \le x) &= P\left(\det H_0(t) \le x \,\Big|\, \|z(t)\|^2 \ge \sqrt{x}\right) \cdot P\left(\|z(t)\|^2 \ge \sqrt{x}\right) + P\left(\Big|\det H_0(t) < x\, \|z(t)\|^2 < \sqrt{x}\Big|\right) P\left(\|z(t)\|^2 < \sqrt{x}\right) \\ &\le P\left(\left(\beta^T(t)\phi(t)\right)^2 \le \sqrt{x} \,\Big|\, \|z(t)\|^2 \ge \sqrt{x}\right) + P\left(\|z(t)\|^2 < \sqrt{x}\right) \end{aligned} \quad (26)$$

where it was assumed that for any positive x: $P(\|z(t)\|^2 < \sqrt{x}) > 0$ so that the corresponding conditional probability is well-defined (if not, a simple modification can be introduced).

Owing to the fact that the regression vector $\phi(t)$ is independent of $z(t) = f(\phi(t-1), \cdots, \phi(t-r+1))$ and that $\|\beta(t)\| = 1$ we get (cf., A5)

$$P\left(\left(\beta^T(t)\phi(t)\right)^2 \le \sqrt{x} \,\Big|\, \|z(t)\|^2 \ge \sqrt{x}\right) \le \sup_{\|\beta\|=1} P\left(\left(\beta^T\phi(t)\right)^2 \le \sqrt{x}\right) \le \gamma x^{\delta/2}. \quad (27)$$

Consider, in turn, the second term on the right-hand side of (26)

$$P\left(\|z(t)\|^2 \le \sqrt{x}\right) = P\left(\sum_{i=1}^{r} z_i^2(t) < \sqrt{x}\right) \le \sum_{i=1}^{r} P\left(z_i^2(t) < \sqrt{x}\right). \quad (28)$$

Now, suppose that the proposition is true for $r-1$, i.e., for all $(r-1)$-dimensional subvectors of $\phi(t)$ there exist constants ξ_{r-1}, r_{r-1} such that

$$P(z_i^2(t) \le x) = P\left(\det\left[\sum_{j=1}^{r-1} \phi_{[i]}(t-j)\phi_{[i]}^T(t-j)\right] \le x\right) \le \xi_{r-1} x^{\tau_{r-1}}, \qquad i = 1, \cdots, r.$$

Combining this with (28) one gets

$$P(\|z(t)\|^2 < \sqrt{x}) \le r\xi_{r-1} x^{\frac{\tau_{r-1}}{2}}$$

and consequently, using (26), (27) and the bound obtained previously, one has

$$P(\det H_0(t) \le x) \le \gamma x^{\frac{\delta}{2}} + r\xi_{r-1} x^{\frac{\tau_{r-1}}{2}} \le \xi_r x^{\tau_r}$$

for $x \in (0, 1]$, $\tau_r = \min\left(\frac{\delta}{2}, \frac{\tau_{r-1}}{2}\right)$ and appropriately chosen ξ_r.

Since the proposition stems immediately from A5 in the case where $r = 1$, it is also true in the general case.

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 29,2025 at 00:49:48 UTC from IEEE Xplore. Restrictions apply.

Finally, observe that

$$P(\det R(t) \leq x) \leq [P(\det H_0(t) \leq x)]^M \leq (\xi_r)^M x^{M\tau_r} = \eta x^L$$

where $L = M\tau_r$ can be made arbitrarily large by increasing M (i.e., N). This concludes the proof of the "if" part of Lemma 3.

b) Necessity: In order to prove the "only if" part of Lemma 3, denote by $\lambda_1(t)$ the minimum eigenvalue of $R(t)$. Observe that $E[R^{-1}(t)] < \infty$ entails $E[\lambda_1^{-1}(t)] < \infty$, i.e.,

$$\int_0^\infty \frac{1}{y}\, dP(\lambda_1(t) \leq y) < \infty$$

and since $P(\lambda_1(t) = 0) = 0$ we have

$$\forall \epsilon > 0\ \exists x_1 > 0 \qquad \text{such that for any } 0 < x < x_1$$

$$\int_0^x \frac{1}{y}\, dP(\lambda_1(t) \leq y) \leq \epsilon.$$

Since $\int_0^x \frac{1}{y}\, dP(\lambda_1(t) \leq y) \geq \frac{1}{x} P(\lambda_1(t) \leq x)$, we finally obtain for any $0 < x < x_1$: $P(\lambda_1(t) \leq x) \leq \epsilon x$.

Since $\lambda_1(t)$ is the minimum eigenvalue of $R(t)$, for any deterministic vector β with $\|\beta\| = 1$ we get

$$\lambda_1(t) \leq \beta^T \left[\sum_{i=0}^{N-1} \phi(t-i)\phi^T(t-i)\right]\beta = \sum_{i=0}^{N-1} (\beta^T\phi(t-i))^2.$$

Consequently, $P(\sum_{i=0}^{N-1}(\beta^T\phi(t-i))^2 \leq x) \leq P(\lambda_1(t) \leq x) \leq \epsilon x$. Observing that $\cap_{i=0}^{N-1}\left\{(\beta^T\phi(t-i))^2 \leq \frac{x}{N}\right\}$ entails $\sum_{i=0}^{N-1}(\beta^T\phi(t-i))^2 \leq x$ and exploiting the i.i.d. property of $\{\phi(t)\}$ we get

$$\left[P\left((\beta^T\phi(t))^2 \leq \frac{x}{N}\right)\right]^N \leq P\left(\sum_{i=0}^{N-1}(\beta^T\phi(t-i)^2 \leq x\right) \leq \epsilon x$$

that is

$$P\left((\beta^T\phi(t))^2 \leq \frac{x}{N}\right) \leq \epsilon^{1/N}x^{1/N}, \qquad 0 < x < x_1.$$

Finally, making the substitution $y = x/N$, we arrive at condition A5 with $\gamma = (N\epsilon)^{1/N}$, $\delta = \frac{1}{N}$ and $x_0 = \frac{x_1}{N}$.

Remark 1: It is known that distribution of any r-dimensional random vector ϕ can be factored as (Lebesgue decomposition theorem)

$$F(\phi) = \mu_c F_c(\phi) + \mu_d F_d(\phi) + \mu_s F_s(\phi)$$

where F_c, F_d, F_s are continuous discrete and singular distributions, respectively, and μ_c, μ_d, μ_s are nonnegative constants such that $\mu_c + \mu_d + \mu_s = 1$.

What A5 effectively says is that $F(\phi)$ should be free of discrete and singular (supported on hyperplanes) components. Additionally, it rules out "almost discrete" and "almost singular" components in the continuous distribution. We note that A5 admits a very large class of continuous distributions, e.g., all ones characterized by bounded probability density functions (such as Gaussian, uniform, etc.).

Remark 2: Note that the dependence structure of $\{\phi(t)\}$ was not used when we derived the moment condition A4. If the input sequence is m-dependent existence of second-order moments (implied by A2) is sufficient to prove boundedness of the second term on the right-hand side of (20).

D. Extension to Weaker Mixing and Covariance Conditions

The requirement that the sequence of regression vectors $\{\phi(t)\}$ should be white (as stated in conditions of Lemma 3) is, quite clearly, very inconvenient. Note, for example, that it is never met for FIR models (2), even if the input sequence is white! (since successive regression vectors share $r-1$ components).

Basically, the results of Lemma 3 can be extended in two different directions: to weaker mixing (asymptotic independence) conditions and weaker covariance (rate of decorrelation) conditions.

As far as mixing is concerned relaxation of i.i.d. assumption to m-dependence (consistent e.g., with (2) under white noise excitation) is straightforward. Suppose that $\{\phi(t)\}$ is an identically distributed and m-dependent sequence obeying A5. Then $\{\phi(tm)\}$ is an i.i.d. sequence and hence, for suitably large N

$$E[R^{-1}(t)] \leq E\left[\left(\sum_i \phi(t-im)\phi^T(t-im)\right)^{-1}\right] < \infty.$$

Extension to weaker mixing conditions is also possible. In particular, denote by $\mathscr{F}_r^s$ the sigma algebra generated by $\{\phi(i);\ t \leq i \leq s\}$. If the following mixing (asymptotic independence) condition is fulfilled:

$$|P(AB) - P(A)P(B)| \leq \psi(n)P(A)P(B)$$

for any events $A \in \mathscr{F}_{-\infty}^t$ and $B \in \mathscr{F}_s^\infty$, where $n = s - t$ and $\psi(n) \to 0$ for $n \to \infty$ the sequence $\{\phi(t)\}$ is called super uniformly mixing or ψ mixing. All previous results can be easily extended to such sequences.

From the practical point of view, much more interesting results can be obtained by means of relaxing covariance conditions imposed on $\{\phi(t)\}$. Actually, consider the case where $\phi(t)$ is the output of the state space model

$$x(t+1) = Ax(t) + B\eta(t)$$
$$\phi(t) = Cx(t) + D\eta(t). \qquad (29)$$

Then we have the following result.

Theorem 1: Lemma 3 holds if the model (29) is output reachable and $\{\eta(t)\}$ is an i.i.d. sequence obeying A5.

Outline of proof:

The proof is based on the following basic inequality valid for output reachable state space models (see e.g., [17], [18])

$$\lambda_{\min}\left[\sum_{i=0}^{N-1} \phi(t-i)\phi^T(t-i)\right] \geq c\lambda_{\min}\left[\sum_{i=0}^{N+v-1} \bar{\eta}(t-i)\bar{\eta}^T(t-i)\right], \quad \forall t \quad (30)$$

where $\bar{\eta}(t) = [\eta^T(t), \cdots, \eta^T(t-v)]^T$, v is the McMillan degree of the system (29) and $c > 0$.

The following example will illustrate the main steps in the proof of Theorem 1.

Example: Let $\{u(t)\}$ be generated from the following

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 29,2025 at 00:49:48 UTC from IEEE Xplore. Restrictions apply.

$AR(p)$ model:

$$u(t) + a_1 u(t-1) + \cdots + a_p u(t-p) = \epsilon(t)$$

where $\{\epsilon(t)\}$ is an i.i.d. sequence satisfying $P(|\epsilon(t)| \le x) \le \gamma x^{\delta}$, for some $\gamma > 0$, $\delta > 0$. Then Lemma 3 also holds with $\phi(t) = [u(t-1), \cdots, u(t-r)]^T$.

Proof: Let us denote

$$A(q^{-1}) = a_0 + a_1 q^{-1} + \cdots + a_p q^{-p}$$

where $a_0 = 1$ and q^{-1} is the backwards shift operator, and define

$$\psi(t) = A(q^{-1})\phi(t), \quad R_1(t) = \sum_{i=0}^{N-p-1} \psi(t-i)\psi^T(t-i).$$

Then by the Schwarz inequality it is seen that for any vector $\alpha \in R^r$

$$\begin{aligned} \alpha^T R_1(t)\alpha &= \sum_{i=0}^{N-p-1} \left[\alpha^T \psi(t-i)\right]^2 \\ &= \sum_{i=0}^{N-p-1} \left[\sum_{j=0}^{p} a_j \alpha^T \phi(t-i-j)\right]^2 \\ &\le \left(\sum_{j=0}^{p} a_j^2\right)\left(\sum_{j=0}^{p}\sum_{i=0}^{N-p-1} \left[\alpha^T \phi(t-i-j)\right]^2\right) \\ &\le (p+1)\sum_{j=0}^{p} a_j^2 \alpha^T R(t)\alpha. \end{aligned}$$

Consequently, by the arbitrariness of α

$$\lambda_{\min}[R(t)] \ge \frac{\lambda_{\min}[R_1(t)]}{(p+1)\sum_{j=0}^{p} a_j^2}.$$

Hence, the desired result follows by observing that

$$\lambda_{\min}[R_1(t)] \ge \frac{\det R_1(t)}{\{\lambda_{\max}[R_1(t)]\}^{r-1}}$$

and that $\psi(t) = [\epsilon(t-1), \cdots, \epsilon(t-r)]$, is an r-dependent sequence. ∎

Theorem 1 extends the invertibility result to a very general class of stationary signals with rational spectra. Extension to a limited class of nonstationary signals is also possible using the same approach (cf. [17]).

Remark: Generally speaking, the higher is dimension r of regression vector and the weaker is mixing condition imposed on $\{\phi(t)\}$, the larger should be N in order to guarantee sufficiently large value of L in (23). We note, however, that the lower bound on L resulting from our analysis is a *deterministic quantity* obtained without referring to any asymptotic arguments.

V. Results for Sliding Window LS Estimators—The Gaussian Case

A. Tighter Bounds for $R^{-1}(t)$

Much stronger results can be obtained if we assume that the sequence $\{\phi(t)\}$ is normally distributed. Assume, for convenience, that $N = mK$. Then we have the following:

Lemma 4: If the sequence $\{\phi(t)\}$ is stationary, Gaussian, and m-dependent then

$$\frac{R_0^{-1}}{N} \le E[R^{-1}(t)] \le \frac{R_0^{-1}}{N - m(r+1)} \qquad \forall t. \quad (31)$$

Proof: Observe that

$$R(t) = \sum_{j=1}^{m} G_j(t) \quad (32)$$

where

$$G_j(t) = \sum_{i=0}^{K-1} \phi(t-j-im+1)\phi^T(t-j-im+1).$$

Since the sequences $\{\phi(t-j-im+1),\ i \ge 0\}$ are i.i.d. and Gaussian, the matrices $G_j(t)$ are Wishart-distributed with K degrees of freedom

$$G_j(t) \sim W(KR_0, K). \quad (33)$$

Hence, using properties of the inverted Wishart distribution [16]

$$E[G_j^{-1}(t)] = \frac{R_0^{-1}}{K-r-1}. \quad (34)$$

Using the inequality (see Appendix I)

$$\left(\sum_{j=1}^{m} G_j(t)\right)^{-1} \le \frac{1}{m^2}\left(\sum_{j=1}^{m} G_j^{-1}(t)\right) \quad (35)$$

and combining it with (32)–(34) one obtains

$$E[R^{-1}(t)] \le \frac{R_0^{-1}}{m(K-r-1)}$$

which is nothing but the upper bound in (31). The lower bound in (31) stems from the fact that (see Appendix II)

$$E[R^{-1}(t)] \ge [E[R(t)]]^{-1} \quad (36)$$

(the matrix variant of the Jensen inequality for inverses).

B. Evaluation of Parameter Tracking Bounds

Several conclusions can be drawn from (31) for the sliding window LS estimators. First, observe that for the rectangular window

$$\tilde{R}(t) = \frac{1}{N} R(t)$$

and hence using (31) and (36) one gets

$$0 \le E[\Delta(t)] \le \frac{m(r+1)}{N - m(r+1)} R_0^{-1}$$

Consequently, for $N > m(r+1)$ we have (cf. Lemma 1)

$$E[\|\hat{\alpha}(t) - \hat{\hat{\alpha}}(t)\|^2 \le \frac{b}{N} \quad (37)$$

where b—we emphasise this fact strongly—is a *deterministic* constant not depending on N and t. We will look for the bound on the mean-square parameter tracking error in the form

$$E\left[\|\hat{\alpha}(t) - \alpha(t)\|^2\right] \le D(N).$$

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 29,2025 at 00:49:48 UTC from IEEE Xplore. Restrictions apply.

First, we consider the case of time-invariant parameters: $\alpha(t) = \alpha_0$. One can show that under A1–A3 $\hat{\alpha}(t)$ is an unbiased estimate of α_0 and

$$\operatorname{cov}[\hat{\alpha}(t)] = \rho_0 E[R^{-1}(t)].$$

Consequently

$$D(N) = \frac{D_1}{N} \tag{38}$$

which parallels (9), the result derived by Macchi and Eweda for LMS filters.

The counterpart of (10) can be derived using the inequality (16). Note that

$$E\left[\|\hat{\alpha}(t) - \alpha(t)\|^2\right] = E\left[\|\hat{\alpha}(t) - \bar{\alpha}(t)\|^2\right] + E\left[\|\bar{\alpha}(t) - \alpha(t)\|^2\right] \tag{39}$$

where $\{\bar{\alpha}(t)\}$ denotes the average path of parameter estimates

$$\bar{\alpha}(t) = E[\hat{\alpha}(t) | A(t)] = \sum_{i=0}^{\infty} w(i)\alpha(t-i) = \frac{1}{N}\sum_{i=0}^{N-1}\alpha(t-i)$$
$$A(t) = \{\alpha(t), \alpha(t-1), \cdots\}$$

and observe that there is no cross-coupling term on the right-hand side of (39) due to orthogonality of $\hat{\alpha}(t) - \bar{\alpha}(t)$ and $\bar{\alpha}(t) - \alpha(t)$.

Assuming that the true parameter trajectory can be modelled as random walk in sufficiently long but finite time interval $T = [t_1, t_2]$, $t_2 - t_1 \gg N$, one can show that ($t \in T$): $E[\|\hat{\alpha}(t) - \bar{\alpha}(t)\|^2] = 0(1)$, $E[\|\bar{\alpha}(t) - \alpha(t)\|^2] = 0(N)$ resulting in

$$D(N) = \frac{D_1}{N} + D_2 N. \tag{40}$$

We note however, that (40) is—unlike (10)—a *local* result, valid for finite, though possibly very long time intervals. Extension to infinite-time intervals is forbidden under A3 (only the mean-square bounded parameter trajectories can be analyzed in the present framework).

C. Extension to the Case of Nonuniform Weighting and Non-Gaussian Regressors

By applying the central limit theorem to the properly normalized elements of the matrix $\tilde{R}(t) - R_0$ and using the appropriate truncation technique one can show that

$$E\left[\|\hat{\alpha}(t) - \hat{\bar{\alpha}}(t)\|^2\right] = 0\left(\frac{1}{\iota}\right)$$

in the case of nonuniform weighting and non-Gaussian regressors. This, however, is based on asymptotic theory we were trying to avoid so far. Hence, it holds *only* as long as $\iota \to \infty$.

A more conservative but nonasymptotic bound can be obtained using the Schwartz inequality. Observe that

$$\operatorname{tr}\{E[\Delta(t)]\} = \operatorname{tr}\{E[\tilde{R}^{-1}(t)(R_0 - \tilde{R}(t))R_0^{-1}]\} = E[\operatorname{tr}\{\Delta_1(t)\,\Delta_2(t)\}]$$

where

$$\Delta_1(t) = R_0^{-1}\tilde{R}^{-1}(t), \qquad \Delta_2(t) = R_0 - \tilde{R}(t).$$

Let $\|\Delta(t)\|^2 = \operatorname{tr}\{\Delta(t)\Delta^T(t)\}$. Using Schwartz inequality and (36) one gets

$$0 \le \operatorname{tr}\{E[\Delta(t)]\} \le \left(E\left[\|\Delta_1(t)\|^2\right]\right)^{\frac{1}{2}}\left(E\left[\|\Delta_2(t)\|^2\right]\right)^{\frac{1}{2}}.$$

By a similar argument to that used in Section V-A $E[\|\Delta_1(t)\|^2] = 0(1)$. In the case of i.i.d. regressors

$$E[\|\Delta_2(t)\|^2 = E\left[\left\|\sum_{i=0}^{\infty} w(i)[\phi(t-i)\phi^T(t-i) - R_0]\right\|^2\right] = d\sum_{i=0}^{\infty} w^2(i)$$

where $d = E[\|\phi(t)\|^4] - (E[\|\phi(t)\|^2])^2$. More generally, one can show that for m-dependent regressors $E[\|\Delta_2(t)\|^2] = 0(1/\iota)$ and hence, combining all the results given previously with (17) one gets

$$E\left[\|\hat{\alpha}(t) - \hat{\bar{\alpha}}(t)\|^2\right] \le \frac{c}{\sqrt{\iota}} \tag{41}$$

where c is a deterministic constant not depending on ι and t. Combining this with (39), a bound analogous to (40) (but expressed in terms of ι) can also be derived for an arbitrary WLS estimator.

The bound (41) was derived for a system with time-varying coefficients. If the system is time-invariant, i.e., $\alpha(t) = \alpha_0$, the problem is much easier to handle. Due to the mutual independence of the processes $\{\phi(t)\}$ and $\{n(t)\}$, implied by assumption A2, one has

$$\operatorname{cov}[\hat{\alpha}(t)] = \rho_0 E\left[\tilde{R}^{-1}(t)\left(\sum_{i=0}^{\infty} w^2(i)\phi(t-i)\phi^T(t-i)\right)\tilde{R}^{-1}(t)\right] \le \frac{\rho_0}{\kappa}E[\tilde{R}^{-1}(t)] \tag{42}$$

where $\kappa = 1/(\max_t w(t))$ is the quantity usually called the effective width of the window (effective number of observations). One can easily show that for a sequence of windows of the same shape but increasing width it holds $\kappa \alpha \iota$, i.e., both measures of the window size differ merely by a constant multiplier.

According to (42), the fluctuations of WLS parameter estimates $E[\|\hat{\alpha}(t) - \alpha_0\|^2] = \operatorname{tr}\{\operatorname{cov}[\hat{\alpha}(t)]\}$ are, under stationary conditions, inversely proportional to the size of the applied window which is a further generalization of (38).

VI. Important Special Case—Exponentially Weighed LS Estimators

Quite clearly, if the window used in the method of WLS is strictly finite-length (i.e., if $w(t) = 0\ \forall t > t_0$) our technical assumption A5 is practically unavoidable. This is also a limitation of all results obtained using the concept of "inscribed" window. Using a slightly different technique, we will show that A5 is not needed any more if exponential weighting is applied. Rewrite the expression for the exponentially weighted LS estimator in the form

$$\hat{\alpha}(t) = R^{-1}(t)S(t) \tag{43}$$

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 29,2025 at 00:49:48 UTC from IEEE Xplore. Restrictions apply.

where now $(0 < \lambda < 1)$ as follows:

$$R(t) = \sum_{i=0}^{t-1} \lambda^i \phi(t-i)\phi^T(t-i),$$

$$S(t) = \sum_{i=0}^{t-1} \lambda^i y(t-i)\phi(t-i).$$

We note that (43) can be recursively updated using (5), (6) (note: $D(t) = R^{-1}(t)$ if the exact initialization is used, i.e., if $D(t_0)$ is set to $R^{-1}(t_0)$ for sufficiently large t_0). Note also that

$$R(t) = \lambda R(t-1) + \phi(t)\phi^T(t). \tag{44}$$

Exact initialization corresponds to taking $R(0) = 0$ in (44). However, in practice, recursion (6) is started using $D(0) = D_0 > 0$ which amounts to taking

$$R(0) > 0 \tag{45}$$

in (44) and which will play a crucial role in our analysis. We are ready to prove the following lemma.

Lemma 5: Suppose that $\{\phi(t)\}$ is an i.i.d. sequence such that $E[\phi(t)\phi^T(t)] = R_0 > 0$. Then the condition (23) of Lemma 2 is fulfilled and the number L in (23) can be made arbitrarily large by increasing $\iota = (1+\lambda)/(1-\lambda)$.

Proof: Note that $R(t) \geq \lambda^r R(t-r) + \lambda^{r-1} Q(t)$ where $Q(t) = \sum_{i=0}^{r-1} \phi(t-i)\phi^T(t-i)$ and

$$\det R(t) \geq \det R'(t) + \det Q'(t) \tag{46}$$

where $R'(t) = \lambda^r R(t-r)$, $Q'(t) = \lambda^{r-1} Q(t)$.

We have

$$\begin{aligned} P\big(\det R(t) \leq x\big) &= P\big(\det R(t) \leq x \big| \det Q'(t) < x_0\big) \\ &\quad \cdot P\big(\det Q'(t) < x_0\big) \\ &\quad + P\big(\det R(t) \leq x \big| \det Q'(t) \geq x_0\big) \\ &\quad \cdot P\big(\det Q'(t) \geq x_0\big). \end{aligned}$$

Using (46) and the fact that the matrices $Q'(t)$ and $R'(t)$ are independent, we obtain $P(\det R(t) \leq x \mid \det Q'(t) \geq x_0) = 0$ for all $x < x_0$, and

$$P\big(\det R(t) \leq x \big| \det Q'(t) < x_0\big) \leq P\big(\det R'(t) \leq x\big)$$

which results in

$$P\big(\det R(t) \leq x\big) \leq P\big(\det R'(t) \leq x\big) p_0' \tag{47}$$

where

$$p_0' = P\big(\det Q'(t) < x_0\big) = \frac{p_0}{\lambda^{r(r-1)}} \tag{48}$$

and

$$p_0 = P\big(\det Q(t) < x_0\big) < 1 \tag{49}$$

(since $E[\phi(t)\phi^T(t)] > 0$).

We will use inductive reasoning. Suppose that our assertion is true for $R(t-r)$, i.e., $P(\det R(t-r) \leq x) \leq \eta x^L$.

Then $\forall t < x_0$

$$\begin{aligned} P\big(\det R'(t) \leq x\big) &= P\Big(\lambda^{r^2} \det R(t-r) \leq x\Big) \\ &\leq \frac{\eta}{\lambda'} x^L, \qquad \lambda' = \lambda^{Lr^2} \end{aligned}$$

and hence, according to (47) $P(\det R(t) \leq x) \leq (p_0'\eta/\lambda') x^L \leq \eta x^L$, i.e., our assertion is true for $R(t)$ provided that $\lambda \geq \lambda_0$ such that

$$\ln \lambda_0 = \frac{\ln p_0}{(L+1)r^2 - r}.$$

Since $R(t) \geq \lambda^r R(0)$, $t = 1, \cdots, r$ we get $P(\det R(t) \leq x) = 0 \leq \eta x^L$ $t = 1, \cdots, r$ for all $x < x_0 = \det(\lambda^r R(0))$ and arbitrarily large L. Our assertion is therefore true for any t.

Finally, note that arbitrarily large value of L can be guaranteed in (23) provided that the forgetting constant λ is sufficiently close to 1. ■

Extension of Lemma 5 to ψ-mixing sequences (which includes m-dependence as a special case) and weaker covariance conditions is straightforward. Therefore, only assumptions A1–A3 are needed to guarantee boundedness of the mean-square parameter tracking error if the method of exponential weighting is used!

VII. Statistical Robustness

On the qualitative level the results obtained in previous sections raise several important issues which can be easily overlooked if a mechanical, "bookkeeping" approach towards certain mathematical details is adopted.

First of all, one should realize that results of Section IV indicate certain nonrobustness properties—as far as statistical analysis of WLS filters is concerned—of strictly finite-length windows. Assumption A5 admits a large class of continuous distributions but rules out all discrete ones. Is it a serious limitation? In a way it is. In the world of computers and digital processing, random variables with continuous distributions belong in mathematical "science fiction." Any form of quantization turns a continuous random variable into a discrete one. Hence, results of Section IV are not robust against quantization. The situation is essentially different if exponential weighting is applied. Let

$$p_{\min} = P\big(\det Q(t) = 0\big).$$

If $p_{\min} = 0$ one can make p_0, given by (49), arbitrarily small by decreasing x_0. This corresponds to the case where there is no discrete or singular component in $F(\phi)$. Presence of such components, however, does not destroy invertibility of $R(t)$ which was the case for finite length windows. Instead, it sets a lower bound on the forgetting constant λ

$$\ln \lambda_{\min} = \frac{\ln p_{\min}}{(L+1)r^2 - r},$$

i.e., the minimum equivalent width of the windows for which invertibility is guaranteed.

Appendix I
Derivation of (35)

The inequality can be easily proved by induction using the following proposition:

Proposition: For any two positive-definite matrices A and B and any integer m

$$m^2 A^{-1} + B^{-1} \geq (m+1)^2 (A+B)^{-1}.$$

Proof: The proof is straightforward in the scalar case. The multivariate case can be converted into the scalar one by performing the simultaneous diagonalization of matrices A and B

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 29,2025 at 00:49:48 UTC from IEEE Xplore. Restrictions apply.

(note: there exists a real matrix Q and a positive-diagonal matrix Λ such that: $Q^T A Q = \Lambda$ and $Q^T B Q = I$). □

Suppose that (35) holds for a certain m. Then, using the result of the aforementioned proposition, one gets

$$\sum_{j=1}^{m+1} G_j^{-1}(t) = \sum_{j=1}^{m} G_j^{-1}(t) + G_{m+1}^{-1}(t)$$
$$\geq m^2 \left(\sum_{j=1}^{m} G_j(t) \right)^{-1} + G_{m+1}^{-1}(t)$$
$$\geq (m+1)^2 \left(\sum_{j=1}^{m+1} G_j(t) \right)^{-1}$$

i.e., (35) is also true for $m + 1$. Since it is also true for $m = 1$ (cf. proposition above) it remains valid for any m.

Appendix II
Proof of (36)

Observe that for $R_0 = E[\tilde{R}(t)]$, $\tilde{R}(t) = R(t)/N$: $E[\tilde{R}^{-1}(t) - R_0^{-1}] = E[(\tilde{R}^{-1}(t) - R_0^{-1})\tilde{R}(t)(\tilde{R}^{-1}(t) - R_0^{-1})] \geq 0$ which is nothing but (36).

Acknowledgment

The first author would like to thank Prof. P. Hall from the Department of Statistics, Australian National University for his comments on the invertibility problem. We would also like to acknowledge the many helpful remarks of Dr. O. Macchi, which improved the readability of this paper.

References

[1] G. Goodwin and K. Sin, *Adaptive Filtering, Prediction and Control.* New York: Prentice-Hall, 1984.

[2] J. Cioffi, "Limited-precision effects in adaptive filtering," *IEEE Trans. Circ. Syst.*, vol. CAS-34, pp. 821-833, 1987.

[3] E. Eleftheriou and D. D. Falconer, "Tracking properties and steady state performance of RLS adaptive filter algorithms," *IEEE Trans. ASSP*, vol. ASSP-34, pp. 1097-1111, 1986.

[4] O. Macchi and E. Eweda, "Second order convergence analysis of stochastic adaptive linear filtering," *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 76-85, 1983.

[5] E. Eweda and O. Macchi, "Tracking error bounds of adaptive nonstationary filtering," *Automatica*, vol. 21, pp. 293-302, 1985.

[6] O. Macchi, "Optimization of adaptive identification for time-varying filters," *IEEE Trans. Automat. Contr.*, vol. AC-31, pp. 283-287, 1986.

[7] E. Eweda and O. Macchi, "Convergence of the RLS and LMS adaptive filters," *IEEE Trans. Circ. Syst.*, vol. CAS-34, pp. 799-803, 1987.

[8] M. Niedźwiecki, "On the localized estimators and generalized Akaike's criteria," *IEEE Trans. Automat. Contr.*, vol. AC-29, pp. 970-983, 1984.

[9] —, "First-order tracking properties of weighted least squares estimators," *IEEE Trans. Automat. Contr.*, vol. AC-33, pp. 94-96, 1988.

[10] —, "On tracking characteristics of weighted least squares estimators applied to nonstationary system identification," *IEEE Trans. Automat. Contr.*, vol. AC-33, pp. 96-98, 1988.

[11] —, "Optimization of the window shape in weighted least squares identification of a class of nonstationary systems," in *Proc. 7th Conf. Analysis Optimiz. Syst.*, Antibes, France, 1986.

[12] A. Benveniste and G. Ruget, "A measurement of tracking capability of recursive algorithms with constant gains," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 639-649, 1982.

[13] A. Benveniste, "Design of adaptive algorithms for the tracking of time-varying systems," *Int. J. Adaptive Contr. Signal Proc.*, vol. 1, pp. 3-29, 1987.

[14] H. J. Kushner and H. Huang, "Asymptotic properties of stochastic approximations with constant gains," *SIAM J. Contr.*, vol. 19, pp. 87-105, 1981.

[15] L. Ljung, "Adaptation and tracking in system identification," in *Proc. IFAC Symp. Ident. Syst. Param. Estimation*, vol. 1, Beijing, People's Republic of China, 1988, pp. 1-10.

[16] S. Das Gupta, "Some aspects of discrimination function coefficients," Sankhyā, vol. A-30, pp. 387-400, 1968.

[17] M. Green and J. B. Moore, "Persistence of excitation in linear systems," *Syst. Contr. Lett.*, vol. 7, pp. 351-360, 1986; see also, *preprints of ACC*, Boston, MA: 1985, pp. 412-417.

[18] H. F. Chen and L. Guo, "Adaptive control via consistent estimation for deterministic systems," *Int. J. Contr.*, vol. 45, no. 6, pp. 2183-2202, 1987.

[19] O. Macchi, and E. Eweda, "Compared speed and accuracy of the RLS and LMS algorithms with constant forgetting factors," *Traitement du signal*, vol. 22, pp. 255-267, 1988.

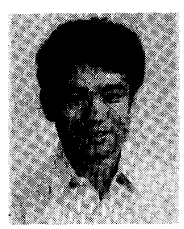
[20] D. C. Farden, "Tracking properties of adaptive signal processing algorithms," *IEEE Trans. ASSP*, vol. ASSP-29, pp. 439-446, 1981.

[21] S. K. Jones, R. K. Cavin, and W. M. Reed, "Analysis of error-gradient adaptive linear estimators for a class of stationary dependent processes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 318-329, 1982.

Maciej Niedźwiecki was born in Poznań, Poland on June 20, 1953. He received the M.Sc. and Ph.D. degrees from the Technical University of Gdańsk, Gdańsk, Poland, in 1977 and 1981, respectively.

He spent three years as a Research Fellow with the Department of Systems Engineering, Australian National University, 1986-1989. He is now with the Institute of Computer Science, Technical University of Gdańsk. His main areas of research interests include system identification, spectrum estimation, and adaptive systems.

Dr. Niedźwiecki is Vice-Chairman of the IFAC Committee on Theory for the triennium 1990-1993.



Lei Guo (M'88) was born in China in 1961. He received the B.S. degree in mathematics from Shandong University, Shandong, R.O.C., in 1982, and the M.S. and Ph.D. degrees in control theory from the Institute of Systems Science, Academia Sinica, Beijing, R.O.C. in 1984 and 1987, respectively.

From June 1987 to June 1989, he was with the Department of Systems Engineering, Australian National University, Canberra. Presently, he is an Associate Professor with the Institute of Systems Science, Academia Sinica.

Dr. Guo has authored or coauthored over 40 research papers in the areas of stochastic systems, adaptive control, system identification, time series analysis, and statistics of random processes.

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 29,2025 at 00:49:48 UTC from IEEE Xplore. Restrictions apply.

Technical Notes and Correspondence

Proportional-Plus-Multiintegral Stabilizing Compensators for a Class of MIMO Feedback Systems with Infinite-Dimensional Plants

Y.-P. Harn and E. Polak

Abstract—**A method is presented for designing proportional-plus-multiintegral stabilizing compensators for a class of feedback systems with exponentially stable infinite-dimensional plants. These compensators enable the feedback system to asymptotically track polynomial inputs and suppress polynomial disturbances of corresponding order.**

I. Introduction

Exponential stability, asymptotic tracking, and disturbance rejection are among the most fundamental requirements in control system design. Not surprisingly, over the years, these requirements have received a considerable amount of attention in the literature. In [5] and [6], Davison presented a characterization of a minimal-order robust error-driven servocompensator which achieves asymptotic tracking and disturbance rejection for finite-dimensional systems. The result was extended to distributed parameter systems in [1], [7] in which, because of the coprime factorization used to obtain it, the compensator turns out to be infinite-dimensional. Since, in practice, one needs to construct a finite-dimensional compensator, the approach in [1] has to be supplemented with cumbersome approximation and order reduction techniques. In [16], [17], [12]–[14], and [10], it is shown that feedback systems with exponentially stable infinite-dimensional plants can be stabilized and regulated by a multivariable proportional-plus-integral compensator of the following form:

$$\frac{1}{s}kK_I + K_p, \qquad 0 < k \le k^* \tag{1.1}$$

where K_I and K_p are real matrices whose dimensions are related to the input and output dimensions of the plant, k^* is some real positive number, and s is the Laplace parameter.

In this note, we present a method for designing finite-dimensional *proportional-plus-multiintegral* stabilizing compensators for a class of feedback systems with exponentially stable infinite-dimensional plants. The resulting feedback systems are internally stable and track, asymptotically, polynomial inputs and suppress, asymptotically, polynomial disturbances. Our analysis makes use of a characteristic function [8] for a class of feedback systems with infinite-dimensional plants, and of the Rouche theorem in complex variable theory [3]. The resulting proofs are quite straightforward.

When used in design, the effectiveness of the construction procedure presented in this note can be increased considerably by combining it with the computational stability criterion presented in [8] and semiinfinite optimization [18]. In this manner, the coefficient matrices of a compensator designed by our method can be further modified to ensure better feedback-system performance with respect to various performance requirements, without destroying stability, tracking, and disturbance rejection properties.

Manuscript received June 24, 1988; revised May 30, 1989. Paper recommended by Associate Editor, R. V. Patel. This work was supported in part by the National Science Foundation under Grant ECS-8713334, by the Air Force Office of Scientific Research under Contract AFOSR-86-0116, and by the State of California MICRO Program under Grant 532410-19900.

Y.-P. Harn was with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720. He is now with Integrated Systems Inc., Santa Clara, CA 95054.

E. Polak is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720.

IEEE Log Number 9040484.

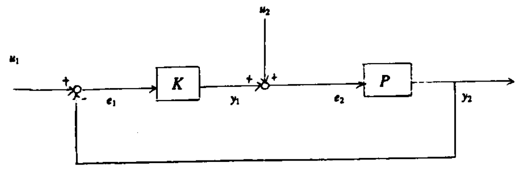


Fig. 1. The feedback system $S(P, K)$.

In Section II, we will introduce the descriptions of the plants that we can deal with and present some preliminary results. The main results will be established in Section III. Finally, we will give a numerical design example in Section IV.

II. Preliminary Results

Consider the feedback system $S(P, K)$ shown in Fig. 1, where the n_i-input and n_o-output plant is described by a linear time-invariant differential equation in a reflexive Banach space Z

$$\dot{x}_p(t) = A_p x_p(t) + B_p e_2(t); \quad y_2(t) = C_p x_p(t) + D_p e_2(t) \tag{2.1}$$

where $x_p(t) \in Z$, $e_2(t) \in \mathbb{R}^{n_i}$, $y_2(t) \in \mathbb{R}^{n_o}$, for $t \ge 0$.

Assumption 2.1: The operators B_p: $\mathbb{R}^{n_i} \to Z$, C_p: $Z \to \mathbb{R}^{n_o}$, and D_p: $\mathbb{R}^{n_i} \to \mathbb{R}^{n_o}$ are assumed to be bounded. The operator A_p from Z to Z, may be an unbounded operator with domain dense in Z, which generates a strongly continuous (C_0) semigroup, $\{e^{A_p t}\}_{t \ge 0}$.[1] ■

We denote the domain and the range of A_p by $D(A_p)$ and $R(A_p)$, respectively. We define the *transfer function* of the plant $G_p(s)$ to be $C_p(sI - A_p)^{-1}B_p + D_p$, $\forall s \in \rho(A_p)$, and we will denote its elements by $g_p^{i,j}(s)$. By [11, Theorem III 6.7], $G_p(s)$ is analytic on $\rho(A_p)$.

Definition 2.1: We say that a semigroup $\{T(t)\}_{t \ge 0}$ is exponentially stable in a Banach space if there exist $\gamma \in (0, \infty)$ and $\alpha > 0$ such that $\|T(t)\| \le \gamma e^{-\alpha t}$, $\forall t \ge 0$. ■

Assumption 2.2: The operator A_p generates an exponentially stable semigroup $\{e^{A_p t}\}_{t \ge 0}$, with $\alpha_0 > 0$ and $M_0 < \infty$ such that

$$\|e^{A_p t}\|_Z \le M_0 e^{-\alpha_0 t}, \qquad \forall t \ge 0. \tag{2.2}$$

■

Assumption 2.3: The matrix $G_p(0)$ has maximum rank. ■

For any $\alpha \ge 0$, we define the *stability region* $D_{-\alpha} \triangleq \{s \in \mathbb{C} \mid \mathrm{Re}(s) < -\alpha\}$, with complement $U_{-\alpha} \triangleq \{s \in \mathbb{C} \mid \mathrm{Re}(s) \ge -\alpha\}$, whose boundary and interior will be denoted by $\partial U_{-\alpha} \triangleq \{s \in \mathbb{C} \mid \mathrm{Re}(s) = -\alpha\}$ and $U^o_{-\alpha} \triangleq \{s \in \mathbb{C} \mid \mathrm{Re}(s) > -\alpha\}$, respectively. Finally, we define $\mathbb{C}^o_- \triangleq \{s \in \mathbb{C} \mid \mathrm{Re}(s) < 0\}$, $\mathbb{C}_+ \triangleq \{s \in \mathbb{C} \mid \mathrm{Re}(s) \ge 0\}$, and $\partial\mathbb{C}_+ \triangleq \{s \in \mathbb{C} \mid \mathrm{Re}(s) = 0\}$.

Definition 2.2: We will say that a function g: $\mathbb{C} \to \mathbb{C}$ *converges at infinity* in a domain $D \subset \mathbb{C}$, if there exists a finite complex number c such that $\lim_{\rho \to \infty} \sup_{|s| \ge \rho,\, s \in D} |g(s) - c| = 0$, and we will denote by $\lim_{|s| \to \infty,\, s \in D} g(s)$ this complex number c. We will say that a matrix function G: $\mathbb{C} \to \mathbb{C}^{m \times n}$ converges at infinity in a domain D if each of its elements converges at infinity in D. ■

[1] When the space Z is obtained via the extension of the domain of definition of an infinitesimal generator A_p, of a C_0 semigroup, then model (2.1) can represent a flexible beam with point actuators and sensors [4], [8], [9].

0018-9286/91/0200-0207$01.00 © 1991 IEEE

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 29,2025 at 00:49:48 UTC from IEEE Xplore. Restrictions apply.

Under Assumption 2.2, $U^o_{-\alpha_0} \subset \rho(A_p)$ [15, p. 11] and hence, $G_p(s)$ is analytic on $U^o_{-\alpha_0}$. It follows from [19] that $\lim_{|s| \to \infty, \mathrm{Re}\, s > -\alpha_0} G_p(s) = D_p$. Therefore, it is easy to show that the following proposition is true.

Proposition 2.1: Suppose that Assumption 2.2 holds. Then there exists $M < \infty$ such that each element of $G_p(s)$ satisfies

$$|g_p^{i,j}(s)| \le M, \forall s \in U^o_{-\alpha_0}, \quad i = 1, 2, \cdots, n_o, j = 1, 2, \cdots, n_i. \tag{2.3}$$

■

We assume that we are required to design a *minimal finite-dimensional proportional-plus-integral* compensator, described by a differential equation of the following form:

$$\dot{x}_c(t) = A_c x_c(t) + B_c e_1(t) \quad y_1(t) = C_c x_c(t) + D_c e_1(t) \tag{2.4}$$

where $x_c(t) \in \mathbb{R}^{n_c}$, $e_1(t) \in \mathbb{R}^{n_o}$, $y_1(t) \in \mathbb{R}^{n_i}$, and A_c, B_c, C_c, and D_c are matrices of appropriate dimension, with all the eigenvalues of A_c equal to zero, for integral action. Since $\sigma(A_c) = \{0\}$, the compensator transfer function is $G_c(s) = C_c(sI - A_c)^{-1}B_c + D_c = \sum_{j=0}^{m} F_j/s^j$, where the $F_j \in \mathbb{R}^{n_i \times n_o}$ and m depends on A_c. To ensure well-posedness of the closed-loop system, we assume that $\det(I_{n_i} + D_c D_p) \neq 0$.

We define the product space H by $H = Z \times \mathbb{R}^{n_c}$. Since $e_1 = u_1 - y_2$ and $e_2 = y_1 + u_2$, we obtain the following state equations for the closed-loop system:

$$\begin{pmatrix} \dot{x}_p \\ \dot{x}_c \end{pmatrix}(t) = A\begin{pmatrix} x_p \\ x_c \end{pmatrix}(t) + B\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}(t) \tag{2.5a}$$

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}(t) = C\begin{pmatrix} x_p \\ x_c \end{pmatrix}(t) + D\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}(t) \tag{2.5b}$$

where

$$A = \begin{pmatrix} A_p - B_p D_c (I_{n_o} + D_p D_c)^{-1} C_p & B_p (I_{n_i} + D_c D_p)^{-1} C_c \\ -B_c (I_{n_o} + D_p D_c)^{-1} C_p & A_c - B_c (I_{n_o} + D_p D_c)^{-1} D_p C_c \end{pmatrix}, \tag{2.6a}$$

$$B = \begin{pmatrix} B_p D_c (I_{n_o} + D_p D_c)^{-1} & B_p (I_{n_i} + D_c D_p)^{-1} \\ B_c (I_{n_o} + D_p D_c)^{-1} & -B_c (I_{n_o} + D_p D_c)^{-1} D_p \end{pmatrix}, \tag{2.6b}$$

$$C = \begin{pmatrix} -(I_{n_o} + D_p D_c)^{-1} C_p & -(I_{n_o} + D_p D_c)^{-1} D_p C_c \\ -D_c (I_{n_o} + D_p D_c)^{-1} C_p & (I_{n_i} + D_c D_p)^{-1} C_c \end{pmatrix},$$

$$D = \begin{pmatrix} (I_{n_o} + D_p D_c)^{-1} & -(I_{n_o} + D_p D_c)^{-1} D_p \\ D_c (I_{n_o} + D_p D_c)^{-1} & (I_{n_i} + D_c D_p)^{-1} \end{pmatrix}. \tag{2.6c}$$

The domain $D(A) = D(A_p) \times \mathbb{R}^{n_c} \subset H$; the operators B, C, and D are easily seen to be bounded. It follows from [15, p. 76] that because, with the exception of A_p, all the operators in the matrix A are bounded, and because $\mathrm{diag}(A_p, 0)$ generates a C_0-semigroup, the operator A also generates a C_0-semigroup $\{e^{At}\}_{t \ge 0}$.

Let $x = [x_p, x_c] \in H$. Then the formula $x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)} Bu(\tau)\, d\tau$ defines a *mild solution* of (2.5a) [15]. We therefore define the *exponential stability* of the feedback system $S(P, K)$ in terms of the semigroup $\{e^{At}\}_{t \ge 0}$.

Definition 2.3: The feedback system $S(P, K)$ is said to be exponentially stable if the semigroup $\{e^{At}\}_{t \ge 0}$ is exponentially stable. ■

The following result relating the exponential stability to the spectrum of the operator A is a special case of Proposition 2.1 in [8].

Proposition 2.2: Suppose that Assumptions 2.1–2.3 hold, then the feedback system is exponentially stable if and only if $\mathbb{C}_+$ is contained in $\rho(A)$. ■

We define the *characteristic function* $\chi(s)$ of the system $S(P, K)$ by

$$\begin{aligned} \chi(s) &= \det(sI_{n_c} - A_c)\det(I_{n_i} + G_c(s)G_p(s)) \\ &= s^{n_c}\det(I_{n_i} + G_c(s)G_p(s)) \\ &= s^{n_c}\det(I_{n_o} + G_p(s)G_c(s)). \end{aligned} \tag{2.7}$$

Next, for any function $f: \mathbb{C} \to \mathbb{C}$, we define $Z(f) \triangleq \{s \in \mathbb{C} \mid f(s) = 0\}$ to be its set of zeros. The following proposition follows directly from Theorem 3.1 in [8].

Proposition 2.3: the system $S(P, K)$ is exponentially stable if and only if $Z(\chi) \subset \mathbb{C}^o_-$. ■

III. Stabilizing Proportional-Plus-Multiintegral Compensators

We will establish the existence of a proportional-plus-multiintegral stabilizing compensator in three steps. First, we will show that we can construct a proportional stabilizing compensator. Then we will show that we can construct an integral stabilizing compensator. Finally, we will combine and extend these two results to show that one can construct proportional-plus-multiintegral stabilizing compensators of arbitrary order. As a corollary to the results in [1], we will obtain the fact that these compensators result in asymptotic error free tracking of polynomial inputs and in asymptotic polynomial output-disturbance suppression.

In the proofs to follow, we will make use of the Rouche theorem, stated as follows [3].

The Rouch Theorem: Let f, g: $\mathbb{C} \to \mathbb{C}$ be functions which are analytic inside and on a positively oriented simple closed contour C in the complex plane. If $|f(s)| > |g(s)|$ at each point s on C, then the functions $f(s)$ and $f(s) + g(s)$ have the same number of zeros, counting multiplicities, inside C. ■

Theorem 3.1: Consider the feedback system $S(P, K)$ in Fig. 1 and suppose that $A_c = 0$, $B_c = 0$, $C_c = 0$, and $n_c = 0$. Then there exists a matrix $D_c \neq 0$ such that the closed-loop system is exponentially stable.

Proof: By Proposition 2.3, the system $S(P, K)$ is exponentially stable if and only if $Z[\det(I_{n_i} + D_c G_p(s))] \subset \mathbb{C}^o_-$. Suppose that $D_c = [d^{i,j}]$. Then

$$\begin{aligned} \det(I_{n_c} + D_c G_p(s)) &= \det\left(\left[\Delta^{i,j} + \sum_{k=1}^{n_0} d^{i,k} g_p^{k,j}(s)\right]_{i,j}\right) \\ &= 1 + \sum_{l=1}^{n_i}\sum_{k=1}^{n_0} d^{l,k} g_p^{k,l}(s) + \cdots \\ &\triangleq 1 + H(s) \end{aligned} \tag{3.1}$$

where $\Delta^{i,j} = 1$ when $i = j$ and $\Delta^{i,j} = 0$ otherwise, and $H(s)$ represents the first and higher order terms in $d^{i,j}$ and $g_p^{i,j}(s)$. It follows from Proposition 2.1 that there exists $M > 0$ such that $|g_p^{i,j}(s)| < M$, for all $s \in \partial U_{-\alpha}$, where $0 < \alpha < \alpha_0$. It is clear that we can always choose a matrix $D_c \neq 0$, with sufficiently small components $d^{i,j}$ to ensure that $|H(s)| < 1$, for all $s \in \partial U_{-\alpha}$. Setting $C = \partial U_{-\alpha}$, $f(s) \equiv 1$, and $g(s) = H(s)$, we obtain from the Rouche theorem that $\det(I_{n_c} + D_c G_p(s)) = 1 + H(s)$ has the same number of zeros in $U^o_{-\alpha}$ as $f(\cdot)$, which is zero. Therefore,

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 29,2025 at 00:49:48 UTC from IEEE Xplore. Restrictions apply.