ON THE STABILITY OF THE INVERSE TIME-VARYING PREDICTION ERROR FILTER OBTAINED WITH THE RWLS ALGORITHM

Roberto López-Valcarce and Soura Dasgupta†*

Department of Electrical and Computer Engineering
University of Iowa, IA-52242, USA
e-mail: valcarce@icaen.uiowa.edu, dasgupta@eng.uiowa.edu

ABSTRACT

This work provides conditions on the input sequence that ensure the exponential asymptotic stability of the inverse of the forward prediction error filter obtained by means of the Recursive Weighted Least Squares algorithm. Note that this filter is in general time varying. Thus this result is a natural extension to the well-known minimum phase property of forward prediction error filters obtained by the autocorrelation method.

1. INTRODUCTION

Linear prediction finds application in such fields as speech modelling, differential pulse code modulation (DPCM), high-precision analog-to-digital converters and parametric spectrum estimation. Consider an m-order forward linear prediction error filter as given by

$$e(k) = u(k) + \sum_{j=1}^{m} a_j(k) u(k-j), \qquad (1)$$

where $u(\cdot)$ is the input data sequence, $e(\cdot)$ is the prediction error, and the filter coefficients $\{a_j(k)\}$ are those minimizing the Weighted Least Squares (WLS) cost function

$$\sum_{i=0}^{k} \lambda^{k-i} \left[u(i) + \sum_{j=1}^{m} a_j(k) u(i-j) \right]^2, \qquad (2)$$

where $0 < \lambda \leq 1$ is the forgetting factor.

In many applications (e.g. speech processing, [5]) the prediction error filter (1) is used to perform data analysis. The inverse filter, given by

$$y(k) = v(k) - \sum_{j=1}^{m} a_j(k) y(k-j), \qquad (3)$$

where $v(\cdot)$ is the input and $y(\cdot)$ the output, is used in data synthesis and modelling. Therefore, it is of importance to ensure that this inverse is exponentially asymptotically stable (e.a.s.) (see section 4 for a definition). In the non-recursive WLS scheme known as the autocorrelation method, a time-invariant prediction error filter is obtained from a *finite* data register and has the property of being minimum phase [2]. However, in recursive implementations (the Recursive Weighted Least Squares algorithm [RWLS]), the input data sequence need not be finite in principle, and the optimum filter is updated on a sample-by-sample basis [2]. *Therefore the inverse system becomes time-varying, rendering its stability analysis nontrivial. The main contribution of this paper is to show that under mild conditions this time varying inverse filter is e.a.s.* Our solution emphasizes the case where the input $u(\cdot)$ is not perfectly predictable by an m-order predictor, i. e. there do not exist constants a_i for which $e(\cdot)$ in (1) is identically zero.

Section 2 reviews the RWLS algorithm and the normal equations that solve the WLS problem. Section 3 gives the state-space description of the inverse of the forward prediction error filter. Sufficient conditions guaranteeing its asymptotic stability are given in section 4. Conclusions are in section 5.

2. RWLS AND THE NORMAL EQUATIONS

The RWLS agorithm uses the information contained in each new data sample $u(k)$ to obtain the filter coefficients $\mathbf{a}(k) = [\ a_1(k) \ \cdots \ a_m(k)\ ]'$ at time k recursively from those at time $k-1$. The algorithm is summarized in Table 1, where a soft-constrained initialization has been assumed. This starting procedure is commonly found in practice; it sets the initial value of the inverse autocorrelation matrix at $\mathbf{P}(0) = \delta^{-1}\mathbf{I}$, with $0 < \delta \ll 1$. See [3] for a discussion on the choice of δ. Due to this initial condition, which is used to avoid a singular autocorrelation matrix, RWLS yields the coefficient vector $\mathbf{a}(k)$ that minimizes the following cost

* Supported by Fundación Pedro Barrié de la Maza under grant no. 340056 and by NSF grant ECS-9350346.

† Supported in part by NSF grant ECS-9350346.

function [6]:

$$J(k) := \lambda^k \delta \mathbf{a}'\mathbf{a} + \sum_{i=0}^{k} \lambda^{k-i} \left[u(i) + \mathbf{a}'\mathbf{u}(j)\right]^2, \qquad (4)$$

with $\mathbf{u}(j)$ defined as in entry 3 of Table 1.

Table 1: RWLS algorithm for linear prediction.

Initialization:

$$\mathbf{a}(0) = \mathbf{0}, \qquad \mathbf{P}(0) = \delta^{-1}\mathbf{I}, \qquad \mathbf{u}(0) = [\ u(0) \quad \mathbf{0}\ ]'.$$

1. Coefficient update: $\mathbf{a}(k) = \mathbf{a}(k-1)$

$$-\frac{\mathbf{P}(k-1)\mathbf{u}(k-1)[u(k) + \mathbf{a}(k-1)'\mathbf{u}(k-1)]}{\lambda + \mathbf{u}(k-1)'\mathbf{P}(k-1)\mathbf{u}(k-1)}.$$

2. Inverse autocorrelation matrix update:

$$\mathbf{P}(k) = \lambda^{-1}\left[\mathbf{I} - \mathbf{g}(k-1)\mathbf{u}(k-1)'\right]\mathbf{P}(k-1).$$

3. Tap vector: $\mathbf{u}(k) = [\ u(k) \quad \cdots \quad u(k-m)\ ]'$.

Let us introduce, for $j \geq 0$, the autocorrelation coefficients

$$r_j(k) := \sum_{i=0}^{k} \lambda^{k-i} u(i) u(i-j),$$

the $m \times m$ positive semidefinite autocorrelation matrix

$$\mathbf{R}(k) := \begin{bmatrix} r_0(k) & r_1(k) & \cdots & r_{m-1}(k) \\ r_1(k) & r_0(k-1) & \cdots & r_{m-2}(k-1) \\ \vdots & \vdots & \ddots & \vdots \\ r_{m-1}(k) & r_{m-2}(k-1) & \cdots & r_0(k-m+1) \end{bmatrix},$$

and the $m \times 1$ autocorrelation vector

$$\mathbf{r}(k) := [\ r_1(k) \quad r_2(k) \quad \cdots \quad r_m(k)\ ]'.$$

The RWLS algorithm provides at every time instant k the vector $\mathbf{a}(k)$ minimizing $J(k)$ as given in (4), which is the solution of the following normal equations [2]:

$$\mathbf{Q}(k)\mathbf{a}(k) = -\mathbf{r}(k), \qquad (5)$$

with $\mathbf{Q}(k) := \lambda^k \delta \mathbf{I} + \mathbf{R}(k-1)$.

3. THE INVERSE SYSTEM

A state-space representation of the inverse prediction error filter (3) is

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{F}(k)\mathbf{x}(k) + \mathbf{e}_1 v(k), & (6) \\ y(k) &= \mathbf{H}(k)'\mathbf{x}(k) + v(k), & (7) \end{aligned}$$

where the matrices $\mathbf{F}(k)$, $\mathbf{e}_1$ and $\mathbf{H}(k)$ are given by

$$\mathbf{F}(k) := \begin{bmatrix} -a_1(k) & -a_2(k) & \cdots & -a_{m-1}(k) & -a_m(k) \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

$$\begin{aligned} \mathbf{e}_1 &:= [\ 1 \quad 0 \quad 0 \quad \cdots \quad 0\ ]', \\ \mathbf{H}(k) &:= [\ -a_1(k) \quad -a_2(k) \quad \cdots \quad -a_m(k)\ ]'. \end{aligned}$$

In this representation, the state vector simply comprises delays of the system output:

$$\mathbf{x}(k) = [\ y(k-1) \quad \cdots \quad y(k-m)\ ]'.$$

4. STABILITY ANALYSIS

We are interested in the eas (definition 1) of (6).

Definition 1 *The linear time-varying system* $\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k)$ *is e.a.s. if there exist* $0 < \beta < 1$ *and* α *such that for all* k_0 *and finite initial condition* $||\mathbf{x}(k_0)||$,

$$||\mathbf{x}(k)|| < \alpha \beta^{k-k_0} ||\mathbf{x}(k_0)|| \qquad \textit{for all} \quad k \geq k_0,$$

In other words, for zero input state response decays to zero at a rate faster than β^k.

Our approach is to obtain a Lyapunov equation involving $\mathbf{F}(k)$ from which sufficient conditions for stability can be derived. The cost function (4) can be rewritten as

$$J(k) = r_0(k) + 2\mathbf{a}'\mathbf{r}(k) + \mathbf{a}'[\lambda^k \delta \mathbf{I} + \mathbf{Q}(k)]\mathbf{a}.$$

Its minimum, $J_*(k)$, (obtained when $\mathbf{a} = \mathbf{a}(k)$ satisfies (5)), takes the value

$$J_*(k) = r_0(k) + \mathbf{a}(k)'\mathbf{r}(k). \qquad (8)$$

Making use of (8) and of the normal equations (5), direct verification shows that

$$\mathbf{Q}(k+1) - \mathbf{F}(k)\mathbf{Q}(k)\mathbf{F}(k)' = \begin{bmatrix} J_*(k) + \delta\lambda^{k+1} & & & \\ & \delta\lambda^k(\lambda - 1) & & \\ & & \ddots & \\ & & & \delta\lambda^k(\lambda - 1) \end{bmatrix} \qquad (9)$$

If $0 < \lambda < 1$, the right-hand side of (9) is not positive semidefinite and thus (9) is not a Lyapunov equation in the usual sense. However, it is still useful. Let us define the dual variables

$$\hat{\mathbf{Q}}(k) := \mathbf{Q}(-k+1),\ \hat{\mathbf{F}}(k) := \mathbf{F}(-k)',\ \hat{J}_*(k) := J_*(-k).$$

Then (9) reads as

$$\hat{\mathbf{Q}}(k) - \hat{\mathbf{F}}(k)'\hat{\mathbf{Q}}(k+1)\hat{\mathbf{F}}(k) = [\hat{J}_*(k) + \delta\lambda^{-k+1}]\mathbf{e}_1\mathbf{e}_1' + \delta\lambda^{-k}(\lambda-1)\mathbf{D}, \qquad (10)$$

with $\mathbf{D} := \mathrm{diag}(\ 0\quad 1\quad \cdots\quad 1\ )$. By the duality theorem [1], the system $\mathbf{x}(k+1) = \mathbf{F}(k)\mathbf{x}(k)$ is e.a.s. if and only if the system $\hat{\mathbf{x}}(k+1) = \hat{\mathbf{F}}(k)\hat{\mathbf{x}}(k)$ is e.a.s.; therefore it suffices to study the Lyapunov-like equation (10). At this point, we need the following result:

Lemma 1 *Suppose that the pair $[\mathbf{A}_k, \mathbf{c}_k]$ is uniformly detectable and bounded, and that there is a bounded nonnegative definite symmetric matrix sequence $\mathbf{P}_k$ satisfying for all k*

$$\mathbf{A}_k'\mathbf{P}_{k+1}\mathbf{A}_k - \mathbf{P}_k = -\mathbf{c}_k\mathbf{c}_k' + a_k\mathbf{B},$$

where a_k is scalar sequence such that $|a_k| \leq \alpha\beta^{|k|}$ for some $\alpha > 0$, $0 < \beta < 1$; and $\mathbf{B}$ is a symmetric nonnegative definite matrix. Then the system $\mathbf{x}_{k+1} = \mathbf{A}_k\mathbf{x}_k$ is e.a.s.

The proof closely resembles that of Theorem 4.2 in [1] (the extended lemma of Lyapunov), and is given in [4]. The precise definition of uniform detectability is not needed for the proof but can be found in [1]. Loosely speaking, a time varying system is uniformly detectable if state trajectories that are not observable at the output are exponentially decaying.

Thus, defining $\hat{\mathbf{v}}(k) := [\hat{J}_*(k) + \delta\lambda^{-k+1}]^{\frac{1}{2}}\mathbf{e}_1$, we see that this result can be readily applied to the system $\hat{\mathbf{x}}(k+1) = \hat{\mathbf{F}}(k)\hat{\mathbf{x}}(k)$ (since the matrix sequence $\hat{\mathbf{F}}(\cdot)$ satisfies (10)), provided that:

1. The matrix sequence $\hat{\mathbf{Q}}(\cdot)$ is bounded,
2. $\hat{\mathbf{F}}(\cdot)$, $\hat{\mathbf{v}}(\cdot)$ are bounded, and
3. The pair $[\hat{\mathbf{F}}(\cdot), \hat{\mathbf{v}}(\cdot)]$ is uniformly detectable.

Note that by the Cauchy-Schwarz inequality, $|r_j(k)|^2 \leq r_0(k)r_0(k-j)$. Therefore the sequence $\hat{\mathbf{Q}}(\cdot)$ is bounded if and only if $r_0(\cdot)$ is bounded. This will hold provided that $0 < \lambda < 1$ and $u(\cdot)$ is bounded in magnitude by M:

$$r_0(k) = \sum_{i=0}^{k} \lambda^{k-i}u^2(i) \leq \frac{M^2}{1-\lambda}.$$

Under these conditions, the vector sequence $\mathbf{r}(\cdot)$ is bounded as well; therefore, the weights $\mathbf{a}(\cdot)$ obtained as solutions of (5), and $J(\cdot)$ (in view of (8)), remain also bounded. Thus conditions 1 and 2 above hold.

We can now state our main result as follows:

Theorem 1 *Suppose that $0 < \lambda < 1$, the input sequence $u(\cdot)$ is bounded and there exist an integer S and a constant $\epsilon > 0$ such that for all k, there exists n_k satisfying:*

1. $n_k \geq k$,
2. $n_k + m - 1 \leq k + S$,
3. $J_*(n_k + i) \geq \epsilon$ *for* $i = 0, 1, \ldots m-1$.

Then the inverse prediction error filter (6) is e.a.s.

The proof is given in the appendix. The conditions 1-3 of the theorem say that in any time window of size S it is possible to find m consecutive time instants in which the forward prediction error variance $J_*(\cdot)$ is bounded away from zero. This is the case if for example $J_*(k) \geq c$ for some constant $c > 0$ and for all k.

To understand this result, suppose first that the data sequence $u(\cdot)$ is perfectly predictable by an m-order predictor. Then $u(\cdot)$ must be a linear combination of sinusoids. In that case the ideal predictor that renders $e(k) = 0$ for all k, in (1), has zeros with magnitude 1, i.e. the inverse predictor is not e.a.s. Hence, indeed one must preclude perfect predictability to achieve an e.a.s. inverse predictor. *The satisfaction of conditions of this Theorem in effect is a strong statement on the lack of perfect predictability.* In practice such perfect predictability is in any case unlikely. There are inevitable modeling and measurement errors that persistently prevent the prediction error from becoming zero.

5. CONCLUSIONS

It has been shown that the RWLS algorithm in the forward linear prediction setting yields a system whose inverse is exponentially asymptotically stable, under the following conditions: i) the input signal is bounded, ii) the forgetting factor is strictly between zero and one, and iii) the minimum of the WLS cost function (the sum of weighted error squares) is bounded away from zero during m consecutive time instants, m being the predictor order, in any time window of arbitrary but fixed size.

6. APPENDIX

Proof of theorem 1: It just remains to show that under the conditions of the theorem, the pair $[\hat{\mathbf{F}}(\cdot), \hat{\mathbf{v}}(\cdot)]$ is uniformly detectable. Introduce the following gain sequence:

$$\hat{\mathbf{h}}(k) := \begin{cases} [\hat{J}_*(k) + \delta\lambda^{-k+1}]^{-\frac{1}{2}}\mathbf{a}(-k), & \text{if} \quad \hat{J}_*(k) \geq \epsilon, \\ \mathbf{0}, & \text{if} \quad \hat{J}_*(k) < \epsilon. \end{cases}$$

Observe that the sequence $\hat{\mathbf{h}}(\cdot)$ is bounded. Thus, with

$$\tilde{\mathbf{F}}(k) := \hat{\mathbf{F}}(k) + \hat{\mathbf{h}}(k)\hat{\mathbf{v}}(k)',$$

the uniform detectability of $[\hat{\mathbf{F}}(\cdot), \hat{\mathbf{v}}(\cdot)]$ is equivalent to the uniform detectability of $[\tilde{\mathbf{F}}(\cdot), \hat{\mathbf{v}}(\cdot)]$ [1]. We will show that the system $\tilde{\mathbf{x}}(k+1) = \tilde{\mathbf{F}}(k)\tilde{\mathbf{x}}(k)$ is e.a.s.; this in turn implies uniform detectability of $[\tilde{\mathbf{F}}(\cdot), \hat{\mathbf{v}}(\cdot)]$ [1], thereby proving the Theorem.

Let $\mathbf{Z}$ be the $m \times m$ shift matrix with ones in the positions $(\mathbf{Z})_{i,i+1}$ and zeros elsewhere. It turns out that $\tilde{\mathbf{F}}(k)$ satisfies

$$\tilde{\mathbf{F}}(k) = \begin{cases} \mathbf{Z}, & \text{if} \quad \hat{J}_*(k) \geq \epsilon, \\ \hat{\mathbf{F}}(k) = \mathbf{F}(-k)', & \text{if} \quad \hat{J}_*(k) < \epsilon. \end{cases}$$

Now let $\Phi(k,l)$ be the state transition matrix for the system $\tilde{\mathbf{x}}(k+1) = \tilde{\mathbf{F}}(k)\tilde{\mathbf{x}}(k)$.

$$\Phi(k,l) = \begin{cases} \mathbf{I}, & k = l, \\ \tilde{\mathbf{F}}(k-1)\tilde{\mathbf{F}}(k-2)\cdots\tilde{\mathbf{F}}(l), & k > l. \end{cases}$$

Now the hypotheses of theorem 1 guarantee that the time window $[k, k+S]$ contains at least m consecutive time instants at which $\tilde{\mathbf{F}}$ reduces to $\mathbf{Z}$. To see this, we proceed as follows: Since for any j there exists n_j as in theorem 1, then in particular for $j = -(k+S)$ one has $n_{-(k+S)}$ such that

$$\begin{aligned} n_{-(k+S)} &\geq -(k+S), \\ n_{-(k+S)} + m - 1 &\leq k, \\ J_*(n_{-(k+S)} + i) &\geq \epsilon, \end{aligned}$$

for $i = 0,1, \ldots m-1$.

Now define $n'_k = -(n_{-(k+S)} + m - 1)$. Then $J_*(-n'_k - i) = \hat{J}_*(n'_k + i) \geq \epsilon$ for $i = 0,1, \ldots m-1$, and accordingly

$$\tilde{\mathbf{F}}(n'_k + i) = \mathbf{Z}, \qquad i = 0, 1, \ldots m-1.$$

Also note that $n'_k \geq k$ and $k + S \geq n'_k + m - 1$. The state transition matrix satisfies then

$$\begin{aligned} \Phi(n'_k + i, k) &= \tilde{\mathbf{F}}(n'_k + i - 1)\cdots\tilde{\mathbf{F}}(n'_k)\Phi(n'_k, k) \\ &= \mathbf{Z}^i \Phi(n'_k, k), \qquad i = 0, 1, \ldots m \end{aligned}$$

(with the convention $\mathbf{Z}^0 = \mathbf{I}$). Since $\mathbf{Z}^m = \mathbf{0}$, it turns out that $\Phi(n,k) = \mathbf{0}$ for all $n \geq k+S$:

$$\begin{aligned} \Phi(n,k) &= \Phi(n, n'_k + m)\Phi(n'_k + m, k) \\ &= \Phi(n, n'_k + m)\mathbf{Z}^m \Phi(n'_k, k) \\ &= \mathbf{0}. \end{aligned}$$

Therefore for all $n \geq k$,

$$\begin{aligned} &\tilde{\mathbf{x}}(n) = \\ &\Phi(n,k)\tilde{\mathbf{x}}(k) = \begin{cases} \mathbf{0}, & n \geq k+S, \\ \tilde{\mathbf{F}}(n-1)\cdots\tilde{\mathbf{F}}(k)\tilde{\mathbf{x}}(k), & n < k+S. \end{cases} \end{aligned}$$

Taking norms,

$$||\tilde{\mathbf{x}}(n)|| \leq \begin{cases} \mathbf{0} & \text{if} \quad n \geq k+S, \\ \gamma ||\tilde{\mathbf{x}}(k)|| & \text{if} \quad n < k+S, \end{cases}$$

with γ existing because $\tilde{\mathbf{F}}(\cdot)$ is bounded. Thus the system $\tilde{\mathbf{x}}(k+1) = \tilde{\mathbf{F}}(k)\tilde{\mathbf{x}}(k)$ is e.a.s., and therefore $[\tilde{\mathbf{F}}(k), \hat{\mathbf{v}}(k)]$ is uniformly detectable, which concludes our proof.

7. REFERENCES

[1] B.D.O. Anderson and J.B. Moore, "Detectability and stabilizability of time-varying discrete-time linear systems", SIAM J. Control and Optimization, vol. 19, pp. 20–32, January 1981.

[2] S. Haykin, *Adaptive Filter Theory, 3rd edition*. Prentice Hall, 1996.

[3] N. E. Hubing and S. T. Alexander, "Statistical analysis of the soft constrained initialization of recursive least squares algorithms", Proc. ICASSP, vol. 3, pp. 1277-1280, 1990.

[4] R. López-Valcarce and S. Dasgupta, "Exponential asymptotic stability of time-varying inverse prediction error filters", under preparation.

[5] J.D. Markel and A.H. Gray, Jr., *Linear prediction of speech*. Springer-Verlag, 1976.

[6] A.H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering", IEEE Signal Processing Mag., vol. 11, no. 3, pp. 18–60, July 1994.