ADAPTIVE RLS FILTERING UNDER THE CYCLO-STATIONARY REGIME

Hanoch Lev-Ari

Center for Communications and Digital Signal Processing
Northeastern University, Boston, MA 02115, USA

ABSTRACT

We present a methodology for adaptive filtering and system identification under the cyclostationary regime. Our technique is based on a deterministic *periodic* least-squares criterion, and gives rise to adaptive periodic recursive-least-squares (P-RLS) algorithms. Furthermore, we show that every adaptive RLS algorithm has a P-RLS counterpart, which has exactly the same architecture and the same performance attributes, and differs only in the length of the delay used in its time-update recursions.

1. INTRODUCTION

Adaptive algorithms have been successfully used in the last few decades in a variety of applications that involve stationary (or close-to stationary) signals, including: echo and interference cancellation, system identification, channel equalization and beamforming [1, 2]. Such signals are associated, either explicitly or implicitly, with linear time-invariant (resp. slowly-varying) systems. Adaptive algorithms are designed to determine (resp. track) the coefficients of such systems. However, the commonly-used adaptive algorithms cannot be used to estimate coefficients of *rapidly-varying* systems, which are encountered in numerous applications, ranging from communications via atmospheric and underwater fading channels, through marine seismography to study of biological signals. This is so because conventional adaptive algorithms use narrowband averaging which suppresses the time variation of rapidly-varying coefficients, and results in highly-biased estimates.

To be more specific, recall that the prototypical adaptive filtering problem involves the identification of the unknown coefficients $w_i(n)$ in the linear multiple-input/single-output relation [1]

$$d(n) = \sum_{i=1}^{M} w_i(n)\, x_i(n) + v(n) = W(n)X(n) + v(n) \tag{1a}$$

where $x_i(n)$ are the inputs, $d(n)$ is the output,

$$W(n) = [w_1(n) \;\; w_2(n) \;\; \ldots \;\; w_{\mathrm{M}}(n)] \tag{1b}$$

$$X(n) = [x_1(n) \;\; x_2(n) \;\; \ldots \;\; x_{\mathrm{M}}(n)]^T \tag{1c}$$

and where $v(\cdot)$ is an additive noise that is uncorrelated with the signals $x_i(\cdot)$. When the coefficients $w_i(\cdot)$ are *time-invariant* (i.e., independent of the time index n) this problem can be solved via the celebrated deterministic least squares method: given the finite signal records

$$\{\, d(n)\,;\; 0 \le n \le N \,\} \quad , \qquad \{\, x_i(n)\,;\; 0 \le n \le N \,\}$$

one forms the cost function

$$J(N) = \sum_{k=0}^{N} \lambda^k \left|\, e(N-k) \,\right|^2 \quad , \qquad 0 < \lambda < 1 \tag{2a}$$

where

$$e(n) = d(n) - \sum_{i=1}^{M} \widehat{w}_i\, x_i(n) = d(n) - \widehat{W}\, X(n) \tag{2b}$$

The unique minimum $\widehat{W}$ of $J(N)$ serves as an estimate of the unknown $W = W(n)$ of (1a). This estimate, when implemented in a time-recursive fashion, gives rise to the family of adaptive recursive least squares (RLS) algorithms [1, 2].

However, when the unknown coefficients $w_i(\cdot)$ are *time-variant*, the conventional RLS method tends to produce low-quality estimates. This is true, in particular, for *periodically-variant* coefficients: while some tracking capability can be retained by lowering the value of the "exponential forgetting factor" λ in (2a), this results in increased sensitivity to the additive noise $v(\cdot)$ of (1), as is evident from the example shown in Fig. 1.

In this paper we extend the deterministic least-squares method to the periodically-variant case: this results in time-recursive algorithms that we call *periodic recursive-least-squares* (P-RLS). The tracking performance of such algorithms under a cyclo-stationary

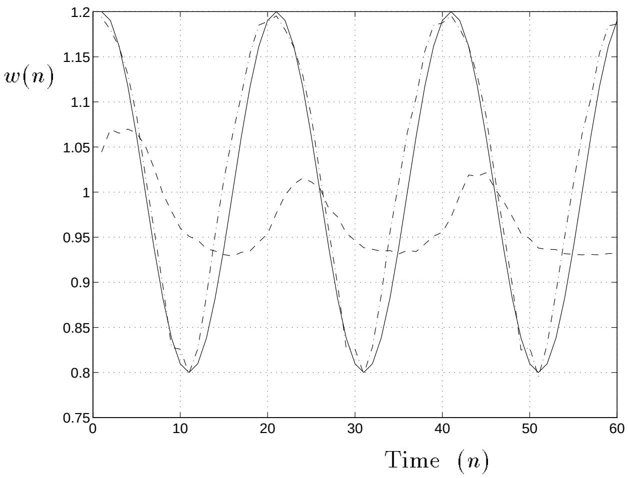

Figure 1: Performance comparison of: (i) standard (exponentially-weighted) RLS, with $\lambda = 0.93$ $(--)$; (ii) Periodic RLS $(-\cdot-)$. The true $w(n)$ is indicated by $—$.

regime is always superior to that of conventional RLS, as demonstrated by Fig. 1. Our main result is the establishment of a one-to-one correspondence between P-RLS and RLS algorithms, so that every adaptive RLS algorithm has a P-RLS counterpart, which has exactly the same architecture and differs only in using delays of length P (instead of unit delays) in its so-called time-update recursions.

In particular, we focus on adaptive fast P-RLS algorithms in *lattice form*. In Sec. 3 we present in detail one such algorithm – the adaptive P-RLS lattice algorithm. We show that it differs from its standard RLS counterpart only in the presence of a P-unit delay in the time-update recursions. The same correspondence exists between all P-RLS algorithms and their RLS counterparts. This means that all available analytical and experimental results concerning performance attributes of specific adaptive RLS algorithms – such as rate of convergence, steady state error, numerical robustness and computational cost – also hold for the P-RLS counterparts of these algorithms.

In Sec. 4 we discuss briefly the equivalence between the P-RLS approach and several existing techniques for identification of periodically-variant linear systems. These include the function series approach [3, 4], the weighted least-squares approach [5, 6], and certain multichannel embedding approaches [7].

2. PERIODIC LEAST-SQUARES

Our starting point for obtaining P-RLS algorithms is the *periodic least squares criterion*

$$J(N) = \sum_{k=0}^{[N/P]} \lambda^{kP} \left| e(N-kP) \right|^2 \quad , \quad 0<\lambda<1 \quad \text{(3a)}$$

where $[N/P]$ denotes the integer part of the fraction N/P, and where

$$e(n) = d(n) - \widehat{W}(n)\, X(n) \quad , \quad \widehat{W}(n+P) = \widehat{W}(n) \quad \text{(3b)}$$

In other words, we replace the fixed (time-invariant) $\widehat{W}$ of (2) by a periodically-variant $\widehat{W}(n)$.

The optimal $\widehat{W}(n)$, i.e., the one that minimizes the periodic least-squares cost $J(N)$ of (3), can be shown to satisfy the *time-variant* Wiener-Hopf equations [8]

$$\widehat{W}(n)\, R_{XX}(n) = R_{dX}(n) \quad , \quad 0 \le n \le N \qquad \text{(4a)}$$

where

$$R_{XX}(n) \;=\; \sum_{l} \lambda^{lP}\, X(n-lP)\, X^*(n-lP) \quad \text{(4b)}$$

$$R_{dX}(n) \;=\; \sum_{l} \lambda^{lP}\, d(n-lP)\, X^*(n-lP) \quad \text{(4c)}$$

and where the asterisk $(*)$ denotes a Hermitian transpose (= complex conjugate for scalars). We observe that by setting $P = 1$, the periodic least squares formulation (3)-(4) reduces to the standard deterministic least squares problem [1, 2]. As we show in Sec. 3, the same holds true for the adaptive P-RLS algorithms we derive: setting $P = 1$ reduces them to standard adaptive RLS algorithms. The starting point for derivation of time-recursive implementations of (4) is a time-recursive equivalent of (4b,c), namely

$$R_{XX}(n) \;=\; \lambda^P\, R_{XX}(n-P) + X(n)\, X^*(n) \quad \text{(5a)}$$

$$R_{dX}(n) \;=\; \lambda^P\, R_{dX}(n-P) + d(n)\, X^*(n) \quad \text{(5b)}$$

Again, when $P = 1$ we recognize (5) as the starting point for all standard RLS algorithms [1].

The characterization (4),(5) of the optimal coefficient vector estimate $\widehat{W}(n)$ can also be interpreted in probabilistic terms. Indeed, the linear relation (1) implies that the true coefficient vector $W(n)$ also satisfies a time-variant Wiener-Hopf equation, viz.,

$$E\left\{\, d(n)X^*(n) \,\right\} = W(n)\, E\left\{\, X(n)X^*(n) \,\right\}$$

because the additive noise $v(n)$ is uncorrelated with $X(n)$. If $d(n)$ and $X(n)$ are *jointly-cyclostationary* with a known period P, then the required probabilistic moments can be efficiently estimated by applying an (optimized) averaging filter with transfer function $H_P(z) = (1-\lambda^P)[1-(\lambda z^{-1})^P]^{-1}$ to the composite signals $d(n)\, X^*(n)$, and $X(n)\, X^*(n)$, respectively [9]. Thus, we recognize $R_{XX}(n)$ and

$R_{dX}(n)$ of (4),(5) as (improperly scaled) *periodically-averaged estimates* of the periodic second-order moments $E\{X(n)X^*(n)\}$ and $E\{d(n)X^*(n)\}$, obtained by using the improperly scaled averaging filter $[1-\lambda^P]^{-1} H_P(z) = [1-(\lambda z^{-1})^P]^{-1}$. The reason for the use of improper scaling is that while it does not affect the estimate $\widehat{W}(n)$ of (4a), it reduces the computational cost of the resulting adaptive RLS algorithms, as compared with algorithms based on the appropriately scaled versions. This fact is well-known for the (standard) case $P=1$ [1].

3. FAST P-RLS ALGORITHMS

As with standard RLS, one obtains fast algorithms (i.e., with $\mathcal{O}(M)$ computations per time instant instead of $\mathcal{O}(M^2)$ computations) by exploiting the so-called "shift property" $x_i(n) = x(n-i+1)$, which holds in many adaptive filtering applications [1]. In this case, the recursive matrix relation (4b) admits a simple geometric interpretation, as described in [10]. Thus, introduce the data (row) vectors

$$\mathbf{d}(n) = [d(n)\ \ d(n-P)\ \ d(n-2P)\ \ \ldots] \quad (6a)$$
$$\mathbf{x}(n) = [x(n)\ \ x(n-P)\ \ x(n-2P)\ \ \ldots] \quad (6b)$$

of some fixed length L (long enough to include all available past data), and define the inner product for any two (row) vectors a, b of length L, as follows,

$$\langle a, b \rangle = a\,\Lambda\, b^* \ , \quad \Lambda = \mathrm{diag}\{1, \lambda^P, \ldots, \lambda^{P(L-1)}\} \quad (6c)$$

We now observe that (5a) reduces to

$$\langle \mathbf{x}(n), \mathbf{x}(k) \rangle = \lambda^P \langle \mathbf{x}(n-P), \mathbf{x}(k-P) \rangle + x(n)x^*(k) \quad (7)$$

and a similar interpretation holds for (5b). In the language of [10] this means that the data vector

$$\mathbf{x}(n)_\pi = [0\ \ x(n-P)\ \ x(n-2P)\ \ \ldots]$$

is isometric (or *congruent*) to the delayed and scaled data vector $\sqrt{\lambda^P}\,\mathbf{x}(n-P)$, where the so-called "pinning vector" $\boldsymbol{\pi} = [1\ \ 0\ \ \ldots]$ is the same as the one used in [10] for standard RLS. Using the technique of [10] in conjunction with the geometric interpretation (6),(7), we can derive a variety of adaptive fast RLS algorithms, both in lattice and transversal form (see also [11]). In the process we observe that the only difference between the geometric set-up of the P-RLS problem and that of the standard RLS is the presence of a P-unit delay in the fundamental time-update relation (7): setting $P=1$ in (7) produces, for instance, eq. (25) of [10].

To illustrate this principle we describe here the periodic counterpart of the best known version of the adaptive RLS lattice algorithm, i.e., the so-called quotient-form with aposteriori residuals [1]. From the geometric interpretation (6),(7), and using essentially the same notation as in Table 15.4 of [1], we conclude that:

(i) The order-update relations remain independent of the value of P, viz.,

$$f_m(n) = f_{m-1}(n) - K_m^f(n)\, b_{m-1}(n-1) \quad (8a)$$
$$b_m(n) = b_{m-1}(n-1) - K_m^b(n)\, f_{m-1}(n) \quad (8b)$$

with

$$K_m^f(n) = \Delta_{m-1}(n)/\mathcal{B}_{m-1}(n-1) \quad (8c)$$
$$K_m^b(n) = \Delta_{m-1}^*(n)/\mathcal{F}_{m-1}(n) \quad (8d)$$

(ii) The time-update relations are modified by the introduction of a P-unit delay, viz.,

$$\Delta_{m-1}(n) = \lambda^P \Delta_{m-1}(n-P) + \frac{f_{m-1}(n)\, b_{m-1}^*(n-1)}{\gamma_{m-1}(n-1)} \quad (8e)$$

$$\mathcal{F}_{m-1}(n) = \lambda^P \mathcal{F}_{m-1}(n-P) + \frac{|\, f_{m-1}(n)\,|^2}{\gamma_{m-1}(n-1)} \quad (8f)$$

$$\mathcal{B}_{m-1}(n-1) = \lambda^P \mathcal{B}_{m-1}(n-1-P) + \frac{|\, b_{m-1}(n-1)\,|^2}{\gamma_{m-1}(n-1)} \quad (8g)$$

(iii) the conversion factor γ is updated via the same order recursion an in standard RLS, viz.,

$$\gamma_m(n-1) = \gamma_{m-1}(n-1) - \frac{|\, b_{m-1}(n-1)\,|^2}{\mathcal{B}_{m-1}(n-1)} \quad (8h)$$

In summary, this version of the adaptive P-RLS lattice algorithm is identical in every respect to its standard RLS counterpart, except for the presence of the P-unit delay in (8d-f), and this difference disappears when we set $P=1$. The same conclusion holds true for every exact implementation of the RLS algorithm, including the standard RLS, the square-root RLS and all lattice and transversal variants of the fast RLS.

4. EQUIVALENCE TO ALTERNATIVE PERIODIC SYSTEM ID TECHNIQUES

In Sec. 2 we have justified the periodic averaging technique (4) by relating it to the method of optimized averaging for time-variant moment estimation [9]. Here we comment briefly on the equivalence of (4) to several other known techniques for identification of periodically time-variant systems.

In the function series expansion approach of [4] one represents the time-variant $w_i(n)$ in terms of their expansion coefficients with respect to a sequence of given (orthonormal) functions. In particular, periodic time variation is characterized by Fourier series expansions [3], viz.,

$$\widehat{w}_i(n) = \sum_{q=0}^{P-1} \nu_{iq}\, e^{j\frac{2\pi}{P}qn} \tag{9a}$$

where P is the underlying period. The cost function $J(N)$ of (2a) is then minimized with respect to the parameters $\{\nu_{iq}\}$, where now

$$e(n) = d(n) - \sum_{i=1}^{M} \widehat{w}_i(n) x_i(n) = d(n) - \widehat{W}(n) X(n) \tag{9b}$$

The solution of this quadratic optimization problem leads to a set of MP Wiener-Hopf equations in the MP unknown Fourier parameters

$$\{\, \nu_{iq}\,;\ 1 \leq i \leq M\ ,\ \ 0 \leq q \leq P-1 \,\}$$

Once the optimal ν_{iq} have been determined, the optimal periodic $\widehat{W}(n)$ is recovered via (9a). Since the constraint imposed by the Fourier series expansion (9a) is equivalent to the periodicity constraint $\widehat{W}(n+P) = \widehat{W}(n)$, it follows that the method of [3, 4] results in the same $\widehat{W}(n)$ as the one obtained by solving the periodic least-squares problem (4).

The periodic least-squares characterization (4) coincides with the weighted least-squares (WLS) approach if one chooses the weighting sequence of WLS to be the impulse response of either the averaging filter $H_P(z)$, or of its improperly scaled version [5, 6].

Since every (discrete-time) cyclostationary signals can be viewed as a multichannel stationary signal, one can bring standard (i.e., stationary) techniques to bear on the periodic system identification problem. One such embedding is discussed in [7], where it is shown that the resulting periodic estimate $\widehat{W}(n)$ satisfies our characterization (4). The most direct embedding, known as "circular" [12], can also be shown to result in the same $\widehat{W}(n)$ as in (4) [8].

In summary, a variety of identification methods for periodically-variant linear systems all produce the same estimate $\widehat{W}(n)$ as the periodic least-squares approach that we described in this paper.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, 1991.

[2] J.G. Proakis, C.M. Rader, F. Ling and C.L. Nikias, *Advanced Digital Signal Processing*, Macmillan, New York, 1992.

[3] R.H. Jones and W.M. Brelsford, "Time series with periodic structure," *Biometrika*, Vol. 54, pp. 403–408, 1967.

[4] Y. Grenier, "Time-Dependent ARMA Modeling of Nonstationary Signals," *IEEE Trans. Acoust. Speech and Signal Processing*, Vol. ASSP-31, pp. 899-911, Aug. 1983.

[5] M. Niedzwiecki, "On Tracking Characteristics of Weighted Least Squares Estimators Applied to Nonstationary System Identification," *IEEE Trans. Automat. Contr.*, Vol. 33, pp. 96-98, Jan. 1988

[6] L. Ljung and S. Gunnarson, "Adaptation and Tracking in System Identification – A Survey," *Automatica*, Vol. 26, pp. 7-21, Jan. 1990.

[7] H. Sakai, "On the Spectral Density Matrix of a Periodic ARMA Process," *J. Time Ser. Anal.*, Vol. 12, No. 1, pp. 73-82, 1991.

[8] H. Lev-Ari, "On Parameter Identification for AR Models for Nonstationary Signals," Research notes (unpublished), Stanford Univ., Feb. 1988.

[9] P.W. Wu and H. Lev-Ari, "Linear Estimation of Moments for Non-Stationary Signals," *IEEE Trans. Sig. Proc.*, vol. 45, pp. 1210-1221, May 1997.

[10] H. Lev-Ari, T. Kailath and J.M. Cioffi, "Least-Squares Adaptive Lattice and Transversal Filters: A Unified Geometric Theory," *IEEE Trans. Inform. Theory*, Vol. IT-30, pp. 222-235, March 1984.

[11] H. Lev-Ari, "Modular Architectures for Adaptive Multichannel Lattice Algorithms," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-35, pp. 543-552, Apr. 1987.

[12] M. Pagano, "Periodic and multiple autoregressions," *Ann. Statist.*, Vol. 6, pp. 1310–1317, 1978.