ON CONVERGENCE OF PROPORTIONATE-TYPE NLMS ADAPTIVE ALGORITHMS

Miloš Doroslovački

The George Washington University
Washington, DC 20052, USA

Hongyang Deng

Acoustic Technologies, Inc.
Mesa, AZ 85204, USA

ABSTRACT

We specify the general form of proportionate-type NLMS adaptive algorithms and show that for sufficiently small adaptation stepsize parameter, the algorithms can be exponentially stable, globally convergent and robust to unmodeled dynamics and measurement noise. Also, we show that for small adaptation stepsize parameter and stationary inputs, behavior of proportionate-type NLMS algorithms can be modeled by proportionate-type steepest descent algorithms. This motivates designing of proportionate-type NLMS adaptive algorithms by looking at the adjoint proportionate-type steepest descent algorithms.

1. INTRODUCTION

Proportionate-type normalized least-mean-square (NLMS) algorithms have been recently proposed and studied in [1]-[8]. Their general form is

$$\hat{\mathbf{w}}(k+1) = \hat{\mathbf{w}}(k) + \frac{\beta \mathbf{G}(k+1)\mathbf{x}(k)e(k)}{\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta} \quad (1)$$

where $\hat{\mathbf{w}}(k)$ is the vector of adaptive filter weights, $\mathbf{x}(k)$ is the input to the adaptive filter and the unknown system to be identified, β is the adaptation stepsize parameter, $\mathbf{G}(k+1)$ is the gain control matrix, $e(k)$ is the output error, i.e. the difference between the measurement of the output of the unknown system and the output of adaptive filter, and δ is small positive real number. The error $e(k)$ can be modeled as

$$e(k) = \mathbf{x}^T(k)\mathbf{w}_0 - \mathbf{x}^T(k)\hat{\mathbf{w}}(k) + v(k) = \mathbf{x}^T(k)\tilde{\mathbf{w}}(k) + v(k) \quad (2)$$

where $\mathbf{w}_0$ is the unknown weight vector of the system to be identified, $\tilde{\mathbf{w}}(k) = \mathbf{w}_0 - \hat{\mathbf{w}}(k)$, and $v(k)$ is the output noise. The gain control matrix $\mathbf{G}(k+1)$ is diagonal positive-definite matrix whose diagonal entries are between a small positive real number ρ and L, where L is the number of adaptive weights. The trace of $\mathbf{G}(k+1)$ is equal to L for every k. $\mathbf{G}(k+1)$ depends only on $\hat{\mathbf{w}}(k)$, i.e. $\mathbf{G}(k+1) = \mathbf{F}(\hat{\mathbf{w}}(k))$. The original proportionate NLMS algorithms [1], [2] are derived heuristically, with the objective to increase the convergence rate when identifying sparse systems described by $\mathbf{w}_0$.

The goal of this paper is twofold. First, the paper will provide general convergence results for proportionate-type NLMS adaptive algorithms. Second, a closeness relation between the proportionate-type NLMS adaptive algorithms and the proportionate-type steepest descent algorithms will be established. The form of the proportinate-type steepest descent algorithms considered here is

$$\tilde{\mathbf{z}}(k+1) = (\mathbf{I} - \mu\mathbf{H}(k+1)\mathbf{R})\tilde{\mathbf{z}}(k) \quad (3)$$

where $\mathbf{R}$ is the autocorrelation matrix of the stationary input vector, μ is the stepsize parameter, $\mathbf{H}(k+1)$ is the stepsize control matrix, and $\tilde{\mathbf{z}}(k) = \mathbf{w}_{opt} - \mathbf{w}(k)$. The matrix $\mathbf{H}(k+1)$ is diagonal and positive definite. Its trace is equal to L. $\mathbf{w}_{opt}$ is the optimal (Wiener) solution for the identification problem and we assume $\mathbf{w}_{opt} = \mathbf{w}_0$ (i.e., the model order is equal to the order of the unknown system). $\{\mathbf{w}(k)\}_{k\in\mathcal{N}_0}$ is the sequence of iterative solutions for the model weights. If the conditions for closeness between (3) and (1) are satisfied, convergence behavior of the proportionate-type NLMS adaptive algorithms can be inferred from the analysis of the proportionate-type steepest descent algorithms.

The existing studies of proportionate-type NLMS adaptive algorithms are missing to address these two fundamental issues for analysis of adaptive algorithms.

2. CONVERGENCE OF PROPORTIONATE-TYPE NLMS ALGORITHMS FOR SMALL ADAPTATION STEPSIZE PARAMETER

We will show that the algorithm (1) provides the weight convergence to the optimal value by [9, Theorem 1] for sufficiently small β. We will assume that there is no output (measurement) noise, i.e. $v(k) \equiv 0$. Now, we can rewrite (1) as

$$\begin{aligned}\tilde{\mathbf{w}}(k+1) &= \tilde{\mathbf{w}}(k) - \frac{\beta \mathbf{G}(k+1)\mathbf{x}(k)\mathbf{x}^T(k)\tilde{\mathbf{w}}(k)}{\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta} \\ &= \left(\mathbf{I} - \beta\frac{\mathbf{G}(k+1)\mathbf{x}(k)\mathbf{x}^T(k)}{\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta}\right)\tilde{\mathbf{w}}(k). \quad (4)\end{aligned}$$

Let us define

$$\mathbf{A}_k = \frac{\mathbf{G}(k+1)\mathbf{x}(k)\mathbf{x}^T(k)}{\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta}. \quad (5)$$

It can be easily shown that $\mathbf{A}_k$ is bounded. More specifically,

$$|[\mathbf{A}_k]_{ij}| \le \|\mathbf{A}_k\| < 1 \quad (6)$$

where $[\mathbf{A}_k]_{ij}$ is the ij-th entry of $\mathbf{A}_k$ and $\|\mathbf{A}_k\|$ is the induced norm of $\mathbf{A}_k$. Now, after choosing $\mathbf{P} = \frac{1}{2}\mathbf{I}$, [9, Theorem 1] claims that there exists β^* such that (4) is uniformly exponentially asymptotically stable for every $0 < \beta < \beta^*$, if $\exists m$ and $\exists \alpha$ such that $\forall k$

$$\mathbf{z}^T \frac{1}{m} \sum_{i=1}^{m} (\mathbf{P}\mathbf{A}_{k+i-1} + \mathbf{A}_{k+i-1}^T \mathbf{P})\mathbf{z} = \mathbf{z}^T \frac{1}{m} \sum_{i=1}^{m} \mathbf{A}_{k+i-1}\mathbf{z} > \alpha, \quad \forall \mathbf{z} : \|\mathbf{z}\| = 1. \tag{7}$$

We are going to show that the condition (7) is satisfied when the weight control matrix is slowly varying, i.e.,

$$\mathbf{G}(k+i) \approx \mathbf{G}(k+1), \quad \forall k, i = 1, 2, \cdots, m \tag{8}$$

and when for the normalized input $\exists m$ such that $\forall k$

$$\frac{1}{m} \sum_{i=1}^{m} \frac{\mathbf{x}(k+i-1)\mathbf{x}^T(k+i-1)}{\mathbf{x}^T(k+i-1)\mathbf{G}(k+1)\mathbf{x}(k+i-1)+\delta} \approx \mathbf{diag}\{d_1(k), \cdots, d_L(k)\} > \alpha' \mathbf{I}, \quad \alpha' > 0. \tag{9}$$

Using (8) and (9) we obtain

$$\begin{aligned}\sum_{i=1}^{m} \mathbf{A}_{k+i-1} &= \sum_{i=1}^{m} \frac{\mathbf{G}(k+i)\mathbf{x}(k+i-1)\mathbf{x}^T(k+i-1)}{x^T(k+i-1)\mathbf{G}(k+i)\mathbf{x}(k+i-1)+\delta} \\ &\approx \mathbf{G}(k+1) \sum_{i=1}^{m} \frac{\mathbf{x}(k+i-1)\mathbf{x}^T(k+i-1)}{x^T(k+i-1)\mathbf{G}(k+1)\mathbf{x}(k+i-1)+\delta} \\ &\approx \mathbf{G}(k+1)\mathbf{diag}\{d_1(k), \cdots, d_L(k)\}. \end{aligned} \tag{10}$$

Since in our case $\mathbf{G}(k+1)$ is a diagonal positive definite matrix with eigenvalues between ρ and L, we have

$$\mathbf{z}^T \frac{1}{m} \sum_{i=1}^{m} \mathbf{A}_{k+i-1}\mathbf{z} \approx \mathbf{z}^T \mathbf{G}(k+1)\mathbf{diag}\{d_1(k), \cdots, d_L(k)\}\mathbf{z} > \rho\alpha' \tag{11}$$

i.e., (7) is satisfied with $\alpha = \rho\alpha'$. Note that $\mathbf{G}(k+1)$ depends on $\hat{\mathbf{w}}(k)$ and therefore a sufficient small β can provide $\mathbf{G}(k+1)$ that is slowly varying (β^* must provide the exponential asymptotic stability and sufficient slowness of $\hat{\mathbf{w}}(k)$). The input $\{x(k)\}_{k\in\mathcal{Z}}$ that randomly fluctuates around zero ("Gaussian zero-mean white noise-like behavior") can provide closeness to the diagonality required in (9). The persistently spanning input, i.e., if $\exists m$ and $\exists \alpha''$ such that $\forall k$

$$\mathbf{z}^T \frac{1}{m} \sum_{i=1}^{m} \frac{\mathbf{x}(k)\mathbf{x}^T(k)}{\mathbf{x}^T(k)\mathbf{x}(k)+\delta}\mathbf{z} > \alpha'', \quad \forall \mathbf{z} : \|\mathbf{z}\| = 1 \tag{12}$$

implies positive definiteness in (9) since

$$\begin{aligned} &\mathbf{z}^T \frac{1}{m} \sum_{i=1}^{m} \frac{\mathbf{x}(k+i-1)\mathbf{x}^T(k+i-1)}{\mathbf{x}^T(k+i-1)\mathbf{G}(k+i)\mathbf{x}(k+i-1)+\delta}\mathbf{z} \\ &= \mathbf{z}^T \frac{1}{m} \sum_{i=1}^{m} \frac{\mathbf{x}(k+i-1)\mathbf{x}^T(k+i-1)}{\mathbf{x}^T(k+i-1)\mathbf{x}(k+i-1)+\delta} \\ &\quad \times \frac{\mathbf{x}^T(k+i-1)\mathbf{x}(k+i-1)+\delta}{\mathbf{x}^T(k+i-1)\mathbf{G}(k+i)\mathbf{x}(k+i-1)+\delta}\mathbf{z} \\ &> \mathbf{z}^T \frac{1}{m} \sum_{i=1}^{m} \frac{\mathbf{x}(k+i-1)\mathbf{x}^T(k+i-1)}{\mathbf{x}^T(k+i-1)\mathbf{x}(k+i-1)+\delta} \\ &\quad \times \frac{\mathbf{x}^T(k+i-1)\mathbf{x}(k+i-1)+\delta}{L\mathbf{x}^T(k+i-1)\mathbf{x}(k+i-1)+\delta}\mathbf{z} \\ &> \frac{1}{L}\mathbf{z}^T \frac{1}{m} \sum_{i=1}^{m} \frac{\mathbf{x}(k+i-1)\mathbf{x}^T(k+i-1)}{\mathbf{x}^T(k+i-1)\mathbf{x}(k+i-1)+\delta}\mathbf{z} \\ &> \frac{1}{L}\alpha'' = \alpha'. \end{aligned} \tag{13}$$

[9, Theorem 1] considers noiseless case, but since the exponential asymptotic stability is claimed, this guarantees "robustness in the presence of nonidealities such as unmodeled dynamics or measurement noise" [9]. The exponential asymptotic stability implies that $\hat{\mathbf{w}}(k)$ is exponentially convergent to $\mathbf{w}_0$. Also, the convergence is global since no linearization is used and the input $\{x(k)\}_{k\in\mathcal{Z}}$ is not assumed bounded [9, p. 397].

3. CLOSENESS OF PROPORTIONATE-TYPE NLMS AND STEEPEST-DESCENT ALGORITHMS

We will now show that for small β and the time-invariant matrix $\mathbf{G}(k+1)$, i.e. $\mathbf{G}(k+1) = \mathbf{G}\ \forall k$, the weight trajectories obtained by a proportionate-type NLMS algorithm can be close to the weight trajectories obtained by a proportionate steepest descent algorithm. [10, Theorems 9.1, 9.3, 9.5] establish closeness between the general primary stochastic system and the first order associate averaged system. They claim closeness of the trajectories of the two systems on the finite interval [10, Theorem 9.1] and infinite interval [10, Theorem 9.5], as well as of the coefficient fluctuations of the two systems on the finite interval [10, Theorem 9.3]. In our case the closeness is established between the primary system

$$\begin{aligned}\tilde{\mathbf{w}}(k+) = \tilde{\mathbf{w}}(k) &- \beta\frac{\mathbf{G}\mathbf{x}(k)\mathbf{x}^T(k)}{\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)+\delta}\tilde{\mathbf{w}}(k) \\ &- \beta\frac{\mathbf{G}\mathbf{x}(k)}{\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)+\delta}v(k)\end{aligned} \tag{14}$$

and the first order associate system

$$\begin{aligned}\tilde{\mathbf{z}}(k+1) = \tilde{\mathbf{z}}(k) &- \beta E\left\{\frac{\mathbf{G}\mathbf{x}(k)\mathbf{x}^T(k)}{\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)+\delta}\right\}\tilde{\mathbf{z}}(k) \\ &- \beta E\left\{\frac{\mathbf{G}\mathbf{x}(k)}{\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)+\delta}v(k)\right\}\end{aligned} \tag{15}$$

where $v(k)$ is the output noise and the corresponding output error is given by (2). Assuming that $v(k)$ is zero-mean noise independent of $\mathbf{x}(k)$, the second expectation in (15) is zero. Regarding the first expectation, we have

$$E\left\{\frac{\mathbf{G}\mathbf{x}(k)\mathbf{x}^T(k)}{\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)+\delta}\right\} = \mathbf{G}E\left\{\frac{\mathbf{x}(k)\mathbf{x}^T(k)}{\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)+\delta}\right\}. \tag{16}$$

Furthermore, we will assume that $\{x(k)\}_{k\in\mathcal{Z}}$ is a zero-mean, white and stationary sequence of Gaussian random variables. Let us consider

$$\frac{1}{L}\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k) = \frac{1}{L}\sum_{i=1}^{L} x_i^2(k)g_i \tag{17}$$

where $\mathbf{G} = \mathbf{diag}\{g_1, \cdots, g_L\}$ and $x_i(k) = [\mathbf{x}(k)]_i = x(k+1-i)$. Note that

$$\begin{aligned} E\left\{\frac{1}{L}\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)\right\} &= \frac{1}{L}\sum_{i=1}^{L} E\{x_i^2(k)\}g_i \\ &= \sigma^2\frac{1}{L}\sum_{i=1}^{L} g_i = \sigma^2. \end{aligned} \tag{18}$$

It is straightforward to find

$$\begin{aligned} \mathrm{VAR}\left\{\frac{1}{L}\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)\right\} &= \frac{1}{L^2}\mathrm{VAR}\{\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)\} \\ &= \frac{1}{L^2}2\sigma^4\sum_{i=1}^{L} g_i^2. \end{aligned} \tag{19}$$

Let us assume that

$$\begin{aligned} \frac{E\left\{\frac{1}{L}\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)\right\}}{\sqrt{\mathrm{VAR}\left\{\frac{1}{L}\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)\right\}}} &= \frac{\sigma^2}{\frac{1}{L}\sigma^2\sqrt{2\sum\limits_{i=1}^{L} g_i^2}} \\ &= \frac{L}{\sqrt{2\sum\limits_{i=1}^{L} g_i^2}} \gg 1 \end{aligned} \tag{20}$$

i.e.,

$$L \gg \sqrt{2\sum_{i=1}^{L} g_i^2}. \tag{21}$$

Now we can write

$$\begin{aligned} \mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k) &= L\frac{1}{L}\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k) \\ &\approx LE\left\{\frac{1}{L}\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)\right\} = L\sigma^2 \end{aligned} \tag{22}$$

and

$$\begin{aligned} E\left\{\frac{\mathbf{G}\mathbf{x}(k)\mathbf{x}^T(k)}{\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)+\delta}\right\} &\approx \mathbf{G}E\left\{\frac{\mathbf{x}(k)\mathbf{x}^T(k)}{L\sigma^2+\delta}\right\} \\ &= \frac{1}{L\sigma^2+\delta}\mathbf{G}\mathbf{R}. \end{aligned} \tag{23}$$

The associate first-order averaged system is

$$\tilde{\mathbf{z}}(k+1) = \tilde{\mathbf{z}}(k) - \frac{\beta}{L\sigma^2+\delta}\mathbf{G}\mathbf{R}\tilde{\mathbf{z}}(k). \tag{24}$$

This is nothing but the steepest descent algorithm (3) with $\mathbf{H}(k+1) = \mathbf{G}$ and $\mu = \beta/(L\sigma^2+\delta)$. The closeness between the two systems can be also established by using the ODE method [11]. For the Duttweiler version of the algorithm (i.e., when normalization is done using $\mathbf{x}^T(k)\mathbf{x}(k)+\delta$ instead of $\mathbf{x}^T(k)\mathbf{G}\mathbf{x}(k)+\delta$) [1], it is not necessary to assume (22). In this case when the input $\{x(k)\}_{k\in\mathcal{Z}}$ is a sequence of zero-mean independent identically distributed random variables and $x(k)$ has symmetric probability density function, the associate system is

$$\tilde{\mathbf{z}}(k+1) = \tilde{\mathbf{z}}(k) - \frac{\beta\kappa}{\sigma^2}\mathbf{G}\sigma^2\mathbf{I}\tilde{\mathbf{z}}(k) \tag{25}$$

where

$$\kappa = E\left\{\frac{x^2(k+1-i)}{\mathbf{x}^T(k)\mathbf{x}(k)+\delta}\right\}, \quad i = 1, \cdots, L. \tag{26}$$

This is just (3) with $\mathbf{H}(k+1) = \mathbf{G}$, $\mathbf{R} = \sigma^2\mathbf{I}$ and $\mu = \beta\kappa/\sigma^2$.

By the established closeness, if the matrix $\mathbf{H}(k+1) = \mathbf{H} = \mathbf{F}(\mathbf{w}_0)$ is the optimal one for the associate steepest descent algorithm in some sense (e.g., it provides the highest convergence rate), among the proportionate-type NLMS adaptive algorithms, the one using this matrix, i.e. the one with $\mathbf{G}(k+1) = \mathbf{F}(\mathbf{w}_0)$, will be optimal as well. Of course, the proportionate-type NLMS adaptive algorithms do not know $\mathbf{w}_0$ and hence $\mathbf{F}(\mathbf{w}_0)$. But if $\mathbf{G}(k+1) = \mathbf{F}(\hat{\mathbf{w}}(k))$, it will converge to the optimal one. How this transition period affects convergence properties and whether the gain control $\mathbf{G}(k+1) = \mathbf{F}(\hat{\mathbf{w}}(k))$ is optimal, are open questions. An analytical approach to analyze this is still missing. Simulations presented in [7] [8] support the choice $\mathbf{G}(k+1) = \mathbf{F}(\hat{\mathbf{w}}(k))$.

4. CONCLUSIONS

The gain control matrix $\mathbf{G}(k+1)$ in the proportionate-type NLMS algorithms is diagonal positive-definite bounded matrix. Assuming sufficiently small adaptation stepsize parameter β and the persistently exciting input, the proportionate-type NLMS algorithms guarantee convergence of the adaptive filter weights to the optimal values [9, Theorem 1]. The convergence is strict in the noiseless case, while in the noisy case small fluctuations around the optimal values are present. Since the matrix $\mathbf{G}(k+1)$ in the proportionate-type NLMS algorithms depends only on the adaptive filter weights, the convergence of weights brings the convergence of the matrix $\mathbf{G}(k+1)$ as well.

When the gain control matrix $\mathbf{H}(k+1)$ of the proportionate-type steepest descent algorithm and the gain control matrix $\mathbf{G}(k+1)$ of the proportionate-type NLMS algorithm are equal to each other and time-invariant (i.e., $\mathbf{H}(k+1) = \mathbf{G}(k+1) = \mathbf{G}$), for sufficiently small adaptation stepsize parameter β, the zero-mean white stationary Gaussian input $x(k)$ and $L \gg \sqrt{2\sum_{i=1}^{L} g_i^2}$ (where L is the adaptive filter order and g_i, $i = 1, \cdots, L$ are the entries of the matrix $\mathbf{G}$), the weight trajectories obtained by the proportionate-type steepest descent algorithm and the weight trajectories obtained by the proportionate-type NLMS algorithm can be very close to each other [9, Theorems 9.1, 9.3, 9.5] [11, Theorem 1]. The above condition involving L and g_i, $i = 1, \cdots, L$ provides that the normalizing factor in the proportionate-type NLMS algorithm can be approximated as a constant.

If the proportionate-type steepest descent algorithm has some special property for a specific value of time-invariant $\mathbf{H}(k+1) = \mathbf{G}$, it will be expected that the proportionate-type NLMS algorithm will have the similar property if its matrix $G(k+1)$ is close to the specific value $\mathbf{G}$. When the specific value depends on the optimal weight values, the matrix $\mathbf{G}(k+1)$ of the realizable proportionate-type NLMS algorithm can converge to the value.

5. REFERENCES

[1] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancellers," *IEEE Trans. Speech, Audio Processing*, vol. 8, pp. 508–518, Sept. 2000.

[2] S. L. Gay, "An efficient, fast converging adaptive filter for network echo cancelletion," in *Proc. Asilomar Conf. Signals, Syst., Comput., vol. 1*, Pacific Grove, CA, Nov. 1998, pp. 394–398.

[3] M. Nekuii and M. Atarodi, "A fast converging algorithm for network echo cancellation," *IEEE Signal Processing Letters*, vol. 11, pp. 427–430, Apr. 2004.

[4] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, vol. 2*, Orlando, FL, May 2002, pp. 1881–1884.

[5] S. Gay and S. Douglas, "Normalized natural gradient adaptive filtering for sparse and non-sparse systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, vol. 2*, Orlando, FL, May 2002, pp. 1405–1408.

[6] Z. Chen, S. Gay, and S. Haykin, "Proportionate adaptation: New paradigms in adaptive filters," in *Least-Mean-Square Adaptive Filters*, S. Haykin and B. Widrow, Eds. Wiley-Interscience, 2003.

[7] H. Deng and M. Doroslovački, "Improving convergence of the PNLMS algorithm for sparse impulse response identification," *IEEE Signal Processing Letters*, vol. 12, pp. 181–184, March 2005.

[8] H. Deng and M. Doroslovački, "Proportionate adaptive algorithms for network echo cancellation," accepted for publication in IEEE Trans. Signal Processing, 2005.

[9] W. A. Sethares, B. D. O. Anderson, and C. R. Johnson, Jr., "Adaptive algorithms with filtered regressor and filtered error," *Math. Contr., Signals, Syst.*, vol. 2, pp. 381–403, 1989.

[10] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*, Prentice Hall, Englewood Cliffs, NJ, 1995.

[11] A. Benveniste, M. Goursat, and G. Ruget, "Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms with applications to data communications," *IEEE Trans. Automatic Control*, vol. 25, no. 6, pp. 1042–1058, Dec. 1980.