Exponential Convergence of Adaptive Algorithms

George V. Moustakides †
Department of Computer Engineering & Informatics
University of Patras, 26 500 Patras, GREECE
Email: moustaki@cti.gr

Abstract — **We introduce a novel method for analyzing a well known class of adaptive algorithms. By combining recent developments from the theory of Markov processes and long existing results from the theory of Perturbations of Linear Operators we study first the behavior and convergence properties of a class of products of random matrices. This in turn allows for the analysis of the first and second order statistics of the adaptive algorithms yielding estimates for the exponential rate of convergence and the covariance matrix of the estimation error.**

I. Introduction

Consider the following regression model

$$y_n = X_n^t W_n + w_n \tag{1}$$

where X_n is the input signal, y_n the desired output, w_n the additive noise and W_n an unknown time varying vector that we like to estimate. Both X_n, W_n are vectors of length N. A very rich class of algorithms for estimating W_n can be defined through the recursion

$$\hat{W}_{n+1} = \hat{W}_n + \mu(y_n - X_n^t \hat{W}_n) Z_n \tag{2}$$

where $\hat{W}_n$ is the estimate of W_n at time n, Z_n a vector of length N known as *regression vector* and μ a positive scalar known as *step size*. We are interested in examining the behavior of the estimation error $\Delta_n = W_n - \hat{W}_n$ for the practically important case $\mu \ll 1$.

II. Main Results

Let us first consider a different problem that will provide us with the necessary mathematical tools. Let $\{\xi_n\}$ denote a stationary Markov process evolving on some general state space, we are then interested in studying the behavior of the following product of random matrices

$$\mathbf{U}_n(\mu) = \prod_{i=1}^{n} \left[\mathbf{I}_K + \sum_{l=1}^{r} \mu^l \mathbf{T}_l(\xi_i)\right] \tag{3}$$

with $\mathbf{T}_l(\xi)$, $l = 1, \ldots, r$, matrix functions of ξ of dimensions $K \times K$, $\mathbf{I}_K$ the identity matrix of the same dimensions and $\mu > 0$ a scalar variable that is to be assumed "small" (i.e. $\mu \ll 1$) and corresponding to the step size μ of (2).

If $\mathbb{E}\{\cdot\}$ denotes expectation under the steady state measure and $\mathbb{E}_\xi\{\cdot\}$ conditional expectation given that the initial state $\xi_0 = \xi$, then for a vector valued function $G(\xi)$ we can write

$$\mathbb{E}_\xi\{\mathbf{U}_n(\mu) G(\xi_n)\} = \mathcal{T}(\mu)^n G(\xi) \tag{4}$$

where $\mathcal{T}(\mu)$ is a linear operator defined as

$$\mathcal{T}(\mu) G(\xi) = \mathbb{E}_\xi\left\{\left[\mathbf{I}_K + \sum_{l=1}^{r} \mu^l \mathbf{T}_l(\xi_1)\right] G(\xi_1)\right\} \tag{5}$$

We thus conclude that to study the behavior of the expectation of $\mathbf{U}_n(\mu)$ it suffices to study the behavior of the n-th power of the operator $\mathcal{T}(\mu)$.

Since $\mu \ll 1$ we can regard $\mathcal{T}(\mu)$ as a perturbed version of the operator $\mathcal{T}(0) = \mathbb{E}_\xi\{\mathbf{I}_K \cdot\}$. With this observation in mind, and combining results from [1] and [2], we can show the following theorems.

Theorem 1: Under suitable assumptions we have

$$\lim_{n\to\infty} \sqrt[n]{\|\mathcal{T}(\mu)^n\|_V} = 1 + \mu \max_i\{\mathrm{Re}(\lambda_i)\} + o(\mu) \tag{6}$$

where λ_i are the eigenvalues of the matrix $\mathbf{T} = \mathbb{E}\{\mathbf{T}_1(\xi_1)\}$.

Theorem 2: Let the matrix $\mathbf{T}$ have eigenvalues with strictly negative real parts, define the matrix $\mathbf{F}(\mu) = \mathbf{I}_K + \mu\mathbf{T}$ then, there exist constants $\alpha > 0$ and $0 \leq \rho < 1$ such that

$$\mathcal{T}(\mu)^n = \mathbf{F}(\mu)^n \mathcal{P} + (1 - \mu\alpha)^n o(1) + \rho^n O(1) \tag{7}$$

$$\mathcal{P}\mathcal{T}(\mu)^n = \mathbf{F}(\mu)^n \mathcal{P} + (1 - \mu\alpha)^n \mathcal{P} o(1) + \mu\rho^n \mathcal{P} O(1) \tag{8}$$

where the operator $\mathcal{P}$ is defined as $\mathcal{P} = \mathbb{E}\{\mathbf{I}_K \cdot\}$.

III. Application to Adaptive Algorithms

Assume $X_n = X(\xi_n), Z_n = Z(\xi_n)$ with $X(\xi), Z(\xi)$ nonlinear vector transformations and $\{\xi_n\}$ a stationary Markov process. Assume that the noise sequence $\{w_n\}$ is white with variance σ_w^2, that $W_{n+1} = W_n + \gamma U_{n+1}$, that $\{U_n\}$ is white independent of $\{\xi_n\}$ and $\{w_n\}$. We can then prove for the adaptive algorithms in (2), based on the results we have developed in the previous section, the next two theorems.

Theorem 3: The estimation error Δ_n converges to zero, in the mean, at an exponential rate equal to $\mu \min_i\{\mathrm{Re}(\lambda_i)\} + o(\mu)$ where λ_i are the eigenvalues of the matrix $\mathbf{A} = \mathbb{E}\{X_1 Z_1^t\}$.

Theorem 4: Let the matrix $\mathbf{A}$ have eigenvalues with strictly positive real parts, define the sequence of matrices $\{\mathbf{\Pi}_n\}$

$$\mathbf{\Pi}_{n+1} = (\mathbf{I}_N - \mu\mathbf{A})\mathbf{\Pi}_n(\mathbf{I}_N - \mu\mathbf{A})^t + \mu^2\sigma_w^2 \mathbf{Q}_Z + \gamma^2 \mathbf{Q}_U \tag{9}$$

where $\mathbf{Q}_Z = \mathbb{E}\{Z_1 Z_1^t\}$, $\mathbf{Q}_Z = \mathbb{E}\{U_1 U_1^t\}$ and $\mathbf{\Pi}_0 = \Delta_0 \Delta_0^t$, then

$$\|\mathbb{E}\{\Delta_n \Delta_n^t\} - \mathbf{\Pi}_n\| = o(1)\mathbf{\Pi}_n \tag{10}$$

IV. Conclusion

Using a novel methodology we arrive in obtaining, for adaptive algorithms, estimates of their exponential convergence rate and their second order statistics without the need of unrealistic or restrictive conditions, as Independence Assumption or essential boundedness of the data.

References

[1] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, 1966.

[2] S.P. Meyn and R.L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, 1993.

†This work was supported by the Greek General Secretariat for Research and Technology under Grant: ΠΕΝΕΔ96:1584. Extended version of the paper can be downloaded using the URL *http://xanthippi.cti.gr/people/moustakides/papers/isit98.ps*