Efficacy of Adaptive Signal Processing Algorithms

George V. Moustakides

Digital Signal & Image Processing Lab., Dept. of Computer Eng., Univ. of Patras, Patras 26500, GREECE

Abstract — **A method for theoretically comparing a class of Adaptive Estimation Algorithms is presented. The method consists in defining for each algorithm in the class a quantity (the efficacy) that is theoretically computable, that depends only on the statistics of the data. The larger the efficacy the better the algorithm performs.**

I. Introduction

In the last years many works in the area of adaptive signal processing algorithms are concentrated in defining methods that can compare in a fair way any two algorithms. The class that is usually considered is all algorithms that can be put under the following recursive formula

$$\theta_{n+1} = \theta_n - \mu H(\theta_n, X_n) \tag{1}$$

where θ_n is the state of the algorithm and X_n the data vector used for time n. This class contains most well known algorithms used in practice. For its analysis there exist basically two directions, the deterministic and the stochastic. Here we will limit ourselves in the second direction, that is, in the stochastic approach. Under this type of analysis, the problem of comparison has led to the need of theoretically describing the transient and steady state behavior of an algorithm. Specifically two algorithms can be compared in a fair way when we compare their transient behavior (speed of convergence) while they are forced to have the same steady state behavior [3].

Works up to now were able to theoretically describe the steady state part of an algorithm using Central Limit Theorem type approaches, while for the transient part they continue to use simulations [1], [2]. Notice also that for the steady state description there was the need to pass from the discrete to the continuous time and properly define an Ornstein-Uhlenbeck process.

II. Main Result

In this work we concentrate on a smaller algorithmic class. Specifically we consider algorithms that can be put under the form

$$\theta_{n+1} = (I - \mu A(X_n))\theta_n + \mu B(X_n) \tag{2}$$

which corresponds to a linear, with respect to θ, general model. This limited class contains many well known and widely used algorithms as RLS, LMS, Leakage LMS, and Signed Regressor LMS. For algorithms satisfying (2) it is possible to theoretically describe their performance for both the transient and the steady state part. Also this can be achieved without passing to the continuous time. The key result is the following theorem.

Theorem 1. Let the data vector X_n be a stationary Markov process. Let also $A = E\{A(X_n)\}$, $B = E\{B(X_n)\}$. Define the following two deterministic systems

$$\bar{\theta}_{n+1} = (I - \mu A)\bar{\theta}_n + \mu B \tag{3}$$

$$\hat{\theta}_{n+1} = (I - \mu A)\hat{\theta}_n - \mu(A(X_n) - A)\bar{\theta}_n + \mu B(X_n) \tag{4}$$

If A is diagonalizable and has eigenvalues on the right complex half plane, then under suitable conditions on the process X_n, for small enough μ, we have UNIFORMLY IN TIME that

$$E\{\|\theta_n - \hat{\theta}_n\|^2\} = O(\mu^2) \tag{5}$$

The practical meaning of Theorem 1 is that the two processes θ_n and $\hat{\theta}_n$ are pointwise close to each other (not only their distributions as is the case with existing results). Thus if we are interested in up to 2-nd order statistics for θ_n, we can approximate them by the corresponding statistics of $\hat{\theta}_n$ with an error $o(\mu)$.

III. Efficacy of an Algorithm

Usually θ_n is considered as an estimate to the Wiener solution $\bar{\theta}_\infty = A^{-1}B$. Thus the steady state performance of the algorithm is defined as the asymptotic error covariance matrix. This definition is not suitable if the data are not i.i.d. [2]. For many signal processing algorithms naturally enters the error $e_n = (\theta_n - \bar{\theta}_\infty)^t X_n$ (for example in cases where a prediction error is used). We can thus use as performance measure its variance (also known as "excess mean square error" for many algorithms). The following theorem gives an approximation to the asymptotic variance σ_e^2.

Theorem 2. Under the assumptions of Theorem 1 the excess mean square error satisfies

$$\sigma_e^2 = \mu\, \mathrm{trace}\{C_X P_X\} + o(\mu) \tag{6}$$

where $C_X = E\{X_n X_n^t\}$, $AP_X + P_X A^t = \sum_{j=-\infty}^{\infty} C_R(j)$, $C_R(j) = E\{R_{n+j} R_n^t\}$ and $R_n = (A(X_n) - A)\bar{\theta}_\infty - (B(x_n) - B)$. Since the statistics of X_n are known we can theoretically obtain the variance σ_e^2. Notice that by specifying a value for the variance actually specifies the step size μ.

In order to theoretically describe the transient behavior of θ_n, notice that this behavior can be approximated by the one of $\bar{\theta}_n$. Since from (4) all transient phenomena decrease to zero as $(I - \mu A)^n$ as measure for speed of convergence can be used the settling time $t_s = \mathrm{trace}\{(I - (I - \mu A))^{-1}\} = \mu^{-1}\mathrm{trace}\{A^{-1}\}$. To compare any two algorithms we set a common σ_e^2 and just compare the corresponding t_s. Since σ_e^2 is the same, we can instead compare the product $t_s \sigma_s^2$ which is independent of μ. This quantity is exactly what we call *efficacy* of the algorithm. Clearly the larger it is, the faster the algorithm converges.

References

[1] A. Benveniste et.al. "Adaptive Algorithms and Stochastic Approximation", New York: Sprienger Verlag, 1990.

[2] J.A. Bucklew et. al. "Weak Convergence and Local Stability Properties of Fixed Step Size Recursive Algorithms", *IEEE Trans. Inform. Theory*, Vol. IT-39, pp.966-978, may 1993.

[3] W.A. Sethares, "The Least Mean Square Family", in *Adaptive System Identification and Signal Processing*, Editors: N. Kalouptsidis, S. Theodoridis, New York: Prentice Hall, 1993.