



On Robust AML Identification Algorithms*

VOJISLAV Ž. FILIPOVIĆ† and BRANKO D. KOVAČEVIĆ‡

Key Words—Identification; recursive algorithms; robustness; convergence.

Abstract—Strong consistency results for a class of nonlinear approximate maximum likelihood algorithms for robust system identification are developed, where the system is assumed to be of the ARMAX form. The analysis uses the Martingale results, and strong consistency is shown to hold under a new assumption, representing a generalization of the strictly positive-real condition. Arguments are also given for using Huber's nonlinearity, in order to reduce the influence of outliers in practice.

1. Introduction

ESTIMATION ALGORITHMS based on the Gaussian model have been found to be especially inefficient when the real distribution belongs to the heavy-tailed variety, occasionally giving rise to very large outliers (Barnet and Levis, 1978). Considerable efforts have been oriented towards the design of robust estimation algorithms possessing a low sensitivity to distribution changes, usually valid locally within a pre-specified distribution class (Ershov, 1978; Hogg, 1979). The fundamental contribution to the field of robust estimation has been given by Huber (1964), who introduced the concept of min-max robust estimation. Further developments of this idea and applications to different types of problems have led to many valuable achievements (Masreliez and Martin, 1977; Poljak and Tsympkin, 1979; Tsympkin, 1984). However, the reported results in the area of robust dynamic system identification are few. An analysis of robustified Gauss-Newton type recursive prediction error algorithms using the ODE approach is given by Ljung and Söderström (1983). Strong consistency results for a class of robustified stochastic approximation (SA) algorithms using the Martingale theory are developed by Filipović and Kovačević (1987). This analysis suffers from a restriction to white and uniformly bounded system disturbances. An application of robustified SA algorithms for identification of linear dynamic systems with correlated disturbances, and a local convergence analysis based on the ODE approach, is discussed in Kovačević and Filipović (1988).

In this paper strong consistency results for a class of robustified approximate maximum likelihood (AML) algorithms of pseudolinear regression (PLR) type are developed, where the system is assumed to be of the ARMAX form. The analysis uses the Martingale results, and strong consistency is shown to hold under a new condition representing a generalization of the strictly positive-real (SPR) condition. Arguments are also given for using Huber's min-max optimal cost function, in order to eliminate the effects of outliers in practice.

* Received 9 August 1991; revised 10 April 1992; revised 29 October 1992; revised 24 May 1993; received in final form 30 November 1993. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor I. Mareels under the direction of Editor P. C. Parks. Corresponding author Professor Branko Kovačević. Tel. +381 11 324 8464; Fax +381 11 324 8681.

† Cellulose Industry 'Viskoza', 15300 Loznica, Serbia.

‡ Faculty of Electrical Engineering, University of Belgrade, Bulevar Revolucije 73, Belgrade, Serbia.

2. Problem formulation

Let the system under consideration be described by a linear single-input/single-output ARMAX model

$$A(q^{-1})y(i) = B(q^{-1})u(i) + C(q^{-1})e(i), \quad (1)$$

where q^{-1} is the backward shift operator: $q^{-1}y(i) = y(i-1)$, and

$$A(q^{-1}) = 1 + \sum_{k=1}^n a_k q^{-k}; \quad B(q^{-1}) = \sum_{k=1}^m b_k q^{-k}; \quad (2)$$

$$C(q^{-1}) = 1 + \sum_{k=1}^{\ell} c_k q^{-k},$$

are characteristic, control and disturbance polynomials, respectively. Here $y(i) \in \mathbb{R}^1$, $u(i) \in \mathbb{R}^1$ and $e(i) \in \mathbb{R}^1$ are the system output, the measurable input and the stochastic disturbance, or noise, respectively. It will be assumed that the sequence $\{e(i)\}$ is a zero-mean white stochastic process, generating an increasing sequence of sub-sigma algebras $\{F_i\}$, and that constants m, n and ℓ are known *a priori*. The problem of recursive identification of the system (1) can be considered as the task of estimating the vector $\theta^T = \{a_1, \dots, a_n, b_1, \dots, b_m, c_1, \dots, c_\ell\}$ in real-time, on the basis of current measurements. Formulation of the PLR or extended least squares (ELS) type identification schemes reduces to the choice of the performance criterion (e.g. Ljung and Söderström, 1983)

$$J_i(\hat{\theta}) = i^{-1} \sum_{k=1}^i F[\varepsilon(k, \hat{\theta})]; \quad \varepsilon(i, \hat{\theta}) = y(i) - Z^T(i)\hat{\theta}, \quad (3)$$

$$Z^T(i) = \{-y(i-1), \dots, -y(i-n), u(i-1), \dots, u(i-m),$$

$$v(i-1), \dots, v(i-\ell)\}, \quad (4)$$

$$v(i) = y(i) - Z^T(i)\hat{\theta}(i),$$

where $\varepsilon(\cdot)$ is the prediction error, $v(\cdot)$ is the residual and $\hat{\theta}(i)$ is the parameter vector estimate at time i . In the AML scheme the cost function is chosen as $F(\cdot) = -\log p(\cdot)$, where $p(\cdot)$ is the noise probability density function (pdf). Particularly for the Gaussian noise, $F(\cdot)$ is the quadratic function, and the resulting algorithm minimizing (3) is the standard linear AML. However, the ML method is very sensitive to deviations of the real noise pdf from the assumed one, and in the presence of outliers it ceases to work (Poljak and Tsympkin, 1979). Thus, one should modify the algorithm to make it more robust. A possible approach is to use the cost function $F(\cdot)$ that is quadratic for small arguments, but increases more slowly for large arguments. The recursive minimization of such a criterion can be done by using the approximate Newton-Raphson type method (Tsympkin, 1984)

$$\hat{\theta}(i) = \hat{\theta}(i-1) - [i \nabla_{\hat{\theta}}^2 J_i(\hat{\theta}(i-1))]^{-1} [i \nabla_{\hat{\theta}} J_i(\hat{\theta}(i-1))];$$

$$\hat{\theta}(0) = \hat{\theta}_0. \quad (5)$$

Moreover, with large i and by virtue of the approximate truth of the optimality conditions, yielding $\nabla J_{i-1}(\hat{\theta}) \approx 0$, one obtains

$$i \nabla_{\hat{\theta}} J_i(\hat{\theta}(i-1)) = -Z(i)\psi(\varepsilon(i, \hat{\theta}(i-1)));$$

$$i \nabla_{\hat{\theta}}^2 J_i(\hat{\theta}(i-1)) = \alpha \sum_{k=1}^i Z(k)Z^T(k), \quad (6)$$

