Statistical Methodology 28 (2016) 18–36

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

Consistency of M-estimators of nonlinear signal processing models

Kaushik Mahata [a], Amit Mitra [b,*], Sharmishtha Mitra [b]

[a] School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW-2308, Australia

[b] Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur-208016, India

ARTICLE INFO

Article history:
Received 18 November 2013
Received in revised form
20 March 2015
Accepted 14 July 2015
Available online 11 August 2015

Keywords:
Asymptotic analysis
M-estimator
Nonlinear signal processing model
Robust statistics
Strong consistency

ABSTRACT

In this paper, we consider the problem of robust M-estimation of parameters of nonlinear signal processing models. We investigate the conditions under which estimators are strongly consistent for convex and non-convex penalty functions and a wide class of noise scenarios, contaminating the actual transmitted signal. It is shown that the M-estimators of a general nonlinear signal model are asymptotically consistent with probability one under different sets of sufficient conditions on loss function and noise distribution. Simulations are performed for nonlinear superimposed sinusoidal model to observe the small sample performance of the M-estimators for various heavy tailed error distributions, outlier contamination levels and sample sizes.

1. Introduction

Estimation of parameters of nonlinear regression models is a fundamental problem in signal processing and time series analysis. In several real life applications, the data signals dealt with can be modeled as:

$$y_t = f_t(\theta) + e_t, \quad t = 1, \ldots, n, \tag{1}$$

where, $f_t(\theta)$ is the noise free nonlinear signal characterized by an unknown parameter vector θ and e_t is a real valued additive noise. The most common technique for estimating θ in (1) is the nonlinear least

* Corresponding author.
E-mail address: amitra@iitk.ac.in (A. Mitra).

http://dx.doi.org/10.1016/j.stamet.2015.07.004

squares (NLS). $\hat{\theta}_n^{NLSE}$, the nonlinear least squares estimator of θ_0, the true parameter vector, is given by

$$\hat{\theta}_n^{NLSE} = \arg\min_{\theta} \sum_{t=1}^{n} |y_t - f_t(\theta)|^2 .$$

Asymptotic (as $n \to \infty$) statistical properties of $\hat{\theta}_n^{NLSE}$ are studied in [9,13,22,23]. It is however well known that $\hat{\theta}_n^{NLSE}$ is sensitive to the presence of heavy tailed noise, outliers in the data and other departures from the underlying distributional assumptions. M-estimators are the most widely used remedy to this problem. Reference may be made to [3,4,2,6–8,25] among others. For signal processing applications, see for example [10,19,20,24]. The M-estimator of the unknown parameter vector is obtained by minimizing

$$l_n(\theta) = \sum_{t=1}^{n} \rho\left(y_t - f_t(\theta)\right). \tag{2}$$

The M-estimator of the signal parameters is thus given by

$$\hat{\theta}_n^{M} = \arg\min_{\theta \in S} l_n(\theta). \tag{3}$$

The set S in (3) is characterized in Assumption 3. Here, $\rho(.)$ is a suitably chosen non-negative penalty function. One popular choice for ρ is the Huber's function:

$$\rho_h(z) = \begin{cases} z^2/2, & \text{if } |z| \le c, \\ c|z| - c^2/2, & \text{if } |z| > c. \end{cases} \tag{4}$$

Taking $c \to \infty$, (4) gives $\hat{\theta}_n^{M} = \hat{\theta}_n^{NLSE}$ and $\hat{\theta}_n^{M}$ becomes the least absolute deviation estimator for $c = 0$. In addition, $\rho_h(z)$ is a convex function of z for any c. However, $\rho(.)$ in (2) need not be convex in general.

In this paper, we prove the strong consistency of $\hat{\theta}_n^{M}$ under very general assumptions on the distribution of noise which contaminates the actual transmitted signal and also on $\rho(.)$.

In particular, we address the following questions:

- Assuming $n \to \infty$, under what condition the true parameter θ_0 is the unique global minimum of limiting $l_n(\theta)$?
- Is there any special advantage of considering $\rho(.)$ to be convex?
- What are the additional requirements for strong consistency if $\rho(.)$ is non-convex?

We give very general answers to the above questions in form of some sufficient conditions. In addition, we provide simple examples and simulation studies to illustrate our results. Compared to the previous attempts [3,5,15], our proof does not assume θ_0 to be the unique global minimum point of $E(l_n(\theta)/n)$, which may not be true.

The rest of the paper is organized as follows. In Section 2, we present the main consistency results of the paper. Section 3 presents illustrative examples and simulation studies. Finally, we conclude in Section 4.

2. Consistency results

Let us define

$$\bar{l}_n(\theta) = E(\ell_n(\theta)/n) = \frac{1}{n}\sum_{t=1}^{n} E\left(\rho\left\{y_t - f_t(\theta)\right\}\right). \tag{5}$$

Our proof of consistency consists of two crucial steps. First, we show that $\frac{1}{n}(l_n(\theta) - E(l_n(\theta)))$ tends to zero with probability one (w.p.1), uniformly for all $\theta \in S$. Next, we show that θ_0 is the unique global minimum of $\bar{l}_n(\theta)$ for all n. Subsequently, we make use of a standard results [9,21,16] to conclude that $\hat{\theta}_n^{M} \to \theta_0$, with probability 1.