Systems & Control Letters 119 (2018) 46–51

Contents lists available at ScienceDirect

Systems & Control Letters

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

Some remarks on the bias distribution analysis of discrete-time identification algorithms based on pseudo-linear regressions

Bernard Vau *, Henri Bourlès

SATIE, Ecole normale supérieure de Paris-Saclay, 61 avenue du président Wilson, 94230 Cachan, France

ARTICLE INFO

Article history:
Received 2 October 2017
Received in revised form 14 June 2018
Accepted 3 July 2018
Available online 27 July 2018

Keywords:
Identification
Recursive identification
Closed-loop identification
Bias analysis

ABSTRACT

In 1998, A. Karimi and I.D. Landau published in this journal an article entitled "Comparison of the closed-loop identication methods in terms of bias distribution". One of its main purposes was to provide a bias distribution analysis in the frequency domain of closed-loop output error identication algorithms that had been recently developed. The expressions provided in that paper are only valid for prediction error identification methods (PEM), not for pseudo-linear regression (PLR) ones, for which we give the correct frequency domain bias analysis, both in open- and closed-loop. Although PLR was initially (and is still) considered as an approximation of PEM, we show that it gives better results at high frequencies.

1. Introduction

In the field of discrete-time identification, the bias distribution analysis over frequency domain is very powerful in order to evaluate the influence of input and noise spectra on an identified model, and to assess qualitatively the model that can be obtained if it has not the same structure as the identified system. This method was first introduced in [1] for open-loop identification; see [2] for more details, especially in the context of closed-loop operations. In these two references, this bias analysis has been developed in the perspective of prediction error methods (PEM), which aim at minimizing a one step further prediction error variance. In [3,4], Karimi and Landau used the same method to infer the bias distribution of closed-loop algorithms. In [3] (section 5) and [5] (section 6), it is claimed that this analysis is valid for CLOE, F-CLOE algorithms which are of PLR type. However, this is not true as already noticed in [4], p. 308, although PLR was initially (and is still) considered as an approximation of PEM (see, e.g., remark 8.4.2 of [6]); rather, PLR algorithms tend to cancel the correlation function between the prediction error and an observation vector, which in general is the regressor of the predictor, possibly filtered (see the correlation approach developed by Ljung in [2]). If the system dynamics is approximately known beforehand, the difference between PEM and PLR can be made quite small, as shown in Section 4, with an appropriate regressor filtering. But this approach presupposes that what one is seeking is already known, a vicious circle which prevents from addressing the core of the problem. It is said in [4] (p. 308) that the bias distribution cannot be computed for CLOE algorithm. Therefore, bias distribution of PLR algorithms, such as CLOE, is an open problem, which is solved here; this bias distribution is determined and we show that it is quite different from that of PEM. To do this we introduce in Section 3, the concept of equivalent prediction error – most of the time a non-measurable signal – whose variance is effectively minimized by the PLR algorithm, even if the identified system is not in the model set. This approach shows that compared to PEM, PLR schemes strongly penalize the model misfit at high frequency, in a way comparable to the classical open-loop least-squares algorithm, whatever the predictor model is (output error, ARMAX, etc.). In Section 5, an example is given, that relies on the Vinnicombe gap, in order to compare the model misfit of a PEM scheme and its corresponding PLR one: It brings to light the discrepancy between the two methods in case of a closed-loop output error (CLOE) identification structure.

2. Optimal prediction error and bias distribution analysis of PEM algorithms

At first, let us recall briefly the model structures used here, both in open- and closed-loop, and the manner to obtain the bias distribution from PEM algorithms. According to Landau et al. ([3], p. 44), we distinguish between equation error models and output error models. Equation error model in open-loop corresponds to the equation:

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})e(t)$$

with $G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})}$, and $W(q^{-1}) = \frac{C(q^{-1})}{A(q^{-1})}$, respectively, the deterministic and stochastic parts of the model; q^{-1} the shift

* Corresponding author.
E-mail address: bernard.vau@satie.ens-cachan.fr (B. Vau).

https://doi.org/10.1016/j.sysconle.2018.07.002
0167-6911/

backward operator, $A(q^{-1})$ is a monic polynomial, $W(q^{-1})$ is the ratio of two monic polynomials, and $u(t), y(t), e(t)$ are the input, the output, and a centered gaussian white noise, respectively.

According to the noise structure, we distinguish between the following cases:

- when $C(q^{-1}) = 1$, corresponding to the ARX model,
- when $C(q^{-1})$ is a monic polynomial of degree strictly greater than 0, corresponding to the ARMAX model.

Other equation error models exist (e.g. ARARMAX, etc.) but they are not treated in this paper. On the other hand, the output error model in open-loop is given by

$$y(t) = G(q^{-1})u(t) + v(t)$$

where $v(t)$ is a centered gaussian noise not necessarily white, but uncorrelated with the input.

Let us call $\widehat{G}(q^{-1})$ and $\widehat{W}(q^{-1})$ the estimations of $G(q^{-1})$ and $W(q^{-1})$, respectively. In the case of an open-loop equation error model, the prediction error is given by ([4], (3.3)), ([4], (9.62):

$$\begin{aligned}\varepsilon(t) = {} & \widehat{W}(q^{-1})\{(G(q^{-1}) - \widehat{G}(q^{-1}))u(t) \\ & + (W(q^{-1}) - \widehat{W}(q^{-1}))e(t)\} + e(t)\end{aligned} \tag{1}$$

whereas the optimal error for the open-loop output error model is simply:

$$\varepsilon(t) = (G(q^{-1}) - \widehat{G}(q^{-1}))u(t) + v(t)$$

The closed-loop case is more complicated, due to the feedback control law. We assume that the controller has an R-S structure, i.e. $S(q^{-1})u(t) = -R(q^{-1})y(t)$; let us define the direct sensitivity function (transfer function from the output noise to the output):

$$S_{yp}(q^{-1}) = \frac{A(q^{-1})S(q^{-1})}{A(q^{-1})S(q^{-1}) + B(q^{-1})R(q^{-1})}$$

In the context of an equation error model, in which the model input is given by $S(q^{-1})\widehat{u}(t) = -R(q^{-1})\widehat{y}(t)$, the optimal predicted output is

$$\widehat{y}(t) = \widehat{G}\widehat{u}(t) + \widehat{\overline{W}}(q^{-1})\varepsilon(t)$$

where $\widehat{\overline{W}}(q^{-1}) = \widehat{W}(q^{-1}) - \widehat{S}_{yp}^{-1}(q^{-1})$ (see [3], eq. (5.7) sq.), thus we have:

$$\varepsilon(t) = \widehat{W}^{-1}(q^{-1})\{(G - \widehat{G})S_{yp}r_u(t) + (WS_{yp}\widehat{S}_{yp}^{-1} - \widehat{W})e(t)\} + e(t)$$

This expression is directly obtained from ([3], (5.12)).

The optimal predicted output is given by $\widehat{y}(t) = \widehat{G}\widehat{u}(t)$, and the corresponding optimal prediction error of output error model by

$$\varepsilon(t) = \widehat{S}_{yp}(G - \widehat{G})S_{yp}r_u(t) + S_{yp}v(t)$$

where $\widehat{S}_{yp} = \frac{\widehat{A}S}{\widehat{A}S + \widehat{B}R}$, $\widehat{A}, \widehat{B}$ being the estimations of A and B, respectively (the dependence in q^{-1} is omitted).

The purpose of PEM algorithms is to minimize the prediction error variance $\mathbf{E}[\varepsilon^2(t)]$, and whatever the algorithm structure is, both in open or closed-loop, one obtains the optimal estimated parameter vector $\widehat{\theta}^*_{PEM}$:

$$\widehat{\theta}^*_{PEM} = Arg\min \mathbf{E}[\varepsilon^2(t)] = Arg\min \int_{-\pi}^{+\pi} \left|\mathcal{Z}\{\varepsilon\}(e^{i\omega})\right|^2 d\omega \tag{2}$$

where $\mathcal{Z}$ is the z-transform. This expression is at the origin of bias analysis for PEM algorithms.

3. Bias distribution of pseudo-linear regression algorithms

3.1. Equivalent prediction error

The *a posteriori* model predicted output $\widehat{y}(t+1)$ is provided by $\widehat{y}(t+1) = \widehat{\theta}(t+1)\phi(t, \widehat{\theta})$, where $\phi(t, \widehat{\theta})$ is the regressor structure, generally depending on $\widehat{\theta}$. The *a posteriori* prediction error is given by the expression: $\varepsilon(t+1) = y(t+1) - \widehat{y}(t+1, \widehat{\theta})$. Most of the PLR identification procedures are solved recursively with the so-called parameter adaptation algorithm (PAA):

$$\widehat{\theta}(t+1) = \widehat{\theta}(t) + F(t)\phi(t)\varepsilon(t+1) \tag{3a}$$

$$F^{-1}(t+1) = \lambda_1 F^{-1}(t) + \lambda_2 \phi(t)\phi^T(t) \tag{3b}$$

$F(t)$ is the adaptation gain (positive definite matrix), $0 < \lambda_1 \leq 1, 0 \leq \lambda_2 < 2$ are forgetting factors. The stationary condition of the PAA is (see [2], p. 224):

$$\mathbf{E}[\varepsilon(t+1)\phi(t)] = 0 \tag{4}$$

Lemma 1. *In general, the stationarity condition of the parameter adaptation algorithm* $\mathbf{E}[\varepsilon(t+1)\phi(t, \widehat{\theta})] = 0$, *is not the one associated with the prediction error variance* $\mathbf{E}[\varepsilon^2(t)]$ *minimization.*

Proof. As a counterexample let us consider the extended least squares algorithm (corresponding to an ARMAX model), for which the predicted output is $\widehat{y}(t+1) = \widehat{\theta}^T\phi(t)$. Let $\varepsilon(t) = y(t) - \widehat{y}(t)$, $\widehat{\theta}^T = [\widehat{a}_1, \ \ldots \ \widehat{b}_1, \ \ldots \ \widehat{c}_1, \ \ldots]$ and $\phi(t)^T = [-y(t), \ \ldots \ u(t), \ \cdots \ \varepsilon(t), \ \ldots]$. Let $\widehat{C}(q^{-1}) = 1 + \widehat{c}_1 q^{-1} + \widehat{c}_2 q^{-2} \cdots$ and assume that $\widehat{C}(q^{-1}) \neq 1$. One has that $\phi(t, \widehat{\theta}) = -\widehat{C}(q^{-1}, \widehat{\theta})\frac{\partial\varepsilon(t+1)}{\partial\widehat{\theta}}$. In this case the stationary condition of the parameter adaptation algorithm is $\mathbf{E}[\varepsilon(t+1)\widehat{C}(q^{-1})\frac{\partial\varepsilon(t+1)}{\partial\widehat{\theta}}] = 0$. Therefore $\frac{\partial}{\partial\widehat{\theta}}\mathbf{E}\left[\varepsilon(t+1)^2\right] = 2\mathbf{E}\left[\varepsilon(t+1)\frac{\partial\varepsilon(t+1)}{\partial\widehat{\theta}}\right]$ cannot be zero unless $\{\varepsilon(t+1)\}$ is white, and that cannot happen if the system is not in the model set. □

Lemma 2. *In PLR schemes, the stationarity condition* $\mathbf{E}\left[\varepsilon\left(t+1\right)\phi\left(t, \hat{\theta}\right)\right] = 0$ *is the stationary condition of the variance minimization problem of the signal* $\varepsilon_E(t+1, \widehat{\theta})$, *called the "equivalent prediction error" in the sequel, and characterized by the following two conditions:*

(1) *If the system is in the model set and if the estimated parameter vector* $\widehat{\theta}^*_{PLR}$ *is equal to the true parameters vector* θ ($\theta = \widehat{\theta}^*_{PLR}$), *then one has:*

- *for the equation error model:*
  $\varepsilon_E(t+1, \widehat{\theta}^*_{PLR}) = \varepsilon_E(t+1, \theta) = \varepsilon(t+1, \theta) = e(t+1)$
- *for the open-loop output error model:*
  $\varepsilon_E(t+1, \widehat{\theta}^*_{PLR}) = \varepsilon_E(t+1, \theta) = \varepsilon(t+1, \theta) = v(t+1)$
- *for the closed-loop output error model:*
  $\varepsilon_E(t+1, \widehat{\theta}^*_{PLR}) = \varepsilon_E(t+1, \theta) = \varepsilon(t+1, \theta) = S_{yp}v(t+1)$

(2) *The vector* $\phi_E(t) = -\frac{\partial\varepsilon_E(t+1)}{\partial\widehat{\theta}}$, *called the "equivalent regressor", is not a function of* $\widehat{\theta}$, *i.e.* $\frac{\partial\phi_E(t)}{\partial\widehat{\theta}} = 0$.

Proof. By (2), $\frac{\partial\phi_E(t)}{\partial\widehat{\theta}} = 0$, so that $\varepsilon_E(t+1) = -\widehat{\theta}^T\phi_E(t) + k$ (k independent of $\widehat{\theta}$). By (1) we get:
$e(t+1) = -\theta^T\phi_E(t) + k$ for the equation error model,
$v(t+1) = -\theta^T\phi_E(t) + k$ for the open-loop output error model,
$S_{yp}v(t+1) = -\theta^T\phi_E(t) + k$ for the closed-loop output error model.

Combining the preceding equations, one gets the following prediction error expressions:
$\varepsilon_E(t+1) = (\theta - \widehat{\theta})^T\phi_E(t) + e(t+1)$ for the equation error model,
$\varepsilon_E(t+1) = (\theta - \widehat{\theta})^T\phi_E(t) + v(t+1)$ for the open-loop output error model,