# Output error identification for multi-input multi-output systems with bounded disturbances

Mathieu Pouliquen, Eric Pigeon, Olivier Gehan
Control Group, GREYC CNRS UMR 6072
ENSICAEN, 06 Bd du Marechal Juin
14050 Caen Cedex, France
mathieu.pouliquen@greyc.ensicaen.fr

*Abstract*— In this paper, we present the adaptation of an Optimal Bounding Ellipsoid (OBE) type algorithm for output error systems with unknown but bounded disturbances. Two identification algorithms are derived in a multi-input multi-output context and the paper proposed stability and convergence analysis. An iterative scheme is introduced and simulation are performed, showing the excellent performance of the proposed approach.

## I. INTRODUCTION

All practical identification algorithms have to deal with measurements corrupted by noise. In some cases (unknown probability distribution of the disturbances, modeling inaccuracy) the noise can't be modeled as a stochastic process, then the bounded noise assumption seems to be more appropriate. In this context, the use of classical identification schemes ([11]) is limited and in contrast Set Membership Identification (SMI) algorithms are interesting alternative approaches. Indeed these approaches are considered to be more appropriate to handle the identification problem in presence of bounded disturbances. Their principe consists in the estimation of a feasible set of parameters which must be consistent with the measurement data and the model structure. Important contributions have been presented in [13], [9], [14], [20], [10], [1] and [4]. Among the SMI type methods, OBE type algorithms represent a very popular class of recursive algorithms ([9], [5], [12], [3], [16], [6], [7], [19], [15], [18], [2]). These methods have been mainly motivated by a low computational complexity and robustness to measurement noise.

To the best of our knowledge, most of the existing works in the OBE algorithms literature aim at designing identification algorithm adapted to equation error models: stability and convergence properties are established in particular for affine-in-parameters model. Less attention has been focused on the output error systems with bounded disturbances ([8]) and our goal in this paper is to adapt an OBE algorithm to this type of systems. The main contribution of this note is the extension of some stability and convergence results ([3], [19], [15], [18], [2]) to output error systems context and the derivation of an algorithm which relaxes the stability and convergence conditions.

The paper is organized as follows: in section II the model structure is presented. In section III, some algorithms are proposed and their convergence analysis are addressed. A geometrical interpretation is also presented in this section. Some simulation results are given in section IV. Finally, section V concludes the paper.

## II. PROBLEM FORMULATION

Consider a discrete-time multi-input multi-output stable system of the form

$$\begin{cases} y_t = G(q)u_t + v_t \\ v_t^T \Delta_{v_t}^{-2} v_t \leq 1 \end{cases} \tag{1}$$

where $u_t \in \mathbb{R}^{n_u}$ and $y_t \in \mathbb{R}^{n_y}$ are respectively the system inputs and outputs vector. $v_t \in \mathbb{R}^{n_y}$ is an unknown and bounded disturbing term (noise measurement, state disturbances, modeling inaccuracy, etc.) and $\Delta_{v_t}$ is a symmetric positive definite known matrix which reflects a known upper bound on $v_t$.

The identification problem treated in this paper is stated as: estimate the parameters of system (1) described by

$$G(q) = (A(q))^{-1} B(q) \tag{2}$$

where $A(q)$ and $B(q)$ are unknown matrix polynomials of the form:

$$\begin{cases} A(q) = I_{n_y} + A_1 q^{-1} + \cdots + A_{n_a} q^{-n_a} \\ B(q) = B_0 + B_1 q^{-1} + \cdots + B_{n_b} q^{-n_b} \end{cases}$$

with $A_i \in \mathbb{R}^{n_y \times n_y}$ and diagonal, $B_i \in \mathbb{R}^{n_y \times n_u}$. Note that this description of the behavior of the system may not be a minimal realization.

This structure of the plant is called an output error model structure. The vector of coefficients of the plant is

$$\Theta^* = \begin{pmatrix} A_1 & \cdots & A_{n_a} & B_0 & \cdots & B_{n_b} \end{pmatrix}$$

and the one-step-ahead output predictor is defined bearing in mind the optimal predictor structure (see [11]), namely

$$\hat{y}_t = y_t - v_t = \Theta^* \varphi_t$$

with $\varphi_t^T = \begin{pmatrix} -\hat{y}_{t-1}^T & \cdots & -\hat{y}_{t-n_a}^T & u_t^T & \cdots & u_{t-n_b}^T \end{pmatrix}$. This can be rewritten as $\hat{y}_t = \phi_t^T \theta^*$ (see [11]) with $\begin{cases} \theta^* = vect(\Theta^*) \\ \phi_t = \varphi_t \otimes I_{n_y} \end{cases}$ where $vect(.)$ denotes the vectorization of a matrix, $\otimes$ the Kronecker product and $\theta^* \in \mathbb{R}^n$ is the unknown parameters vector to be identified with $n = (n_a n_y + (n_b + 1)n_u)n_y$ the number of parameters.

A key observation is that the output predictor $\hat{y}_t$ is not linear in the system parameters. This will have a strong impact on the form of the estimation algorithm.

## III. IDENTIFICATION ALGORITHMS

### A. The OE-OBE (Output Error - OBE) algorithm.

Our aim in this paper is to design an iterative identification algorithm for the system described by (1) and (2). Before introducing the parameter adaptation algorithm, let us define the well known a priori and a posteriori predictors as

$$\begin{cases} \hat{y}_{t/t-1}=\hat{\phi}_t^T\hat{\theta}_{t-1} \\ \hat{y}_{t/t}=\hat{\phi}_t^T\hat{\theta}_t \end{cases}$$

$\hat{\theta}_t$ represents the estimation of the parameters vector at the actual time $t$ while $\hat{\phi}_t$ is an estimate of the prediction data vector $\phi_t$ which is simply obtained by replacing the unknown components $\hat{y}(t-i)$ by their a posteriori estimates $\hat{y}(t-i/t-i)$: $\hat{\phi}_t = \hat{\phi}_t \otimes I_{n_y}$ with

$$\hat{\phi}_t^T=\begin{pmatrix} -\hat{y}_{t-1/t-1}^T & \cdots & -\hat{y}_{t-n_a/t-n_a}^T & u_t^T & \cdots & u_{t-n_b}^T \end{pmatrix}$$

Here again, the predictor $\hat{y}_{t/t}$ is not linear in $\hat{\theta}_t$.

The estimated parameters vector has to maintain the output error below a bound defined from the upper bound on the disturbance $v_t$. Generalization of some results performed in [3], [19], [18] and [2], the following Output Error - OBE algorithm provides such estimation:

$$\begin{cases} \hat{\theta}_t=\hat{\theta}_{t-1}+\Gamma_t\varepsilon_{t/t-1} \\ \Gamma_t=P_{t-1}\hat{\phi}_t\sigma_t\left(\lambda I_{n_y}+\hat{\phi}_t^T P_{t-1}\hat{\phi}_t\sigma_t\right)^{-1} \\ P_t=\frac{1}{\lambda}\left(I_n-\Gamma_t\hat{\phi}_t^T\right)P_{t-1} \\ \varepsilon_{t/t-1}=y_t-\hat{\phi}_t^T\hat{\theta}_{t-1} \end{cases} \quad (3)$$

where $0 < \lambda \leq 1$ is a design parameter forgetting factor that will be used to monitor the parameter adaptation dynamics. $\sigma_t$ is a switching flag given by:

$$\sigma_t=\begin{cases} \lambda\left(\hat{\phi}_t^T P_{t-1}\hat{\phi}_t\right)^{-1}\left((\varepsilon_{t/t-1}^T\Delta_t^{-2}\varepsilon_{t/t-1})^{1/2}-1\right) \\ \quad \text{if } \left(\varepsilon_{t/t-1}^T\Delta_t^{-2}\varepsilon_{t/t-1}>1\right) \text{ and } \left(\hat{\phi}_t^T P_{t-1}\hat{\phi}_t>0\right) \\ 0 \text{ otherwise} \end{cases} \quad (4)$$

$\Delta_t$ is a user defined symmetric positive definite matrix whose role will be specify below.

$\varepsilon_{t/t-1}$ is the a priori prediction error. From (3) the a posteriori prediction error $\varepsilon_{t/t} = y_t - \hat{\phi}_t^T\hat{\theta}_t$ can be written as:

$$\varepsilon_{t/t}=\lambda\left(\lambda I_{n_y}+\hat{\phi}_t^T P_{t-1}\hat{\phi}_t\sigma_t\right)^{-1}\varepsilon_{t/t-1} \quad (5)$$

Using the value of $\sigma_t$ for $\sigma_t \neq 0$ (5) yields:

$$\varepsilon_{t/t}^T\Delta_t^{-2}\varepsilon_{t/t}=1$$

This clearly shows that the OE-OBE algorithm ensures the following key property:

$$\sigma_t\neq 0 \implies \varepsilon_{t/t}^T\Delta_t^{-2}\varepsilon_{t/t}=1$$

$\Delta_t$ is then a bound on the a posteriori adaptation error which has to be specified taking into account the bound $\Delta_{v_t}$.

*Remark 1:* From the expression (3) it is apparent that the parameter adaptation is frozen when $\sigma_t = 0$. This occurs whenever the a priori prediction error is less than the threshold $\Delta_t$ (i.e. the algorithm doesn't have to update the parameters vector) or the observation vector is no longer persistently exciting (i.e. the algorithm doesn't have enough informations to realize an update).

Before providing a first result, let consider the following observation on the a posteriori prediction error $\varepsilon_{t/t} = y_t - \hat{y}_{t/t}$. Bearing in mind structures of $\phi_t^T$ and $\hat{\phi}_t^T$, $\varepsilon_{t/t}$ becomes

$$\varepsilon_{t/t}=(A(q))^{-1}\hat{\phi}_t^T\tilde{\theta}_t+v_t \quad (6)$$

with $\tilde{\theta}_t = \theta^* - \hat{\theta}_t$.

This equation has the typical form encountered in pseudo-linear regression. The stability analysis with an equation of the form $\varepsilon_{t/t} = \phi_t^T\tilde{\theta}_t + v_t$ has been already investigated in [3], [19], [18] and [2]. The following result presents a stability analysis for the form (6).

*Result 1:* Consider the class of systems defined in section II and the OE-OBE algorithm given by (3) and (4). Assume that
- $A(q)$ is such that

$$\left\|I_{n_y}-A(q)\right\|_1 < 1 \quad (7)$$

  where $\|.\|_1$ is the $l_1$ induced norm;
- $\Delta_t$ is such that:

$$\delta_{t/min}\geq\frac{\|A(q)\|_1}{1-\left\|I_{n_y}-A(q)\right\|_1}\delta_{v_t/min} \quad (8)$$

  where $\delta_{t/min}$ is the minimal singular value of $\Delta_t$ and $\delta_{v_t/min}$ the minimal singular value of $\Delta_{v_t}$.

  then for all initial conditions
- 

$$|\tilde{\theta}_t|^2\leq\gamma_1 |\tilde{\theta}_0|^2 \quad (9)$$

  with $\gamma_1 = \frac{\lambda_{max}\left(P_0^{-1}\right)}{\lambda_{min}\left(P_0^{-1}\right)}$.

where $\lambda_{max}\left(P_0^{-1}\right)$ and $\lambda_{min}\left(P_0^{-1}\right)$ are respectively the maximum and the minimum eigenvalues of $P_0^{-1}$.

If, furthermore, $\{\hat{\phi}_t\}$ is a persistently exciting sequence of order $o_e \geq n$, i.e there exist $\alpha > 0$ and $\beta > 0$ such that for all $t$

$$\alpha I_n \leq \sum_{i=0}^{o_e-1}\hat{\phi}_{t+i}\sigma_{t+i}\hat{\phi}_{t+i}^T \leq \beta I_n \quad (10)$$

Then the following properties hold:
- for all $t \geq o_e+1$

$$|\tilde{\theta}_t|^2 \leq \gamma_2 \lambda^t |\tilde{\theta}_0|^2 \quad (11)$$

  with $\gamma_2=\begin{cases} \frac{\lambda_{max}\left(P_0^{-1}\right)}{\alpha}\left(\frac{\lambda^{-o_e}-1}{\lambda^{-1}-1}\right) & \text{if } \lambda<1 \\ \frac{\lambda_{max}\left(P_0^{-1}\right)}{\alpha} & \text{if } \lambda=1 \end{cases}$ .
- For $\lambda < 1$ one has

$$\lim_{t\to\infty}\varepsilon_{t/t-1}^T\Delta_t^{-2}\varepsilon_{t/t-1} \leq 1 \quad (12)$$

■

*Proof:*
- Consider the following Lyapounov function:

$$V_t=\tilde{\theta}_t^T P_t^{-1}\tilde{\theta}_t$$

In a first time we are going to show that $V_t \leq \lambda V_{t-1}$ if conditions (7) and (8) hold. From (3) and (5) we have:

$$\tilde{\theta}_{t-1} = \tilde{\theta}_t + P_{t-1}\hat{\phi}_t \frac{\sigma_t}{\lambda}\varepsilon_{t/t}$$

then

$$V_{t-1} = \tilde{\theta}_t^T P_{t-1}^{-1}\tilde{\theta}_t + \frac{2}{\lambda}\tilde{\theta}_t^T \hat{\phi}_t \sigma_t \varepsilon_{t/t} + \frac{1}{\lambda^2}\varepsilon_{t/t}^T \sigma_t \hat{\phi}_t^T P_{t-1}\hat{\phi}_t \sigma_t \varepsilon_{t/t}$$

After a few lines of calculation we show that $V_t = \lambda V_{t-1} + q_t$ with

$$\begin{aligned} q_t &= \tilde{\theta}_t^T \hat{\phi}_t \sigma_t \hat{\phi}_t^T \tilde{\theta}_t - 2\tilde{\theta}_t^T \hat{\phi}_t \sigma_t \varepsilon_{t/t} \\ &\quad - \frac{1}{\lambda}\varepsilon_{t/t}^T \sigma_t \hat{\phi}_t^T P_{t-1}\hat{\phi}_t \sigma_t \varepsilon_{t/t} \end{aligned}$$

- If $\sigma_t = 0$, then $q_t = 0$ and

$$V_t = \lambda V_{t-1}$$

- Let consider the case where $\sigma_t \neq 0$ and let define $b_t = \hat{\phi}_t^T \tilde{\theta}_t$. After a few lines of calculation we show that $q_t \leq 0$ provided that:

$$\left(\varepsilon_{t/t} - b_t\right)^T \Delta_t^{-2}\left(\varepsilon_{t/t} - b_t\right) \leq 1 \tag{13}$$

$\varepsilon_{t/t}$ is such that $\varepsilon_{t/t} = (A(q))^{-1}b_t + v_t$, thus Condition (13) becomes:

$$\left(\left(I_{n_y} - A(q)\right)\varepsilon_{t/t} + A(q)v_t\right)^T \left(\left(I_{n_y} - A(q)\right)\varepsilon_{t/t} + A(q)v_t\right) \leq \delta_{t/min}^2$$

where $\delta_{t/min}$ is the minimal singular value of $\Delta_t$. From triangular inequality, one obtains the following condition:

$$\begin{aligned} &\left(\left(\left(I_{n_y} - A(q)\right)\varepsilon_{t/t}\right)^T \left(\left(I_{n_y} - A(q)\right)\varepsilon_{t/t}\right)\right)^{1/2} \\ &\quad + \left((A(q)v_t)^T(A(q)v_t)\right)^{1/2} \leq \delta_{t/min} \end{aligned} \tag{14}$$

Let define $\|.\|_1$ the $l_1$ induced norm and suppose condition (7) to be true. Equation (14) holds if:

$$\left\|I_{n_y} - A(q)\right\|_1 (\varepsilon_{t/t}^T \varepsilon_{t/t})^{1/2} + \|A(q)\|_1 (v_t^T v_t)^{1/2} \leq \delta_{t/min} \tag{15}$$

We know that $\varepsilon_{t/t}^T \Delta_t^{-2}\varepsilon_{t/t} \leq 1$ and $v_t^T \Delta_{v_t}^{-2}v_t \leq 1$. It can be easily shown that inequality (15) holds if:

$$\left\|I_{n_y} - A(q)\right\|_1 \delta_{t/min} + \|A(q)\|_1 \delta_{v_t/min} \leq \delta_{t/min}$$

where $\delta_{v_t/min}$ is the minimal singular value of $\Delta_{v_t}$. Finally one obtains the following inequality:

$$\delta_{t/min} \geq \frac{\|A(q)\|_1}{1 - \left\|I_{n_y} - A(q)\right\|_1}\delta_{v_t/min}$$

ensuring $q_t \leq 0$ and then $V_t \leq \lambda V_{t-1}$.

- Henceforth we know that in each cases ($\sigma_t = 0$ or $\sigma_t \neq 0$) one has $V_t \leq \lambda V_{t-1}$ (if conditions (7) and (8) are satisfied). This gives:

$$V_t \leq \lambda^t V_0$$

For lack of space we do not present the proof of the results (9), (11) and (12). These demonstrations are based on the previous result. ∎

*Remark 2:* The conditions (7) and (8) are only sufficient conditions and since they are conservative we have observed the algorithm to work well in some cases where these conditions are not satisfied. In general, more poles are slow more the condition (7) might not be verified, this leads to an increased risk of instability of the algorithm.

*Remark 3:* The ability to seek the true parameters vector depends on the threshold $\Delta_t$, thus its specification proves to be particularly crucial. If $\Delta_t$ is too large, the OE-OBE algorithm will be stable but the update will be frozen too early. If $\Delta_t$ is too small, stability conditions will not be satisfied. Actually the choice of the threshold depends on the system throughout the matrix $A(q)$ (which is an unknown object) and the disturbances effects throughout the weighting matrix $\Delta_{v_t}$. A dichotomy-based procedure could be used to get an appropriate value for $\Delta_t$ using all the a priori knowledge on the system. In subsection III-C a modified algorithm will be proposed to relax conditions (7) and (8).

### B. Geometrical interpretation.

In order to provide some useful insights on the parameters estimation algorithm, let us define for each time $t$:

- the observation set $\mathscr{S}_t$

$$\mathscr{S}_t = \left\{\theta \in \mathbb{R}^n, \left(y_t - \hat{\phi}_t^T \theta\right)^T \Delta_t^{-2}\left(y_t - \hat{\phi}_t^T \theta\right) \leq 1\right\}$$

- the ellipsoid $\mathscr{E}_{t-1}$

$$\mathscr{E}_{t-1} = \left\{\theta \in \mathbb{R}^n, (\theta - \hat{\theta}_{t-1})^T P_{t-1}^{-1}(\theta - \hat{\theta}_{t-1}) \leq \rho_{t-1}^2\right\}$$

where $\rho_{t-1}^2$ is a scalar such that $\theta^*$ belongs to the ellipsoid $\mathscr{E}_{t-1}$.

Given $(y_t, \hat{\phi}_t)$, $\mathscr{S}_t$ is the set of all possible $\theta$ which are consistent with the chosen threshold $\Delta_t$. An essential property of that observation set is given in the following result.

*Result 2:* If condition (7) and (8) in result 1 hold then

$$\theta^* \in \mathscr{S}_t \tag{16}$$

∎

*Proof:* If $\theta^* \in \mathscr{S}_t$ then it means that

$$\left(y_t - \hat{\phi}_t^T \theta^*\right)^T \Delta_t^{-2}\left(y_t - \hat{\phi}_t^T \theta^*\right) \leq 1 \tag{17}$$

$y_t$ is such that $y_t = \phi_t^T \theta^* + v_t$, thus (17) is true if:

$$\left((\phi_t - \hat{\phi}_t)^T \theta^* + v_t\right)^T \Delta_t^{-2}\left((\phi_t - \hat{\phi}_t)^T \theta^* + v_t\right) \leq 1$$

Bearing in mind structures of $\phi_t$ and $\hat{\phi}_t$ this can be rewritten as:

$$\left(\left(I_{n_y} - A(q)\right)\varepsilon_{t/t} + A(q)v_t\right)^T \Delta_t^{-2}\left(\left(I_{n_y} - A(q)\right)\varepsilon_{t/t} + A(q)v_t\right) \leq 1$$

In the proof of result 1 it is shown that this last relation is satisfied provided condition (7) and (8) hold. ∎

The following result provides a geometrical interpretation of the adaptation algorithm. It shows that the parameters vector $\hat{\theta}_t$ estimated at the actual time $t$ is included in both the subset $\mathscr{S}_t$ and the ellipsoid $\mathscr{E}_{t-1}$. Generally, $(\mathscr{S}_t \bigcap \mathscr{E}_{t-1})$ is not a regular convex set, the result provides another ellipsoid
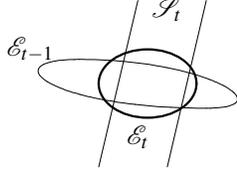
Fig. 1. 2-Dimensional example: $(\mathscr{S}_t \cap \mathscr{E}_{t-1}) \subset \mathscr{E}_t$

$\mathscr{E}_t$ which contains the intersection $(\mathscr{S}_t \bigcap \mathscr{E}_{t-1})$ (figure 1) and the parameters vector $\theta^*$ too.

*Result 3:* Consider the class of systems defined in section II and the OE-OBE algorithm given by (3) and (4). Assume (7), (8) and

$$\theta^* \in \mathscr{E}_{t-1} \tag{18}$$

then for all initial conditions
- An outer bounding ellipsoid of $(\mathscr{S}_t \cap \mathscr{E}_{t-1})$ is given by the following ellipsoid $\mathscr{E}_t$:

$$\mathscr{E}_t = \left\{ \theta \in \mathbb{R}^n, (\theta - \hat{\theta}_t)^T P_t^{-1} (\theta - \hat{\theta}_t) \leq \rho_t^2 \right\} \tag{19}$$

with

$$\rho_t^2 = \lambda \rho_{t-1}^2 + \frac{\varepsilon_{t/t-1}^T \sigma_t \varepsilon_{t/t-1}}{\varepsilon_{t/t-1}^T \Delta_t^{-2} \varepsilon_{t/t-1}} - \lambda \varepsilon_{t/t-1}^T \left( \lambda I_{ny} + \hat{\phi}_t^T P_{t-1} \hat{\phi}_t \sigma_t \right)^{-1} \sigma_t \varepsilon_{t/t-1}$$

-

$$\theta^* \in \mathscr{E}_t \tag{20}$$

Moreover if condition (10) in result 1 holds then we have:

- For $\lambda < 1$, there exists an ellipsoid $\mathscr{E}$ such that

$$\lim_{t \to \infty} \mathscr{E}_t = \mathscr{E} \tag{21}$$

∎

*Proof:*
- Let $\theta$ such that $\theta \in \mathscr{E}_{t-1}$ then we have:

$$\lambda (\theta - \hat{\theta}_{t-1})^T P_{t-1}^{-1} (\theta - \hat{\theta}_{t-1}) \leq \lambda \rho_{t-1}^2 \tag{22}$$

- $\theta$ is such that $\theta \in \mathscr{S}_t$, it follows:

$$\left( y_t - \hat{\phi}_t^T \theta \right)^T \Delta_t^{-2} \left( y_t - \hat{\phi}_t^T \theta \right) \leq 1$$

If $\sigma_t = 0$, then $\left( y_t - \hat{\phi}_t^T \theta \right)^T \sigma_t \left( y_t - \hat{\phi}_t^T \theta \right) \leq \frac{\varepsilon_{t/t-1}^T \sigma_t \varepsilon_{t/t-1}}{\varepsilon_{t/t-1}^T \Delta_t^{-2} \varepsilon_{t/t-1}}$. For $\sigma_t \neq 0$ one gets for all $(y_t - \hat{\phi}_t^T \theta)$:

$$\left( y_t - \hat{\phi}_t^T \theta \right)^T \left( \Delta_t^{-2} \left( y_t - \hat{\phi}_t^T \theta \right) \left( y_t - \hat{\phi}_t^T \theta \right)^T \sigma_t - \sigma_t \right) \left( y_t - \hat{\phi}_t^T \theta \right) \leq 0$$

It follows:

$$\left( y_t - \hat{\phi}_t^T \theta \right)^T \sigma_t \left( y_t - \hat{\phi}_t^T \theta \right) \leq \frac{\varepsilon_{t/t-1}^T \sigma_t \varepsilon_{t/t-1}}{\varepsilon_{t/t-1}^T \Delta_t^{-2} \varepsilon_{t/t-1}} \tag{23}$$

- Therefore if $\theta \in (\mathscr{S}_t \cap \mathscr{E}_{t-1})$, then it comes from (22) and (23):

$$\lambda (\theta - \hat{\theta}_{t-1})^T P_{t-1}^{-1} (\theta - \hat{\theta}_{t-1}) + \left( y_t - \hat{\phi}_t^T \theta \right)^T \sigma_t \left( y_t - \hat{\phi}_t^T \theta \right)$$
$$\leq \lambda \rho_{t-1}^2 + \frac{\varepsilon_{t/t-1}^T \sigma_t \varepsilon_{t/t-1}}{\varepsilon_{t/t-1}^T \Delta_t^{-2} \varepsilon_{t/t-1}} \tag{24}$$

After a few lines of calculation this gives

$$(\theta - \hat{\theta}_t)^T P_t^{-1} (\theta - \hat{\theta}_t)$$
$$\leq \lambda \rho_{t-1}^2 + \frac{\varepsilon_{t/t-1}^T \sigma_t \varepsilon_{t/t-1}}{\varepsilon_{t/t-1}^T \Delta_t^{-2} \varepsilon_{t/t-1}} - \lambda \varepsilon_{t/t-1}^T \left( \lambda I_{ny} + \hat{\phi}_t^T P_{t-1} \hat{\phi}_t \sigma_t \right)^{-1} \sigma_t \varepsilon_{t/t-1}$$

This corresponds to the ellipsoid $\mathscr{E}_t$ described by equation (19).
- From result 2 we have $\theta^* \in \mathscr{S}_t$ Together with (18) it gives: $\theta^* \in (\mathscr{S}_t \cap \mathscr{E}_{t-1}) \subset \mathscr{E}_t$.
- $\mathscr{E}_t$ is described by $\mathscr{E}_t = \left\{ \theta \in \mathbb{R}^n, (\theta - \hat{\theta}_t)^T P_t^{-1} (\theta - \hat{\theta}_t) \leq \rho_t^2 \right\}$. From (12) in result 1 $\lim_{t \to \infty} \sigma_t = 0$, thus for $t \gg 1$ $\hat{\theta}_t = \hat{\theta}_{t-1}$, $P_t^{-1} = \lambda P_{t-1}^{-1}$ and $\rho_t^2 = \lambda \rho_{t-1}^2$ and $\mathscr{E}_t$ becomes

$$\mathscr{E}_t = \left\{ \theta \in \mathbb{R}^n, (\theta - \hat{\theta}_{t-1})^T P_{t-1}^{-1} (\theta - \hat{\theta}_{t-1}) \leq \rho_{t-1}^2 \right\} = \mathscr{E}_{t-1}$$

This shows convergence on $\mathscr{E}_t$ and concludes the proof. ∎

*Remark 4:* This result is a generalization to the output error systems framework and the MIMO case of some results established in the linear prediction framework and SISO case. More precisely, if $n_u = n_y = 1$ and $A(q) = 1$, then our result is equivalent to results in [19] and [15].

### C. The F-OE-OBE (Filtered - OE-OBE) algorithm.

In a number of case it may be interesting to remove the condition (7) required for stability. This needs a modification of the parameter adaptation algorithm. In [8] an estimation algorithm adapted to output error systems is derived from the substitution of the observation set $\mathscr{S}_t$ by an extended observation set $\mathscr{S}_t' \supset \mathscr{S}_t$. However this leads to a conservative algorithm.

Here we use an adaptation filter so as to relax stability condition. Let first define the a priori and a posteriori adaptation errors as

$$\begin{cases} \eta_{t/t-1} = \varepsilon_{t/t-1} + \left( F(q) - I_{ny} \right) \varepsilon_{t/t} \\ \eta_{t/t} = F(q) \varepsilon_{t/t} \end{cases}$$

where $F(q)$ is the adaptation filter designed by the user. These definitions allow us to propose a filtered parameter adaptation algorithm by simply substituting in (3) and (4):
- $\varepsilon_{t/t-1}$ by $\eta_{t/t-1}$ and $\varepsilon_{t/t}$ by $\eta_{t/t}$;
- $y_t$ by $y_t^F$ such that $F(q) y_t^F = y_t$ and $\hat{\phi}_t$ by $\widehat{\phi^F}_t$ such that $F(q) \widehat{\phi^F}_t = \hat{\phi}_t$.

The idea is to compensate the effect of $A(q)$ in (7) and (8).

Taking into account these adjustments, the two following equations sets summarize the proposed Filtered OE-OBE algorithm:

$$\begin{cases} \hat{\theta}_t = \hat{\theta}_{t-1} + \Gamma_t \eta_{t/t-1} \\ \Gamma_t = P_{t-1} \widehat{\phi^F}_t \sigma_t \left( \lambda I_{ny} + \widehat{\phi^F}_t^T P_{t-1} \widehat{\phi^F}_t \sigma_t \right)^{-1} \\ P_t = \frac{1}{\lambda} \left( I_n - \Gamma_t \widehat{\phi^F}_t^T \right) P_{t-1} \\ \varepsilon_{t/t-1} = y_t^F - \widehat{\phi^F}_t^T \hat{\theta}_{t-1} \end{cases} \tag{25}$$

with

$$\sigma_t=\begin{cases} \lambda\left(\widehat{\phi^F}_t^T P_{t-1}\widehat{\phi^F}_t\right)^{-1}\left((\eta_{t/t-1}^T\Delta_t^{-2}\eta_{t/t-1})^{1/2}-1\right) \\ \quad \text{if } \left(\eta_{t/t-1}^T\Delta_t^{-2}\eta_{t/t-1}>1\right) \text{ and } \left(\widehat{\phi^F}_t^T P_{t-1}\widehat{\phi^F}_t>0\right) \\ 0 \text{ otherwise} \end{cases} \quad (26)$$

It can easily be established that the following key property holds: $\sigma_t \neq 0 \implies \eta_{t/t}^T\Delta_t^{-2}\eta_{t/t}=1$ bearing in mind that $\Delta_t$ is a now a bound on the a posteriori adaptation error $\eta_{t/t}$.

Here the equation of the a posteriori prediction error is:

$$\varepsilon_{t/t}=(A(q))^{-1}\widehat{\phi^F}_t^T\tilde{\theta}_t+(F(q))^{-1}v_t$$

which gives $\eta_{t/t}=F(q)(A(q))^{-1}\widehat{\phi^F}_t^T\tilde{\theta}_t+v_t$. Using this last equation, the following result presents an analysis of the proposed F-OE-OBE algorithm.

*Result 4:* Consider the class of systems defined in section II and the F-OE-OBE algorithm given by (25) and (26). Assume that

- $F(q)$ is such that

$$F(q)=I_{n_y}+F_1q^{-1}+\cdots+F_{n_f}q^{-n_f} \quad (27)$$

  with $F_i \in \mathbb{R}^{n_y \times n_y}$ and diagonal.
- $A(q)$ is such that

$$\left\|I_{n_y}-A(q)(F(q))^{-1}\right\|_1 < 1 \quad (28)$$

  where $\|.\|_1$ is the $l_1$ induced norm;
- $\Delta_t$ is such that:

$$\delta_{t/min}\geq\frac{\left\|A(q)(F(q))^{-1}\right\|_1}{1-\left\|I_{n_y}-A(q)(F(q))^{-1}\right\|_1}\ \delta_{v_t/min} \quad (29)$$

then for all initial conditions

- 
$$|\tilde{\theta}_t|^2\leq\gamma_1\ |\tilde{\theta}_0|^2 \quad (30)$$

If, furthermore, $\{\widehat{\phi^F}_t\}$ is a persistently exciting sequence of order $o_e \geq n$, then the following properties hold:

- for all $t \geq o_e+1$

$$|\tilde{\theta}_t|^2 \leq \gamma_2\ \lambda^t\ |\tilde{\theta}_0|^2 \quad (31)$$

- For $\lambda < 1$ one has

$$\lim_{t\to\infty}\eta_{t/t-1}^T\Delta_t^{-2}\eta_{t/t-1} \leq 1 \quad (32)$$

∎

*Proof:* The proof is similar to the proof of result 1. ∎

It can be notice, of course, that the ideal filter $F(q)$ is $F(q)=A(q)$. Thus, conditions (28) and (29) are much milder than conditions (7) and (8) if a reasonable estimated model is available. This remark is the corner stone of an iterative scheme proposed here. The strategy is the following:

1) Choose a high threshold $\Delta_t$ and apply the OE-OBE algorithm to get $\widehat{G}(q)$;
2) Given this first estimation, design a filter $F(q)=\widehat{A}(q)$ and choose a lower $\Delta_t$;
3) Apply the F-OE-OBE algorithm to get a new $\widehat{G}(q)$;
4) Repeat steps 2 and 3 until convergence of step 3.

In the next section we illustrate this iterative scheme with simulation studies.

*Remark 5:* Note that if $F(q) \simeq A(q)$ it is possible to choose $\Delta_t = \Delta_{v_t}$. In that case, from (32) we have

$$\lim_{t\to\infty}\hat{\theta}_t=\hat{\theta}$$

where $\hat{\theta}$ is such that $\eta_t^T\Delta_{v_t}^{-2}\eta_t \leq 1$ with $\eta_t = F(q)\varepsilon_t = F(q)\left(y_t^F-\widehat{\phi^F}_t^T\hat{\theta}\right) = y_t-\widehat{\phi}_t^T\hat{\theta}$. Then the contribution of the filter $F(q)$ is twofold: it relaxes stability condition of the algorithm and it allows the estimation of a model $\widehat{G}(q)$ such that:

$$\left(y_t-\widehat{G}(q)u_t\right)^T\Delta_{v_t}^{-2}\left(y_t-\widehat{G}(q)u_t\right)\leq1 \quad (33)$$

which is coherent with (1).

## IV. NUMERICAL EXAMPLE

Numerical data have been generated according to (1). The system is the following two inputs/two outputs system:

$$\begin{cases} A(q)=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}+\begin{pmatrix} -1.6 & 0 \\ 0 & -1 \end{pmatrix}q^{-1}+\begin{pmatrix} 0.6175 & 0 \\ 0 & 0.41 \end{pmatrix}q^{-2} \\ B(q)=\begin{pmatrix} 0.0175 & 0.014 \\ 0.082 & 0.41 \end{pmatrix}+\begin{pmatrix} 0 & -0.0168 \\ 0 & 0 \end{pmatrix}q^{-1} \end{cases}$$

The two inputs $u_t^1$ and $u_t^2$ are uncorrelated random binary sequences of length $N = 2000$. The noise components $v_t^1$ and $v_t^2$ are generated as in [17] in the following manner:

$$\begin{pmatrix} v_t^1 \\ v_t^2 \end{pmatrix} = \Delta_{v_t}\begin{pmatrix} \frac{1}{2}(e_t^1+sin(\pi t/10)) \\ \frac{1}{2}(e_t^2+cos(\pi t/15)) \end{pmatrix}$$

where $e_t^1$ and $e_t^2$ are white noises uniformly distributed in $[-1;1]$ and $\Delta_{v_t}=\begin{pmatrix} 0.15 & 0 \\ 0 & 0.25 \end{pmatrix}$.

It can easily be shown that condition (7) in result 1 is not satisfied in this example. However three identification procedures have been tested:

- Procedure 1: the development of (1) and (2) gives

$$A(q)y_t=B(q)u_t+A(q)v_t=B(q)u_t+v_t'$$

  The first identification procedure consists in identifying this ARX model using linear OBE algorithm proposed in [2] and [3].
- Procedure 2: even if condition (7) is not satisfied, we can try to use the OE-OBE algorithm to identify the system. This is the second identification procedure.
- Procedure 3: this third identification procedure is our iterative scheme. This scheme uses the F-OE-OBE algorithm and has been applied over 20 iterations.

For each procedure we have adjusted the threshold $\Delta_t$ so as to have the lowest threshold and in the same time a kind of stability on the $N$ available data. This has lead to the following choices:

- Procedure 1: $\Delta_t = \Delta_{v_t}$.
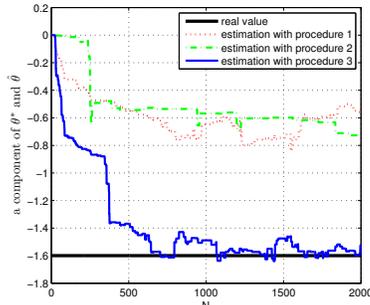- Procedure 2: $\Delta_t = 2\Delta_{v_t}$.
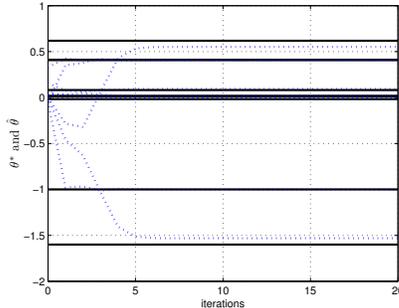
Fig. 2. Simulation results on the sixth component of $\theta^*$



Fig. 3. Improvement after each iteration with procedure 3

- Procedure 3: $\Delta_t$: we have chosen a decreasing threshold. At iteration $i$ we choose $\Delta_t^{(i)} = \left(\Delta_t^{(ini)} - \Delta_t^{(fin)}\right) e^{-i} + \Delta_t^{(fin)}$ with $\Delta_t^{(ini)} = 5\Delta_{v_t}$ and $\Delta_t^{(fin)} = \Delta_{v_t}$.

The simulation results for procedure 1, 2 and 3 are shown in Fig. 2. In this figure only the sixth component of $\theta^*$ and its estimations appear. The other components have similar behavior. Fig. 3 presents the improvement of the estimated vector at the end of each iteration in the iterative scheme. It is clear that the iterative scheme proposed in this paper works well. It increases estimation quality compared with the first estimated model and compared with the other procedures. Fig. 4 presents thresholds $\pm\Delta_{v_t}$ and the output errors $y_t - \widehat{G}(q)u_t$ for each final model (final model for procedure 1, final model for procedure 2 and final model for procedure 3 at the $20^{th}$ iteration). It appears that the model obtained with the iterative scheme using the F-OE-OBE is the only one satisfying (33).

## V. CONCLUSION

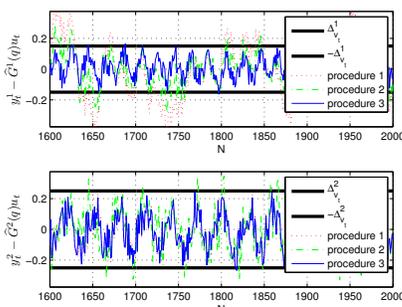In this paper we have considered the identification problem of an output error system with bounded disturbances. Two identification algorithms which belong to the class of OBE type algorithm are presented. Sufficient conditions for stability and convergence of the first algorithm (conditions related to a condition on the system) have been established. These conditions can be relaxed by introducing an adaptation filter which gives the second algorithm. This second algorithm has been used in an iterative scheme which has lead to a significant improvement of the estimation. In terms of future research, it seems necessary to refine the process for achieving relaxation $F(q) \simeq A(q)$. We also believe that the results developed in the paper can be extended to the closed loop identification problem.

## REFERENCES

[1] A. Vicino A. Garulli and G. Zappa. Conditional central algorithms for worst-case set membership indentification and filtering. *IEEE Transactions on automatic control*, 45, 2000.

[2] M. Boutayeb, Y. Becis, and M. Darouach. Recursive identification of linear multivariable systems with bounded disturbances. *Proceedings 15$^{th}$ IFAC World Congress-Barcelona*, 2002.

[3] C. Canudas-De-Wit and J. Carrillo. A modified EW-RLS algorithm for systems with bounded noise. *Automatica*, 26(3):599–606, 1990.

[4] M. Casini, A. Garulli, and A. Vicino. On input design in l1 conditional set membership identification. *Automatica*, 42(5):815–823, 2006.

[5] S. Dasgupta and Y-F Huang. Asymptotically convergent modified recursive least square with data-dependent updating and forgetting factor for systems with bounded noise. *IEEE Transactions on information theory*, 33(3):383–392, 1987.

[6] C. Durieu, B. Polyak, and E. Walter. Trace versus determinant in ellipsoidal outer bounding, with application to state estimation. *13$^{th}$ IFAC worl congress-San Francisco*, 1996.

[7] G. Favier and L.V.R. Arruda. *Review and comparison of ellipsoidal algorithms*. In bounding approaches to system identification - Plenium Press, 1996.

[8] G. Ferreres and M. MSaad. Estimation of output error models in the presence of unknown but bounded disturbances. *International Journal of Adaptive Control and Signal Processing*, 11, 1997.

[9] E. Fogel and Y.F. Huang. On the value of informaton in system identification - bounded noise case. *Automatica*, 18(2):229–238, 1982.

[10] A. Garulli, B.Z. Kacewicz, A. Vicino, and G. Zappa. Error bounds for conditional algorithms in restricted complexity set membership identification. *IEEE Transactions on automatic control*, 45, 2000.

[11] L. Ljung. *System identification: theory for the user*. Prentice Hall, 1999.

[12] R. Lozano-Leal and R. Ortega. Reformulation of the parameter identification problem for systems with bounded disturbances. *Automatica*, 23(2):247–251, 1987.

[13] M. Milanese and G. Belforte. Estimaton theory and uncertainty intervals evaluation in presence of unknown but bounded errors. linear families of models and estimators. *IEEE Transactions on automatic control*, 27, 1982.

[14] S.H. Mo and J.P. Norton. Fast and robust algorithm to compute exact polytope parameters bounds. *Mathematics and Computers in Simulation*, 32, 1990.

[15] S. Nagaraj, S. Gollamudi, S. Kapoor, and Y.F. Huang. Bounded error estimation: set theoric and least squares formulations. *Conference on information sciences and systems-Baltimore*, march 1997.

[16] M. Nayeri, M.S. Liu, and J.R. Deller. An interpretable and converging setmembership algorithm. *Proceeding IEEE of International Conference on Acoustics, Speech and Signal Processing-Minneapolis*, 1993.

[17] A.K. Rao and Y.F. Huang. Recent developments in optimal bounding ellipsoidal parameter estimation. *Mathematics and Computers in simulation*, 32, 1990.

[18] X-F. Sun and Y-Z. Fan. Comments on "identification for systems with bounded noise". *IEEE Transactions on automatic control*, 46(5):808–809, 2001.

[19] G. Tan, C. Wen, and Y.C. Soh. Identification for systems with bounded noise. *IEEE Transactions on automatic control*, 42(7):998–1001, 1997.

[20] A. Vicino and G. Zappa. Sequential approximation of parameter sets for identification with parametric and nonparametric uncertainty. *32$^{th}$ IEEE Control and Decision Conference-New York*, 1993.

Fig. 4. Output errors and threshold $\Delta_{v_t}$