Technical communique

On dynamic regressor extension and mixing parameter estimators: Two Luenberger observers interpretations

Romeo Ortega a, *, Laurent Praly b, Stanislav Aranovskiy c, d, Bowen Yi e, Weidong Zhang e

a Laboratoire des Signaux et Systèmes, CNRS, CentraleSupélec, Plateau du Moulon, 91192 Gif-sur-Yvette, France
b MINES ParisTech, PSL Research University, Centre for Automatic Control and Systems Theory, 75272 Paris, France
c CentraleSupélec – IETR, Avenue de la Boulaie, 35576 Cesson-Sévigné, France
d Department of Control Systems and Informatics, ITMO University, 197101 Saint Petersburg, Russia
e Department of Automation, Shanghai Jiao Tong University, 200240 Shanghai, China

A R T I C L E   I N F O

Article history:
Received 28 March 2017
Received in revised form 29 March 2017
Accepted 5 June 2017
Available online 25 June 2018

Keywords:
Parameter estimation
Observer design theory
Adaptive systems
Nonlinear systems

A B S T R A C T

Dynamic regressor extension and mixing is a new technique for parameter estimation with guaranteed performance improvement – with respect to classical gradient or least-squares estimators – that has proven instrumental in the solution of several open problems in system identification and adaptive control. In this brief note we give two interpretations of this parameter estimator in terms of the recent extensions, to the cases of nonlinear systems and observation of linear functionals for time-varying systems, of the classical Luenberger’s state observers.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

A new procedure to design parameter estimators for linear and nonlinear regressions, called dynamic regressor extension and mixing (DREM), was recently proposed in Aranovskiy, Bobtsov, Ortega, and Pyrkin (2017). For linear regressions DREM estimators clearly outperform classical gradient or least-squares estimators and its convergence is established without the usual, restrictive requirement of regressor persistency of excitation (PE) (Ioannou & Sun, 1996; Sastry & Bodson, 1989). Instead of PE a non-square integrability condition on the determinant of a designer-dependent extended regressor matrix is imposed. As discussed in the paper, a key feature of DREM is that it ensures the monotonicity of the individual estimation errors. The technique has been successfully applied in a variety of identification and adaptive control problems (Aranovskiy, Bobtsov, Ortega, & Pyrkin, 2016; Gerasimov, Ortega, & Nikiforov, 2018; Pyrkin, Mancilla, Ortega, Bobtsov, & Aranovskiy, 2017).

Pursuing the work reported in Praly (2016b), in this note we prove that the DREM parameter estimator for linear regressions can be derived following the classical Luenberger state observer design procedure (Luenberger, 1964) for linear time-invariant (LTI) systems and its recent extension to linear time-varying (LTV) systems (Rotella & Zambettakis, 2013; Shafai & Carroll, 1986; Trumpf, 2007). See Shoshtaiashvili (1992) for some early developments of this theory, Andrieu and Praly (2006), Kazantzis and Kravaris (1998) for its extension to nonlinear systems, Astolfi, Karagiannis, and Ortega (2008) for a related observer design based on ideas of immersion and invariance and the recent paper (Afri, Andrieu, Bako, & Dufour, 2017) for an elegant solution of the far more challenging problem of simultaneous state and parameter estimation of LTI systems using these observers.

2. Dynamic regressor extension and mixing parameter estimators

In this section we briefly review the application of DREM for linear regressions—referring the interested reader to Aranovskiy et al. (2017) for additional details.

Consider the basic problem of on-line estimation of the constant parameters of the q-dimensional linear regression

$$y(t) = \phi^\top(t)\theta,$$  

(1)
where $\gamma \in \mathbb{R}$ and $\phi \in \mathbb{R}^q$ are known, bounded functions of time and $\theta \in \mathbb{R}^q$ is the vector of unknown parameters. The standard gradient estimator
\[
\dot{\hat{\theta}} = \Gamma \phi(y - \phi^T \hat{\theta}),
\]
with a positive definite adaptation gain $\Gamma \in \mathbb{R}^{q \times q}$ yields the error equation
\[
\dot{\hat{\theta}} = -\Gamma \phi \phi^T \theta,
\]
where $\hat{\theta} := \hat{\theta} - \theta$ are the parameter estimation errors. Taking the derivative of the function $|\hat{\theta}|^2$ it is easy to show that
\[
|\hat{\theta}(0)| \geq |\hat{\theta}(t)|, \quad \forall \ t \geq 0.
\]
That is, the norm of the parameter error is non-increasing.

It is also well-known (Ioannou & Sun, 1996; Sastry & Bodson, 1989) that the zero equilibrium of the LTV system (2) is globally exponentially stable if and only if the regressor vector $\phi$ is PE, that is, if
\[
\int_0^{+T} \phi(s)\phi^T(s)ds \geq \delta I_q,
\]
for some $T, \delta > 0$ and for all $t \geq 0$, which will be denoted as $\phi \in \text{PE}$. In many practical circumstances $\phi \not\in \text{PE}$. To overcome this problem recent studies (Barabanov & Ortega, 2016; Pralay, 2016a) have derived strictly weaker conditions for global (but non-uniform) asymptotic stability of (2) – hence for convergence of the parameter errors to zero – which is sufficient in many applications. Unfortunately, these new conditions are non-robust and remain hard to verify.

To overcome the limitation imposed by the PE condition and improve the transient performance of the estimator the DREM procedure, introduced in Aranovskiy et al. (2017), generates new, one-dimensional, regression models to independently estimate each of the parameters.

The first step in DREM is to introduce a linear, single-input $q$-output, $L_\infty$-stable operator $\mathcal{H} : L_\infty \to L_\infty$, and define the vector $Y \in \mathbb{R}^q$ and the matrix $\Phi \in \mathbb{R}^{q \times q}$ as
\[
Y := \mathbf{h}y, \quad \Phi := \mathbf{h}\phi^T.
\]
Clearly, because of linearity and $L_\infty$ stability, these signals satisfy
\[
Y = \Phi \theta + \epsilon_1,
\]
with $\epsilon_1$ a vector of exponentially decaying terms neglected, without loss of generality, in the sequel—see Remark 3 in Aranovskiy et al. (2017) where the effect of these terms is rigorously analyzed.

The elements of the operator $\mathcal{H}$ may be simple, exponentially stable LTI filters of the form
\[
\mathcal{H}_i(p) = \frac{\alpha_i}{p^2 + \beta_i}, \quad i \in \mathbb{Q} := \{1, 2, \ldots, q\}
\]
with $p := \frac{d}{dt}$ and $\alpha_i \neq 0, \beta_i > 0$; in this case $\epsilon_1$ accounts for the effect of the initial conditions of the filters. Another option of interest are delay operators, that is, $[\mathcal{H}_i(\cdot)\mathbf{1}](t) := \mathbf{1}(t - d_i)$, where $d_i \in \mathbb{R}$. See Section 4 for the case of general LTV operators.

Premultiplying (6) by the adjoint matrix of $\Phi$, denoted adj($\Phi$), we get $q$ scalar regressors of the form
\[
\mathbf{y}(t) = \Delta(t)\theta_i,
\]
where we defined the scalar function $\Delta \in \mathbb{R}$
\[
\Delta := \det(\Phi),
\]
and the vector $\mathbf{y} \in \mathbb{R}^q$
\[
\mathbf{y} := \text{adj}(\Phi)\mathbf{y}.
\]
The estimation of the parameters $\hat{\theta}_i$ from the scalar regression form (7) can be easily carried out via
\[
\hat{\theta}_i = \gamma_i \Delta(\mathbf{y}_i - \Delta \hat{\theta}_i),
\]
with adaptation gains $\gamma_i > 0$. From (7) it is clear that the latter equations are equivalent to
\[
\hat{\theta}_i = -\gamma_i \Delta^2 \hat{\theta}_i,
\]
A first important advantage of DREM is that the individual parameter errors satisfy
\[
|\hat{\theta}_i(0)| \geq |\hat{\theta}_i(t)|, \quad \forall \ t \geq 0,
\]
that is, strictly stronger than the monotonicity property (3). Moreover, solving the simple scalar differential equation (11) we conclude that
\[
\lim_{t \to \infty} \hat{\theta}_i(t) = 0 \iff \Delta(t) \not\in L_2,
\]
that is, parameter convergence is established without the restrictive PE assumption. In Aranovskiy et al. (2017) the relationship between the condition $\Delta \not\in L_2$ and $\phi \in \text{PE}$ is thoroughly discussed—see also Section 5.

3. DREM as a gradient descent of a Kazantzis–Kravaris–Luenberger observer

In this section we derive DREM following the Luenberger’s observer design procedure proposed in Luenberger (1964) see also the extension to nonlinear systems, originally proposed in Kazantzis and Kravaris (1998) and refined in Andrieu and Pralay (2006). First, we select $\ell$ negative real (or complex with negative real part) numbers $\lambda_j$ and consider the system
\[
\dot{z}_j = \lambda_j \{-z - \gamma_j\}, \quad j \in \mathbb{Q} := \{1, \ldots, \ell\}.
\]
Defining
\[
T := \begin{bmatrix} T_1 & \cdots & T_\ell \end{bmatrix}, \quad z := \text{col}(z_1, \ldots, z_\ell), \quad \Lambda := \text{diag}(\lambda_1, \ldots, \lambda_\ell)
\]
we have that
\[
\frac{d}{dt}(z - T^T \dot{\theta}) = -\Lambda(z - T^T \dot{\theta}),
\]
from which it is clear that $z(t) \to T^T(t)\theta$. It follows that, if the matrix $T$ is left-invertible, we can generate an asymptotically convergent estimate of $\theta$ as
\[
\hat{\theta} = (T^T)\Lambda z,
\]
where $(T^T)\Lambda \in \mathbb{R}^{\ell \times \ell}$ is the left-inverse of $T^T$.

It should be underscored that the construction above is exactly the one proposed in Luenberger (1964) for LTI systems adapted to the particular problem when the “state” $\theta$ verifies $\theta = 0$ and the output matrix is time varying, that is, $\phi^T(t)$.  

\footnote{When clear from the context, in the sequel the arguments of the functions are omitted.}

\footnote{In the sequel the clarification $i \in \mathbb{Q}$ is omitted for brevity.}
The crucial issues of left-invertibility of $T^T$ and the actual computation of $(T^T)_+$ first discussed in Kazantzis and Kravaris (1998) – are topics of current research extensively studied in the observers literature. The key question here is the extension to non-stationary systems of the sufficient conditions reported for the stationary case in Andrieu and Praly (2006). See also Section 5 for some further discussion.

Instead of “solving” for $\hat{\theta}$ the equation $z = T^T \hat{\theta}$ an alternative route to generate the parameter estimate is to apply a gradient descent method to solve the optimization problem

$$\min_{\hat{\theta}(t)}[\{z(t) - T^T(t)\hat{\theta}(t)\}^\top Q(t)\{z(t) - T^T(t)\hat{\theta}(t)\}]$$

(15)

where the matrix $Q : R_+ \rightarrow R^q \otimes q$ is positive definite. This yields the on-line estimator

$$\hat{\theta} = \Gamma T \Phi(z - T^T \hat{\theta}).$$

(16)

We are in position to present the first result of the note.

**Proposition 1.** Consider the linear regression (1). The DREM estimator (5), (8), (9), (10) exactly coincides with the Luenberger observer (14) with the gradient descent (16) selecting

$$\ell = q$$

$$H_i(p) = \frac{-\lambda_i}{p - \lambda_i}$$

(17)

$$Q(t) = \text{adj}(T(t)\text{adj}(T^T(t))$$

$$\Gamma = \text{diag}\{\gamma_1, \ldots, \gamma_q\}.$$ 

**Proof.** The proof follows directly noting that, for the choices of the DREM parameters (17), we have that $\Phi = T^T$, and the fact that $T\text{adj}(T) = \Delta I_q$.

4. DREM as a gradient descent of a functional observer for LTV Systems

The problem of functional observers for LTV systems is formulated as follows. Given a classical LTV system

$$\dot{x} = A(t)x + B(t)u$$

$$y = C(t)x,$$

(18)

where $x \in R^q, u \in R^p, y \in R^m$ and a linear functional

$$v = M(t)x$$

(19)

with $v \in R^l$ we want to design an observer for the signal $v$. In Trumpf (2007) the following result is established.

**Proposition 2.** Define a completely observable $\ell$-dimensional system

$$\dot{z} = F(t)z + G(t)u + H(t)y$$

$$w = P(t)z,$$

(20)

with all the solutions of $\dot{x} = F(t)x$ converging to zero. The system (20) is a global asymptotic observer\(^3\) of the linear functional (19) for the system (18) if and only if there exists a continuously differentiable $n \times q$ matrix $T(t)$ solution of the equations

$$G(t) = T(t)B(t)$$

$$\dot{T}(t) = F(t)T(t) - T(t)A(t) + H(t)C(t)$$

$$M(t) = P(t)T(t).$$

(21)

To apply this result for the estimation of the parameters of the linear regression (1) we make the assignments

$$p = m = 1$$

$$x = \theta, u = 0$$

$$A(t) = 0, B(t) = 0, C(t) = \Phi^T(t).$$

The observer (20) and the conditions (21) become then

$$\dot{z} = F(t)z + H(t)y,$$

$$w = P(t)z$$

(22)

and

$$\dot{T}(t) = F(t)T(t) + H(t)\Phi^T(t)$$

$$M(t) = P(t)T(t).$$

(23)

respectively. Now, we select the dimension of the observer equal to the dimension of the systems state $\theta$, that is $\ell = q$, and choose

$$M(t) = \text{det}(T(t))I_q.$$ 

(24)

The linear functional to be observed (19) takes a decoupled form

$$v_i = \text{det}(T(t))\theta_i.$$ 

Moreover, the conditions for existence of the functional observer (23) are satisfied with the choice

$$P(t) = \text{adj}(T(t)).$$

Invoking Proposition 2 we have the following result.

**Proposition 3.** Consider the linear regression (1). For any observable pair $(F(t), H(t))$, with all the solutions of $\dot{x} = F(t)x$ converging to zero, the functional observer

$$\dot{z} = F(t)z + H(t)y$$

$$\dot{T}(t) = F(t)T(t) + H(t)\Phi^T(t)$$

$$w = \text{adj}(T(t))z$$

(25)

ensures

$$\lim_{t \to \infty}(w_i(t) - \text{det}(T(t))\theta_i) = 0$$

for all $T(t) \in R^{q \otimes q}$ and all $z(t) \in R^q$.

If $\text{det}(T(t)) \neq 0$ we can obtain an estimate of $\theta$ as

$$\hat{\theta} = \frac{1}{\text{det}(T(t))}w.$$ 

(27)

To avoid the possibility of a singularity in the calculation above we can proceed as done in Section 3 to apply a gradient descent method to solve the optimization problem

$$\min_{\hat{\theta}(t)}\{(w_i(t) - \text{det}(T(t))\hat{\theta}_i(t))^2\}. $$

(26)

This yields the on-line estimator

$$\dot{\hat{\theta}_i} = \gamma_i \text{det}(T(t))(w_i - \text{det}(T(t))\hat{\theta}_i).$$

(27)

The derivations above established the following result.

**Proposition 4.** The functional observer (25) with the gradient descent (27) exactly coincides with the DREM estimator of Section 2 with the LTV operator

$$\mathcal{H} = [p I_q - F(t)]^{-1}H(t)$$

and the identities

$$Y = z, \ \Phi = T, \ \gamma = w, \ \Delta = \text{det}(T(t)).$$

---

\(^3\) That is, for all $x(0) \in R^q, z(0) \in R^l$ and all continuous, bounded inputs $u$ we have $\lim_{t \to \infty}(v(t) - w(t)) = 0$. 

---

R. Ortega et al. / Automatica 95 (2018) 548–551
5. Concluding remarks

It has been shown that the DREM parameter estimator can be interpreted as a Luenberger observer with a gradient descent search in two different ways. First, as a “standard” full state observer (14) where the gradient search is done for the quadratic criterion (15) with the weighting matrix

\[ Q(t) = \text{adj}(T(t))\text{adj}(T^\top(t)). \]

Second, as a functional observer for LTV systems of the form (25) with the weighting matrix of the state functional to be observed (19) selected as (24) and the standard quadratic criterion to be minimized (26).

For the case when the dimension of the observer state \( z \) is equal to the dimension of the unknown vector \( \theta \), that is, when \( \ell = q \), the problem of left invertibility of the matrix \( T \) discussed in Section 3 reduces to the full rank condition of \( T \) mentioned in Section 4, and it is hardly verified in practice. On the other hand, in DREM the necessary and sufficient convergence condition (13) of the observers gradient descent search, which is related with the previous full rank assumption, remains an essentially open question.

One interesting possibility that has been considered in the literature (Afri et al., 2017; Andrieu & Praly, 2006) is to take the dimension of \( z \) larger than the one of \( \theta \), that is, \( \ell > q \). Propositions 3 and 4 apply verbatim selecting

\[ Q(t) = T^\top(t)\text{adj}(T(t))T^\top(t) \]

in the former, and

\[ w = \text{adj}(T^\top(t)T(t))T^\top(t)z \]

\[ P(t) = \text{adj}(T^\top(t)T(t))T^\top(t), \]

in the latter. Unfortunately, the existing results for observers – without the gradient descent step – reported in Andrieu and Praly (2006), Kazantzis and Kravaris (1998), Rotella and Zambettakis (2013) and Trumpf (2007) do not seem applicable for the investigation of the non square-integrability condition (13) for convergence of DREM. Current research is under way along this direction.

It should be pointed out that an important component of DREM is the flexibility in the selection of the operator \( \mathcal{H} \). It has been shown in several applications (Aranovskiy et al., 2016; Gerasimov et al., 2018; Pyrkin et al., 2017) that choosing them as LTI filters and/or delays, with different values of the coefficients \( \alpha_i, \beta_i \) and \( d_i \), is essential to guarantee the performance improvement of the estimator. The selection of these coefficients is done invoking considerations of signals bandwidth and frequency content. Although little is known on the effect of time-varying operators on a signals spectrum, choosing them to be time-varying –as suggested in Proposition 4– widens our options and opens up a new interesting area of research. In this respect, the optimality results reported for the Extended Instrumental Variable method (in the stochastic framework) might prove useful also for DREM, see Gilson (2015) for a recent survey of this method.

References


Barabanov, N. E., & Ortega, R. (2016). On global asymptotic stability of \( \dot{x} = -\phi(t)^\top (t)x \) with \( \phi(t) \) bounded and not persistently exciting. In LSS-centralesupelec Int. report, December 20.


Gilson, M. (2015). What has Instrumental Variable method to offer for system identification? In 8th IFAC international conference on mathematical modelling, MATHMOD 2015, Vienna, Austria, February. (See also hal-01242758).


