Adaptive Observer with Enhanced Gain to Address Deficient Excitation

Qinghua Zhang, Fouad Giri

Adaptive Observer with Enhanced Gain to Address Deficient Excitation

Qinghua Zhang * **Fouad Giri** **

* *Univ. Gustave Eiffel, Inria, 35042 Rennes, France (qinghua.zhang@inria.fr).*
** *Normandie University, UNICAEN, LAC, 14032 Caen, France (fouadgiri@unicaen.fr)*

Abstract: For joint estimation of state variables and unknown parameters, adaptive observers usually assume some persistent excitation (PE) condition. In practice, the PE condition may not be satisfied, because the underlying recursive estimation problem is ill-posed. To remedy the lack of PE condition, inspired by the ridge regression, this paper proposes a regularized adaptive observer with enhanced parameter adaptation gain. Like in typical ill-posed inverse problems, regularization implies an estimation bias, which can be reduced by using prior knowledge about the unknown parameters.

Keywords: Adaptive observer, regularization, persistent excitation, linear time varying (LTV) systems, joint state-parameter estimation.

1. INTRODUCTION

For most dynamic systems encountered in various engineering fields, only part of state variables are directly observed through sensor instruments, because of limited availability of sensors in practice. State estimation is thus a common task for different engineering purposes. Moreover, some model parameters may be unknown a priori, due to production variations of system components, or because of parameter evolution reflecting aging or degradation of the underlying dynamic system. *Adaptive observers* are recursive algorithms for joint estimation of both state variables and unknown parameters, based on available sensor measurements. Early reseraches about linear time invariant (LTI) systems go back to the seventies (Kreisselmeier, 1977; Ioannou and Kokotovic, 1983), then some classes of nonlinear systems are studied (see e.g. (Bastin and Gevers, 1988; Marino and Tomei, 1995; Cho and Rajamani, 1997; Besançon et al., 2006; Farza et al., 2009)) while the results about LTI and linear time varying (LTV) systems continue to be completed (Zhang, 2018).

In most methods for adaptive observer design, in addition to the observability condition for state estimation, a *persistent excitation* (PE) condition is required to ensure the convergence of the algorithms. It is well known in classical adaptive estimation problems that PE is essential for parameter estimation (Narendra and Annaswamy, 1987; Shimkin and Feuer, 1987; Narendra and Annaswamy, 1989; Astrom and Wittenmark, 1994).

However, in practice, it may happen that the PE conditions required by adaptive observers are not satisfied, because the considered recursive estimation problem is *ill-posed*. In this case, it is not possible to ensure the convergence of adaptive observers in the usual sense if they are applied. On the other hand, it is well known that the so-called *inverse problems* in various engineering fields are often ill-posed (Chavent, 2010), yet practical solutions are frequently implemented and applied. In (Self et al., 2017), a finite time excitation condition is considered instead of the usual PE condition. In (Bobtsov et al., 2021; Ortega et al., 2021), the lack of PE is addressed with a dynamic regressor extension approach.

Inspired by typical techniques solving ill-posed inverse problems (Chavent, 2010), in (Zhang et al., 2019) some simple regularization terms have been introduced into an adaptive observer. In this paper, this regularization will be revised, following some hints from the *ridge regression* (Hoerl and Kennard, 1970), well known in statistics. This new regularization leads to an enhanced gain of the adaptive observer, improving the results of (Zhang et al., 2019). Regularization is also used in robust adaptive control (Ioannou and Sun, 1996).

The introduction of regularization in estimation problems generally leads to biased estimates. This is the price to pay to solve the considered ill-posed problems. Nevertheless, *prior knowledge* about model parameters, if available, can be used to reduce the bias.

2. PROBLEM FORMULATION

In this paper, an adaptive observer will be designed for continuous-time multiple-input multiple-output (MIMO) LTV systems, formulated as

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + \Phi(t)\theta + w(t) \tag{1a}$$

$$y(t) = C(t)x(t) + v(t) \tag{1b}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^q$ the input (control), $y(t) \in \mathbb{R}^m$ the output, $\dot{x}(t) = dx(t)/dt$, $\theta \in \mathbb{R}^p$ the unknown constant parameter vector, $A(t), B(t), C(t), \Phi(t)$ are appropriate size matrix-valued functions of the time t, $w(t) \in \mathbb{R}^n$ and $v(t) \in \mathbb{R}^m$ represent uncertainties in the state and output equations. At the initial instant t_0, the initial state $x(t_0)$ is unknown. In order to keep the problem formulation and analysis within the framework of ordinary differential equations (ODE), it is assumed that the uncertainties $w(t)$ and $v(t)$ are unknown arbitrary *bounded* functions of t.

The purpose of this paper is to jointly estimate the state $x(t)$ and the parameter θ from the input $u(t)$, the output $y(t)$ and the time varying matrices $A(t), B(t), E(t), C(t), \Phi(t)$, despite deficient excitations.

3. A CONVENTIONAL ADAPTIVE OBSERVER

This section *recalls* an adaptive observer for MIMO LTV systems, whose convergence has been established under a PE condition. It will serve as the basis for the results presented in the next sections.

For general LTV systems as formulated in (1), the adaptive observer initially proposed in (Zhang, 2002), and improved with an exponential forgetting factor in (Zhang and Clavel, 2001; Li et al., 2011), is

$$\begin{aligned}\dot{\hat{x}}(t) &= A(t)\hat{x}(t) + B(t)u(t) + \Phi(t)\hat{\theta}(t)\\ &\quad + K(t)[y(t) - C(t)\hat{x}(t)]\\ &\quad + \Upsilon(t)\Gamma(t)\Upsilon^T(t)C^T(t)[y(t) - C(t)\hat{x}(t)] \end{aligned}\tag{2a}$$

$$\dot{\hat{\theta}}(t) = \Gamma(t)\Upsilon^T(t)C^T(t)[y(t) - C(t)\hat{x}(t)] \tag{2b}$$

$$\dot{\Upsilon}(t) = [A(t) - K(t)C(t)]\Upsilon(t) + \Phi(t) \tag{2c}$$

$$\dot{\Gamma}(t) = \lambda\Gamma(t) - \Gamma(t)\Upsilon^T(t)C^T(t)C(t)\Upsilon(t)\Gamma(t), \tag{2d}$$

with the initialization

$$\hat{x}(t_0) = \hat{x}_0 \tag{3a}$$

$$\hat{\theta}(t_0) = \hat{\theta}_0 \tag{3b}$$

$$\Upsilon(t_0) = 0_{n\times p} \tag{3c}$$

$$\Gamma(t_0) = \Gamma_0, \tag{3d}$$

where $\hat{x}(t) \in \mathbb{R}^n$ is the state estimate, $\hat{\theta}(t) \in \mathbb{R}^p$ the parameter estimate, $\Upsilon(t) \in \mathbb{R}^{n\times p}$ a matrix of auxiliary variables, $K(t) \in \mathbb{R}^{n\times m}$ the state estimation gain matrix, $\Gamma(t) \in \mathbb{R}^{p\times p}$ the parameter gain matrix, $\hat{x}_0 \in \mathbb{R}^n$, $\hat{\theta}_0 \in \mathbb{R}^p$, $\Gamma_0 \in \mathbb{R}^{p\times p}$ are initial values of $\hat{x}(t)$, $\hat{\theta}(t)$, $\Gamma(t)$, $0_{n\times p}$ is the $n \times p$ zero matrix initializing the auxiliary matrix $\Upsilon(t)$, and $\lambda > 0$ is a scalar forgetting factor.

In general, the state estimation gain $K(t)$ is chosen such that, if the term $\Phi(t)\theta$ was omitted in the state equation (1a), $K(t)$ would lead to a convergent state observer in the form of

$$\dot{\hat{x}}(t) = A(t)\hat{x}(t) + B(t)u(t) + K(t)[y(t) - C(t)\hat{x}(t)]. \tag{4}$$

Various methods for designing such observer gains are available. In this paper, the Kalman gain is chosen as the observer gain $K(t)$, and accordingly computed as

$$\begin{aligned}\dot{P}(t) &= A(t)P(t) + P(t)A^T(t) + Q(t)\\ &\quad - P(t)C^T(t)R^{-1}(t)C(t)P(t)\end{aligned}\tag{5}$$

$$K(t) = P(t)C^T(t)R^{-1}(t) \tag{6}$$

with a positive definite initial matrix $P(t_0)$. In the Kalman filter literature, usually $P(t) \in \mathbb{R}^{n\times n}$ is known as the covariance matrix of the state estimate, $Q(t) \in \mathbb{R}^{n\times n}$ and $R(t) \in \mathbb{R}^{m\times m}$ are respectively the state and output noise covariance matrices. In this paper, the considered problem is formulated in a deterministic framework, therefore $Q(t), R(t)$ *are treated as tuning parameters*, typically chosen as constant matrices, both symmetric positive definite.

It was assumed in (Zhang, 2002; Zhang and Clavel, 2001; Li et al., 2011) that the matrix $\Phi(t)$ contains sufficient variations such that the matrix $\Upsilon(t)$, which is driven by $\Phi(t)$ through (2c), satisfies

$$\int_t^{t+T} \Upsilon^T(\tau)C^T(\tau)C(\tau)\Upsilon(\tau)d\tau \geq \rho I_p, \tag{7}$$

for some positive constants T and ρ, and for all $t \geq t_0$.

The auxiliary matrix $\Upsilon(t)$ contains signals obtained by linearly filtering $\Phi(t)$ through the linear filter (2c). Therefore, the above assumption assumed in (Zhang, 2002; Zhang and Clavel, 2001; Li et al., 2011) is indeed about the properties of $\Phi(t)$, which is usually filled with exogenous signals (inputs), sometimes with injected outputs. Stated in this way, it appears that the PE condition (7) concerns only $\Phi(t)$ in the considered system model (1), regardless of the system input $u(t)$. In some applications, the term $\Phi(t)\theta$ is a reformulation of the input term $Bu(t)$, with the parameter vector θ involving entries of B. In this particular case, the PE condition does involve the input $u(t)$.

In the remaining part of this paper, modifications to the adaptive observer (2) will be proposed to address the case where (7) *is not satisfied* due to lack of excitation.

4. SOME HINTS FROM RIDGE REGRESSION

Ridge regression is a well known solution to linear regression problems when the regressors are highly correlated, a situation similar to the lack of excitation for adaptive observers. Reviewing ridge regression (Hoerl and Kennard, 1970) will be instructive for the main topic of this paper.

Consider the simple linear regression

$$z(t) = \phi^T(t)\theta + e(t) \tag{8}$$

with the output $z(t) \in \mathbb{R}$, the regressor vector $\phi(t) \in \mathbb{R}^p$ the unknown parameter vector $\theta \in \mathbb{R}^p$, and the noise $e(t) \in \mathbb{R}$. With data collected over a finite time interval, for $t \in [t_0, t_f]$, the classical least squares estimation of θ is

$$\hat{\theta}_{\text{LS}} = \left(\int_{t_0}^{t_f} \phi(t)\phi^T(t)dt\right)^{-1} \int_{t_0}^{t_f} \phi(t)z(t)dt. \tag{9}$$

This solutions assumes that the involved matrix integral

$$\int_{t_0}^{t_f} \phi(t)\phi^T(t)dt \tag{10}$$

is invertible. If it is not the case, *i.e.*, this matrix is rank deficient, then a remedy is the ridge regression. Usually formulated in discrete time, the ridge regression solution for the continuous time linear regression problem (8) is given by

$$\hat{\theta}_{\text{Ridge}} = \left(\int_{t_0}^{t_f} \phi(t)\phi^T(t)dt + \alpha_0 I_p\right)^{-1} \int_{t_0}^{t_f} \phi(t)z(t)dt \tag{11}$$

where α_0 is some chosen positive real number ensuring the existence of the involved matrix inverse. Clearly, ridge regression leads to a biased solution (Hoerl and Kennard, 1970). Anyway no unbiased solution exists when the matrix integral in (10) is not invertible.

For any $t \geq t_0$, let

$$L(t) \triangleq \int_{t_0}^{t} \phi(s)\phi^T(s)ds + \alpha_0 I_p \tag{12}$$

$$\xi(t) \triangleq \int_{t_0}^{t} \phi(s)z(s)ds, \tag{13}$$

then $\hat{\theta}_{\text{Ridge}} = L^{-1}(t_f)\xi(t_f)$.

Now consider the "recursive" algorithm

$$L(t_0) = \alpha_0 I_p \tag{14a}$$
$$\xi(t_0) = 0 \tag{14b}$$
$$\dot{L}(t) = \phi(t)\phi^T(t) \tag{14c}$$
$$\dot{\xi}(t) = \phi(t)z(t) \tag{14d}$$
$$\hat{\theta}_R(t) = L^{-1}(t)\xi(t), \tag{14e}$$

it is then trivial to check that $\hat{\theta}_R(t_f) = \hat{\theta}_{\text{Ridge}}$, as given in (11).

By differentiating $\hat{\theta}_R(t)$ as defined in (14e) and by eliminating $\xi(t)$, the above algorithm is then more compactly written as

$$L(t_0) = \alpha_0 I_p \tag{15a}$$
$$\hat{\theta}_R(t_0) = 0 \tag{15b}$$
$$\dot{L}(t) = \phi(t)\phi^T(t) \tag{15c}$$
$$\dot{\hat{\theta}}_R(t) = L^{-1}(t)\phi(t)[z(t) - \phi^T(t)\hat{\theta}_R(t)]. \tag{15d}$$

For a *given time interval* $[t_0, t_f]$, the ridge regression solution $\hat{\theta}_{\text{Ridge}}$ as first expressed in (11) is indeed equivalent to $\hat{\theta}_R(t_f)$ computed by algorithm (15). However, for any (arbitrarily large) t, the application of algorithm (15) may be troublesome, since in the solution

$$L(t) = \int_{t_0}^{t} \phi(s)\phi^T(s)ds + \alpha_0 I_p, \tag{16}$$

the term $\alpha_0 I_p$ would represent a smaller and smaller portion when t increases, leading to ill-conditioned $L(t)$. For this reason, equation (15c) is then modified as

$$\dot{L}(t) = \phi(t)\phi^T(t) + \alpha I_p \tag{17}$$

with some chosen positive constant α, so that the modified

$$L(t) = \int_{t_0}^{t} \phi(s)\phi^T(s)ds + \alpha_0 I_p + (t - t_0)\alpha I_p \tag{18}$$

contains a linearly increasing regularization term.

Finally, let

$$\Gamma(t) \triangleq L^{-1}(t), \tag{19}$$

and introduce an exponential forgetting factor $\lambda > 0$, then algorithm (15) becomes

$$\Gamma(t_0) = \frac{1}{\alpha_0} I_p \tag{20a}$$
$$\hat{\theta}_R(t_0) = 0 \tag{20b}$$
$$\dot{\Gamma}(t) = \lambda\Gamma(t) - \Gamma(t)[\phi(t)\phi^T(t) + \alpha I_p]\Gamma(t). \tag{20c}$$
$$\dot{\hat{\theta}}_R(t) = \Gamma(t)\phi(t)[z(t) - \phi^T(t)\hat{\theta}_R(t)] - \alpha\Gamma(t)\hat{\theta}_R(t). \tag{20d}$$

Of course, by setting $\alpha = 0$ and appropriately choosing $\Gamma(t_0)$, algorithm (20) would reduce to the continuous time RLS with an exponential forgetting factor:

$$\Gamma(t_0) = \frac{1}{\alpha_0} I_p \tag{21a}$$
$$\hat{\theta}_R(t_0) = 0 \tag{21b}$$
$$\dot{\Gamma}(t) = \lambda\Gamma(t) - \Gamma(t)\phi(t)\phi^T(t)\Gamma(t). \tag{21c}$$
$$\dot{\hat{\theta}}_R(t) = \Gamma(t)\phi(t)[z(t) - \phi^T(t)\hat{\theta}_R(t)]. \tag{21d}$$

The extra terms in (20) involving α introduces the regularization effect, as in the basic ridge regression. These facts will inspire the modification of the adaptive observer (2) for a similar regularization effect, in order to address the lack of excitation.

5. ENHANCED GAIN ADAPTIVE OBSERVER

It is easy to recognize the similarity between the parameter gain adaptation equation (2d) of the conventional adaptive observer and equation (21c) of the recursive least squares algorithm with an exponential forgetting factor. The parameter update equation (2b) is also similar to (21d). Another similarity to be noticed is between the two integral matrices in (7) and (10), respectively involved in the PE condition of the conventional adaptive observer and in the condition for solving the basic least squares problem.

Inspired by the relationship between the ridge regression recursive algorithm (20) and the recursive least squares algorithm (21), the conventional adaptive observer (2) is then modified as

$$\begin{aligned}\dot{\hat{x}}(t) = {} & A(t)\hat{x}(t) + B(t)u(t) + \Phi(t)\hat{\theta}(t) \\ & + K(t)[y(t) - C(t)\hat{x}(t)] \\ & + \Upsilon(t)\Gamma(t)\Upsilon^T(t)C^T(t)[y(t) - C(t)\hat{x}(t)] \\ & - \alpha\Upsilon(t)\Gamma(t)\hat{\theta}(t)\end{aligned} \tag{22a}$$

$$\begin{aligned}\dot{\hat{\theta}}(t) = {} & \Gamma(t)\Upsilon^T(t)C^T(t)[y(t) - C(t)\hat{x}(t)] \\ & - \alpha\Gamma(t)\hat{\theta}(t)\end{aligned} \tag{22b}$$

$$\dot{\Upsilon}(t) = [A(t) - K(t)C(t)]\Upsilon(t) + \Phi(t) \tag{22c}$$

$$\begin{aligned}\dot{\Gamma}(t) = {} & \lambda\Gamma(t) \\ & - \Gamma(t)[\Upsilon^T(t)C^T(t)C(t)\Upsilon(t) + \alpha I_p]\Gamma(t),\end{aligned} \tag{22d}$$

with some regularization parameter $\alpha > 0$, and with the same initial conditions as in (3) and the Kalman gain $K(t)$ as in (6). The *only modifications* w.r.t. the adaptive observer (2) are the 3 extra terms involving α in equations (22a), (22b) and (22d).

The similarity between equation (22d) and the corresponding equation (20c) of the recursive ridge regression algorithm is obvious, with $\phi(t)$ corresponding to $\Upsilon^T(t)C^T(t)$. While the last term of (22b) is as in (20d), the last term of (22a) has been accordingly added so that the last two terms of this equation sum to $\Upsilon(t)\dot{\hat{\theta}}(t)$. This simplification plays an important role in the analysis of the error system of the adaptive observer. Finally, note that in (Zhang et al., 2019) a simple constant gain matrix Γ was used instead of the time varying $\Gamma(t)$ of this paper, therefore, only the two regularization terms as in (22a) and (22b) were added.

6. NUMERICAL SIMULATION EXAMPLE

For a simple example, consider the LTI system

$$\begin{bmatrix}\dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3\end{bmatrix} = \begin{bmatrix}-1 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & -1\end{bmatrix}\begin{bmatrix}x_1 \\ x_2 \\ x_3\end{bmatrix} + \begin{bmatrix}-1 \\ 0 \\ 0\end{bmatrix}u + \begin{bmatrix}\theta_1 \\ \theta_2 \\ \theta_3\end{bmatrix} + \begin{bmatrix}w_1 \\ w_2 \\ w_3\end{bmatrix}$$

$$\begin{bmatrix}y_1 \\ y_2\end{bmatrix} = \begin{bmatrix}1 & 0 & 0 \\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_1 \\ x_2 \\ x_3\end{bmatrix} + \begin{bmatrix}v_1 \\ v_2\end{bmatrix}.$$

Intuitively, the corresponding constant matrix $\Phi(t) = I_3$ is likely not to be sufficiently exciting. To confirm this intuition, let $\sigma \in \mathbb{R}$ be *any* constant value. Make a state variable change with $z_2 = x_2 + \sigma$ replacing x_2, whereas x_1 and x_2 remain unchanged. After this variable change, the state and output equations are exactly as before, except that θ_1 and θ_3 are replaced respectively by $(\theta_1-\sigma)$ and $(\theta_3+\sigma)$. This result implies that all parameter pairs θ_1, θ_3 corresponding to the same sum value $\theta_1+\theta_3$ lead to the same input-output relationship! It is thus impossible to uniquely determine θ_1 and θ_3 from input-output data.

For the simulation example presented below, the input $u(t)$

$$u(t) = \sin(t) + \cos(\sqrt{7}t), \tag{23}$$

the parameter vector $\theta = [1; 0.7; 0.5]$, and the initial state $x(0) = [2; 2; 2]$. The uncertainty terms are simulated as

$$w(t) = 0.1[\Delta(t); \Delta(t/2); \Delta(t/3)] \tag{24}$$

$$v(t) = 0.01[\Delta(t); \Delta(t/2)] \tag{25}$$

with the triangular wave function

$$\Delta(t) \triangleq 4\,|t - \lfloor t + 0.5\rfloor| - 1,$$

and the notation $\lfloor x \rfloor$ denoting the largest integer less than or equal to $x \in \mathbb{R}$.

For the regularized adaptive observer, the state estimation gain $K(t)$ is computed using (5) and (6) with the initial condition $P(0) = I_3$, $Q(t) = 0.1I_3$, $R(t) = 0.01I_2$, the initial parameter estimation gain $\Gamma(0) = 10^3 I_3$, the exponential forgetting factor $\lambda = 0.5$, and the regularization parameter $\alpha = 0.0004$. The initial state estimate $\hat{x}_0$ and parameter estimate $\hat{\theta}_0$ are set to zero values.

As shown in Figures 1 and 2, with the enhanced gain adaptive observer (22), despite the deficient excitation and the simulated uncertainties, the state and parameter estimates converge toward their true values with some biases, after a transient period of about 2 seconds. The two parameter estimates $\hat{\theta}_1(t)$ and $\hat{\theta}_3(t)$ have similar bias values, but of opposite signs. This fact is due to the indetermination of the two parameters: recall that all parameter pairs (θ_1, θ_3) corresponding to the same sum value $\theta_1+\theta_3$ lead to the same input-output relationship.

To compare with the result of the conventional adaptive observer, the above simulation is run again by setting $\alpha = 0$, so that the enhanced gain adaptive observer (22) is equivalent to the conventional adaptive observer (2). The obtained state and parameter estimates are plotted in Figures 3 and 4. At the beginning, before $t = 1$ sec., the behaviors of state estimation and parameter estimation are similar to those of the enhanced gain adaptive observer. However, the second component of the estimated state $\hat{x}_2(t)$ maintains its bias, so do the first and the third components of parameter estimation $\hat{\theta}_1(t)$ and $\hat{\theta}_3(t)$. This example clearly illustrates the advantage of the enhanced gain adaptive observer.

7. CONCLUSION

The convergence of adaptive observers is usually ensured by some persistent excitation condition. In order to apply such algorithms when this condition is not satisfied, a regularized adaptive observer inspired by the usual practice for solving ill-posed inverse problems is proposed in this paper. In future studies, regularization will be studied in a stochastic framework in order to guide the tuning of the regularization parameter.

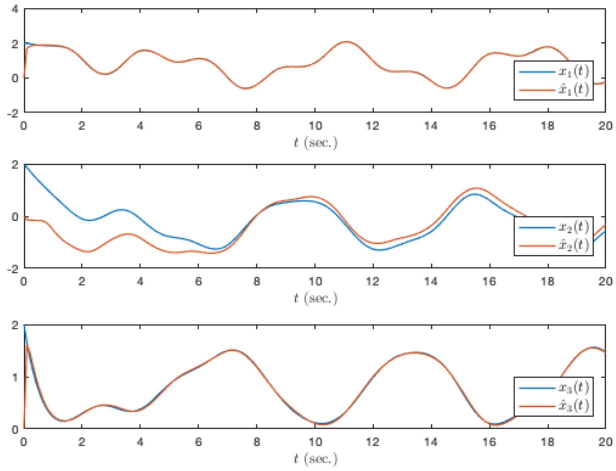

Fig. 1. Enhanced gain adaptive observer: simulated states (in blue) and their estimates (in red).

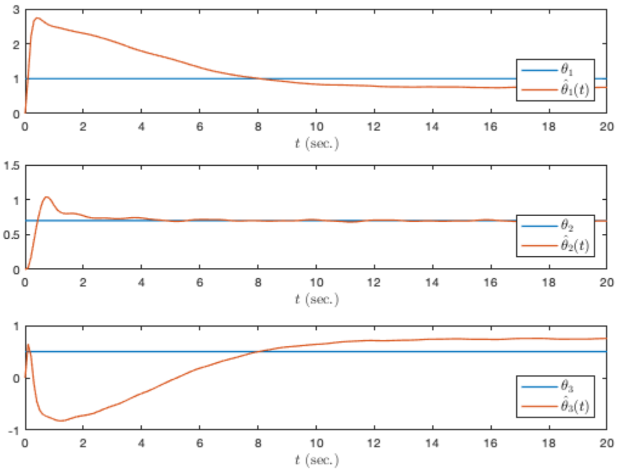

Fig. 2. Enhanced gain adaptive observer: simulated parameter values (in blue) and their estimates (in red).

REFERENCES

Astrom, K. and Wittenmark, B. (1994). *Adaptive control*. Addison-Wesley, 2nd edition edition.

Bastin, G. and Gevers, M. (1988). Stable adaptive observers for nonlinear time varying systems. *IEEE Trans. on Automatic Control*, 33(7), 650–658.

Besançon, G., De Leon-Morales, J., and Huerta-Guevara, O. (2006). On adaptive observers for state affine systems. *International Journal of Control*, 79(6), 581–591.

Bobtsov, A., Nikolaev, N., Ortega, R., and Efimov, D. (2021). State observation of LTV systems with delayed measurements: A parameter estimation-based approach with fixed convergence time. *Automatica*, 131.

Chavent, G. (2010). *Nonlinear Least Squares for Inverse Problems*. Springer.

Cho, Y.M. and Rajamani, R. (1997). A systematic approach to adaptive observer synthesis for nonlinear systems. *IEEE Trans. on Automatic Control*, 42(4), 534–537.

Farza, M., M'Saada, M., Maatouga, T., and Kamounb, M. (2009). Adaptive observers for nonlinearly parameterized class of nonlinear systems. *Automatica*, 45(10), 2292–2299.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.

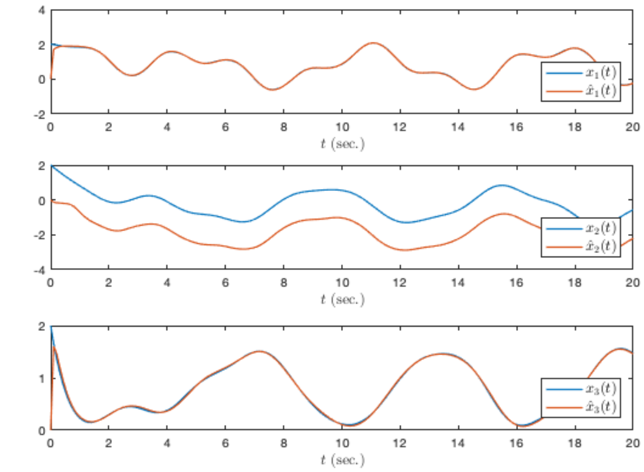

Fig. 3. Conventional adaptive observer: simulated states (in blue) and their estimates (in red).

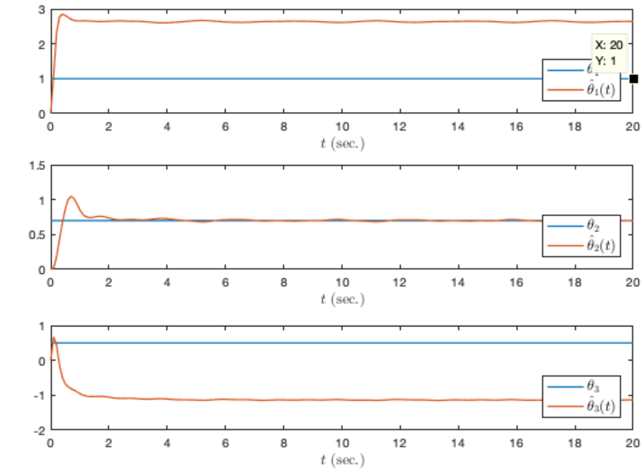

Fig. 4. Conventional adaptive observer: simulated parameter values (in blue) and their estimates (in red).

Ioannou, P. and Sun, J. (1996). *Robust Adaptive Control*. Prentice Hall.

Ioannou, P.A. and Kokotovic, P.V. (1983). *Adaptative systems with reduced models*, volume 47 of *Lecture Notes in Control and Information Sciences*. Springer, Berlin.

Kreisselmeier, G. (1977). Adaptive observers with exponential rate of convergence. *IEEE Trans. on Automatic Control*, 22(1), 2–8.

Li, X., Zhang, Q., and Su, H. (2011). An adaptive observer for joint estimation of states and parameters in both state and output equations. *International Journal of Adaptive Control and Signal Processing*, 25(9), 831–842.

Marino, R. and Tomei, P. (1995). *Nonlinear control design*. Information and system sciences. Prentice Hall, London, New York.

Narendra, K.S. and Annaswamy, A.M. (1989). *Stable adaptive systems*. Prentice Hall, Boston.

Narendra, K.S. and Annaswamy, A.M. (1987). Persistent excitation in adaptive systems. *International Journal of Control*, 45(1), 127–160.

Ortega, R., Bobtsov, A., Nikolaev, N., Schiffer, J., and Dochain, D. (2021). Generalized parameter estimation-based observers: Application to power systems and chemical biological reactors. *Automatica*, 129.

Self, R., Abudia, M., Mahmud, N., and Kamalapurkar, R. (2017). Online simultaneous state and parameter estimation. arXiv:1703.07068.

Shimkin, N. and Feuer, A. (1987). Persistency of excitation in continuous-time systems. *Systems & Control Letters*, 9(3), 225–233.

Zhang, Q. (2002). Adaptive observer for multiple-input-multiple-output (MIMO) linear time varying systems. *IEEE Trans. on Automatic Control*, 47(3), 525–529.

Zhang, Q. (2018). Adaptive Kalman filter for actuator fault diagnosis. *Automatica*, 93, 333–342.

Zhang, Q. and Clavel, A. (2001). Adaptive observer with exponential forgetting factor for linear time varying systems. In *IEEE Conference on Decision and Control (CDC'2001)*, 3886–3891. Orlando.

Zhang, Q., Giri, F., and Ahmed-Ali, T. (2019). Regularized Adaptive Observer to Address Deficient Excitation. In *13th IFAC Workshop on Adaptive and Learning Control Systems (ALCOS 2019)*. Winchester, United Kingdom. URL `https://hal.archives-ouvertes.fr/hal-02414207`.