Received: 24 May 2020 | Revised: 1 August 2020 | Accepted: 15 September 2020

DOI: 10.1002/rnc.5270

RESEARCH ARTICLE

WILEY

Attractive ellipsoid design for robust sliding-mode observation error in stochastic nonlinear discrete-time systems

David Velázquez | **Alex Poznyak**

Department of Automatic Control, CINVESTAV-IPN, Mexico City, Mexico

Correspondence
David Velazquez, Department of Automatic Control, CINVESTAV-IPN, Mexico City, Mexico.
Email: dvelazquez@ctrl.cinvestav.mx

Summary

In this paper the observation process of stochastic discrete-time nonlinear system is analyzed. The system to be observed is assumed to be uncertain, but fulfilling the global "quasi-Lipschitz" condition and is subjected to stochastic input and output disturbances of a white noise type. The combination of a traditional Luenberger residual term with a discontinuous one is considered. The designing of the best observer gain matrices is realized by using the Robust Attractive Ellipsoid Method for the analysis of the averaged observation error. The construction of this attractive ellipsoid is based on the numerical solution of some matrix optimization problem under specific constraints of Bilinear and Linear Matrix Inequalities (BMI's and LMI's) type applied to improve the attractiveness zone estimation. Two numerical xamples illustrate the effectiveness of the suggested approach.

KEYWORDS

attractive ellipsoid method, bilinear and linear matrix inequalities, discrete-time stochastic process, Quasi-Lipschitz conditions, sliding mode observer

1 | INTRODUCTION

The problem of state variables estimating of a dynamic system, given observations of the output variables, is of fundamental importance in control theory since many feedback control designs require the availability of the state of the controlled plant.

1.1 | Complete information on dynamics and conditional probabilities

1.1.1 | Linear estimations

Considering the class of *linear systems*, there are two approaches available:

- if the output variables can be measured exactly and if there are no stochastic disturbances acting on the system, then one can use a Luenberger observer to reconstruct the state[1-3]

Abbreviations: ANA, anti-nuclear antibodies; APC, antigen-presenting cells; IRF, interferon regulatory factor

- on the other hand, if all the output variables are corrupted by additive white noise then one can use a Kalman filter[4,5] for a state estimation.

The relation between the two approaches for linear systems was clarified in Reference 6.

1.1.2 | Nonlinear estimations

There is extensive past work on estimation of *nonlinear stochastic continuous systems* using various approaches by Stratonovich,[7-12] and others; however, unlike the linear case the solution to the nonlinear estimation problem is given as a nonexplicit representation. To obtain a practical solution to the nonlinear estimation problem, that is, a suitable filter, one can find algorithms for the evaluation of the representation,[13,14] (these have severe limitations due to the extensive computational work required) or use suboptimal filters[15,16] based on the assumption that the conditional density can be adequately characterized by its low-order moments. For stochastic discrete nonlinear systems, the situation is even more unsatisfactory. Only scattered results based on approximation theory exist. Several authors have suggested filtering algorithms based on the Gram-Charlier expansion of the conditional density,[17,18] or on the related Edgeworth expansion. Kizner[19] has suggested basing on approximation scheme on the Hermite expansion. Alspach and Sorenson[20] have suggested using a weighted sum of Gaussian densities to approximate the conditional density. Center[14] suggested a new class of filtering algorithms based on a generalization of standard least square approximation. Bucy and Senne[21] have used a recursive algorithm which is related to generalized least square approximation by step functions. Observers for nonlinear stochastic systems, constructed based on a Lyapnov-like method, were presented in Reference 22.

Summarizing, the usefulness of nonlinear filters is limited by the complexity and the quantity of online operations required. As a result, before 90th of the last century nonlinear filters have not yet become practical. Moreover, all references mentioned above supposed that the exact information on a considered dynamics is completely known.

1.2 | Observers for nonlinear uncertain systems

But beginning from these years, a very promising approach dealing with Sliding Mode (SM) technique generated a large number of publications concerning the observer design for *nonlinear systems containing uncertainties*. This technique is one of the most practical techniques for the stabilization of linear or nonlinear, continuous or discrete systems.[23-26] Several perspectives have been addressed, discrete-time quasi-sliding-mode control systems are considered in Reference 27 where a new definition describe the quasi-sliding mode as a motion of the system, such that its state always remains in a certain band around the sliding hyperplane. However, the SM application becomes difficult when not all the states of the system are available for feedback. And even more, when a few states available for measurement contain noises of any kind, the observation problems becomes to be very actual for designing a feedback controllers based on state estimates obtained online. Such SM observers have been applied in systems with deterministic bounded perturbations.[28-30]

Research in discrete-time control has been intensified in recent years. A primary reason is that most control strategies nowadays are implemented in discrete time.[31] This also necessitated a rework in the SM control strategy for sampled-data systems. In such systems, the switching frequency in control variables is limited by T^{-1} ; where T is the sampling period. This has led researchers to approach discrete-time SM control from two directions. The first is the emulation that focuses on how to map continuous-time SM control to discrete-time and the switching term can be preserved,[32,33] or how to improve the performance of Sliding Mode Control(SMC) as[34] that shows increased robustness can be achieved for discrete time SMC systems by choosing the sliding variable, or the output. The second is based on the equivalent control design and disturbance observer.[35,36] The observation of deterministic linear and nonlinear systems with bounded perturbations has been extensively studied as can be seen in References 37 and 38. However, in the most papers there was considered the situation when the available states as well as the output were free of measurement noise.

1.3 | Observers for linear and nonlinear uncertain stochastic systems

Stochastic systems have not been so lucky as deterministic ones. The number of works, in which the methodology of SMs is applied to observe or control the Discrete Time Stochastic Systems, seems to be very limited,[39-42] basically dealing with linear models (eg, Reference 43). The stability analysis of nonlinear discrete time stochastic systems can be found in

Reference 44. The most advanced studies, concerning the SM observers design for discrete time systems, can be found in References 45,46 and 47. In Reference 48 both linear extended state observer (ESO) and nonlinear ESO with homogeneous weighted functions are proposed for a class of multi-input multi-output (MIMO) nonlinear systems composed of coupled subsystems with large stochastic uncertainties. The stochastic uncertainties in each subsystem including internal coupled unmodeled dynamics and external stochastic disturbance without known statistical characteristics are lumped together as the stochastic total disturbance (extended state) of each subsystem. The linear ESO and nonlinear ESO are designed separately for real-time estimation of not only the unmeasured state but also the stochastic total disturbance of each subsystem.

1.4 | Main contribution of the paper

The **novelty** of the paper consists in the following:

- This paper is focused on the study of *nonlinear models* given in discrete time, which contain the stochastic perturbations (white noise type) *exerting on both the state behavior and model outputs*.
- The estimate are made using the Luenberger-like filter, containing the linear residual term, with *the additional discontinuous term*. Usually, the discontinuous term is used in Sliding Mode Theory (SMT) for bounded perturbations, but the noise has a stochastic nature (Gaussian white noise type) and, hence, is unbounded. So, here *we explain how to apply SMT for unbounded disturbances*.
- A special procedure, which is based on the *Attractive Ellipsoid Method (AEM)*,[49,50] is used to find the quasi-optimal gains matrices for the suggested observer (filter).
- Two examples (one academic and one application) are presented to illustrate the workability of suggested approach.

2 | PLANT DESCRIPTION AND PROBLEM FORMULATION

The following the symbol notations are used below:

- $\{\mathcal{F}_k\}_{k\geq 0}$ is a flow of the σ -algebras $\mathcal{F}_k$, which for each $k=0,1,\dots$ is a minimal sigma-algebra, generated by the prehistory of the process, that is,

$$\mathcal{F}_k=\sigma\{x(0),u(0),\xi_y(0);\dots;x(k),u(k),\xi_x(k),\xi_y(k)\};\tag{1}$$

- $\mathrm{E}\{\cdot|\mathcal{F}_k\}$ and $\mathrm{E}\{\cdot\}$ represent the operators of conditional and complete mathematical expectation;
- *a.s.* means "almost sure" or, equivalently, "with probability one".

2.1 | Model of the plant

Let us consider the stochastic discrete-time processes $\{x(k)\}_{k\geq 0}$, $\{y(k)\}_{k\geq 0}$ and $\{u(k)\}_{k\geq 0}$, referred hereafter as to states, outputs and measurable inputs (controls) of certain plant, which dynamics is generated by the following recurrent difference equation

$$\left.\begin{array}{r}x(k+1)=f(x(k),k)+Bu(k)+\xi_x(k+1),\\ y(k)=Cx(k)+\xi_y(k).\end{array}\right\}.\tag{2}$$

Here $x(k)\in\mathbb{R}^n$, $y(k)\in\mathbb{R}^m$, $u(k)\in\mathbb{R}^l$ and $k=0,1,2\dots$; $\{\xi_x(k)\}_{k\geq 0}$ and $\{\xi_y(k)\}_{k\geq 0}$ are random sequences associated with external "noise" (unmeasurable) inputs, applied to the current state dynamics $x(k+1)$ and output $y(k)$, respectively. The output signals $\{y(k)\}_{k\geq 0}$ are available during the process, but $\{x(k)\}_{k\geq 0}$ not.

2.2 | Main assumptions

Suppose that the following assumptions hold:

1. All random sequences are defined on the probability space $(\Omega, \{\mathcal{F}_k\}_{k\geq 0}, P)$.
2. Random variables $\xi_x(k+1)$ and $\xi_y(k)$ are independent martingale-differences, namely,

$$\left.\begin{aligned} \mathrm{E}\{\xi_x(k+1)|\mathcal{F}_k\} &\overset{a.s.}{=} 0, \\ \mathrm{E}\{\xi_y(k)|\mathcal{F}_k\} &\overset{a.s.}{=} 0, \\ \mathrm{E}\{\xi_x(k+1)\xi_y^{\intercal}(k)|\mathcal{F}_k\} &\overset{a.s.}{=} 0 \end{aligned}\right\}, \tag{3}$$

with bounded conditional covariations

$$\left.\begin{aligned} \mathrm{E}\{\xi_x(k+1)\xi_x^{\intercal}(k+1)|\mathcal{F}_k\} &\overset{a.s.}{\leq} \Xi_x, \\ \mathrm{E}\{\xi_y(k)\xi_y^{\intercal}(k)|\mathcal{F}_k\} &\overset{a.s.}{\leq} \Xi_y, \end{aligned}\right\} \tag{4}$$

that is, the noise terms have "*zero-mean conditional average value* ", "*conditionally mean-value bounded covariations*" and not obligatory should have Gaussian distribution.
3. The considered plant (2) is assumed to be *mean-square BIBO-stable*, that is,

$$\left.\begin{aligned} \limsup_{k\geq 0} \mathrm{E}\{\|u(k)\|^2\} &\leq U_+, \\ \limsup_{k\geq 0} \mathrm{E}\{\|x(k)\|^2\} &\leq X_+, \end{aligned}\right\} \tag{5}$$

namely, the plant has a mean-square bounded state dynamics if the control actions are also mean-square bounded.
4. The nonlinear mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ is supposed to be *a* priory *unknown* but belonging to the class $C(A, f_0, f_1)$ of *quasi-Lipchitz functions* (see Reference 50 on AEM), satisfying

$$\|f(x(k), k) - Ax(k)\|^2 \leq f_0 + f_1\|x(k)\|^2, \tag{6}$$

globally on $\mathbb{R}^n$, which means that the right-hand side of the considered class on nonlinear models deviates from some *reference* object not more than some linear (with respect to the norma of state) upper estimate and admits to have some discontinuity ($f_0 > 0$), including in the origin such elements as hysteresis, deed zone, switching elements and etc.
5. The matrices $B \in \mathbb{R}^{n\times l}$ and $C \in \mathbb{R}^{m\times n}$ are assumed to be known such that the pair (A, C) is observable.

2.3 | Problem formulation

The problem, which we are interested in, can be formulated as follows: based on measurable sequences $\{y(k)\}_{k\geq 0}$ and $\{u(k)\}_{k\geq 0}$ to design a sequence $\{\hat{x}(k)\}_{k\geq 0}$, referred to as a sequence of *state estimates*, which in some probabilistic sense is closed to the unobservable sequence $\{x(k)\}_{k\geq 0}$, generated by (2).

3 | ROBUST OBSERVER STRUCTURE

To estimate the states $\{x(k)\}_{k\geq 0}$ of the process (2) let us use the following recursive scheme (filter):

$$\left.\begin{aligned} \hat{x}(k+1) &= A\hat{x}(k) + Bu(k) + L\sigma(k) + L_a(k)\mathrm{Sign}(\sigma(k)), \\ \sigma(k) &= y(k) - C\hat{x}(k), \end{aligned}\right\}, \tag{7}$$

which, in fact, is a Luenberger-type observer mixed with an additional discontinuous term in the right-hand side. Here

$$\mathrm{Sign}(\sigma) = (\mathrm{sign}(\sigma_1), .., \mathrm{sign}(\sigma_m))^\intercal,$$

$$\mathrm{sign}(\sigma_i) = \begin{cases} 1 & \text{if} \quad \sigma_i > 0 \\ -1 & \text{if} \quad \sigma_i < 0 \\ [-1, 1] & \text{if} \quad \sigma_i = 0 \end{cases}. \tag{8}$$

In this observer structure (7) only two matrices L and $L_a(k)$ are available for a designer to provide a small enough mean-squared value of state estimation error within the class $C(A, f_0, f_1)$ of admissible nonlinearities f. Exactly speaking, for each admissible matrix parameters L and $L_a(k)$ we wish to find an upper estimate $\varkappa(L, L_a(\cdot))$ for mean-squared estimation error, namely,

$$\lim_{k\to\infty} \sup \sup_{f\in C(A,f_0,f_1)} \mathrm{E}\{\|\hat{x}(k) - x(k)\|^2\} \leq \varkappa(L, L_a(\cdot)), \tag{9}$$

and, then, to find optimal matrix parameters L^* and $L_a^*(k)$ of the robust-observer (7) such the

$$(L^*, L_a^*(\cdot)) = \arg\min_{(L,L_a(\cdot))\in\mathcal{P}_{adm}} \varkappa(L, L_a(\cdot)), \tag{10}$$

where $\mathcal{P}_{adm}$ is the set of admissible parameters L and $L_a(k)$, for which $\varkappa(L, L_a(\cdot)) < \infty$.

4 | ESTIMATION ERROR DYNAMICS AND ATTRACTIVE ELLIPSOID METHOD

4.1 | State estimated error dynamics

Define the estimation error $e(k)$ at time k as

$$e(k) = \hat{x}(k) - x(k), \tag{11}$$

which satisfies the following recurrence

$$e(k+1) = \hat{x}(k+1) - x(k+1) = \tilde{A}(L)e(k) + L_a(k)\mathrm{Sign}(\sigma(k)) + \eta(k), \tag{12}$$

where

$$\tilde{A}(L) = A - LC, \tag{13}$$

and the "uncertain term" $\eta(k)$ is defined as follow

$$\eta(k) := L\xi_y(k) - \xi_x(k+1) - [f(x(k), k) - Ax(k)]. \tag{14}$$

4.2 | Robust attractive ellipsoid

Definition 1. The state estimation error $\{e(k)\}_{k\geq 0}$ belongs asymptotically to the **Robust (in the class** $C(A, f_0, f_1)$**) Attractive Ellipsoid** $\varepsilon(0, P_e)$

$$\varepsilon(0, P_e) := \{e(k) \in \mathbb{R}^n : e^\intercal(k)\, P_e\, e(k)\}, \tag{15}$$

in **mean-square sense**, if any admissible nonlinearity $f \in C(A, f_0, f_1)$, defining the dynamics (2), the following inequality holds

$$\limsup_{k\to\infty} \sup_{f\in C(A,f_0,f_1)} \mathrm{E}\{e^{\intercal}(k)\ P_e\ e(k)\} \leq 1. \tag{16}$$

4.3 | Recurrence analysis for "energetic function"

Theorem 1. *If for the observer (7), containing two fixed gain-matrices L and L_s, there exist a positive definite matrix $P \in \mathbb{R}^{n\times n}$, positive scalars ε, μ and $\alpha \in [0, 1)$ such that the following matrix inequality holds*

$$W_{\alpha,\varepsilon}(P, L) < 0 \in \mathbb{R}^{(2n+m)\times(2n+m)}, \tag{17}$$

with

$$L_a(k) = L_s\|\sigma(k)\|,\ \ L_s := \mu(\tilde{A}(L)^{\intercal}P)^{-1}C^{\intercal}, \tag{18}$$

and

$$W_{\alpha,\varepsilon,\mu}(P, L) \triangleq \begin{bmatrix} \tilde{A}(L)^{\intercal}P\tilde{A}(L) - \alpha P - 2\mu I_{n\times n} & 0 & \tilde{A}(L)^{\intercal}P \\ 0 & L_s^{\intercal}PL_s - \varepsilon I_{m\times m} & L_s^{\intercal}P \\ P\tilde{A}(L) & PL_s & P - \varepsilon I_{n\times n} \end{bmatrix}, \tag{19}$$

then for the quadratic "energetic" function

$$V(k) = e^{\intercal}(k)Pe(k), \tag{20}$$

the following inequality holds

$$\mathrm{E}\{V(k+1)\} \leq \alpha\mathrm{E}\{V(k)\} + \varepsilon\beta, \tag{21}$$

implying

$$\limsup_{k\to\infty} \sup_{f\in C(A,f_0,f_1)} \mathrm{E}\{V(k)\} \leq \frac{\beta}{1-\alpha}, \tag{22}$$

where

$$\beta(L) := (m + d_0 + d_1X_+ + \mathrm{tr}[L^{\intercal}\Xi_yL] + \mathrm{tr}[\Xi_x]). \tag{23}$$

The proof of this theorem can be found in Appendix A.

Corollary 1. *From (22) it follows directly that*

$$\limsup_{k\to\infty} \sup_{f\in C(A,f_0,f_1)} \mathrm{E}\{e^{\intercal}(k)\ P_e\ e(k)\} \leq 1,$$

with the attractive ellipsoid matrix P_e equal to

$$P_e = \frac{P}{\phi(\varepsilon, \alpha, \mu)}. \tag{24}$$

where

$$\phi(L,\alpha) = \frac{\beta(L)}{1-\alpha}. \tag{25}$$

This means that all trajectory errors $\{e(k)\}_{k\geq 0}$ converge asymptotically in mean-square sense to the described ellipsoid for all nonlinear plants from the given class $C(A, f_0, f_1)$.

5 | OPTIMAL GAIN MATRICES SELECTION

5.1 | Optimization problem

Considering (19), if one wishes to reduce the estimation error $e(k)$ it makes sense to *maximize* $\frac{P}{(L,\alpha)}$ with respect to P, L and the others scalar parameters $\alpha \in [0,1)$, $\varepsilon > 0$, $\mu > 0$. It is suggested, as in Reference 50, to find the matrix gain by solving the following optimization problem

$$\frac{\mathrm{tr}(P)}{\phi(L,\alpha)} \to \sup_{P>0, L, \alpha\in[0,1), \varepsilon>0, \mu>0}, \tag{26}$$

subject to the matrix constraints

$$W_{\alpha,\varepsilon}(P,L) < 0. \tag{27}$$

The aim now is to find the observation matrix gain L which provides a *good enough* estimation of the system states for a wide class of nonlinear systems in $C(f_0, f_1, A)$.

5.2 | Transformation of the optimization problem into one with LMI constrains

To transform the matrix (19) such that we deal only with linear ones let us use the Schur's complement and the definitions

$$X := P, Y := PL, \tag{28}$$

and represent $W_{\alpha,\varepsilon,\mu}(P,L)$ in the form

$$W_{\alpha,\varepsilon,\mu}(P,L) := \begin{bmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{bmatrix}. \tag{29}$$

(a) First, start with

$$W_{11} = [A - LC]^\intercal P [A - LC] - \alpha P - 2\mu I_{n\times n}. \tag{30}$$

Expanding the first term in (30) results

$$\begin{aligned}
&[A - LC]^\intercal P[A - LC] = \\
&A^\intercal PA - C^\intercal (L^\intercal P)A - A^\intercal (PL)C + (PLC)^\intercal P^{-1} PLC = \\
&A^\intercal XA - C^\intercal Y^\intercal A - A^\intercal YC + C^\intercal Y^\intercal X^{-1} YC < \\
&A^\intercal XA - C^\intercal Y^\intercal A - A^\intercal YC + Q_Y,
\end{aligned} \tag{31}$$

with Q_Y satisfying

$$C^{\intercal}Y^{\intercal}X^{-1}YC < Q_Y, \tag{32}$$

which, by Schur's complement, can be expressed as

$$\begin{bmatrix} Q_Y & C^{\intercal}Y^{\intercal} \\ YC & X \end{bmatrix} > 0. \tag{33}$$

So, finally, we have

$$W_{11} < A^{\intercal}XA - C^{\intercal}Y^{\intercal}A - A^{\intercal}YC + Q_Y - \alpha P - 2\mu I_{n\times n}, \tag{34}$$

where Q_Y satisfies (33).

(b) Now consider the term

$$\begin{aligned}
&W_{22} = L_s^{\intercal}PL_s - \varepsilon I_{m\times m} = \\
&\mu^2 C(P^{-1}[\tilde{A}(L)^{\intercal}]^{-1})^{\intercal}P(\tilde{A}(L)^{\intercal}P)^{-1}C^{\intercal} - \varepsilon I_{m\times m} = \\
&\mu^2 C[\tilde{A}^{-1}(L)]P^{-1}[\tilde{A}^{-1}(L)]^{\intercal}C^{\intercal} - \varepsilon I_{m\times m} = \\
&\mu^2 C([A-LC]^{\intercal}P[A-LC])^{-1}C^{\intercal} - \varepsilon I_{m\times m} = \\
&\mu^2 C(A^{\intercal}XA - A^{\intercal}YC - C^{\intercal}Y^{\intercal}A + C^{\intercal}Y^{\intercal}X^{-1}YC)^{-1}C^{\intercal} - \varepsilon I_{m\times m}.
\end{aligned} \tag{35}$$

Suppose that there exists matrix $Q_{Ls} > 0$ satisfying

$$W_{22} < -Q_{Ls}. \tag{36}$$

Again applying the Schur's complement we may conclude that the matrix inequality is equivalent the following one

$$\begin{bmatrix} -Q_{Ls} + \varepsilon I_{m\times m} & \mu C \\ \mu C^{\intercal} & A^{\intercal}XA - A^{\intercal}YC - C^{\intercal}Y^{\intercal}A + \gamma I_{n\times n} \end{bmatrix} > 0, \tag{37}$$

where γ satisfies

$$C^{\intercal}Y^{\intercal}YC > \frac{\varepsilon\gamma}{2} I_{n\times n}. \tag{38}$$

due to inequality $X < \frac{\varepsilon}{2} I_{n\times n}$.

(c) In view of the following inequality

$$\begin{aligned}
W_{2\div 3} := \begin{bmatrix} W_{22} & W_{23} \\ W_{32} & W_{33} \end{bmatrix} &< \begin{bmatrix} -Q_{Ls} & L_s^{\intercal}X \\ XL_s & X - \varepsilon I_{n\times n} \end{bmatrix} \\
&< \begin{bmatrix} -Q_{Ls} + \varepsilon I_{m\times m} & 0 \\ 0 & X - \frac{\varepsilon}{2} I_{n\times n} \end{bmatrix} < 0,
\end{aligned} \tag{39}$$

the block matrix $W_{2\div 3}$ can be replaced by the right-hand block matrix in (19) preserving inequality. This implies

$$\begin{aligned}
&\begin{bmatrix} -Q_{Ls} & L_s^{\intercal}X \\ XL_s & X - \varepsilon I_{n\times n} \end{bmatrix} - \begin{bmatrix} -Q_{Ls} + \varepsilon I_{m\times m} & 0 \\ 0 & X - \frac{\varepsilon}{2} I_{n\times n} \end{bmatrix} \\
&= \begin{bmatrix} -\varepsilon I_{m\times m} & L_s^{\intercal}X \\ XL_s & -\frac{\varepsilon}{2} I_{n\times n} \end{bmatrix} < 0,
\end{aligned}$$

and again by the Schur's complement result

$$\varepsilon I_{m\times m} - L_s^{\mathsf{T}} X \left(\frac{2}{\varepsilon} I_{n\times n} \right) X L_s > 0.$$

Taking into account that $X - \frac{\varepsilon}{2} I_{n\times n} < 0$, the last inequality is equivalent to

$$\begin{bmatrix} \varepsilon I_{m\times m} & \mu C \\ \mu C^{\mathsf{T}} & A^{\mathsf{T}} X A - A^{\mathsf{T}} Y C - C^{\mathsf{T}} Y^{\mathsf{T}} A + \gamma I_{n\times n} \end{bmatrix} > 0, \tag{40}$$

with γ satisfying (38).

(d) Finally, noticing that the term $\mathrm{tr}[L^{\mathsf{T}} \Xi_y L]$ in (23) can be estimated by a number $\theta > 0$ as

$$\mathrm{tr}[L^{\mathsf{T}} \Xi_y L] < \theta, \tag{41}$$

we may confirm that

$$\beta(L) < \hat{\beta} := (m + d_0 + d_1 X_+ + \theta + \mathrm{tr}[\Xi_x]),$$

which in view of (41) implies

$$\mathrm{tr}[L^{\mathsf{T}} \Xi_y L] = \mathrm{tr}[Y^{\mathsf{T}} (X^{-1} \Xi_y X^{-1}) Y] < \theta.$$

Now, taking into account that scalar θ can be represented as

$$\theta = \mathrm{tr} \left[\frac{\theta}{m} I_{m\times m} \right],$$

the inequality

$$\mathrm{tr}[Y^{\mathsf{T}} (X^{-1} \Xi_y X^{-1}) Y] < \theta$$

is equivalent to

$$0 < \mathrm{tr} \left[\frac{\theta}{m} I_{m\times m} \right] - \mathrm{tr}[Y^{\mathsf{T}} (X^{-1} \Xi_y X^{-1}) Y].$$

Using the linearity property of the operator $\mathrm{tr}(\cdot)$ results to

$$\mathrm{tr} \left[\frac{\theta}{m} I_{m\times m} - Y^{\mathsf{T}} (X^{-1} \Xi_y X^{-1}) Y \right] > 0.$$

Again, by the Schur's complement the last inequality holds if

$$\begin{bmatrix} \frac{\theta}{m} I_{m\times m} & \frac{2}{\varepsilon} Y^{\mathsf{T}} \\ \frac{2}{\varepsilon} Y & \Xi_y^{-1} \end{bmatrix} > 0. \tag{42}$$

Notice that for the fixed scalar parameters α, ε, μ, θ and after some equivalent transformations the matrix inequality (19) can be estimated from above by LMI (43) with the symmetric extra restriction (38):

$$\begin{aligned} & W_{\alpha,\varepsilon,\mu}(P, L) < W(P, L) \\ & := \begin{bmatrix} A^{\mathsf{T}} X A - C^{\mathsf{T}} Y^{\mathsf{T}} A - A^{\mathsf{T}} Y C + Q_Y - \alpha P - 2\mu I_{n\times n} & 0 & A^{\mathsf{T}} X - C^{\mathsf{T}} Y^{\mathsf{T}} \\ 0 & -Q_{Ls} + \varepsilon I_{m\times m} & 0 \\ XA - YC & 0 & X - \frac{\varepsilon}{2} I_{n\times n} \end{bmatrix}, \end{aligned} \tag{43}$$

They can be solved using the MATLAB toolboxes LMItoolbox, SeDuMi, and Yalmip. Our main optimization problem can be also solved using the following two-step procedure:

First: We fix the scalar parameters α, ε, μ, and θ and solve our problem with respect to the matrix variables which satisfy LMI-constraints.

Second: For the found matrix variables Y and X we solve our optimization problem only with respect to scalar parameters α, ε, μ, and θ.

Iterating this process we finally find the solution α^*, ε^*, μ^*, θ^* and $X^* = P^*$, $Y^* = P^*L$ for which the optimal gain matrices

$$L^* = (X^*)^{-1}Y^* \text{ and } L_s^* = \mu^*(A^\intercal X - C^\intercal Y^\intercal)^{-1}C^\intercal,$$

can be found.

6 | NUMERICAL EXAMPLES

6.1 | Illustrative example

Consider the dynamic system which dynamics is governed by the following recurrent equation

$$\begin{aligned} x(k+1) &= \begin{bmatrix} \frac{x_2(k)}{1+x_2(k)} \\ x_1(k) \end{bmatrix} + \begin{bmatrix} 0.2 \\ 1 \end{bmatrix} u_k + \xi_x(k+1), \\ y(k) &= \begin{bmatrix} 0 & 1 \end{bmatrix} x(k) + \xi_y(k). \end{aligned} \tag{44}$$

In the quasi-linear format it may be represented as

$$\begin{aligned} x(k+1) &= Ax(k) + Bu_k + \hat{\xi}_x(k+1), \\ y(k) &= Cx(k) + \xi_y(k+1), \end{aligned}$$

where

$$\hat{\xi}_x(k+1) = \xi_x(k+1) + f(x(k), k) - Ax(k),$$

and

$$f(x(k), k) = \begin{bmatrix} \frac{x_2(k)}{1+x_2(k)} \\ x_1(k) \end{bmatrix}, \quad A = \begin{bmatrix} 0.1 & 0.5 \\ -0.5 & 0.1 \end{bmatrix}.$$

It is easy to see that $\|f(x(k), k) - Ax(k)\|^2 \le f_0 + f_1\|x(k)\|^2$ with $f_0 = 0.5^2$ and $f_1 = 1.5^2$. The restrictions of the covariance matrices are given by

$$\Xi_y = 0.1 I_m, \Xi_x = 0.1 I_n.$$

Notice that the pair (A, C) is observable. Then, the optimization process, described in the previous section, can be apply, resulting

$$\alpha^* = 0.1989, \quad \mu^* = 0.0001, \quad \varepsilon^* = 0.00021, \quad \theta^* = 1. \tag{45}$$

The matrix gains L y L_s of the observer are:

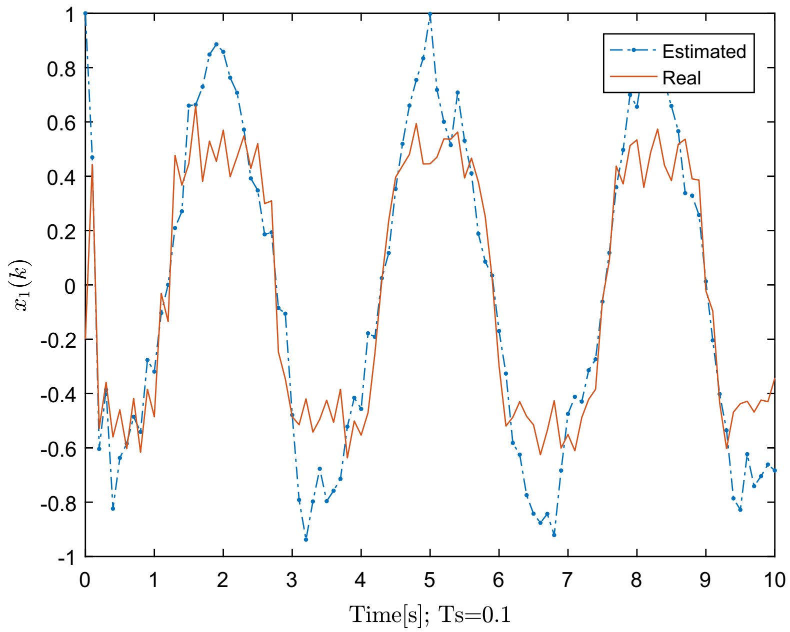

FIGURE 1 The trajectories of the state x_1 and its estimate $\hat{x}_1$ [Colour figure can be viewed at wileyonlinelibrary.com]

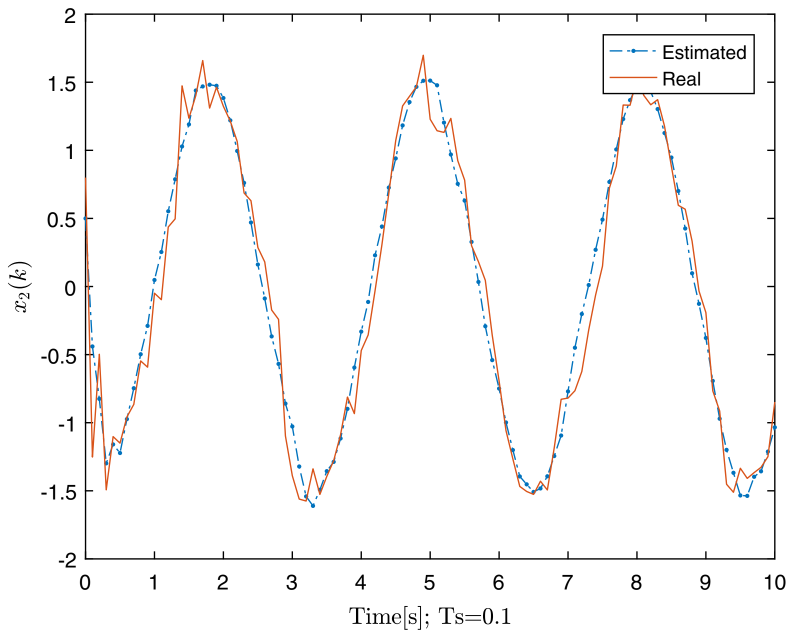

FIGURE 2 The trajectories of the state x_2 and its estimate $\hat{x}_2$ [Colour figure can be viewed at wileyonlinelibrary.com]

$$L^* = \begin{bmatrix} 0.2393 \\ 0.0836 \end{bmatrix}, \quad L_s^* = \begin{bmatrix} 0.2427 \\ -0.0436 \end{bmatrix}. \tag{46}$$

and the matrix P_e^*, defining the attraction ellipsoid, is:

$$P_e^* = \begin{bmatrix} 1.2659 & -0.0344 \\ -0.0344 & 0.9707 \end{bmatrix}. \tag{47}$$

Figures 1 and 2 shows the real states and their estimates. The error convergence into the found invariant ellipsoid is shown in Figure 3.

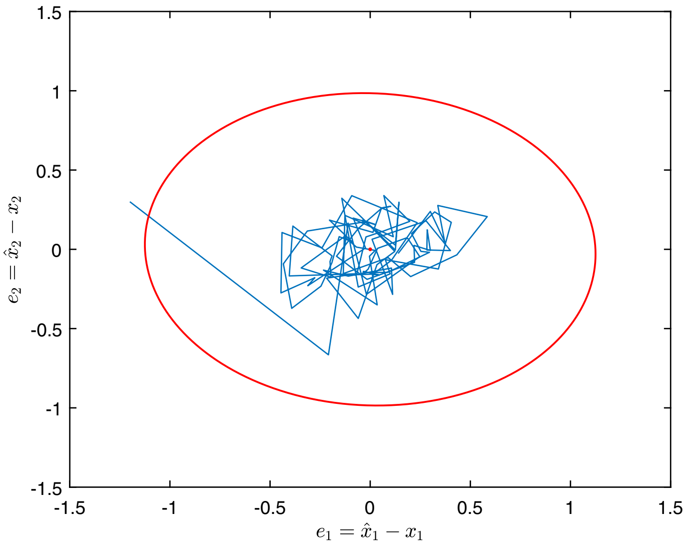

FIGURE 3 The invariant ellipsoid for the errors [Colour figure can be viewed at wileyonlinelibrary.com]

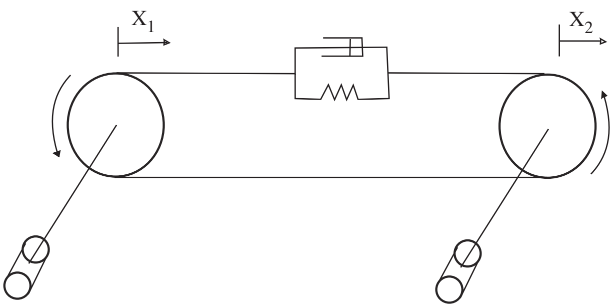

FIGURE 4 Magnetic tape drive system

6.2 | Application example

In Reference 46 the following model of magnetic tape drive, Figure 4, was considered:

$$x(k+1) = \begin{bmatrix} 0.599 & 0.0401 & -0.4861 & 0.0139 \\ 0.0401 & 0.7599 & -0.0139 & 0.4861 \\ 0.1566 & -0.1566 & 0.8321 & -0.0679 \\ 0.1566 & -0.1566 & 0.0679 & 0.8321 \end{bmatrix} x(k) + \begin{bmatrix} -0.1049 & 0.0017 \\ -0.0017 & 0.1049 \\ 0.4148 & -0.0118 \\ -0.0118 & 0.4148 \end{bmatrix} u_k + \xi_x(k+1),$$

$$y(k) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} x(k) + \xi_y(k). \tag{48}$$

Here the state vector has four components, that is,

$$x(k) = [x_1(k) \;\; x_2(k) \;\; x_3(k) \;\; x_4(k)]^\intercal,$$

where $x_1(k)$, $x_2(k)$ represents positions and $x_3(k)$ and $x_4(k)$ are angular velocities.

Here we have considered $\Xi_y = 0.1I_m$ and $\Xi_x = 0.1I_n$. The optimization process give us

$$\alpha^* = 0.5, \quad \mu^* = 0.00021, \quad \varepsilon^* = 0.001, \quad \theta^* = 0.1.$$

The robust optimal attractive ellipsoid matrix $P^*_{attr} = P^*_e$ is defined by

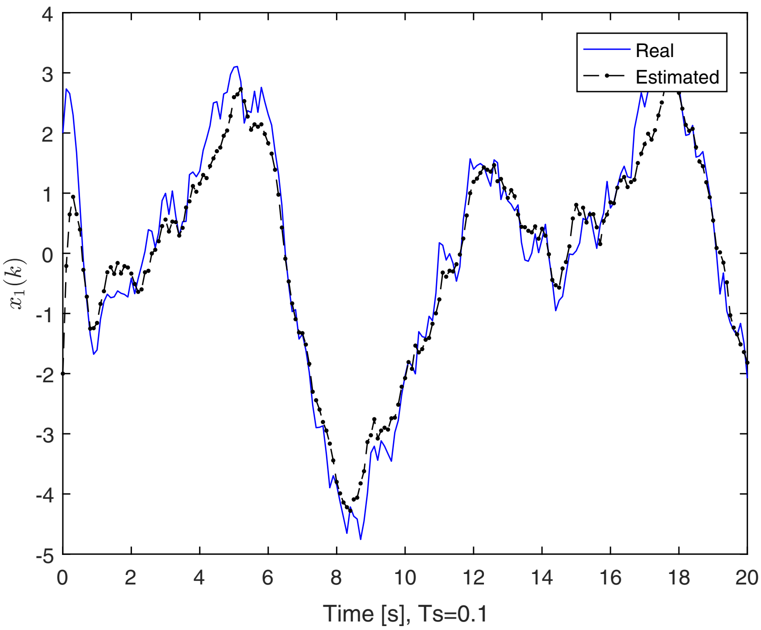

FIGURE 5 The trajectories of the state x_1 and its estimate $\hat{x}_1$ [Colour figure can be viewed at wileyonlinelibrary.com]

$$P_e^* = \begin{bmatrix} 4.1062 & -0.6738 & -1.6197 & 0.3593 \\ -0.6738 & 4.9032 & -0.8041 & 1.1857 \\ -1.6197 & -0.8041 & 4.6222 & 0.8927 \\ 0.3593 & 1.1857 & 0.8927 & 4.9609 \end{bmatrix}. \tag{49}$$

The matrix parameters L^* and L_s^* (18) are

$$L^* = \begin{bmatrix} -0.0574 & -0.2635 & 0.0177 \\ 0.3022 & 0.0009 & 0.2259 \\ -0.0976 & 0.3519 & 0.0133 \\ -0.0697 & 0.0630 & 0.3672 \end{bmatrix}; \; L_s^* = \begin{bmatrix} -0.0351 & -0.1319 & -0.0281 \\ 0.2003 & 0.0204 & 0.1169 \\ -0.0527 & 0.2245 & 0.0259 \\ -0.0700 & 0.0595 & 0.2302 \end{bmatrix}. \tag{50}$$

Figures 5 to 8 represent a comparison between the real signal and the estimations by the suggested observer (with additional sliding term purposed in this paper). Figures 9 and 10 show that the trajectories of tracking errors asymptotically converge to their corresponding centered ellipsoids.

7 | CONCLUSIONS

The paper

- proposes the observer to realize the current state estimation for a wide class of quasi-Lipschitz systems with input and output noises which not obligatory be Gaussian;
- the robust analysis, performed by attractive ellipsoids technique, gives us an area where the state error converges asymptotically;
- the algorithm design for the "best" selection of the observer parameters is based on attractive ellipsoid method and the solution of a matrix optimization problem subject some non-linear matrix constrains.
- the coordinate transformation method of the nonlinear matrix constrains to LMI's is presented.
- two numerical examples illustrate a "good" workability of the suggested approach.

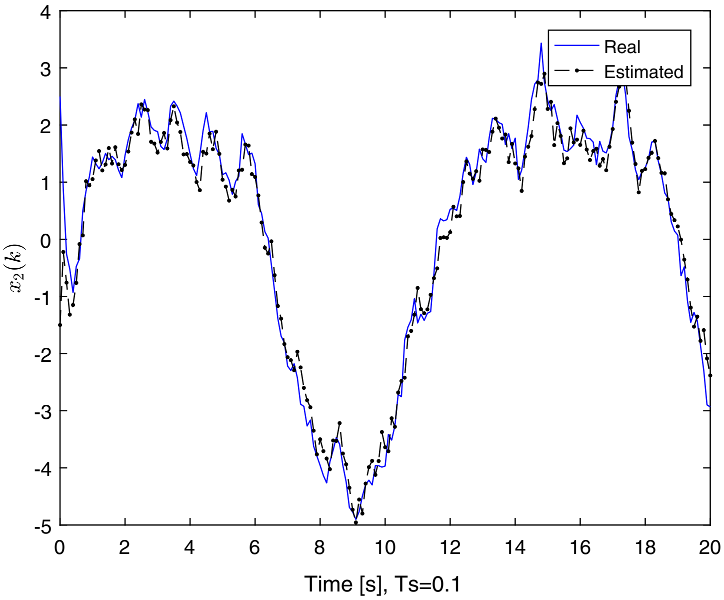

FIGURE 6 The trajectories of the state x_2 and its estimate $\hat{x}_2$ [Colour figure can be viewed at wileyonlinelibrary.com]

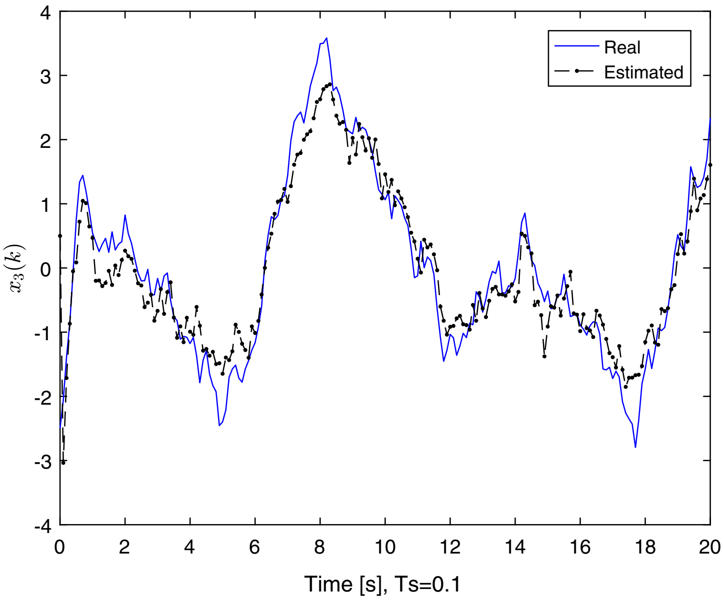

FIGURE 7 The trajectories of the state $x_3$and its estimate $\hat{x}_3$ [Colour figure can be viewed at wileyonlinelibrary.com]

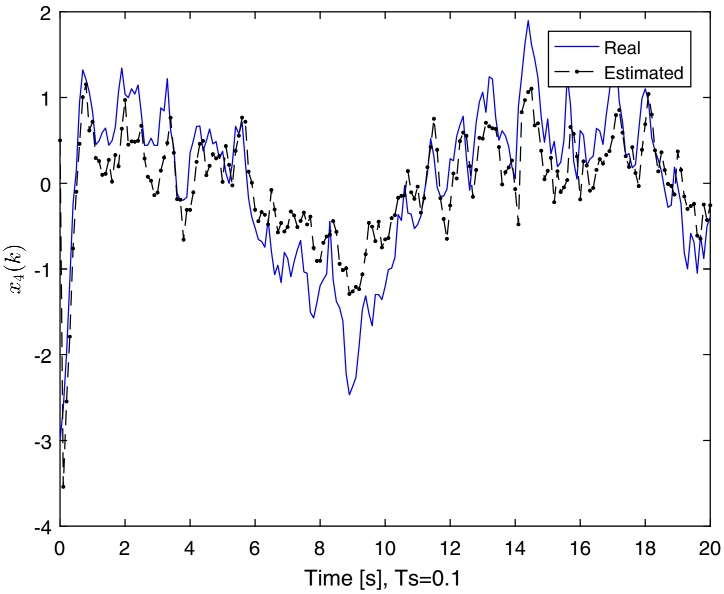

FIGURE 8 The trajectories of the state $x_4$and its estimate $\hat{x}_4$ [Colour figure can be viewed at wileyonlinelibrary.com]

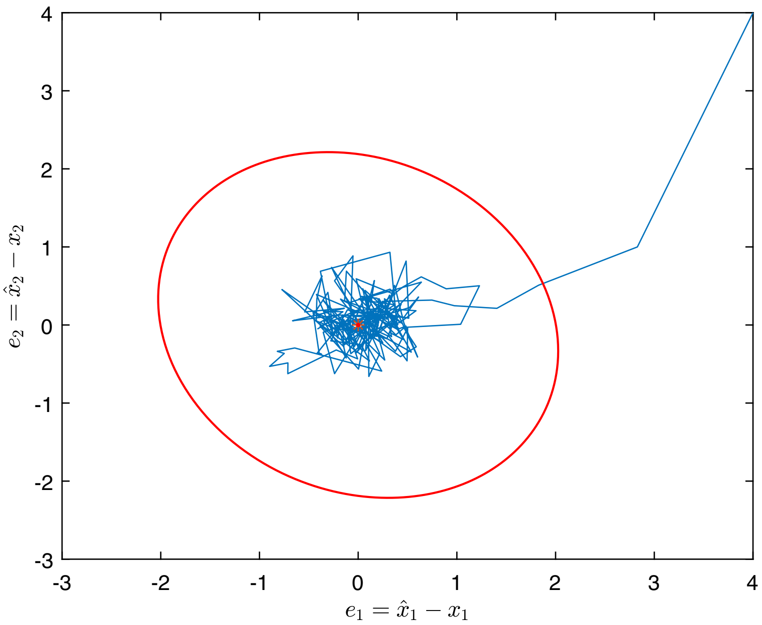

FIGURE 9 The invariant ellipsoid for the tracking errors e_1 and e_2 [Colour figure can be viewed at wileyonlinelibrary.com]

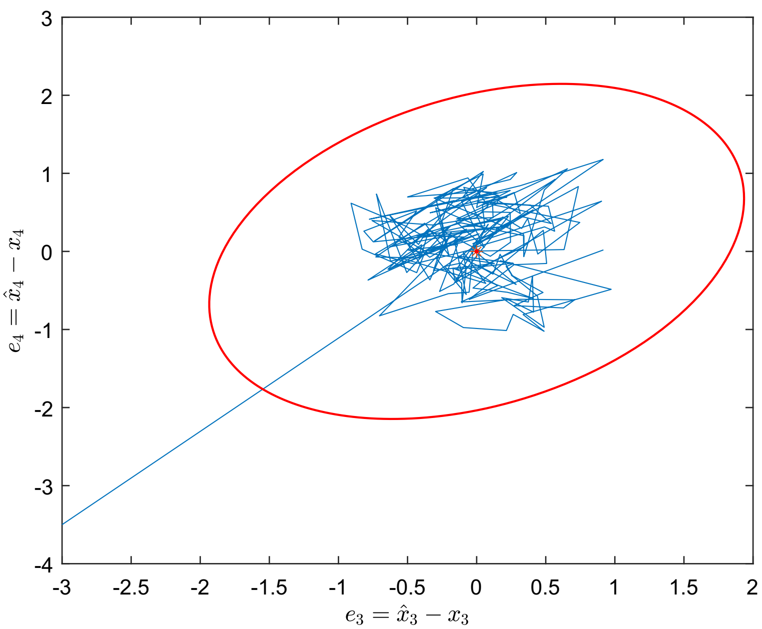

FIGURE 10 The invariant ellipsoid for the tracking errors e_3 and e_4 [Colour figure can be viewed at wileyonlinelibrary.com]

Future research may be related with the designing of SM Controller for discrete-time stochastic systems based on the state estimates in the feedback control, using the stochastic observer suggested in this paper.

ORCID

David Velázquez https://orcid.org/0000-0002-4712-8410

REFERENCES

1. Luenberger DG. Observing the state of a linear system. *IEEE Trans Military Electron*. 1964;8(2):74-80.
2. Luenberger D. Observers for multivariable systems. *IEEE Trans Autom Control*. 1966;11(2):190-197.
3. Bongiarno Y. On observer in multivariable control systems. *Int J Control*. 1968;8(3):221-243.
4. Kalman RE. A new approach to linear filtering and prediction problems. *J Basic Eng*. 1960;82(1):35-45.
5. Kalman RE, Bucy RS. New results in linear filtering and prediction theory. *J Basic Eng*. 1961;83(1):95-108.
6. Tse E, Athans M. Optimal minimal-order observer-estimators for discrete linear time-varying systems. *IEEE Trans Autom Control*. 1970;15(4):416-426.

7. Stratonovich R L'e. Conditional Markov processes. *Non-linear Transformations of Stochastic Processes*. Amsterdam, Netherlands: Elsevier; 1965:427-453.
8. Kushner HJ. On the differential equations satisfied by conditional probablitity densities of Markov processes, with applications. *J Soc Ind Appl Math Ser A Control*. 1964;2(1):106-119.
9. Wonham WM. Some applications of stochastic differential equations to optimal nonlinear filtering. *J Soc Ind Appl Math Ser A Control*. 1964;2(3):347-369.
10. Bucy RS. Nonlinear filtering theory. *IEEE Trans Autom Control*. 1965;10(2):198-198.
11. Bucy RS, Joseph PD. *Filtering for Stochastic Processes with Applications to Guidance*. Vol 326. New York, NY: American Mathematical Society; 2005.
12. Frost P, Kailath T. An innovations approach to least-squares estimation-Part III: nonlinear estimation in white Gaussian noise. *IEEE Trans Autom Control*. 1971;AC-16:217-226.
13. Hecht C. Digital realization of nonlinear filters. Paper presented at: Proceedings of the 2nd Symposium on Nonlinear Estimation Theory and Its Application, San Diego, CA; 1971:152-158.
14. Bucy RS, Senne KD. Digital synthesis of non-linear filters. *Automatica*. 1971;7(3):287-298.
15. Pearson JB. On nonlinear least-squares filtering. *Automatica*. 1967;4(3):97-105.
16. Jazwinski AH. *Stochastic Processes and Filtering Theory*. Dunkirk, Maryland: Courier Corporation; 2007.
17. Fisher JR. Optimal nonlinear filtering. *Advances in Control Systems*. Amsterdam, Netherlands: Elsevier; 1967.
18. Srinivasan K. State estimation by orthogonal expansion of probability distributions. *IEEE Trans Autom Control*. 1970;15(1):3-10.
19. Kizner W. Optimal nonlinear estimation based on orthogonal expansions. *Optimal Nonlinear Estimation Based on Orthogonal JPL Technical Report; 32-1366*. Pasadena, CA: JET Propulsion Laboratory, California Institute of Technology; 1969.
20. Alspach DL, Sorenson HW. Approximation of density functions by a sum of Gaussians for nonlinear Bayesian estimation. Proceedings of the Symposium Nonlinear Estimation Theory and Its Application; 1970.
21. Bucy Richard S, Senne Kenneth D. *Realization of Optimum Discrete-Time Nonlinear Estimators*. Los Angeles: University of Southern California, Department of Aerospace Engineering; 1971.
22. Tarn T-J, Rasis Y. Observers for nonlinear stochastic systems. *IEEE Trans Autom Control*. 1976;21(4):441-448.
23. Utkin VI. *Sliding Modes in Control and Optimization*. Berlin, Germany: Springer Science & Business Media; 2013.
24. Edwards C, Spurgeon S. On the development of discontinuous observers. *Int J Control*. 1994;59(5):1211-1229.
25. Shtessel Y, Edwards C, Fridman L, Levant A. *Sliding Mode Control and Observation*. Control Engineering Series. New York, NY: Birkhauser; 2014.
26. Yang J, Zhu F, Wang X, Bu X. Robust sliding-mode observer-based sensor fault estimation, actuator fault detection and isolation for uncertain nonlinear systems. *Int J Control Autom Syst*. 2015;13(5):1037-1046.
27. Bartoszewicz A. Discrete-time quasi-sliding-mode control strategies. *IEEE Trans Ind Electron*. 1998;45(4):633-637.
28. Bejarano FJ, Fridman L, Poznyak A. Exact state estimation for linear systems with unknown inputs based on hierarchical super-twisting algorithm. *Int J Robust Nonlinear Control IFAC-Affiliat J*. 2007;17(18):1734-1753.
29. Moreno JA, Osorio M. A Lyapunov approach to second-order sliding mode controllers and observers. Paper presented at: Proceedings of the 2008 47th IEEE Conference on Decision and Control. Cancún, México: IEEE; 2008:2856-2861.
30. Davila J, Fridman L, Poznyak A. Observation and identification of mechanical systems via second order sliding modes. *Int J Control*. 2006;79(10):1251-1262.
31. Abidi K, Xu JX. *Advanced Discrete-Time Control, Studies in Systems, Decision and Control*. Berlin, Germany: *Springer Science and Business Media*; 2015.
32. Gao W, Wang Y, Homaifa A. Discrete-time variable structure control systems. *IEEE Trans Ind Electron*. 1995;42(2):117-122.
33. Golo G, Milosavljevic C. Robust discrete-time chattering free sliding mode control. *Syst Control Lett Electron*. 2000;41(1):19-28.
34. Chakrabarty S, Bartoszewicz A. Improved robustness and performance of discrete time sliding mode control systems. *ISA Trans*. 2016;65:143-149.
35. Su WC, Drakunov SV, Ozguner U. An O(T2) boundary layer in sliding mode for sampled-data systems. *Transactions on Automatic Control*. 2000;45(3):482-485.
36. Utkin V. Chapter 5 Sliding mode control in discrete-time and difference systems. *Variable Structure and Lyapunov Control*. Vol 193. Berlin, Heidelberg/Germany: Springer; 1994:87-107.
37. Grobov Igor D. Input disturbance estimation for linear systems in the bounded noise context. *Asian J Control*. 2001;3(4):300-308.
38. Koshkouei Ali J, Zinober Alan SI. Sliding mode state observation for non-linear systems. *Int J Control*. 2004;77(2):118-127.
39. Wu L, Shi P, Gao H. State estimation and sliding-mode control of Markovian jump singular systems. *IEEE Trans Autom Control*. 2010;55(5):1213-1219.
40. Abidi K, Xu J-X, Xinghuo Y. On the discrete-time integral sliding-mode control. *IEEE Trans Autom Control*. 2007;52(4):709-715.
41. Wu L, Gao H. Dissipativity-based sliding mode control of switched stochastic systems. *IEEE Trans Autom Control*. 2013;1(1):1.
42. Kailath T. *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall; 1980.
43. Singh S, Janardhanan S. Sliding mode control-based linear functional observers for discrete-time stochastic systems. *Int J Syst Sci*. 2017;48(15):3246-3253.
44. Bensoubaya M, Ferfera A, Iggidr A. Stabilization of nonlinear stochastic discrete systems. Paper presented at: Proceedings of the 38th IEEE Conference on Decision and Control (Cat. No. 99CH36304). Phoenix, Arizona: IEEE; 1999:3180-3181.

45. Janardhanan S, Bandyopadhyay B. Multirate output feedback based robust quasi-sliding mode control of discrete-time systems. *IEEE Trans Autom Control*. 2007;52(3):499-503.
46. Bandyopadhyay B, Janardhanan S. *Discrete-Time Sliding Mode Control: A Multirate Output Feedback Approach*. Berlin, Germany: Springer Science & Business Media; 2005.
47. Alazki H, Poznyak A. Constraint robust stochastic discrete-time tracking: attractive ellipsoids technique. Paper presented at: Proceedings of the 2010 7th International Conference on Electrical Engineering Computing Science and Automatic Control. Tuxtla Gutierrez, México: IEEE; 2010:99-104.
48. Wu Z-H, Guo B-Z. Extended state observer for MIMO nonlinear systems with stochastic uncertainties. *Int J Control*. 2020;93(3):424-436.
49. Poznyak A, Nazin A, Murano D. Observer matrix gain optimization for stochastic continuous time nonlinear systems. *Syst Control Lett*. 2004;52(5):377-385.
50. Poznyak A, Polyakov A, Azhmyakov V. *Attractive Ellipsoids in Robust Control*. Berlin, Germany: Springer International Publishing; 2014.

How to cite this article: Velázquez D, Poznyak A. Attractive ellipsoid design for robust sliding-mode observation error in stochastic nonlinear discrete-time systems. *Int J Robust Nonlinear Control*. 2021;31:69–86. https://doi.org/10.1002/rnc.5270

APPENDIX A. PROOF OF THEOREM 1

Proof. For the function

$$V(k) = e^{\intercal}(k)Pe(k), \tag{A1}$$

in view of (12) and using the identity

$$Ce(k) = -\sigma(k) + \xi_y(k),$$

we have

$$\left.\begin{array}{c} V(k+1) = e(k+1)Pe(k+1) = \\ e^{\intercal}(k)\tilde{A}(L)^{\intercal}P\tilde{A}(L)e(k) + \|\sigma(k)\|\mathrm{Sign}^{\intercal}(\sigma(k))L_s^{\intercal}PL_s\mathrm{Sign}(\sigma(k))\|\sigma(k)\|+ \\ \eta^{\intercal}(k)P\eta(k) + 2e^{\intercal}(k)\tilde{A}(L)^{\intercal}PL_s(k)\|\sigma(k)\|\mathrm{Sign}(\sigma(k)) + 2\mathrm{Sign}^{\intercal}(\sigma(k))\|\sigma(k)\|L_s^{\intercal}(k)P\eta(k)+ \\ 2e^{\intercal}(k)\tilde{A}(L)^{\intercal}P\eta(k) = e^{\intercal}(k)\tilde{A}(L)^{\intercal}P\tilde{A}(L)e(k) + 2\mu\|\sigma(k)\|[-\sigma(k)+\xi_y(k)]^{\intercal}\mathrm{Sign}(\sigma(k))+ \\ 2e^{\intercal}(k)\tilde{A}(L)^{\intercal}P\eta(k) + [L_s\|\sigma(k)\|\mathrm{Sign}(\sigma(k))]^{\intercal}PL_s\|\sigma(k)\|\mathrm{Sign}(\sigma(k))+ \\ +2[L_s\|\sigma(k)\|\mathrm{Sign}(\sigma(k))]^{\intercal}P\eta(k) + \eta^{\intercal}(k)P\eta(k) \end{array}\right\}. \tag{A2}$$

Then using the property

$$\sigma^{\intercal}(k)\mathrm{Sign}(\sigma(k)) = \sum_{i=1}^{n}|\sigma_i(k)| \geq \sqrt{\sum_{i=1}^{n}\sigma_i^2(k)} = \|\sigma(k)\|. \tag{A3}$$

the inequality (A2) can be represented as:

$$\begin{aligned} V(k+1) &\leq 2\mu\|\sigma(k)\|\xi_y^{\intercal}(k)\mathrm{Sign}(\sigma(k)) \\ &+ z^{\intercal}(k)\begin{bmatrix} [A-LC]^{\intercal}P[A-LC] - \alpha P - 2\mu I_{n\times n} & 0 & [A-LC]^{\intercal}P \\ 0 & L_s^{\intercal}PL_s & L_s^{\intercal}P \\ P[A-LC] & PL_s & P \end{bmatrix} z(k) \end{aligned}$$

$$= +2\mu\|\sigma(k)\|\xi_y^{\intercal}(k)\mathrm{Sign}(\sigma(k)) + z^{\intercal}(k)W_{\alpha,\varepsilon,\mu}(P,L)\ \ z(k)$$
$$+ \alpha V(k) + \varepsilon\|\mathrm{Sign}(\sigma(k))\|^2 + \varepsilon\|\eta(k)\|^2, \tag{A4}$$

where the extended vector $z(k)$ is defined as

$$z(k) := \begin{bmatrix} e^{\intercal}(k) & \|\sigma(k)\|\mathrm{Sign}^{\intercal}(\sigma(k)) & \eta^{\intercal}(k,k+1) \end{bmatrix}^{\intercal} \in \mathbb{R}^{2n+m}. \tag{A5}$$

By the property (19) $W_{\alpha,\varepsilon}(P,L) < 0$, and expanding the term $\|\eta(k)\|^2$, the right-hand side of the inequality (A4) can be estimated from above as

$$\left.\begin{array}{c} V(k+1) \leq 2\mu\|\sigma(k)\|\xi_y^{\intercal}(k)\mathrm{Sign}(\sigma(k)) + \alpha V(k) + \varepsilon\|\mathrm{Sign}(\sigma(k))\|^2 + \varepsilon\|\eta(k)\|^2 = \\ +2\mu\|\sigma(k)\|\xi_y^{\intercal}(k)\mathrm{Sign}(\sigma(k)) + \alpha V(k) + \varepsilon m + \varepsilon\|[f(x(k),k) - Ax(k)]\|^2 + \varepsilon\|L\xi_y(k) - \xi_x(k+1)\|^2 - \\ 2\varepsilon[L\xi_y(k+1) - \xi_x(k+1)]^{\intercal}[f(x(k),k) - Ax(k)] \leq \alpha V(k) + 2\mu\|\sigma(k)\|\xi_y^{\intercal}(k)\mathrm{Sign}(\sigma(k)) + w(k) + \varepsilon m \\ +\varepsilon(f_0 + f_1\|x(k)\|^2) + \varepsilon\|L\xi_y(k)\|^2 + \varepsilon\|\xi_x(k+1)\|^2, \end{array}\right\} \tag{A6}$$

where

$$w(k) := -2\varepsilon[L\xi_y(k) - \xi_x(k+1)]^{\intercal}[f(x(k),k) + Ax(k)] - 2\varepsilon\xi_y^{\intercal}(k)L^{\intercal}\xi_x(k+1) \tag{A7}$$

$$+ 2\mu\mathrm{E}\{\|\sigma(k)\|\xi_y^{\intercal}(k)\mathrm{Sign}(\sigma(k)) \mid \mathcal{F}_k\}. \tag{A8}$$

satisfies

$$\mathrm{E}\{w(k) \mid \mathcal{F}_k\} \overset{a.s.}{=} 0. \tag{A9}$$

Taking the conditional mathematical expectation of both sides of (A6) leads to

$$\left.\begin{array}{c} \mathrm{E}\{V(k+1) \mid \mathcal{F}_k\} \overset{a.s.}{\leq} \alpha\mathrm{E}\{V(k) \mid \mathcal{F}_k\} \\ \varepsilon[m + (d_0 + d_1\|x(k)\|^2)] + \\ \varepsilon\mathrm{E}\{\|L\xi_y(k+1)\|^2 \mid \mathcal{F}_k\} + \varepsilon\mathrm{E}\{\|\xi_x(k+1)\|^2 \mid \mathcal{F}_k\} \end{array}\right\}. \tag{A10}$$

Then, applying the relations

$$\mathrm{E}\{\|L\xi_y(k+1)\|^2 \mid \mathcal{F}_k\} \overset{a.s.}{=} \mathrm{tr}[L^{\intercal}L\mathrm{E}\{[\xi_y(k+1)\xi_y^{\intercal}(k+1)] \mid \mathcal{F}_k\}] \leq \mathrm{tr}\{L^{\intercal}\Xi_y L\},$$

and

$$\mathrm{E}\{\|\xi_x(k+1)\|^2 \mid \mathcal{F}_k\} = \mathrm{tr}\{\Xi_x\},$$

to the form (A10) implies

$$\left.\begin{array}{c} \mathrm{E}\{V(k+1) \mid \mathcal{F}_k\} \overset{a.s.}{\leq} \alpha V(k) + \\ \varepsilon(m + d_0 + d_1\|x(k)\|^2 + \mathrm{tr}[L^{\intercal}\Xi_y L] + \mathrm{tr}[\Xi_x]) \leq \\ \alpha V(k) - 2\mu\|\sigma(k)\| + \varepsilon(m + d_0 + d_1\|x(k)\|^2 + \mathrm{tr}[L^{\intercal}\Xi_y L] + \mathrm{tr}[\Xi_x]). \end{array}\right\} \tag{A11}$$

Finally, applying the operator $\mathrm{E}\{\cdot\}$ to both sides of (A11), we obtain (21). Theorem is proven. ∎