Active disturbance rejection control: Theoretical perspectives

BAO-ZHU GUO AND ZHI-LIANG ZHAO

This paper articulates, from a theoretical perspective, a new emerging control technology, known as active disturbance rejection control to this day. Three cornerstones toward building the foundation of active disturbance rejection control, namely, the tracking differentiator, the extended state observer, and the extended state observer based feedback are expounded separately. The paper tries to present relatively comprehensive overview about origin, idea, principle, development, possible limitations, as well as some unsolved problems for this viable PID alternative control technology.

KEY WORDS AND PHRASES: Active disturbance rejection control, disturbance, tracking differentiator, extended state observer, feedback.

MATHEMATICS SUBJECT CLASSIFICATION: 93C15, 93B52, 34D20, 93D15, 93B51.

1. Introduction

Disturbance rejection is a different paradigm in control theory since the inception of the modern control theory in the later years of 1950s, seeded in [43] where it is stated that the control operation "must not be influenced by internal and external disturbances" [43, p.228]. The tradeoff between mathematical rigor by model-based control theory and practicability by model-free engineering applications has been a constantly disputed issue in control community. On the one hand, we have mountains of papers, books, monographes published every year, and on the other hand, the control engineers are nowhere to find, given the difficulty of building (accurate) dynamic model for the system to be controlled, a simple, model free, easy tuning, better performance control technology more than proportional-integral-derivative

---

(PID) control. Actually, it is indicated in [42] (see also [5]) that "98% of the control loops in the pulp and paper industries are controlled by single-input single output PI controllers and that in process control applications, more than 95% of the controllers are of the PID type". This awkward coexistence of huge modern control theories on the one hand and a primitive control technology that has been dominating engineering applications for one century on the other pushed Jingqing Han, a control scientist at the Chinese Academy of Sciences to propose active disturbance rejection control (ADRC), as an alternative of PID. This is because PID has the advantage of model free nature whereas most parts of modern control theory are based on mathematical models. By model-based control theory, it is hard to cross the boundaries such as time variance, nonlinearity, and uncertainty created mainly by the limitations of mathematics. However, there are four basic limitations for PID in practice to accommodate the liability in the digital processors according to [34]: a) Setpoint is often given as a step function, not appropriate for most dynamics systems; b) PID is often implemented without the D part because of the noise sensitivity; c) weighted sum of the three terms in PID may not be the best control law based on the current and the past of the error and its rate of change; d) The integral term introduces problems such as saturation and reduced stability margin.

To address this problem, Han would seek solution from the seed idea of disturbance rejection imbedded in [43]. Consider stabilization for the following second order Newton system:

(1.1)
$$\begin{cases} \dot{x}_1(t) = x_2(t), \\ \dot{x}_2(t) = f(x_1(t), x_2(t), d(t), t) + u(t), \\ y(t) = x_1(t), \end{cases}$$

where $u(t)$ is the control input, $y(t)$ is the measured output, $d(t)$ is the external disturbance, and $f(\cdot)$ is an unknown function which contains unmodelled dynamics of the system or most possibly, the internal and external disturbance discussed in [43]. Introducing ADRC is by no means saying that there is no way other than ADRC to deal with system (1.1). For simplicity, we just take state feedback instead of output feedback control to stabilize system (1.1) by sliding model control (SMC), a powerful robust control in modern control theory. Suppose that $|f(\cdot)| \leq M$ which is standard because in any sense $f(\cdot)$ represents somehow a "perturbation", otherwise, any stabilizing controller needs infinite gain which is not realistic. In SMC, we need first choose a manifold called "sliding surface" which is chosen here

as $S = \{(x_1, x_2) \in \mathbb{R}^2 | ax_1 + x_2 = 0\}, a > 0$, a closed subspace in the state space $\mathbb{R}^2$. On the sliding surface, since $\dot{x}_1(t) = x_2(t)$, we have

$$\dot{x}_1(t) = -ax_1(t) \Rightarrow x_1(t) \to 0 \text{ and hence } x_2(t) = -ax_1(t) \to 0 \text{ as } t \to \infty,$$

where $a > 0$ is seen to regulate the convergence speed on the sliding surface: the larger a is, the faster the convergence. Design the state feedback as

$$(1.2) \quad u(t) = -ax_2(t) - K\text{sign}(s(t)), \quad s(t) = ax_1(t) + x_2(t), \quad K > M,$$

where $s(t) = ax_1(t) + x_2(t)$ is, by abuse of terminology, also the sliding surface. As a result, $s(t)$ satisfies the "finite reaching condition":

$$\dot{s}(t)s(t) \leq -(K - M)|s(t)|,$$

and therefore $|s(t)| \leq |s(0)| - (K - M)t$. This means that $s(t) = 0$ for some $t > t_0$ where t_0 depends on the initial condition.

At the moment, everything proceeds beautifully: The control law (1.2) meets the standard of mathematical "beauty" where the control is so robust and the convergence proof is a piece of cake even for beginners. However, subject the control law (1.2) to scrutiny, we find that the control gain K must satisfy $K > M$, that is, the control law (1.2) focuses on the worst case scenario which makes the controller design rather conservative.

Now it is Han who came to the scene. Han just let $a(t) = f(x_1(t), x_2(t), d(t), t)$ and system (1.1) becomes

$$(1.3) \qquad \begin{cases} \dot{x}_1(t) = x_2(t), \\ \dot{x}_2(t) = a(t) + u(t), \\ y(t) = x_1(t). \end{cases}$$

A flash of insight arises ([29]): system (1.3) is exactly observable because it is trivially seen that $(y(t), u(t)) \equiv 0, t \in [0, T] \Rightarrow a(t) = 0, t \in [0, T]; (x_1(0), x_2(0)) = 0$ (see, e.g., [6, p.5, Definition 1.2]). This means that $y(t)$ contains all information of $a(t)$! *Why not use $y(t)$ to estimate $a(t)$?*, was perhaps the question in Han's mind. If we can, for instance, $y(t) \Rightarrow \hat{a}(t) \approx a(t)$, then we can cancel $a(t)$ by designing $u(t) = -\hat{a}(t) + u_0(t)$ and system (1.3) amounts

to, approximately of course,

$$
(1.4) \qquad
\begin{cases}
\dot{x}_1(t) = x_2(t), \\
\dot{x}_2(t) = u_0(t), \\
y(t) = x_1(t).
\end{cases}
$$

The nature of the problem is therefore changed now. System (1.4) is just a linear time invariant system for which we have many ways to deal with it. This is likewise feedforward control yet to use output to "transform" the system first. In a different point of view, this part is called the "rejector" of disturbance ([11]). It seems that a further smarter way would be hardly to find anymore because the control $u(t) = -\hat{a}(t) + u_0(t)$ adopts a strategy of estimation/cancellation, much alike our experience in dealing with uncertainty in daily life. One can imagine and it actually is, one of the most energy saving control strategies as confirmed in [57].

This paradigm-shift is revolutionary for which Han wrote in [29] that "to improve accuracy, it is sometimes necessary to estimate $a(t)$ but it is not necessary to know the nonlinear relationship between $a(t)$ and the state variables". The idea breaks down the garden gates from time varying dynamics (e.g., $f(x_1, x_2, d, t) = g_1(t)x_1 + g_2(t)x_2$), nonlinearity (e.g., $f(x_1, x_2, d, t) = x_1^2 + x_2^3$), and "internal and external disturbance" (e.g., $f(x_1, x_2, d, t) = x_1^2 + x_2^2 + \Delta f(x_1, x_2) + d$). The problem now becomes: how can we realize $y(t) \Rightarrow \hat{a}(t) \approx a(t)$?

Han told us in [31] that it is not only possible but also realizable systematically. This is made possible by the so called extended state observer (ESO). Firstly, Han considered $a(t)$ to be an extended state variable and changed system (1.3) to

$$
(1.5) \qquad
\begin{cases}
\dot{x}_1(t) = x_2(t), \\
\dot{x}_2(t) = a(t) + u(t), \\
\dot{a}(t) = a'(t), \\
y(t) = x_1(t).
\end{cases}
$$

A linear observer for system (1.5), or equivalently linear ESO for system (1.3) can be designed as

$$
(1.6) \qquad
\begin{cases}
\dot{\hat{x}}_1(t) = \hat{x}_2(t) + a_1(\hat{x}_1(t) - y(t)), \\
\dot{\hat{x}}_2(t) = \hat{x}_3(t) + u(t) + a_2(\hat{x}_1(t) - y(t)), \\
\dot{\hat{x}}_3(t) = a_3(\hat{x}_1(t) - y(t)),
\end{cases}
$$

where we can choose $a_i, i = 1, 2, 3$ so that

$$
(1.7) \qquad \hat{x}_1(t) \to x_1(t), \quad \hat{x}_2(t) \to x_2(t), \quad \hat{x}_3(t) \to a(t) \quad \text{as } t \to \infty.
$$

It is seen that we have obtained estimation $\hat{x}_3(t) \approx a(t)$ from $y(t)$! Perhaps a better way to avoid aesthetic fatigue for readers is to leave the verification of (1.7) for (1.6) to give a much simpler example than (1.5). Consider the one-dimensional system

$$
(1.8) \qquad
\begin{cases}
\dot{x}(t) = a(t) + u(t), \\
y(t) = x(t),
\end{cases}
$$

where for first order system we have no other choice more than output to be identical to state. Likewise (1.6), a linear ESO can be designed as

$$
(1.9) \qquad
\begin{cases}
\dot{\hat{x}}(t) = \hat{a}(t) + u(t) + a_1(\hat{x}(t) - x(t)), \\
\dot{\hat{a}}(t) = a_2(\hat{x}(t) - x(t)).
\end{cases}
$$

Set $\tilde{x}(t) = \hat{x}(t) - x(t)$ and $\tilde{a}(t) = \hat{a}(t) - a(t)$ to be the errors. Then

$$
(1.10) \qquad
\begin{cases}
\dot{\tilde{x}}(t) = \tilde{a}(t) + a_1 \tilde{x}(t), \\
\dot{\tilde{a}}(t) = a_2 \tilde{x}(t) - \dot{a}(t).
\end{cases}
$$

We solve system (1.10) to obtain

$$
(1.11) \qquad (\tilde{x}(t), \tilde{a}(t)) = e^{At}(\tilde{x}(0), \tilde{a}(0)) + \int_0^t e^{A(t-s)} B \dot{a}(s) ds,
$$

where

$$
(1.12) \qquad A = \begin{pmatrix} a_1 & 1 \\ a_2 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.
$$

Definition 1.1. The ESO (1.9) is said to be convergent, if for any given $\delta > 0$, there exists a $T_\delta > 0$ such that

$$|\tilde{x}(t)| = |\hat{x}(t) - x(t)| \leq \delta \quad \text{and} \quad |\tilde{a}(t)| = |\hat{a}(t) - a(t)| \leq \delta, \, \forall \, t > T_\delta.$$

Now we investigate when ESO (1.9) is convergent. Choose a_1 and a_2 so that e^{At} decays as fast as desired with conjugate pair roots:

$$(1.13) \qquad |\lambda - A| = \lambda^2 - a_1 \lambda - a_2 = (\lambda + \omega + \omega_0 i)(\lambda + \omega - \omega_0 i)$$
$$\Rightarrow \quad a_1 = -2\omega, a_2 = -\omega^2 - \omega_0^2,$$

where ω_0 and ω are positive numbers. An obvious fact that

$$(1.14) \qquad \omega \to \infty \text{ if and only if } a_1 \to -\infty, \, a_2 \to -\infty$$

will be used for other purpose later. With $\omega_0 \neq 0$, we have

(1.15)
$$e^{At} = \frac{1}{2\omega_0 i} \begin{pmatrix} 1 & 1 \\ \omega - \omega_0 i & \omega + \omega_0 i \end{pmatrix} \begin{pmatrix} e^{-\omega t - \omega_0 i t} & 0 \\ 0 & e^{-\omega t + \omega_0 i t} \end{pmatrix} \begin{pmatrix} \omega + \omega_0 i & -1 \\ -\omega + \omega_0 i & 1 \end{pmatrix}$$
$$= \frac{1}{2\omega_0 i} \begin{pmatrix} (\omega + \omega_0 i)e^{-\omega t - \omega_0 i t} - (\omega - \omega_0 i)e^{-\omega t + \omega_0 i t} & -e^{-\omega t - \omega_0 i t} + e^{-\omega t + \omega_0 i t} \\ (\omega^2 + \omega_0^2)e^{-\omega t - \omega_0 i t} - (\omega^2 + \omega_0^2)e^{-\omega t + \omega_0 i t} & -(\omega - \omega_0 i)e^{-\omega t - \omega_0 i t} + (\omega + \omega_0 i)e^{-\omega t + \omega_0 i t} \end{pmatrix},$$
$$e^{At} B = \frac{1}{2\omega_0 i} \begin{pmatrix} -e^{-\omega t - \omega_0 i t} + e^{-\omega t + \omega_0 i t} \\ -(\omega - \omega_0 i)e^{-\omega t - \omega_0 i t} + (\omega + \omega_0 i)e^{-\omega t + \omega_0 i t} \end{pmatrix}.$$

Choosing particularly that

$$(1.16) \qquad\qquad\qquad\qquad \omega_0 \geq \omega,$$

we arrive at

$$(1.17) \qquad\qquad \|e^{At}\| \leq L\omega e^{-\omega t} \quad \text{and} \quad \|e^{At} B\| \leq L e^{-\omega t}$$

for some constant $L > 0$ independent of ω. When $\omega = \omega_0 = 1/\varepsilon$, it leads to the high gain design

$$(1.18) \qquad\qquad\qquad a_1 = -\frac{2}{\varepsilon}, \quad a_2 = -\frac{2}{\varepsilon^2}.$$

When we choose two different real roots,

$$(1.19) \qquad |\lambda - A| = \lambda^2 - a_1 \lambda - a_2 = (\lambda + \omega_1)(\lambda + \omega_2)$$
$$\Rightarrow \quad a_1 = -\omega_1 - \omega_2, a_2 = -\omega_1 \omega_2,$$

where real $\omega_i, i = 1, 2$ are positive numbers. It also leads to (1.14) alike:

(1.20) $\omega = \min\{\omega_1, \omega_2\} \to \infty$ if and only $a_1 \to -\infty,\ a_2 \to -\infty,$

and to high gain kind design (1.18) when $\omega_1 = \omega_2 = 1/\varepsilon$. Now
(1.21)
$$e^{At} = -\frac{\omega_1 + \omega_2}{\omega_2 - \omega_1} \begin{pmatrix} \frac{\omega_2}{\omega_1+\omega_2}e^{-\omega_1 t} + \frac{\omega_1}{\omega_1+\omega_2}e^{-\omega_2 t} & e^{-\omega_1 t} + e^{-\omega_2 t} \\ -\frac{\omega_1\omega_2}{\omega_1+\omega_2}e^{-\omega_1 t} + \frac{\omega_1\omega_2}{\omega_1+\omega_2}e^{-\omega_2 t} & +\frac{\omega_1}{\omega_1+\omega_2}e^{-\omega_1 t} + \frac{\omega_2}{\omega_1+\omega_2}e^{-\omega_2 t} \end{pmatrix},$$

$$e^{At}B = -\frac{\omega_1 + \omega_2}{\omega_2 - \omega_1} \begin{pmatrix} -e^{-\omega_1 t} - e^{-\omega_2 t} \\ -\frac{\omega_1}{\omega_1+\omega_2}e^{-\omega_1 t} + \frac{\omega_2}{\omega_1+\omega_2}e^{-\omega_2 t} \end{pmatrix}.$$

We also have (1.17) in most of the cases where $|\omega_2 - \omega_1| = c$ with c being a fixed constant restraining the possibility like $\omega_2 = \omega_1 + e^{-\omega_1}$. Now suppose that

(1.22) $|\dot{a}(t)| \leq M, \forall\, t \geq 0.$

Then the solution of (1.11) is estimated as

(1.23) $\|(\tilde{x}(t), \tilde{a}(t))\| \leq \omega e^{-\omega t}\|(\tilde{x}(0), \tilde{a}(0))\| + \dfrac{L}{\omega}M.$

The first term of (1.23) tends to zero as $t \to \infty$ and the second term can be as small as desired by setting ω to be large. In other words, to make ESO (1.9) be convergent, ω must be chosen large. This is the meaning of the high gain. The sufficiency is obvious from (1.23), the necessity can be checked directly from (1.11) for special signal like

$$a(t) = nt - n\left(n + \frac{i-1}{n}\right), \quad t \in \left[n + \frac{i-1}{n}, n + \frac{i}{n}\right),$$
$$i = 1, 2, \ldots, n, \quad n = 1, 2, \ldots,$$

for which $|a(t)| \leq 1$ but $\dot{a}(t) = n$ in $t \in (n, n+1)$. We leave reader for this verification as an exercise.

Remark 1.1. The boundedness of derivative $\dot{a}(t)$ obviously seen from (1.11) can be relaxed up to $a^{(N)}(t)$ to be bounded for some finite positive integer

N. In this case the ESO (1.9) can be changed into

(1.24)
$$\begin{cases} \dot{\hat{x}}(t) = \hat{a}_1(t) + u(t) + a_1(\hat{x}(t) - x(t)), \\ \dot{\hat{a}}_i(t) = \hat{a}_{i+1}(t) + a_i(\hat{x}(t) - x(t)), i = 2, 3, \cdots, N - 1. \end{cases}$$

We can also prove that $\hat{x}(t) - x(t) \to 0, \hat{a}_i(t) - a^{(i-1)}(t) \to 0, i = 1, 2, \ldots,$ $N - 1$ by properly choosing $a_i, i = 1, 2, \cdots N - 1$.

To sum up, we can say now that under condition (1.22), the ESO (1.9) is convergent, that is,

(1.25)
$$\hat{x}(t) \to x(t), \ \hat{a}(t) \to a(t) \text{ as } t \to \infty, \ \omega \to \infty.$$

A remarkable finding beyond what is usual from (1.23) is that the peaking phenomena occurs at $t = 0$ only with large ω and nonzero $(\tilde{x}(0), \tilde{a}(0))$. Finally, to stabilize system (1.8), we simply cancel the disturbance by using the ESO-based feedback:

(1.26)
$$u(t) = -\hat{a}(t) - x(t),$$

where the first term is used to cancel (compensate) the disturbance and the second term is a standard stabilizing controller for the "transformed" system $\dot{x}(t) = u(t)$. The closed-loop of (1.8) under the feedback (1.26) becomes

(1.27)
$$\begin{cases} \dot{x}(t) = -x(t) - \hat{a}(t) + a(t), \\ \dot{\hat{x}}(t) = -x(t) + a_1(\hat{x}(t) - x(t)), \\ \dot{\hat{a}}(t) = a_2(\hat{x}(t) - x(t)), \end{cases}$$

which is equivalent, by setting $\tilde{x}(t) = \hat{x}(t) - x(t)$ and $\tilde{a}(t) = \hat{a}(t) - a(t)$, to

(1.28)
$$\begin{cases} \dot{x}(t) = -x(t) - \tilde{a}(t), \\ \dot{\tilde{x}}(t) = \tilde{a}(t) + a_1\tilde{x}(t), \\ \dot{\tilde{a}}(t) = a_2\tilde{x}(t) - \dot{a}(t). \end{cases}$$

Since $(\tilde{x}(t), \tilde{a}(t)) \to 0$ as $\omega \to \infty$ and $t \to \infty$, we have immediately that

$$x(t) \to 0 \text{ as } t \to \infty, \omega \to \infty,$$

or equivalently

(1.29)
$$x(t) \to 0, \ \hat{x}(t) \to 0, \ \hat{a}(t) - a(t) \to 0 \text{ as } t \to \infty, \omega \to \infty.$$

This is the well known separation principle in linear system theory. So, the whole idea not only works but also works in an extremely wise way of estimating and cancelling the disturbance in real time.

Certainly, as any other methods, some limitations likely exist in an otherwise perfect setting of ESO in the sense:

- The high gain is resorted in ESO as shown by (1.14) to suppress the effect of the derivative of disturbance;

- the derivative of disturbance as shown in (1.22) is supposed to be bounded (or some finite order of derivative is bounded as explained in Remark 1.1) as well as from (1.5) where $a(t)$ is regarded as an extended state variable.

The second problem can be resolved by time-varying gain instead of constant high gain. Suppose for instance that the derivative of $a(t)$ satisfies

$$|\dot{a}(t)| \leq B_0 + Be^{bt}, \quad b > 0.$$

Then let a_1 and a_2 be replaced by $a_1 r(t)$ and $a_2 r(t)$ with $r(t) = e^{\beta t}, \beta > b$ in (1.27) and let

$$\begin{cases} \eta_1(t) = r(t)(\hat{x}(t) - x(t)), \\ \eta_2(t) = \hat{a}(t) - a(t). \end{cases}$$

Then

(1.30)
$$\begin{cases} \dot{\eta}_1(t) = r(t)(\eta_2(t) - a_1\eta_1(t)) + 2a\eta_1(t), \\ \dot{\eta}_2(t) = -a_2 r(t)\eta_1(t) - \dot{d}(t). \end{cases}$$

Let $V(x_1, x_2) = (x_1, x_2)P(x_1, x_2)^\top$ with P being the positive definite matrix solution of the Lyapunov function $PA + A^\top P = -I_{2\times 2}$ where $I_{2\times 2}$ is the 2-dimensional identity matrix. We have

$$\frac{dV(\eta_1(t), \eta_2(t))}{dt} \leq -e^{at}\|(\eta_1(t), \eta_2(t))\|^2 + 4a\lambda_{\max}(P)\|(\eta_1(t), \eta_2(t))\|^2$$
$$+ 2\lambda_{\max}(P)\|(\eta_1(t), \eta_2(t))\|(B_0 + Be^{bt}),$$

by which we have arrived at $\lim_{t\to\infty} V(\eta_1(t), \eta_2(t)) = 0$ and therefore once again (1.29). Notice that although the gain is increasing to infinity in ESO (1.30), the control (1.26) is bounded: $u(t) = -\hat{a}(t) - x(t) \to a(t)$ as $t \to \infty$. This is very different from another well known control method known as high gain control where the high gain is also required in control ([39, 44]) as well as in observer, and will be explained in subsequent sections.

One may argue at this moment that the derivative $\dot{a}(t)$ in (1.28) is an manmade problem because in ESO (1.27), we have no $\dot{a}(t)$. An alternative way is to change (1.11) as

(1.31) $(\tilde{x}(t), \tilde{a}(t))$

$$= e^{At}(\tilde{x}(0), \tilde{a}(0)) + \int_0^t e^{A(t-s)} B\dot{a}(s)ds$$

$$= e^{At}(\tilde{x}(0), \tilde{a}(0)) + Ba(t) - e^{At} Ba(0) + \int_0^t Ae^{A(t-s)} Ba(s)ds$$

$$\approx Ba(t) + \int_0^t Ae^{A(t-s)} Ba(s)ds \quad \text{as } t \to \infty,$$

from which, however, we are not able to obtain convergence.

The first problem is possibly resolved by designing a different type of ESO because in the final analysis when we scrutinize the whole process, ESO (1.9) is nothing more than one of such devices, developed by Han himself only aiming at estimating disturbance from observable measured output which is the ultimate goal of ADRC. It is not, and should not be, a unique way for this purpose. To explain this point, let us go back to other methods of modern control theory where we can find expectedly the similar idea as economical as ADRC on the basis of estimation/cancellation strategy yet no high gain is utilized.

The popular and powerful method to deal with unknown parameter in the system is the adaptive control. Consider again stabilization of one dimensional system:

(1.32) $$\begin{cases} \dot{x}(t) = \theta f(x(t)) + u(t), \\ y(t) = x(t), \end{cases}$$

where θ is an unknown parameter. Suppose that we have parameter estimation:

(1.33) $\hat{\theta}(t) \to \theta.$

Then the stabilizing control is certainly designed by estimation/cancellation way

(1.34) $u(t) = -x(t) - \hat{\theta}(t)f(x(t)),$

where the second term is also used to cancel the effect of the disturbance. When it is substituted into (1.32), we have the closed-loop:

(1.35)
$$\begin{cases} \dot{x}(t) = \tilde{\theta}(t)f(x(t)) - x(t), \ \tilde{\theta}(t) = \theta - \hat{\theta}(t), \\ y(t) = x(t). \end{cases}$$

Since system has added additional variable $\hat{\theta}(t)$, the Lyapunov function should include $\hat{\theta}(t)$, which is assigned as

$$V(x(t), \tilde{\theta}(t)) = \frac{1}{2}x^2(t) + \frac{1}{2}\tilde{\theta}^2(t)$$

for system (1.35). Finding the derivative of $V(x(t), \tilde{\theta}(t))$ along the solution of (1.35), we obtain

$$\frac{dV(x(t), \tilde{\theta}(t))}{dt} = -x^2(t) + \tilde{\theta}(t)[\dot{\tilde{\theta}}(t) + x(t)f(x(t)] = -x^2(t)$$

provided $\dot{\tilde{\theta}}(t) = -x(t)f(x(t)$. The closed-loop now reads

(1.36)
$$\begin{cases} \dot{x}(t) = \tilde{\theta}(t)f(x(t)) - x(t), \\ \dot{\tilde{\theta}}(t) = -x(t)f(x(t)). \end{cases}$$

By Lasalle's invariance principle, we obtain immediately that

(1.37) $x(t) \to 0$ as $t \to \infty$

for closed-loop (1.36). The left problem is: does $\hat{\theta}(t) \to \theta$ or equivalently $\tilde{\theta}(t) \to 0$ in the end? The answer is not necessarily. By Lasalle's invariance principle further, when $\dot{V} = 0$ we obtain $x = 0$ and hence $\tilde{\theta} = \tilde{\theta}_0$ is a constant and $\tilde{\theta}_0 f(0) = 0$. So we come up two cases a) $\tilde{\theta}_0 = 0$ when $f(0) \neq 0$; b) $(x(t), \tilde{\theta}(t)) = (0, \tilde{\theta}_0)$ is a solution to (1.36). The second case means that $\tilde{\theta}(t) \to 0$ is not necessarily occurring whereas the first case leads to the well known persistence exciting (PE) condition which is $f(0) \neq 0$ for our case. However, in both cases, we always have

$$\tilde{\theta}(t)f(x(t)) \to 0 \text{ as } t \to \infty$$

no matter the parameter update law which is $\dot{\hat{\theta}}(t) = x(t)f(x(t))$ is convergent or not. In other words, the unknown term $\theta f(x(t))$ in system (1.32) is cancelled anyway by the feedback control (1.33). Very accidentally, the

order of (1.32) is increased from one to two in (1.36). People are surprised in ADRC configuration (1.5) that the disturbance is regarded as an extended state variable which is not actually an state but increases the order of system while in control theory, the reduced order state observer is to reduce the number of states for the sake of simplification, as indicated in [11]. But here we first meet the order increasing in a matured control theory. What is more, the unknown constant θ comes from internal disturbance and we do not use high gain in both parameter estimation and feedback.

Next, we come to another control method known as internal model principle [8, 37], a robust control method in dealing with external disturbance. Once gain we consider system (1.8) where $a(t) = \theta \sin \omega t$ with θ being unknown but usually $a(t)$ represents an external disturbance. Since $a(t)$ satisfies $\ddot{a}(t) = -\omega^2 a(t)$, we couple this with (1.8) to obtain

(1.38)
$$\begin{cases} \dot{x}(t) = a(t) + u(t), \\ \ddot{a}(t) = -\omega^2 a(t), \\ y(t) = x(t), \end{cases}$$

We write (1.38) in the matrix form that we are familiar with:

(1.39)
$$\begin{cases} \dot{X}(t) = AX(t) + Bu(t), \\ y(t) = CX(t), \end{cases}$$

where

$$X(t) = (x(t), a(t), \dot{a}(t))^\top,$$

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -\omega^2 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad C = (1, 0, 0).$$

A simple computation shows that

$$\text{rank} \begin{pmatrix} C \\ CA \\ CA^2 \end{pmatrix} = 3.$$

So system (1.38) or (1.39) is exactly observable and certainly it is not exactly controllable (even not stabilizable). This point is explained from another point for system (1.3). So we can naturally design a Luenberger observer as $\dot{\hat{X}}(t) = A\hat{X}(t) + Bu(t) + L(C\hat{X}(t) - x(t))$ with $L = \ell_1, \ell_2, \ell_3)$ to

make $A + LC$ be Hurwitz. For instance when $L = (-3, \omega^2 - 3, 3\omega^2 - 1)^\top$ the eigenvalues of $A + LC$ are identical to negative one. Explicitly,

(1.40)
$$\begin{cases} \dot{\hat{x}}(t) = a(t) + u(t) + \ell_1(\hat{x}(t) - x(t)), \\ \dot{\hat{a}}(t) = -\omega^2 \hat{a}(t) + \ell_2(\hat{x}(t) - x(t)), \\ \dot{z}(t) = -\omega^2 \hat{a}(t) + \ell_3(\hat{x}(t) - x(t)). \end{cases}$$

Surprisedly, (1.40) is very likewise an ESO for (1.38) where we also consider $a(t)$ as an extended state variable. This is our second time from a matured conventional modern control method to regard the external disturbance as an extended state variable and the order of system is certainly increased from first order to even third order which is additionally implicitly extravagant because we do not need estimation of $\dot{a}(t)$ anyway. In addition, slightly different to adaptive control, we estimate the disturbance before the feedback. In this sense, ESO is no more than a kind of observer in control theory. The stabilizing feedback control is the same as (1.26). Once again, here we do not use high gain.

If we let the matter drop here, the ADRC seems not very new idea in control theory. But when we go further to have a comparison, we find from (1.1) that ADRC regards both internal and external disturbance $a(t) = f(x_1(t), x_2(t), d(t), t)$ together as a signal of time which can be estimated by the output. This spans significantly the concept of the disturbance where in adaptive control they are some internal unknown parameters and in internal model principle they are some "almost known" external disturbance produced from a dynamical exosystem. The ADRC's major component ESO provides a systematical feasible way to estimate total disturbance from measured output. It opens another gate so that we can get rid of mathematical brunt like $a(t) = f(x_1(t), x_2(t), d(t), t)$ to be state dependent or state free, time invariant or time variant, linear or nonlinear and whatever. This is an almost model free control method, carrying PID control forward.

Up to now, we have introduced ESO and ESO based feedback control which are two key parts of ADRC and exemplified stabilization for introduction of ADRC. Actually, ADRC's configuration is for output tracking which contains stabilization as a special case when the reference signal is identical to zero. As mentioned, in PID, "the setpoint is often given as a step function, not appropriate for most dynamics systems because it amounts to asking the output and, therefore, the control signal, to make a sudden jump" ([34]), whereas in output tracking, the reference, i.e. the desired output trajectory, should be generated to make it physically feasible for the output to

374 B.-Z. Guo and Z.-L. Zhao

follow and to keep the tracking error small. Han was aware of this practical concern and suggested a simple generator called tracking differentiator. In the linear case, Han's tracking differentiator is shown to be the high gain tracking differentiator ([26]). The tracking differentiator, shorthanded as TD, is included as a first part of ADRC but is relatively independent part of ADRC. In the subsequent sections, we will concern ourselves with TD, ESO, and ESO based feedback control.

Before we end this opening story, we indicate that the possible improvement of ADRC lies in ESO as what we have seen in adaptive control and internal model principle where the inherent estimation/cancellation is kept yet no high gain is used.

The remaining part is organized as follows. In the next section, section 2, we introduce tracking differentiator. The extended state observer is introduced in section 3. Section 4 is devoted to extended state observer based feedback control, followed by a short summary in section 5.

2. Tracking differentiator

In three parts of ADRC, tracking differentiator (TD) is the most obvious part inherited from PID directly. Because in PID, D is usually not physically implementable, which makes PID most often PI control actually. Commonly, when D is used in PID, the implementation that a differentiation of a signal $v(t)$ is obtained approximately as

$$(2.1) \qquad y(t) \approx \frac{1}{\tau}(v(t) - v(t - \tau)) \approx \dot{v}(t),$$

where τ is a small positive constant. Let delay signal $v(t - \tau)$ be denoted by $y_1(t)$. Then

$$(2.2) \qquad \hat{y}_1(s) = e^{-\tau s}\hat{v}(s).$$

where $\hat{\phi}(s)$ represents the Laplace transform which for a given function $\phi(t)$, is

$$(2.3) \qquad \hat{\phi}(s) = \int_0^\infty \phi(t)e^{-st}dt.$$

Approximating $e^{s\tau}$ by its first order in Taylor expansion, which is $1 + s\tau$ as τ small, we have

$$(2.4) \qquad \hat{y}_2(s) = \frac{1}{1 + \tau s}\hat{v}(s).$$

The state-space realization of (2.4) is

$$(2.5) \qquad \dot{y}_2(t) = -\frac{1}{\tau}(y_2(t) - v(t)) \Rightarrow y_2(t) = \frac{1}{\tau} \int_0^t e^{-\frac{1}{\tau}(t-\rho)} v(\rho) d\rho.$$

This obtained delayed signal can filter the high-frequency noise ([33, p.50-51]). So when the input $v(t)$ is contaminated by high-frequency noise let's say, $n(t)$ with zero expectation, that is, in (2.2), the input is $v(t) + n(t)$ instead of $v(t)$, $y_2(t)$ can also approximate $v(t - \tau)$ satisfactorily. However,

$$(2.6) \qquad y(t) \approx \frac{1}{\tau}(v(t) + n(t) - y_2(t) - n(t - \tau)) \approx \dot{v}(t) + \frac{1}{\tau}n(t)$$

is still quite sensitive to the noise in $v(t)$ because it is amplified by a factor of $1/\tau$. To address this difficulty, Han proposed a noise tolerant tracking differentiator:

$$(2.7) \qquad \dot{v}(t) \approx \frac{v(t - \tau_1) - v(t - \tau_2)}{\tau_2 - \tau_1}, \quad 0 < \tau_1 < \tau_2.$$

Since the first term in (2.7) involves a time delay, it can also filter high-frequency noise as explained. Analogously from (2.2) to (2.4), we have

$$(2.8) \qquad \hat{y}(s) = \frac{1}{\tau_2 - \tau_1} \left(\frac{1}{\tau_1 s + 1} - \frac{1}{\tau_2 s + 1} \right) \hat{v}(s)$$

and its state-space realization is

$$(2.9) \qquad \begin{cases} \dot{x}_1(t) = x_2(t), \\ \dot{x}_2(t) = -\dfrac{1}{\tau_1 \tau_2}(x_1(t) - v(t)) - \dfrac{\tau_2 - \tau_1}{\tau_1 \tau_2}x_2(t), \\ y(t) = x_2(t). \end{cases}$$

System (2.9) is a special form of the nonlinear TD proposed first time by Han in [29] as follows:

$$(2.10) \qquad \begin{cases} \dot{x}_1(t) = x_2(t), \\ \dot{x}_2(t) = r^2 f\left(x_1(t) - v(t), \dfrac{x_2(t)}{r} \right), \end{cases}$$

where r is the tuning parameter, $f(\cdot)$ is an appropriate nonlinear function. In [29], it is declared, based on numerous numerical simulations, that the TD (2.10) is convergent.

Conjecture 2.1 (Han's TD). *If all the solutions (in the sense of Filippov) of the following reference free system:*

$$(2.11) \qquad \begin{cases} \dot{x}_1(t) = x_2(t), \\ \dot{x}_2(t) = f(x_1(t), x_2(t)), \end{cases}$$

is globally asymptotically stable: $\lim_{t\to\infty} x_1(t) = 0,\ \lim_{t\to\infty} x_2(t) = 0,$ *then for any bounded measurable function $v(t)$ and any constant $T > 0$, the solution of system following*

$$(2.12) \qquad \begin{cases} \dot{z}_{1R}(t) = z_{2R}(t), \\ \dot{z}_{2R}(t) = R^2 f\left(z_{1R}(t) - v(t), \dfrac{z_{2R}(t)}{R}\right) \end{cases}$$

satisfies

$$(2.13) \qquad \lim_{R\to\infty} \int_0^T |z_{1R}(t) - v(t)|\,dt = 0$$

for any $T > 0$.

The convergence of (2.13) implies that $z_{2R}(t)$ is (weakly) convergent to $\dot{v}(t)$ as $R \to \infty$ in the sense of generalized function which is defined as a functional of $C_0^\infty(0,T)$ for any $T > 0$ as follows:

$$(2.14) \qquad v^{(i-1)}(\varphi) = (-1)^{(i-1)} \int_0^T v(t)\varphi^{(i-1)}(t)\,dt,$$

where $\varphi \in C_0^\infty(0,T), i > 1$. The definition (2.14) is a standard definition for the generalized derivative [1]. From this definition, we see that any order of the generalized derivative $v^{(i)}(t)$ always exists provided that $v(t)$ is bounded measurable. Suppose that (2.13) holds. Then considering $z_{iR}(t)$ as a functional of $C_0^\infty(0,T)$, we have

$$(2.15)\quad \lim_{R\to\infty} z_{iR}(\varphi) = \lim_{R\to\infty} \int_0^T z_{iR}(t)\varphi(t)\,dt = \lim_{R\to\infty} \int_0^T z_{1R}^{(i-1)}(t)\varphi(t)\,dt$$

$$= \lim_{R\to\infty} (-1)^{(i-1)} \int_0^T z_{1R}(t)\varphi^{(i-1)}(t)\,dt$$

$$= (-1)^{(i-1)} \int_0^T v(t)\varphi^{(i-1)}(t)\,dt, \quad \varphi \in C_0^\infty(0,T),\ i > 1.$$

Comparing the right-hand sides of (2.14) and (2.15), we see that

$$z_{iR}(t) \rightharpoonup v^{(i-1)}(t), \quad R \to \infty.$$

In what follows, we say that if (2.13) is satisfied, then $z_{2R}(t)$ converges weakly to $\dot{v}(t)$ as time goes to infinity, and we say that $z_{2R}(t)$ is strongly convergent to $\dot{v}(t)$ if $\lim_{R\to\infty} |z_{2R}(t) - \dot{v}(t)| = 0$ uniformly on $[a, \infty)$ for some $a > 0$.

From above process, we see that Han's proposal for (2.10) is motivated entirely from classical derivative extraction for reference signal and is relatively independent of high-gain observer ([7]) although in linear case, (2.9) happens to be high-gain tracking differentiator under coordinate transformation as shown in [22].

As mentioned in previous section, TD (2.10) severs not only the derivative extraction, but also a transient profile that the output of plant can reasonably follow to avoid setpoint jump in PID. In other words, in engineering applications, the designed trajectory to be tracked by output of the plant is $x_1(t)$ produced by (2.10) rather than $v(t)$ which could be jumping like step function. This makes the reference smooth. By making the reference smooth, the control signal can be made smooth as well, which has significant engineering values. Han suggested transient profile specifically in [34].

Now we come to the convergence of Conjecture 2.1. Although a first proof is reported in [30], it is later on shown to be true only for constant signal $v(t)$. Nevertheless, the effectiveness of a tracking differentiator (2.10) has been witnessed by countless numerical experiments and control practices [35, 36]. An indirect but relevant evidence is that the linear TD (LTD) (2.9), which is neatly written as follows, is shown early in [12] to be indeed convergent:

$$(2.16) \quad \begin{cases} \dot{z}_{1R}(t) = z_{2R}(t), \\ \dot{z}_{2R}(t) = -k_1 R^2((z_{1R}(t) - v(t)) - k_2 R z_{2R}(t), \end{cases}$$

where $k_1 > 0$ and $k_2 > 0$ are constants, and $R > 0$ is the tuning parameter.

Theorem 2.1. *Suppose that $k_1, k_2 > 0$, and $v : [0, \infty) \to \mathbb{R}$ satisfies* $\sup_{t\in[0,\infty)}[|v(t)| + |\dot{v}(t)|] = M < \infty$ *for constant $M > 0$. Then the linear tracking differentiator* (2.16) *is convergent in the sense that for any $a > 0$,*

$$\lim_{R\to\infty} |z_{1R}(t) - v(t)| = 0, \quad \lim_{R\to\infty} |z_{2R}(t) - \dot{v}(t)| = 0$$

uniformly for $t \in [a, \infty)$.

In many control practices, we also need high-order derivatives of reference signal. To do this, we can use the above differentiator repeatedly. The linear high order tracking differentiator can be designed as follows:

(2.17)
$$
\begin{cases}
\dot{z}_{1R}(t) = z_{2R}(t), \ z_{1R}(0) = z_{10}, \\[4pt]
\dot{z}_{2R}(t) = z_{3R}(t), \ z_{2R}(0) = z_{20}, \\[4pt]
\quad \vdots \\[4pt]
\dot{z}_{(n-1)R}(t) = z_{nR}(t), \ z_{(n-1)R}(0) = z_{(n-1)0}, \\[4pt]
\dot{z}_{nR}(t) = R^n \bigg(a_1(z_{1R}(t) - v(t)) + \dfrac{a_2 z_{2R}(t)}{R} \\[6pt]
\qquad\qquad + \cdots + \dfrac{a_n z_{nR}(t)}{R^{n-1}} \bigg), \ z_{nR}(0) = z_{n0}.
\end{cases}
$$

The high-order linear TD (2.17) is constructed in [25] and the following strong convergence results are also proved.

Theorem 2.2. *Suppose that the matrix of the following*

(2.18)
$$
A = \begin{pmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
a_1 & a_2 & a_3 & \cdots & a_n
\end{pmatrix}
$$

is Hurwitz, and $v : [0,\infty) \to \mathbb{R}$ satisfies $\sup_{t\in[0,\infty),1\le k\le n} |v^{(k)}(t)| < \infty$ for some constant $M > 0$. Then the linear tracking differentiator (2.17) is convergent in the sense that for any $a > 0$,

(2.19) $\displaystyle \lim_{R\to\infty} |z_{iR} - v^{(i-1)}(t)| = 0$ *unoformly on $[a,\infty)$, $i = 1, 2, \ldots, n$,*

where $(z_{10}, z_{20}, \ldots, z_{n0})$ is any given initial value.

We are now coming to nonlinear TD for which we also have convergence ([25]).

Theorem 2.3. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a locally Lipschtz continuous function, $f(0,0) = 0$. Suppose that the equilibrium point $(0,0)$ of the following system*

is globally asymptotically stable:

(2.20)
$$\begin{cases} \dot{x}_1(t) = x_2(t), & x_1(0) = x_{10}, \\ \dot{x}_2(t) = f(x_1(t), x_2(t)), & x_2(0) = x_{20}, \end{cases}$$

where (x_{10}, x_{20}) is any given initial value. If the signal $v(t)$ is differentiable and satisfies $\sup_{t \in [0,\infty)} |\dot{v}(t)| < \infty$, then the solution of the following tracking differentiator:

(2.21)
$$\begin{cases} \dot{z}_{1R}(t) = z_{2R}(t), & z_{1R}(0) = z_{10}, \\ \dot{z}_{2R}(t) = R^2 f\left(z_{1R}(t) - v(t), \dfrac{z_{2R}(t)}{R}\right), & z_{2R}(0) = z_{20}, \end{cases}$$

is convergent in the sense that for every $a > 0$,

(2.22)
$$\lim_{R \to \infty} |z_{1R} - v(t)| = 0 \quad \text{unoformly on } [a, \infty),$$

where (z_{10}, z_{20}) is any given initial value.

Compared with Han's Conjecture 2.1, we require in Theorem 2.3 additionally the Lyapunov stability for the reference free system, which is slightly restrictive than Han's Conjecture 2.1 but it is natural, one cannot image a controlled physical system is asymptotically stable but not Lyapunov stable (yet it is mathematically possible). What is more, the convergence is also stronger. Theorem 2.3 can also be generalized to include high-order nonlinear TD.

Theorem 2.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschtz continuous function, $f(0, \ldots, 0) = 0$. Suppose that the equilibrium point $(0, \ldots, 0)$ of the following reference free system is globally asymptotically stable:*

(2.23)
$$\begin{cases} \dot{x}_1(t) = x_2(t), & x_1(0) = x_{10}, \\ \dot{x}_2(t) = x_3(t), & x_2(0) = x_{20}, \\ \quad \vdots \\ \dot{x}_n(t) = f(x_1(t), x_2(t), \ldots, x_n(t)), & x_n(0) = x_{n0}, \end{cases}$$

where $(x_{10}, x_{20}, \ldots, x_{n0})^\top$ is a given initial value. If the signal $v(t)$ is differentiable and satisfies $\sup_{t \in [0,\infty)} |v^{(n+1)}(t)| < \infty$, then the solution of the

following tracking differentiator:

$$(2.24) \quad \begin{cases} \dot{z}_{1R}(t) = z_{2R}(t), \ z_{1R}(0) = z_{10}, \\ \dot{z}_{2R}(t) = z_{R3}(t), \ z_{R2}(0) = z_{20}, \\ \quad \vdots \\ \dot{z}_{nR}(t) = R^n f\left(z_{1R}(t) - v(t), \dfrac{z_{2R}(t)}{R}, \ldots, \dfrac{z_{nR}(t)}{R^{(n-1)}} \right), \ z_{nR}(0) = z_{n0}, \end{cases}$$

is convergent in the sense that for every $a > 0$,

$$(2.25) \qquad \lim_{R \to \infty} |z_{1R} - v(t)| = 0 \ \ \text{unoformly on } [a, \infty),$$

where $(z_{10}, z_{20}, \ldots, z_{n0})$ is any given initial value.

For general nonlinear TD, we are not able to give an explicit error estimation, similarly with the LTD in Theorems 2.1 and 2.2. However, for some special nonlinear TD, we do have error estimation.

Theorem 2.5. *Suppose that*
(i) $\sup_{t \in [0, \infty}|v^{(i)}(t)| < \infty, \ i = 1, 2, \ldots, n$;
(ii) the nonlinear function $f(x)$ in (2.24) satisfies that

$$(2.26) \qquad |f(x) - f(\overline{x})| \le \sum_{j=1}^{n} k_j \|x_j - \overline{x}_j\|^{\theta_j} \ \text{for some } k_j > 0, \ \theta_j \in (0, 1];$$

(iii) there exists a continuous, positive definite function $V : \mathbb{R}^n \to \mathbb{R}$, with all continuous partial derivatives in its variables, satisfying

$$(2.27) \qquad L_h V(x) \le -c(V(x))^\alpha,$$

with $c > 0$, $\alpha \in (0, 1)$, $\gamma = \frac{1-\alpha}{\alpha}$, and $h(x)$ is the vector field: $h(x) = (x_2, x_3, \ldots, x_{n-1}, f(x))^\top$. Then for any initial value of (2.24) and constant $a > 0$, there exists $R_0 > 0$ such that for every $R > R_0$,

$$(2.28) \qquad |z_{iR}(t) - v^{(i-1)}(t)| \le L \left(\frac{1}{R} \right)^{\theta\gamma - i + 1}, \ \forall \, t > a,$$

where L is some positive constant, $\theta = \min\{\theta_2, \theta_3, \ldots, \theta_n\}$, and $z_{iR}(t)$ is the solution of (2.24), $i = 1, 2, \ldots, n$.

Theorem 2.5 is presented in [22]. To be easily practically implementable, an concrete nonlinear TD in terms of Theorem 2.5 is constructed.

Theorem 2.6. *If the signal $v(t)$ satisfies $\sup_{t \in [0,\infty)} |v^{(i)}(t)| < \infty$ for $i = 1, 2$, then the second order finite-time stable differentiator following:*

(2.29)
$$
\begin{cases}
\dot{z}_{1R}(t) = z_{2R}(t), \\
\dot{z}_{2R}(t) = R^2 \left(-k_1 [z_{1R}(t) - v(t)]^\alpha - k_2 \left[\dfrac{z_{2R}(t)}{R} \right]^\beta \right),
\end{cases}
$$

is convergent in the sense that for any initial value of (2.29) and $a > 0$, there exists a constant $R_0 > 0$ such that for all $R > R_0$ and $t > a$,
(2.30)
$$
|z_{1R}(t) - v(t)| \leq M_1 \left(\frac{1}{R} \right)^{\beta \frac{1-\gamma}{\gamma}}, \quad |z_{2R}(t) - \dot{v}(t)| \leq M_2 \left(\frac{1}{R} \right)^{\beta \frac{1-\gamma}{\gamma} - 1},
$$

where $\gamma = \frac{l-1}{l}$, $l > \max\{1, a, b\}$, and the parameters in (2.29) are selected to satisfy

(2.31) $\quad k_1, k_2 > 0, \quad \alpha = \dfrac{b-1}{a}, \quad \beta = \dfrac{b-1}{b}, \quad a = b+1, \quad b > 1.$

If $\beta \frac{1-\gamma}{\gamma} > 1$ then $z_{2R}(t) \to \dot{v}(t)$ in classical sense, while $\beta \frac{1-\gamma}{\gamma} \leq 1$, $z_{2R}(t) \to \dot{v}(t)$ in the sense of weak derivative (weak convergence).

To have a visual comparison, we compute LTD and nonlinear (NLTD), in terms of (2.29), numerically. Here LTD is based on Theorem 2.1:

(2.32)
$$
\begin{cases}
\dot{z}_{1R}(t) = z_{2R}(t), \\
\dot{z}_{2R}(t) = R^2 \left(-(z_{1R}(t) - v(t)) - \dfrac{z_{2R}(t)}{R} \right).
\end{cases}
$$

The NLTD is based on Theorem 2.6:

(2.33)
$$
\begin{cases}
\dot{z}_{1R}(t) = z_{2R}(t), \\
\dot{z}_{2R}(t) = R^2 \left(-[z_{1R}(t) - v(t)]^{\frac{1}{2}} - \left[\dfrac{z_{2R}(t)}{R} \right]^{\frac{2}{3}} \right).
\end{cases}
$$

The Matlab with Eular method is programmed. Here we choose the same zero initial value, step $h = 0.001$, and $v(t) = \sin t$ in all simulations.

The results by LTD are plotted in Figure 1 where in Figure 1(a), $R = 10$ is taken, and in Figure 1(b) $R = 20$. The results by FTSTD are plotted in
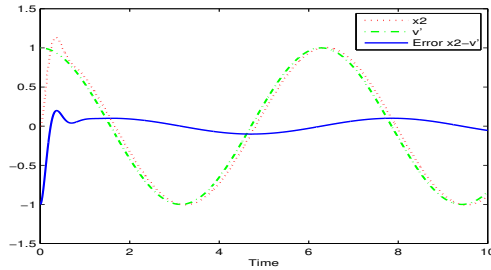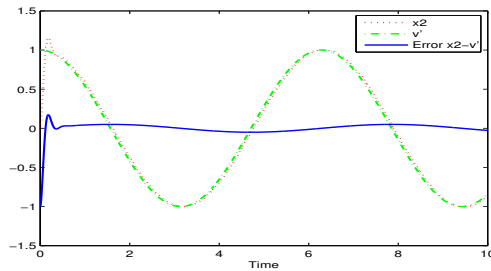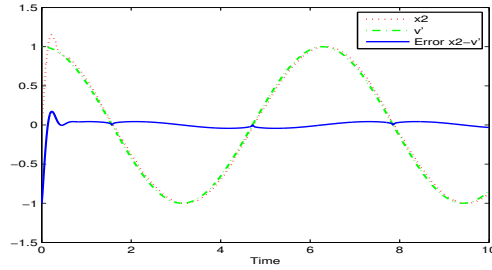

(a) $R = 10$

(b) $R = 20$

Figure 1: Derivative tracking for $v(t) = \sin t$ by LTD.

Figure 2 where in Figure 2(a), $R = 10$ is taken, while in Figure 2(b) $R = 20$. It is seen from Figures 1 and 2 that both TDs can recover effectively the derivatives of reference signal. In addition, for both TDs, the larger R is, the more accurate estimation. For the same tuning parameter, it is seen from comparisons of Figure 1(a) with 2(a), Figure 1(b) with 2(b) to find that NLTD is more accurate than LTD.

Now we illustrate the performance of TD in the presence of measurement noise. Suppose that the reference signal $v(t)$ is contaminated by its 1% Gaussian noise, that is, the input of TD is $v(t) + 0.01v(t)$ instead of $v(t)$. The results by LTD are plotted in Figure 3 where in Figure 3(a), $R = 10$ is taken, and in Figure 3(b) $R = 20$. The results by FTSTD are plotted in Figure 4 where in Figure 4(a), $R = 10$ is taken, while in Figure 4(b) $R = 20$. It is seen from Figures 3 and 4 that both TD can filter the noise to some extent. So in the presence of noise, the performance is getting worse as R is

(a) $R = 10$
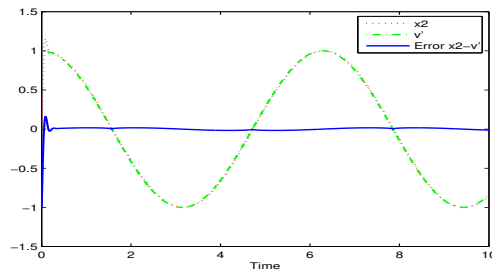
(b) $R = 20$

Figure 2: Derivative tracking for $v(t) = \sin t$ by NLTD.

increasing. Therefore, there is always a tradeoff between accuracy and noise filtering in applications.

Comparing Figure 3 over Figure 4, we find that for the same tuning parameter R, LTD is better than NLTD in the presence of measurement noise. This is another tradeoff in choosing LTD and NLTD in practice.

There are many theoretical issues to be addressed further. The first problem is convergence for the time optimal control system based tracking differentiator. For the second order system:

$$(2.34) \qquad \begin{cases} \dot{x}_1(t) = x_2(t), \\ \dot{x}_2(t) = u(t), \ |u(t)| \leq R, \ R > 0. \end{cases}$$

The time-optimal feedback control is

$$(2.35) \qquad u(x_1(t), x_2(t)) = -R\,\mathrm{sign}\left(x_1(t) + \frac{x_2(t)|x_2(t)|}{2R}\right).$$
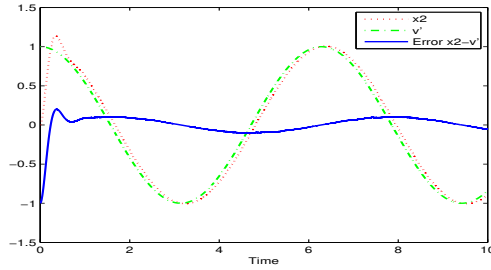
(a) $R = 10$
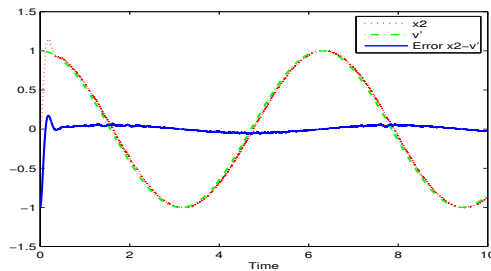
(b) $R = 20$

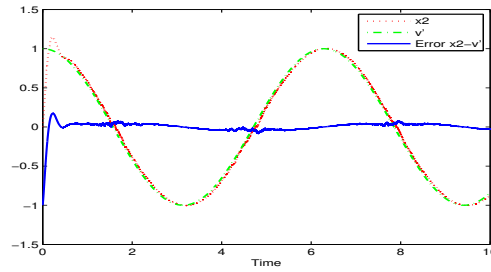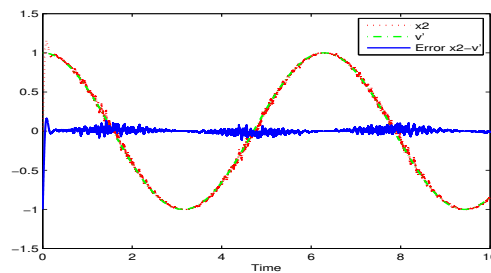Figure 3: Derivative tracking for $v(t) = \sin t$ by LTD.

It is this state feedback control (2.35) that drives the system from the initial state to zero equilibrium in shortest time. Based on the time-optimal feedback control, we can construct the following TD:

$$(2.36) \qquad \begin{cases} \dot{z}_{1R}(t) = x_{2R}(t), \\[2mm] \dot{x}_{2R}(t) = -R\,\mathrm{sign}\left(z_{1R}(t) - v(t) + \dfrac{z_{2R}(t)|z_{2R}(t)|}{2R}\right). \end{cases}$$

The numerical results show that TD (2.36) is convergent and quite fast. However, the convergence of TD (2.36) remains open. In [34], a desired transient profile of type (2.36) is specially emphasized.

The second problem is that whether a tracking differentiator based on an attractive system only is also convergent. In this section, the zero equilibrium state of the reference free system of TD, that is, the TD with reference zero and parameter identical to one, is assumed to be asymptotically stable. However, we do have systems that are attracting yet not Lyapunov stable

(a) $R = 10$

(b) $R = 20$

Figure 4: Derivative tracking for $v(t) = \sin t$ by NTD.

(hard to imagine physically). It is interesting to know if an attractive based TD is also convergent.

Another problem is that except LTD, convergence for other tracking differentiators is in the sense of distribution which is weaker than uniform convergence. It is interesting to know whether a nonlinear weak convergent tracking differentiator is also convergent uniformly.

3. Extended state observer

As streamlined in section 1, the second key part of ADRC is the extended state observer (ESO). ESO is an extension of the state observer in modern control theory. A state observer is an auxiliary system that provides an estimation of the internal state of a given real system from its input and

output. For the linear system of the following:

(3.1)
$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ y(t) = Cx(t), \end{cases}$$

where $x(t) \in \mathbb{R}^n (n \geq 1)$ is the state, $u(t) \in \mathbb{R}^m$ is the control (input), and $y(t) \in \mathbb{R}^l$ is the output (measurement), when $n = l$, the whole state is measured and the state observer is unwanted, while if $n > l$, the classical Luenberger observer can be designed in the following way to recover the whole state by the input and output:

(3.2)
$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + L(y(t) - C\hat{x}(t)),$$

where the matrix L is chosen so that $A - LC$ is Hurwitz. This "copy of the plant plus injection of the output" is the general principle in designing an observer. It is readily shown that the observer error $x(t) - \hat{x}(t) \to 0$ as $t \to \infty$. The existence of the gain matrix L is guaranteed by the detectability of system (3.1). In addition, when $\Sigma(A, B)$ is stabilizable with stabilizing feedback back $u(t) = Fx(t)$, then an observer-based output feedback $u(t) = F\hat{x}(t)$ also stabilizes system (3.1). This is the well-known separation principle for linear systems, which is, however, not always true for nonlinear systems.

The observer design is a relatively independent topic in control theory. There are huge works dedicated to observer design. For more details of the state observer we refer to recent monograph [4].

ESO is a ground breakthrough of state observer where not only state but also "total disturbance" are estimated. The "total disturbance" can come from un-modeled system dynamics, unknown coefficient of control, and external disturbance. Since in ESO, the "external disturbance" and state of the system are estimated simultaneously, we can design an output feedback control law which is not critically reliant to the mathematical model. A first ESO was proposed by Han in [31] as

(3.3)
$$\begin{cases} \dot{\hat{x}}_1(t) = \hat{x}_2(t) - \alpha_1 g_1(\hat{x}_1(t) - y(t)), \\ \dot{\hat{x}}_2(t) = \hat{x}_3(t) - \alpha_2 g_2(\hat{x}_1(t) - y(t)), \\ \qquad \vdots \\ \dot{\hat{x}}_n(t) = \hat{x}_{n+1}(t) - \alpha_n g_n(\hat{x}_1(t) - y(t)) + u(t), \\ \dot{\hat{x}}_{n+1}(t) = -\alpha_{n+1} g_{n+1}(\hat{x}_1(t) - y(t)), \end{cases}$$

for the following single-input single-output (SISO) system

$$
(3.4) \quad
\begin{cases}
\dot{x}_1(t) = x_2(t), \\
\dot{x}_2(t) = x_3(t), \\
\quad \vdots \\
\dot{x}_n(t) = f(t, x_1(t), x_2(t), \ldots, x_n(t)) + w(t) + u(t), \\
y(t) = x_1(t),
\end{cases}
$$

where $u(t) \in C(\mathbb{R}, \mathbb{R})$ is the control (input), $y(t)$ is the output (measurement), $f \in C(\mathbb{R}^n, \mathbb{R})$ is the system function which is possibly unknown, $w \in C(\mathbb{R}, \mathbb{R})$ is an unknown external disturbance. $f(\cdot, t) + w(t)$ is called the "total disturbance" or "extended state", and $\alpha_i \in \mathbb{R}, i = 1, 2, \ldots, n + 1$ are the tuning parameters. By appropriately choosing the nonlinear functions $g_i \in C(\mathbb{R}, \mathbb{R})$, and tuning the parameters α_i, we expect that the states $\hat{x}_i(t), i = 1, 2, \ldots, n + 1$ of the ESO (3.3) can approximately recover the states $x_i(t), i = 1, 2, \ldots, n$ and the extended state $f(\cdot, t) + w(t)$.

In order to be easy of use for engineers, in [10], Gao simplified ESO (3.3) by introducing, in terms of bandwidth, one parameter tuning linear ESO (LESO) for second order nonlinear systems, which is extended to the n-th order nonlinear system (3.4) in [56], as follows:

$$
(3.5) \quad
\begin{cases}
\dot{\hat{x}}_1(t) = \hat{x}_2(t) + \dfrac{\alpha_1}{\varepsilon}(y(t) - \hat{x}_1(t)), \\[2mm]
\dot{\hat{x}}_2(t) = \hat{x}_3(t) + \dfrac{\alpha_2}{\varepsilon^2}(y(t) - \hat{x}_1(t)), \\[2mm]
\quad \vdots \\[2mm]
\dot{\hat{x}}_n(t) = \hat{x}_{n+1}(t) + \dfrac{\alpha_n}{\varepsilon^n}(y(t) - \hat{x}_1(t)) + u(t), \\[2mm]
\dot{\hat{x}}_{n+1}(t) = \dfrac{\alpha_{n+1}}{\varepsilon^{n+1}}(y(t) - \hat{x}_1(t)),
\end{cases}
$$

where $\alpha_i, i = 1, 2, \ldots, n + 1$ are pertinent constants chosen by virtue of pole assignment, and ε which is ω in [56], is the bandwidth. This is likewise in (1.13), $\omega = -1/(2\varepsilon)$, $\omega_0 = \sqrt{3}/(2\varepsilon)$. Generally, if the total disturbance changes fast, in order the states of ESO track the states and "total disturbance", the tuning parameter ε should be tuned to be large accordingly. Essentially, this parameter design method is similar with parameter design in high-gain observer, but in classical high-gain observer, only system state

is estimated and there is no "total disturbance" estimation ([39]). The convergence of LESO is investigated in [56]. As a consequence of convergence of nonlinear ESO (NLESO), the convergence of LESO (3.3) can also be found in [24]. The succeeding Theorem 3.1 is on convergence of linear ESO (3.3) presented in [24].

The following Assumption (H1) is about prior assumption on the unknown nonlinear function $f(t, x)$ and the external disturbance $w(t)$.

Assumption (H1). The possibly unknown functions $f(t, x)$ and $w(t)$ are continuously differentiable with respect to their variables, and

$$(3.6) \qquad |u(t)| + |f(t, x)| + |\dot{w}(t)| + \left| \frac{\partial f(t, x)}{\partial t} \right| + \left| \frac{\partial f(t, x)}{\partial x_i} \right|$$

$$\leq c_0 + \sum_{j=1}^{n} c_j |x_j|^k, \quad \forall\, t \geq 0,\; x = (x_1, x_2, \ldots, x_n),$$

for some positive constants $c_j, j = 0, 1, \ldots, n$ and positive integer k.

The Assumption (H2) is a priori assumption about solution.

Assumption (H2). The solution of (3.4) and the external disturbance $w(t)$ satisfy $|w(t)| + |x_i(t)| \leq B$ for some constant $B > 0$ and all $i = 1, 2 \ldots, n$, and $t \geq 0$.

Theorem 3.1. *If the matrix E defined by (3.7) below:*

$$(3.7) \qquad E = \begin{pmatrix} -\alpha_1 & 1 & 0 & \cdots & 0 \\ -\alpha_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\alpha_n & 0 & 0 & \cdots & 1 \\ -\alpha_{n+1} & 0 & 0 & \cdots & 0 \end{pmatrix},$$

is Hurwitz and Assumptions (H1)–(H2) are satisfied, then
(i). For every positive constant $a > 0$,

$$\lim_{\varepsilon \to 0} |x_i(t) - \hat{x}_i(t)| = 0 \text{ uniformly in } t \in [a, \infty).$$

(ii). For any $\varepsilon > 0$ there exists $t_\varepsilon > 0$ depending on initial value such that

$$|x_i(t) - \hat{x}_i(t)| \leq \Gamma_i \varepsilon^{n+2-i}, \quad \forall\, t \geq t_\varepsilon,$$

where Γ_i is an ε and initial value independent constant, $x_i(t)$ and $\hat{x}_i(t)$ are solutions of (3.4) and (3.5) respectively, $i = 1, 2, \ldots, n + 1$, and $x_{n+1}(t) = f(t, \cdot) + w(t)$ is the extended state for system (3.4).

The one parameter tuning NLESO proposed in [24] is as follows:

(3.8)
$$\begin{cases} \dot{\hat{x}}_1(t) = \hat{x}_2(t) + \varepsilon^{n-1} g_1 \left(\dfrac{y(t) - \hat{x}_1(t)}{\varepsilon^n} \right), \\[2ex] \dot{\hat{x}}_2(t) = \hat{x}_3(t) + \varepsilon^{n-2} g_2 \left(\dfrac{y(t) - \hat{x}_1(t)}{\varepsilon^n} \right), \\[1ex] \quad\vdots \\[1ex] \dot{\hat{x}}_n(t) = \hat{x}_{n+1}(t) + g_n \left(\dfrac{y(t) - \hat{x}_1(t)}{\varepsilon^n} \right) + u(t), \\[2ex] \dot{\hat{x}}_{n+1}(t) = \dfrac{1}{\varepsilon} g_{n+1} \left(\dfrac{y - \hat{x}_1(t)}{\varepsilon^n} \right), \end{cases}$$

where ε is constant tuning gain and $g_i(\cdot), i = 1, 2, \ldots, n + 1$ are pertinent chosen functions. The NLESO (3.8) is a special case of (3.3) and a nonlinear extension of the LESO (3.5) for gain ε.

The nonlinear functions $g_i(\cdot)(i = 1, 2, \ldots, n + 1)$ are chosen such that the following Assumption (H3) holds.

Assumption (H3). There exist constants $\lambda_i(i = 1, 2, 3, 4)$, α, β, and positive definite, continuous differentiable functions $V, W : \mathbb{R}^{n+1} \to \mathbb{R}$ such that

- $\lambda_1 \|y\|^2 \leq V(y) \leq \lambda_2 \|y\|^2$, $\lambda_3 \|y\|^2 \leq W(y) \leq \lambda_4 \|y\|^2$,

- $\displaystyle\sum_{i=1}^{n} \frac{\partial V(y)}{\partial y_i} (y_{i+1} - g_i(y_1)) - \frac{\partial V(y)}{\partial y_{n+1}} g_{n+1}(y_1) \leq -W(y)$,

- $\left| \dfrac{\partial V(y)}{\partial y_{n+1}} \right| \leq \beta \|y\|$,

where $y = (y_1, y_2, \ldots, y_{n+1})$.

Theorem 3.2. *Suppose that Assumption (H1)–(H3) are satisfied. Then*

(i) *for every positive constant $a > 0$,*

$$\lim_{\varepsilon \to 0} |x_i(t) - \hat{x}_i(t)| = 0 \ \text{uniformly in } t \in [a, \infty);$$

(ii) $\overline{\lim\limits_{t \to \infty}} |x_i(t) - \hat{x}_i(t)| \leq O\left(\varepsilon^{n+2-i}\right),$

where $x_i(t)$ and $\hat{x}_i(t)$ are the solutions of (3.4) and (3.8) respectively, $i = 1, 2, \ldots, n+1$, and $x_{n+1}(t) = f(t, \cdot) + w(t)$ is the extended state variable of system (3.4).

The further investigation shows that Assumption (H3) can be weakened by Assumption (H4). ([24]):

Assumption (H4). There exist constants $R, \alpha > 0$, and positive definite, continuous differentiable functions $V, W : \mathbb{R}^{n+1} \to \mathbb{R}$ such that for $y = (y_1, y_2, \ldots, y_{n+1})$,

- $\{y | V(y) \le d\}$ is bounded for any $d > 0$,

- $\displaystyle\sum_{i=1}^{n} \frac{\partial V(y)}{\partial y_i}(y_{i+1} - g_i(y_1)) - \frac{\partial V(y)}{\partial y_{n+1}} g_{n+1}(y_1) \le -W(y),$

- $\left| \dfrac{\partial V(y)}{\partial y_{n+1}} \right| \le \alpha W(y)$ for $\|y\| > R$.

Under this weaker condition, the convergence obtained in [24] is accordingly slightly weaker than Theorem 3.2. However, the class of nonlinear functions $g_i(\cdot)$ for constructing the nonlinear ESO is also broadened.

Theorem 3.3. *Under Assumptions (H1),(H2), and (H4), the nonlinear extended state observer (3.8) is convergent in the sense that for any $\sigma \in (0, 1)$, there exists $\varepsilon_\sigma \in (0, 1)$ such that for any $\varepsilon \in (0, \varepsilon_\sigma)$,*

$$|x_i(t) - \hat{x}_i(t)| < \sigma, \quad \forall\, t \in (T_\varepsilon, \infty),$$

where $T_\varepsilon > 0$ depends on ε and initial value, $x_i(t)$ and $\hat{x}_i(t)$ are the solutions of (3.4) and (3.8) respectively, $i = 1, 2, \ldots, n+1$, and $x_{n+1}(t) = f(t, \cdot) + w(t)$ is the extended state variable of system (3.4).

It is verified in [24] that the following nonlinear ESO satisfies Assumption (H4):

$$(3.9) \quad \begin{cases} \dot{\hat{x}}_1(t) = \hat{x}_2(t) + 3\varepsilon \left[\dfrac{y(t) - \hat{x}_1(t)}{\varepsilon^2} \right]^{\alpha}, \\[2ex] \dot{\hat{x}}_2(t) = \hat{x}_3(t) + 3 \left[\dfrac{y(t) - \hat{x}_1(t)}{\varepsilon^2} \right]^{2\alpha - 1} + u(t), \\[2ex] \dot{\hat{x}}_3(t) = \dfrac{1}{\varepsilon} \left[\dfrac{y(t) - \hat{x}_1(t)}{\varepsilon^2} \right]^{3\alpha - 2}, \end{cases}$$

for some $\alpha \in (0,1)$.

Now we use numerical methods to illustrate the effectiveness of LESO and NLESO (3.9). The considered system is the following system with large uncertainty:

(3.10)
$$\begin{cases} \dot{x}_1(t) = x_2(t), \\ \dot{x}_2(t) = f(x_1(t), x_2(t)) + w(t) + u(t), \\ y(t) = x_1(t), \end{cases}$$

where $f(x_1(t), x_2(t))$ is the system unknown function, $w(t)$ is the external disturbance, and $x_3(t) \triangleq f(x_1(t), x_2(t)) + \sin(x_1(t) + x_2(t)) + w(t)$ is the total disturbance. The LESO is designed as follows:

(3.11)
$$\begin{cases} \dot{\hat{x}}_1(t) = \hat{x}_2(t) + \dfrac{3}{\varepsilon}(y(t) - \hat{x}_1(t)), \\ \dot{\hat{x}}_2(t) = \hat{x}_3(t) + \dfrac{3}{\varepsilon^2}(y(t) - \hat{x}_1(t)) + u(t), \\ \dot{\hat{x}}_3(t) = \dfrac{1}{\varepsilon^3}(y(t) - \hat{x}_1(t)). \end{cases}$$

In numerical simulation, we use $f(x_1(t), x_2(t)) = -x_1(t) - x_2(t) + \sin(x_1(t) + x_2(t))$, $u(t) = \sin(t)$, $w(t) = 1 + \cos(t) + \sin 2t$, $\alpha = 0.8$ and $\varepsilon = 1/20$. The numerical results by the LESO (3.11) is plotted in Figure 5, and NLESO (3.9) in Figure 6.

It is seen from Figures 5 and 6 that both LESO and NLESO track the system state and total disturbance satisfactorily. Comparing Figure 5 with Figure 6 we can find that for the same tuning parameter, the NLESO takes advantage of accurate estimating and much smaller peaking value near the initial stage. On the other hand the tracking speed of LESO is faster than NLESO.

We should point out that there are two drawbacks in constant gain ESO. The first one is the peaking problem and the other is that only practical convergence can be achieved for general uncertainty. Although the NLESO has smaller peaking value near the initial stage, but it may arise again as ε increases. To tackle this problem, we proposed a time-varying gain ESO in
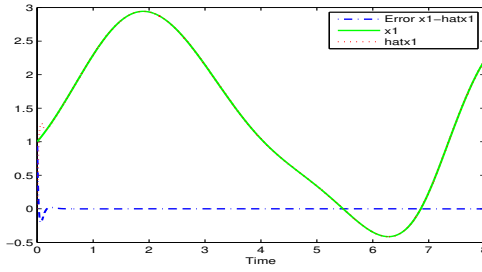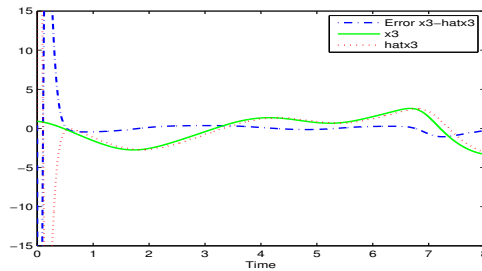
(a) $x_1(t)$(green), $\hat{x}_1(t)$(red), $x_1(t) - \hat{x}_1(t)$(blue)
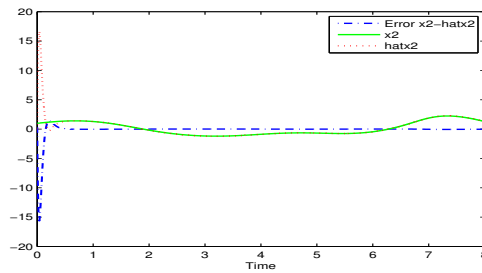
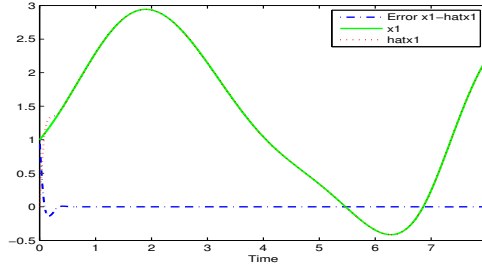(b) $x_2(t)$(green), $\hat{x}_2(t)$(red), $x_2(t) - \hat{x}_2(t)$(blue)
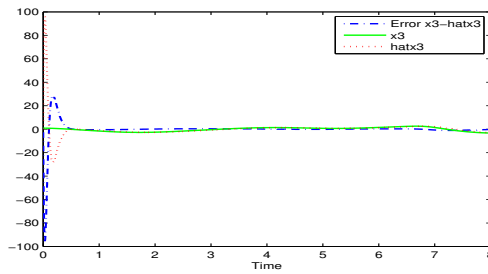
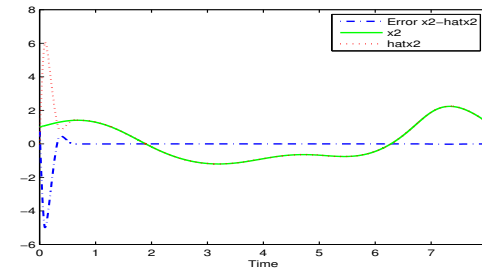(c) $x_3(t)$(green), $\hat{x}_3(t)$(red), $x_3(t) - \hat{x}_3(t)$(blue)

(d) Magnification of Figure 5(c)
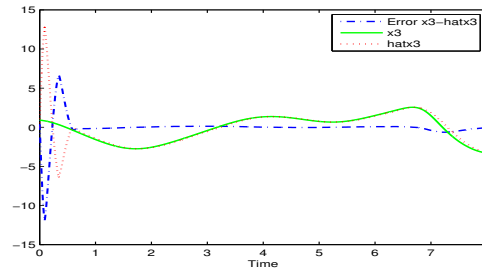
Figure 5:   Linear ESO (3.11) for system (3.10).

(a) $x_1(t)$(green), $\hat{x}_1(t)$(red), $x_1(t) - \hat{x}_1(t)$(blue)

(b) $x_2(t)$(green), $\hat{x}_2(t)$(red), $x_2(t) - \hat{x}_2(t)$(blue)

(c) $x_3(t)$(green), $\hat{x}_3(t)$(red), $x_3(t) - \hat{x}_3(t)$(blue)

Figure 6: NLESO (3.9) for system (3.10).

[51] as follows:

$$(3.12) \quad \begin{cases} \dot{\hat{x}}_1(t) = \hat{x}_2(t) + \frac{1}{r^{n-1}(t)}g_1(r^n(t)(y(t) - \hat{x}_1(t))), \\ \dot{\hat{x}}_2(t) = \hat{x}_3(t) + \frac{1}{r^{n-2}(t)}g_2(r^n(t)(y(t) - \hat{x}_1(t))), \\ \quad\quad \vdots \\ \dot{\hat{x}}_n(t) = \hat{x}_{n+1}(t) + g_n(r^n(t)(y(t) - \hat{x}_1(t))) + b_0 u(t), \\ \dot{\hat{x}}_{n+1}(t) = r(t)g_{n+1}(r^n(t)(y(t) - \hat{x}_1(t))), \end{cases}$$

which is used also to estimate the state $(x_1(t), x_2(t), \ldots, x_n(t))$ and the total disturbance, where $r(t)$ is the time-varying gain to be increasing gradually. When $r(t) \equiv 1/\varepsilon$, (3.12) is reduced to the constant gain ESO (3.8). The suitable time varying gain can reduce dramatically the peaking value of the ESO near the initial time. To achieve the asymptotical convergence of ESO, we assume the time-varying gain $r(t)$ in ESO (3.12) to satisfy Assumption (A1).

Assumption (A1). $r(t) \in C^1([0, \infty), [0, \infty))$, $r(t), r'(t) > 0$, $\lim_{t \to +\infty} r(t) = +\infty$, and there exists a constant $M > 0$ such that $\overline{\lim}_{t \to +\infty} \frac{\dot{r}(t)}{r(t)} \leq M$.

The following Assumption (A2) is on functions $g_i(\cdot)$ in the ESO (3.12).

Assumption (A2). There exist positive constants R and $N > 0$, and radially unbounded, positive definite functions $V \in C^1(\mathbb{R}^{n+1}, [0, \infty))$ and $W \in C(\mathbb{R}^{n+1}, [0, \infty))$ such that

- $\displaystyle\sum_{i=1}^{n} (x_{i+1} - g_i(x_1)) \frac{\partial V(x)}{\partial x_i} - g_{n+1}(x_1) \frac{\partial V(x)}{\partial x_{n+1}} \leq -W(x);$

- $\displaystyle\max_{\{i=1,\ldots,n\}} \left\{ \left| x_i \frac{\partial V(x)}{\partial x_i} \right| \right\} \leq NW(x), \text{ and } \left| \frac{\partial V(x)}{\partial x_{n+1}} \right| \leq NW(x),$

for all $x = (x_1, x_2, \ldots, x_{n+1}), \|x\| \geq R$.

Theorem 3.4. *Assume Assumptions (H1), (A1), and (A2). Then the states of ESO* (3.12) *converge to the states and the extended state of (3.4) in the sense that*

$$(3.13) \qquad \lim_{t \to +\infty} |\hat{x}_i(t) - x_i(t)| = 0, \quad i = 1, 2, \ldots, n+1.$$

In practice, we can design $r(t)$ to be increasing continuously at first and then keep a large constant thereafter. By this idea, we modify the constant gain ε in (3.9) to time-varying $r(t)$ as follows:

$$(3.14) \qquad r(t) = \begin{cases} e^{5t}, & 0 \leq t < \frac{1}{5} \ln 100, \\ 100, & t \geq \frac{1}{5} \ln 100. \end{cases}$$

The numerical results for NLESO (3.9) with time-varying gain are plotted in Figure 7. It is seen from Figure 7 that time-varying gain ESO has the merits of accurate estimation and smaller peaking value in the initial time.

In what follows, we give a brief discussion on the notorious peaking value problem in high gain method, both by constant high gain and time varying gain. To bring the object into focus, we consider the linear ESO (3.5) with
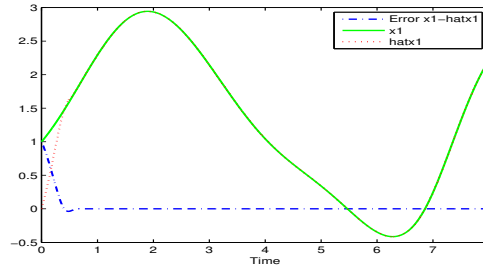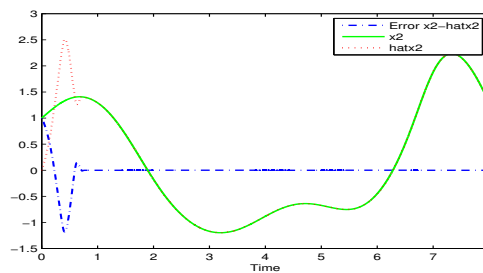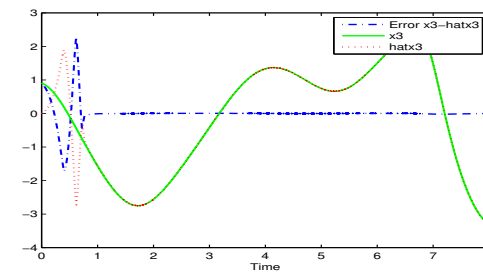
(a) $x_1(t)$(green), $\hat{x}_1(t)$(red), $x_1(t) - \hat{x}_1(t)$(blue)

(b) $x_2(t)$(green), $\hat{x}_2(t)$(red), $x_2(t) - \hat{x}_2(t)$(blue)

(c) $x_3(t)$(green), $\hat{x}_3(t)$(red), $x_3(t) - \hat{x}_3(t)$(blue)

Figure 7: Time-varying gain nonlinear ESO (3.9) for system (3.10).

the Hurwitz matrix E given by (3.7). Suppose that E has $n + 1$ different negative real eigenvalues $\lambda_1, \ldots, \lambda_{n+1}$. First, by constant high gain $1/\varepsilon = r$, the solution of ESO (3.24) is:

$$(3.15) \qquad \hat{x}_i(t) = \frac{1}{r^{n+1-i}} \varepsilon_i(t) + x_i(t),$$

where

(3.16)
$$\begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \vdots \\ \varepsilon_{n+1}(t) \end{pmatrix} = e^{-rEt} \begin{pmatrix} r^n(x_1(0) - \hat{x}_1(0)) \\ r^{n-1}(x_2(0) - \hat{x}_2(0)) \\ \vdots \\ x_{n+1}(0) - \hat{x}_{n+1}(0) \end{pmatrix}$$
$$+ \int_0^t e^{-rE(t-s)} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \dot{x}_{n+1}(s) \end{pmatrix} ds.$$

The peaking value is mainly caused by large initial value of $\varepsilon(t) = (\varepsilon_1(t), \ldots, \varepsilon_{n+1}(t))$:

(3.17)
$$\varepsilon_i(t) = \sum_{j=1}^{n+1} \sum_{l=1}^{n+1} d_{ij}^l e^{rt\lambda_l} r^{n+1-i}(x_i(0) - \hat{x}_i(0))$$
$$+ \sum_{l=1}^{n+1} \int_0^t \dot{x}_{n+1}(s) d_{i(n+1)}^l e^{rt\lambda_l} ds,$$

where d_{ij}^l are real numbers determined by the matrix E. It is seen that the peaking value occurs only at $t = 0$ since for any $a > 0$, $\varepsilon_i(t) \to 0$ as $r \to \infty$ uniformly in $t \in [a, \infty)$. On the other hand, in the initial time stage, however, $e^{rt\lambda_i}$ is very close to 1. This is the reason behind for the peaking value problem by constant high gain. Actually, the peaking values for $\hat{x}_2(t), \ldots, \hat{x}_{n+1}(t)$ are the orders of $r, r^2, \ldots, r^n$, respectively. The larger r is, the larger the peaking values.

Next, when we replace the time varying gain $r(t)$ by $1/\varepsilon$ and let the gain be relatively small in the initial stage, the initial value of error $\eta(t)$ is

(3.18)
$$\left(r(0)^n(x_1(0) - \hat{x}_1(0)), \; r^{n-1}(0)(x_2(0) - \hat{x}_2(0)), \right.$$
$$\left. \ldots, x_{n+1}(0) - \hat{x}_{n+1}(0) \right)^\top,$$

which is also small. Actually the initial value of error $\eta(t)$ is

(3.19)
$$\left((x_1(0) - \hat{x}_1(0)), (x_2(0) - \hat{x}_2(0)), \ldots, x_{n+1}(0) - \hat{x}_{n+1}(0) \right)^\top.$$

Since the gain function $r(t)$ is small in the initial stage, when $\|\varepsilon(t)\|$ increases with increasing of eigenvalues to some given value, $\|\varepsilon(t)\|$ stops increasing at some value which is determined by the system functions and the external

disturbances, but does not rely on the maximal value of $r(t)$. Actually, let $\mathcal{V} : \mathbb{R}^{n+1} \to [0, \infty)$ be $\mathcal{V}(\nu) = \langle P\nu, \nu \rangle, \nu \in \mathbb{R}^{n+1}$ and let the gain function $r(t)$ be chosen as in (3.14), we can prove that the derivative of $\mathcal{V}(\varepsilon(t))$ satisfies

$$(3.20) \qquad \frac{d\mathcal{V}(\varepsilon(t))}{dt} \leq -(r(t) - N_{12})\langle \eta(t), \eta(t) \rangle,$$

where N_{12} is the upper bound of the derivative of total disturbance. When $r(t)$ increases to N_{12}, then $\mathcal{V}(\varepsilon(t))$ stops increasing. This together with

$$(3.21) \qquad \|\varepsilon(t)\| \leq \frac{1}{\lambda_{\max}(P)} \mathcal{V}(\varepsilon(t))$$

shows that $\|\varepsilon(t)\|$ does not increase any more although $\rho(t)$ increases continuously to a large number r or ∞. If $N_{12} \leq 1$, then $\mathcal{V}(\varepsilon(t))$ decreases from the beginning. It then follows from (3.21) and (3.15) that the peaking values become much small.

The ESO is extended to lower-triangular systems with large uncertainty in [52] as follows:

$$(3.22) \qquad \begin{cases} \dot{x}_1(t) = x_2(t) + g_1(u(t), x_1(t)), \\ \dot{x}_2(t) = x_3(t) + g_2(u(t), x_1(t), x_2(t)), \\ \qquad \vdots \\ \dot{x}_n(t) = f(t, x(t), w(t)) + g_n(u(t), x(t)), \\ y(t) = x_1(t), \end{cases}$$

where $g_i \in C(\mathbb{R}^{i+m}, \mathbb{R})$ is the known nonlinear function, $f \in C(\mathbb{R}^{n+s+1}, \mathbb{R})$ is usually an unknown nonlinear function, $x(t) = (x_1(t), x_2(t), \dots, x_n(t))$ is the state of system, $u \in \mathbb{R}^m$ is the input (control), $y(t)$ is the output (measurement), and $w \in C([0, \infty), \mathbb{R})$ is the external disturbance.

The constant high gain ESO is designed to recover both state of system (3.22) and its extended state

$$(3.23) \qquad x_{n+1}(t) \triangleq f(t, x(t), w(t))$$

as follows:

(3.24)   **ESO:**
$$
\begin{cases}
\dot{\hat{x}}_1(t) = \hat{x}_2(t) + \dfrac{1}{r^{n-1}} h_1(r^n(y(t) - \hat{x}_1(t))) \\
\qquad\qquad + g_1(u(t), \hat{x}_1(t)), \\
\qquad\qquad \vdots \\
\dot{\hat{x}}_n(t) = \hat{x}_{n+1}(t) + h_n(r^n(y(t) - \hat{x}_1(t))) \\
\qquad\qquad + g_n(u(t), \hat{x}_1(t), \ldots, \hat{x}_n(t)), \\
\dot{\hat{x}}_{n+1}(t) = r h_{n+1}(r^n(y(t) - \hat{x}_1(t))),
\end{cases}
$$

where r is the constant high gain parameter and $h_i \in C(\mathbb{R}, \mathbb{R})$, $i = 1, 2, \ldots$, $n + 1$ are the design functions.

To achieve convergence of ESO (3.24), some mathematical assumptions are required. The following Assumptions B1 and B2 are on $g_i(\cdot)$ and $f(\cdot)$ in system (3.22).

Assumption B1. $g_i : \mathbb{R}^{i+1} \to \mathbb{R}$ satisfies

(3.25)
$$
|g_i(u, \nu_1, \ldots, \nu_i) - g_i(u, \tilde{\nu}_1, \ldots, \tilde{\nu}_i)|
$$
$$
\leq \Gamma(u) \|(\nu_1 - \tilde{\nu}_1, \ldots, \nu_i - \tilde{\nu}_i)\|^{\theta_i}, \quad \Gamma \in C(\mathbb{R}^m, \mathbb{R}),
$$

where $\theta_i \in ((n-i)/(n+1-i),\ 1]$, $i = 1, 2, \ldots, n$.

The condition (3.25) means that $g_i(\cdot), i = 1, 2, \ldots, n$ are Hölder continuous. For triangular systems, the widely assumed Lipschitz continuity is just a special case of the Hölder continuity with the exponents $\theta_i = 1$. Some systems with appropriate Hölder continuous functions have merit of finite-time stable, and these kinds of functions can be used for feedback control design ([3]).

Assumption B2. $f \in C^1(\mathbb{R}^{n+2}, \mathbb{R})$ satisfies

$$
|f(t, x, w)| + \left| \frac{\partial f(t, x, w)}{\partial t} \right| + \left| \frac{\partial f(t, x, w)}{\partial x_i} \right| + \left| \frac{\partial f(t, x, w)}{\partial w} \right| \leq \varpi_1(x) + \varpi_2(w),
$$

where $i = 1, 2, \ldots, n$, $\varpi_1 \in C(\mathbb{R}^n, [0, \infty))$, $\varpi_2 \in C(\mathbb{R}, [0, \infty))$ are two known functions.

The succeeding Assumption B3 is on the control input $u(t)$ and external disturbance $w(t)$.

Assumption B3. $\sup_{t \in [0, \infty)}(|w(t)| + |\dot{w}(t)| + |u(t)|) < \infty$.

The Assumption B4 below is on functions $h_i(\cdot)$ in ESO (3.24). It proposes a principle of choosing $h_i(\cdot)$.

Assumption B4. All $h_i \in C(\mathbb{R}, \mathbb{R})$ satisfy the following Lyapunov conditions: There exist positive constants $R, N > 0$, and continuous, radially unbounded, positive definite functions $V, W \in C(\mathbb{R}^{n+1}, [0, \infty))$ such that

1) $\displaystyle\sum_{i=1}^{n} (\nu_{i+1} - h_i(\nu_1))\frac{\partial V(\nu)}{\partial \nu_i} - h_{n+1}(\nu_1)\frac{\partial V(\nu)}{\partial \nu_n} \le -W(\nu), \quad \forall \nu = (\nu_1, \nu_2,$
$\dots, \nu_{n+1}) \in \mathbb{R}^{n+1}$;

2) $\displaystyle\max_{i=1,\dots,n} \left\{ \|(\nu_1, \dots, \nu_i)\|^{\theta_i} \left| \frac{\partial V(\nu)}{\partial \nu_i} \right| \right\} \le NW(\nu), \left| \frac{\partial V(\nu)}{\partial \nu_{n+1}} \right| \le NW(\nu), \nu \in$
$\mathbb{R}^{n+1}, \|\nu\| \ge R.$

The Assumption B4 guarantees that the zero equilibrium of the following system

$$\dot{\nu}(t) = \big(\nu_2(t) - h_1(\nu_1(t)),$$
$$\dots, \nu_{n+1}(t) - h_n(\nu_1(t)), -h_{n+1}(\nu_1(t))\big)^\top, \quad \nu \in \mathbb{R}^{n+1}$$

is asymptotically stable.

Theorem 3.5. *Assume that Assumptions B1-B4 and suppose that the solution of (3.22) is globally bounded. Then the states of ESO (3.24) converge practically to the states and extended state of system (3.22): For any $\sigma > 0$, there exists a positive constant $r_0 > 0$ such that*

(3.26) $\qquad |\hat{x}_i(t) - x_i(t)| < \sigma, \quad \forall t > t_r, \ r > r_0, \ i = 1, 2, \dots, n+1,$

where t_r is an r and initial value dependent constant.

The first concrete ESO of Theorem 3.5 is certainly the LESO, that is, the nonlinear functions $h_i(\cdot)(i = 1, 2, \dots, n+1)$ in (3.24) are linear functions: $h_i(\nu) = \alpha_i \nu, \ \nu \in \mathbb{R}$, where α_i are constants to be specified. Let the matrix

(3.27) $$E = \begin{pmatrix} -\alpha_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\alpha_n & 0 & \cdots & 1 \\ -\alpha_{n+1} & 0 & \cdots & 0 \end{pmatrix}$$

be Hurwitz and assume Assumptions B2 and B3, and Assumption B1 with all $\theta_i = 1$. Then there exists $r_0 > 0$ such that

$$(3.28) \quad |x_i(t) - \hat{x}_i(t)| \leq D_1 \left(\frac{1}{r}\right)^{n+2-i} \quad, \quad r > r_0, \ t > t_r, \ i = 1, 2, \ldots, n + 1,$$

where t_r is an r and initial value dependent constant and D_1 is an r-independent constant only.

The second concrete ESO of Theorem 3.5 is homogeneous ESO (HESO), that is, the nonlinear functions $h_i(\cdot)$ are chosen as (3.29),

$$(3.29) \qquad h_i(\nu) = [\nu]^{i\beta-(i-1)} \triangleq \text{sign}(\nu)|\nu|^{i\beta-(i-1)}, \quad \nu \in \mathbb{R}, \ \beta \in (0,1).$$

Assume that Assumptions B1-B3 hold true with $\theta_i \in (0, 1]$ and the matrix in (3.27) is Hurwitz. Then there exists $\beta^* \in (0, 1)$ such that for any $\beta \in (\beta^*, \ 1)$, there exists an $r_0 > 0$ such that for all $r > r_0, i \in \{1, 2, \ldots, n + 1\}$,

$$(3.30) \qquad \overline{\lim_{t \to \infty}} |x_i(t) - \hat{x}_i(t)| \leq D_2 \left(\frac{1}{r}\right)^{n+1-i+\frac{(i-1)\beta-(i-2)}{(n+1)\beta-n}(1-\Lambda)},$$
$$r > r_0, \ t > r_r, \ i = 1, 2, \ldots, n + 1,$$

where $\Lambda = \max(n + 1 - i)(1 - \theta_i)$, t_r is an r-dependent constant, D_2 is an r-independent.

In [27], ESO is extended to the following MIMO systems with large uncertainty:

$$(3.31) \quad \begin{cases} x_1^{(n_1)}(t) = f_1\left(x_1(t), \ldots, x_1^{(n_1-1)}(t), \ldots, x_m^{(n_m-1)}(t), w_1(t)\right) \\ \qquad \qquad +g_1(u_1(t), \ldots, u_k(t)), \\ x_2^{(n_2)}(t) = f_2\left(x_1(t), \ldots, x_1^{(n_1-1)}(t), \ldots, x_m^{(n_m-1)}(t), w_2(t)\right) \\ \qquad \qquad +g_2(u_1(t), \ldots, u_k(t)), \\ \qquad \vdots \\ x_m^{(n_m)}(t) = f_m\left(x_1(t), \ldots, x_1^{(n_1-1)}(t), \ldots, x_m^{(n_m-1)}(t), w_m(t)\right) \\ \qquad \qquad +g_m(u_1(t), \ldots, u_k(t)), \\ y_i(t) = x_i(t), \quad i = 1, 2, \ldots, m, \end{cases}$$

where $n_i \in \mathbb{Z}$, $f_i \in C(\mathbb{R}^{n_1+n_2+\cdots+n_m+1})$ represents the system function, $w_i \in C([0, \infty), \mathbb{R})$ the external disturbance, $u_i \in C([0, \infty), \mathbb{R})$ the control (input), $y_i(t)$ the observations (output), $g_i \in C(\mathbb{R}^k, \mathbb{R})$.

We have presented some convergence results of ESO for open-loop systems. It is fairly to say that ESO is a systematical design to realize ultimate

goal of estimating total disturbance from the measured output, which covers much wide class of disturbances both internally and externally. It is the ESO that estimates the first time and in large scale, the unknown internal uncertainty while other modern control methods estimate mainly the external disturbance. ESO constitutes the major key part toward ADRC technology. By cancellation, it transforms essentially, beforehand the feedback taking place, the control system that is seemly mathematically formidable due to variant forms of total disturbance to be a canonical system (linear system most often) for which we have countless ways to deal with. Moreover, some sophisticated part of the plant which is hardly to be dealt with by practitioner can also be treated as part of total disturbance, which changes significantly the concept of disturbance and spans the applicability of ESO. In this sense, ESO shows the characteristic of feedforward control although it does utilize the measurement of the state.

As a completely new strategy, there are many theoretical problems needed to be further investigated. The first problem is the tuning parameter. Basically, our design is based typically on a high gain approach:

(3.32)
$$\begin{cases} \dot{\hat{x}}_1(t) = \hat{x}_2(t) - \beta_1 g_1(\hat{x}_1(t) - y(t)), \\ \dot{\hat{x}}_2(t) = \hat{x}_3(t) - \beta_2 g_2(\hat{x}_1(t) - y(t)), \\ \quad\vdots \\ \dot{\hat{x}}_n(t) = \hat{x}_{n+1}(t) - \beta_n g_n(\hat{x}_1(t) - y(t)) + u(t), \\ \dot{\hat{x}}_{n+1}(t) = -\beta_{n+1} g_{n+1}(\hat{x}_1(t) - y(t)), \end{cases}$$

where

(3.33)
$$\beta_1 = \frac{1}{\varepsilon}, \quad \beta_2 = \frac{1}{\varepsilon^2}, \quad \ldots, \quad \beta_{n+1} = \frac{1}{\varepsilon^{n+1}}.$$

Although as pointed out in section 1, high gain is unavoidable in ESO design, there are many other ways to choose the gain to improve the performance. It is indicated in [33], based on numerical simulations, that ESO with $n + 1$ tuning parameters $\beta_1, \beta_2, \ldots, \beta_{n+1}$ can be chosen based on Fibnacci sequence:

(3.34)
$$\beta_1 = \frac{1}{h}, \quad \beta_2 = \frac{1}{3h^2}, \quad \beta_3 = \frac{2}{8^2 h^3}, \quad \beta_4 = \frac{5}{13^3 h^4}, \ldots,$$

which can achieve more satisfactory estimation both for system state and total disturbances, where h is the sample step length. But a mathematical justification is required for this choice. In addition, more tuning parameters

in (3.32) are expected to be more effective than one parameter high gain tuning.

In [33], it indicates, based on numerical experiments again, that the following nonlinear ESO:

$$
(3.35) \qquad
\begin{cases}
\dot{\hat{x}}_1(t) = \hat{x}_2(t) - \beta_1(\hat{x}_1(t) - x_1(t)), \\
\dot{\hat{x}}_2(t) = \hat{x}_3(t) - \beta_2 \mathrm{fal}(\hat{x}_1(t) - x_1(t), 1/2, \delta), \\
\dot{\hat{x}}_3(t) = -\beta_3 \mathrm{fal}(\hat{x}_1(t) - x_1(t), 1/4, \delta),
\end{cases}
$$

can also recover state and total disturbance simultaneously for the second order control systems with uncertainty, where

$$
(3.36) \qquad
\mathrm{fal}(e, \alpha, \delta) =
\begin{cases}
\dfrac{e}{\delta^{\alpha-1}}, & |e| \le \delta, \\
|e|^{\alpha}\mathrm{sign}(e), & |e| > \delta,
\end{cases}
\qquad \delta > 0, \ \alpha > 0,
$$

with $\beta_{1,2,3}$ being chosen as (3.34). However convergence for this special nonlinear ESO is still open. A recent progress for convergence of ESO with nonlinear function of type (3.36) has been made in [55].

4. Active disturbance rejection control: ESO based feedback

The last key part of ADRC is TD and ESO based feedback control. In the feedback loop, one component is to cancel (compensate) the "total disturbance" using its estimation obtained from ESO. For simplicity, we start from ADRC for the following SISO system:

$$
(4.37) \qquad
\begin{cases}
\dot{x}_1(t) = x_2(t), \\
\dot{x}_2(t) = x_3(t), \\
\qquad \vdots \\
\dot{x}_n(t) = f(t, x(t), \zeta(t), w(t)) + b(w(t))u(t), \\
\dot{\zeta}(t) = F_0(x(t), \zeta(t), w(t)), \\
y(t) = x_1(t),
\end{cases}
$$

where $x(t) = (x_1(t), x_2(t), \dots, x_n(t))^\top \in \mathbb{R}^n$, $\zeta \in \mathbb{R}^m$ are states, $y(t) \in \mathbb{R}$ is the output, $u(t) \in \mathbb{R}$ is the input, $f \in C^1(\mathbb{R}^{n+m+2}, \mathbb{R})$ and $F_0 \in C^1(\mathbb{R}^{n+m+1}, \mathbb{R}^m)$ are unknown system functions, $b \in C^1(\mathbb{R}^1, \mathbb{R})$ is control magnifying

coefficient which also contains some uncertainty, but there is a known nominal value b_0 close to $b(\cdot)$, $w \in C^1(\mathbb{R}, \mathbb{R})$ is the external disturbance. In standard control formulation, $\zeta(t)$ represents the zero dynamics.

The control objective of ADRC is to design a state and disturbance observer-based output feedback control so that the output $y(t)$ tracks a given reference signal $v(t)$, and at the same time $x_i(t)$ tracks $v^{(i-1)}(t)$ for every $i = 2, 3, \ldots, n$ provided that the latter exist in some sense.

To have a holistic feeling about what ADRC is about, let us restate the three key components of ADRC. The first component is to recover all $v^{(i-1)}(t), i = 2, \ldots, n+1$ through the reference $v(t)$ itself. This is realized by tracking differentiator (TD):

$$(4.38) \quad \begin{cases} \dot{z}_{1R}(t) = z_{2R}(t), \\ \quad \vdots \\ \dot{z}_{nR}(t) = z_{(n+1)R}(t), \\ \dot{z}_{(n+1)R}(t) = R^n \psi \left(z_{1R}(t) - v(t), \dfrac{z_{2R}(t)}{R}, \ldots, \dfrac{z_{(n+1)R}(t)}{R^n} \right), \\ \psi(0, 0, \ldots, 0) = 0, \end{cases}$$

where R is the tuning parameter and $\psi \in C(\mathbb{R}^{n+1}, \mathbb{R})$. The system (4.38) also serves as a transient profile for output tracking where $y(t)$ tracks $z_{1R}(t)$ instead of $v(t)$ to avoid setpoint jump.

The second component of ADRC is ESO which is used to recover the state $(x_1(t), \ldots, x_{n+1}(t))^\top$ of system (4.37), and the total disturbance $x_{n+1}(t) = f(t, x(t), \zeta(t), w(t)) + (b(w(t)) - b_0)u(t)$:

$$(4.39) \quad \begin{cases} \dot{\hat{x}}_1(t) = \hat{x}_2(t) + \varrho^{n-1}(t)g_1(\theta(t)), \\ \dot{\hat{x}}_2(t) = \hat{x}_3(t) + \varrho^{n-2}(t)g_2(\theta(t)), \\ \quad \vdots \\ \dot{\hat{x}}_n(t) = \hat{x}_{n+1}(t) + g_n(\theta(t)) + b_0 u(t), \\ \dot{\hat{x}}_{n+1}(t) = \dfrac{1}{\varrho(t)} g_{n+1}(\theta(t)), \end{cases}$$

where $\theta(t) = (y(t) - \hat{x}_1(t))/\varrho^n(t)$, $y(t) = x_1(t)$, $g_i \in C(\mathbb{R}, \mathbb{R})$, and $\varrho \in C([0, \infty), \mathbb{R}^+)$ is the gain function which is chosen here as a time-varying

gain to satisfy:

$$\text{(4.40)} \qquad \varrho(0) = 1; \ \dot{\varrho}(t) = -a\varrho(t), \ a > 0, \ \text{if } \varrho(t) > \varepsilon;$$
$$\text{and } \dot{\varrho}(t) = 0, \ \text{otherwise,}$$

where $\varepsilon > 0$ is a positive constant.

The third and also the last component of ADRC is to design ESO based output feedback control law:

$$\text{(4.41)} \qquad \text{ADRC}: \ u(t) = \frac{1}{b_0} \left[\varphi(\hat{x}(t) - z_R(t)) + z_{(n+1)R}(t) - \hat{x}_{n+1}(t) \right],$$

where $(\hat{x}(t) = (\hat{x}_1(t), \hat{x}_2(t), \ldots, \hat{x}_n(t)), \hat{x}_{n+1}(t))$ is the solution of (4.39) and $(z_R(t) = (z_{1R}(t), z_{2R}(t), \ldots, z_{nR}(t)), z_{(n+1)R}(t))$ is the solution of (4.38). The role played by $\hat{x}_{n+1}(t)$ is to cancel the total disturbance $x_{n+1}(t)$. The nonlinear function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is chosen so that zero equilibrium state of the following target error profile system is asymptotically stable:

$$\text{(4.42)} \qquad \begin{cases} \dot{x}_1^*(t) = x_2^*(t), \\ \dot{x}_2^*(t) = x_3^*(t), \\ \quad \vdots \\ \dot{x}_n^*(t) = \varphi(x_1^*(t), \ldots, x_n^*(t)), \ \varphi(0, 0, \ldots, 0) = 0. \end{cases}$$

Roughly speaking, under ADRC (4.41), the trajectory of the error $(x_1(t) - v(t), \ldots, x_n(t) - v^{(n-1)}(t))$ converge to $(0, \ldots, 0)$ similarly with the trajectory of (4.42) tending zero. In other words, we also regulate the way of convergence for error profile.

The block of active disturbance rejection control is depicted as Figure 8.

Definition 4.1. We say that the ADRC closed-loop composed by (4.37), (4.38), (4.39), and (4.41) is practically convergent, if for any given initial values of (4.37), (4.38), (4.39), there exits a constant $R_0 > 0$ such that for all $R > R_0$,

$$\text{(4.43)} \qquad \begin{aligned} \lim_{\substack{t \to \infty \\ \varepsilon \to 0}} [x_i(t) - \hat{x}_i(t)] &= 0, \quad 1 \le i \le n+1, \\ \lim_{\substack{t \to \infty \\ \varepsilon \to 0}} [x_i(t) - z_{iR}(t)] &= 0, \quad 1 \le i \le n. \end{aligned}$$
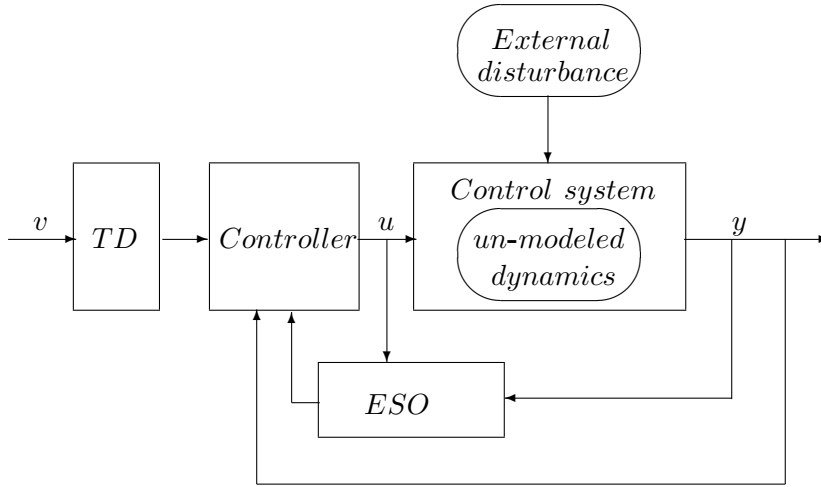
Figure 8: Block of active disturbance rejection control.

Moreover, for any given $\tau > 0$, $\lim_{R\to\infty} |z_{1R}(t) - v(t)| = 0$ uniformly in $t \in [\tau, \infty)$. In particular, when $v(t) \equiv 0$, then $z_{iR}(t) = 0$, we say that the ADRC closed-loop is practically stable.

From Definition 4.1, stabilization is a special case of output tracking with $v(t) \equiv 0$.

The convergence of ADRC is reached under some mathematical assumptions. The Assumption A1 is made for system (4.37) itself and the external disturbance.

Assumption A1. Both $w(t)$ and $\dot{w}(t)$ are bounded in $\mathbb{R}$, and there exist positive constants $\overline{C}_1, \overline{C}_2$, and function $\varpi \in C(\mathbb{R}, [0, \infty))$ such that

$$\left|\frac{\partial f(t,x,w)}{\partial x_i}\right| \le \overline{C}_1 + \varpi(w), \quad i = 1, 2, \ldots, n, \quad |f(t,0,w)| \le \overline{C}_1 + \varpi(w),$$

$$\left|\frac{\partial f(t,x,w)}{\partial w}\right| + \left|\frac{\partial f(t,x,w)}{\partial t}\right| \le \overline{C}_1 + \overline{C}_2 \|x\| + \varpi(w).$$

The Assumption A2 is for ESO (4.39) and the unknown parameter b.

Assumption A2. $|g_i(r)| \le \Lambda_i |r|$ for some positive constants Λ_i for all $i = 1, 2, \ldots, n+1$. There exist constants $\lambda_{1i}(i = 1, 2, 3, 4)$, β_1, and positive definite continuous differentiable functions $V_1, W_1 : \mathbb{R}^{n+1} \to \mathbb{R}$ such that

- $\lambda_{11}\|\nu\|^2 \le V_1(\nu) \le \lambda_{12}\|\nu\|^2$, $\lambda_{13}\|\nu\|^2 \le W_1(\nu) \le \lambda_{14}\|\nu\|^2$, $\forall \nu \in \mathbb{R}^{n+1}$;

- $\sum_{i=1}^{n}(\nu_{i+1} - g_i(\nu_1))\dfrac{\partial V_1(\nu)}{\partial \nu_i} - g_{n+1}(\nu_1)\dfrac{\partial V_1(\nu)}{\partial \nu_{n+1}} \leq -W_1(\nu),\ \forall\ \nu \in \mathbb{R}^{n+1};$

- $\left|\dfrac{\partial V_1(\nu)}{\partial \nu_{n+1}}\right| \leq \beta_1 \|\nu\|,\ \forall\ \nu = (\nu_1, \nu_2, \ldots, \nu_{n+1}) \in \mathbb{R}^{n+1}.$

Moreover, the parameter b satisfies $B \triangleq \left|\dfrac{b-b_0}{b_0}\right| \Lambda_{n+1} < \dfrac{\lambda_{13}}{\beta_1}$. Here and throughout the paper, we always use $\|\cdot\|$ to denote the corresponding Euclidian norm.

The Assumption A3 is for nonlinear function $\varphi(\nu)$ in (4.41).

Assumption A3. $\varphi(\nu)$ is continuously differentiable, $\varphi(0) = 0$, and Lipschitz continuous with Lipschitz constant L: $|\varphi(\nu) - \varphi(\hat{\nu})| \leq L\|\nu - \hat{\nu}\|$ for all $\nu, \hat{\nu} \in \mathbb{R}^n$. There exist constants λ_{2i} $(i = 1, 2, 3, 4)$, β_2, and positive definite continuous differentiable functions $V_2, W_2 : \mathbb{R}^n \to \mathbb{R}$ such that

- $\lambda_{21}\|\nu\|^2 \leq V_2(\nu) \leq \lambda_{22}\|\nu\|^2, \quad \lambda_{23}\|\nu\|^2 \leq W_2(\nu) \leq \lambda_{24}\|\nu\|^2;$

- $\sum_{i=1}^{n-1} \nu_{i+1}\dfrac{\partial V_2(\nu)}{\partial \nu_i} + \varphi(\nu_1, \nu_2, \ldots, \nu_n)\dfrac{\partial V_2(\nu)}{\partial \nu_n} \leq -W_2(\nu);$

- $\left|\dfrac{\partial V_2(\nu)}{\partial \nu_n}\right| \leq \beta_2 \|\nu\|, \forall\ \nu = (\nu_1, \nu_2, \ldots, \nu_n) \in \mathbb{R}^n.$

In above assumption, the continuous differentiability and Lipschitz continuity of $\varphi(\nu)$ imply that

$$(4.44) \qquad \left|\dfrac{\partial \varphi(\nu)}{\partial \nu_i}\right| \leq L, \quad \nu \in \mathbb{R},\ i = 1, 2, \ldots, n.$$

The Assumption A4 is for TD (4.38).

Assumption A4. Both $v(t)$ and $\dot{v}(t)$ are bounded in $[0, \infty)$, and $\psi(\cdot)$ is locally Lipschitz continuous, and system (4.38) with $v(t) \equiv 0, R = 1$ is globally asymptotically stable.

The Assumption A5 is for the zero dynamics.

Assumption A5. There exist constants $\bar{k}_1, \bar{k}_2$, and function $\varpi \in C(\mathbb{R}, \bar{\mathbb{R}}^+)$ such that

$$\|F_0(x, \zeta, w)\| \leq \bar{k}_1 + \bar{k}_2\|x\| + \varpi(w), \forall\ x \in \mathbb{R}^n, \zeta \in \mathbb{R}^m, w \in \mathbb{R}.$$

Theorem 4.1. *Assume Assumptions A1-A5. Then under ADRC* (4.41) *based on* (4.38) *and* (4.39) *with time varying gain given in* (4.40), *the closed-loop is convergent in the sense of Definition 4.1.*

Since the initial state is arbitrary, we say feedback control law (4.41) to be global ADRC. If the bound of initial states is available, we can design what so called semi-global ADRC.

Assumption A6. There exist positive constants C_1, C_2, and C_3 such that $\sup_{t \in [0,\infty)} \|(v(t), \ldots, v^{(n)}(t))\| < C_1$, $\|x(0)\| < C_2$, $\|(w(t), \dot{w}(t))\| < C_3$ for all $t \in [0, \infty)$.

The Assumption A1 can be weakened by Assumption A1*.

Assumption A1*. There exist positive constant N and functions $\varpi_1 \in C(\mathbb{R}^n, \mathbb{R}^+)$, $\varpi_2 \in C(\mathbb{R}^m, \mathbb{R}^+)$, and $\varpi_3 \in C(\mathbb{R}, \mathbb{R}^+)$ such that

$$\sum_{i=1}^{n} \left| \frac{\partial f(t, x, \zeta, w)}{\partial x_i} \right| + \sum_{i=1}^{m} \left| \frac{\partial f(t, x, \zeta, w)}{\partial \zeta_i} \right| + \left| \frac{\partial f(t, x, \zeta, w)}{\partial t} \right| + |f(t, x, \zeta, w)|$$
$$\leq N + \varpi_1(x) + \varpi_2(\zeta) + \varpi_3(w).$$

Assumption A5 can also be replaced by Assumption A5*.

Assumption A5*. There exist positive definite functions $V_0, W_0 : \mathbb{R}^m \to \mathbb{R}$ such that $L_{F_0} V_0(\zeta) \leq -W_0(\zeta)$ for all $\zeta : \|\zeta\| > \chi(x, w)$, where $\chi : \mathbb{R}^{n+1} \to \mathbb{R}$ is a wedge function, and $L_{F_0} V_0(\zeta)$ denotes the Lie derivative along the zero dynamics of system (4.37).

The Assumption A7 is for the control coefficient $b(\cdot)$ and its nominal $b_0(\cdot)$.

Assumption A7. The nominal control magnifying coefficient $b_0 \neq 0$ and

$$(4.45) \qquad \left| \frac{b(w(t)) - b_0}{b_0} \right| < \min \left\{ \frac{1}{2}, \ \lambda_{13} \left(\beta_1 \Lambda_{n+1} \left(M_1 + \frac{1}{2} \right) \right)^{-1} \right\}.$$

In addition, the third condition of Assumption A2 is changed into

$$\left| \frac{\partial V_1(\nu_1, \nu_2, \ldots, \nu_{n+1})}{\partial \nu_n} \right| + \left| \frac{\partial V_1(\nu_1, \nu_2, \ldots, \nu_{n+1})}{\partial \nu_{n+1}} \right| \leq \beta_1 \|\nu\|$$
$$\forall \ \nu = (\nu_1, \nu_2, \ldots, \nu_{n+1}) \in \mathbb{R}^{n+1}.$$

In what follows, we construct a semi-global ADRC with the notations:

$$C_1^* \geq \max_{\{y\in\mathbb{R}^n,\ \|y\|\leq C_1+C_2+1\}} V_2(y),$$

$$C_4 \geq \max\left\{\sup_{\|x\|\leq C_1+(C_1^*+1)/\lambda_{23}+1,\|w\|\leq C_3,} |\chi(x,w)|,\ \|\zeta(0)\|\right\},$$

(4.46)
$$M_1 \geq 2\bigg(1 + M_2 + C_1$$
$$+ \sup_{\|x\|\leq C_1+(C_1^*+1)/\lambda_{23}+1,\|w\ \|\leq C_3,\|\zeta\|\leq C_4} \Big[N + \varpi_1(x)$$
$$+ \varpi_2(\zeta) + \varpi_3(w)\Big]\bigg),$$

$$M_2 \geq \max_{\|x\|\leq C_1+(C_1^*+1)/\lambda_{23}+1} |\varphi(x)|.$$

Let $\mathrm{sat}_M : \mathbb{R} \to \mathbb{R}$ be an odd continuous differentiable saturation function given by

(4.47)
$$\mathrm{sat}_M(r) = \begin{cases} r, & 0 \leq r \leq M, \\ -\frac{1}{2}r^2 + (M+1)r - \frac{1}{2}M^2, & M < r \leq M+1, \\ M + \frac{1}{2}, & r > M+1, \end{cases}$$

where $M > 0$ is some constant.

Using the same tracking differentiator (4.38), the extended state observer (4.39) with the constant tuning gain $\varrho(t) \equiv \varepsilon$, the observer based feedback control is designed as follows:

(4.48) ADRC(S) : $u(t) = \dfrac{1}{b_0(\hat{x})}\Big[\mathrm{sat}_{M_2}(\varphi(\hat{x}(t) - z_R(t)))$
$$- \mathrm{sat}_{M_1}(\hat{x}_{n+1}(t)) + \mathrm{sat}_{C_1+1}(z_{(n+1)R}(t))\Big].$$

Theorem 4.2. *Assume Assumptions A2-A4, A1*, A5*, A6, and A7, and in ESO (4.39), $\varrho(t) \equiv \varepsilon$, with ε being a positive constant. Then under the semi-global ADRC (4.48), the closed-loop is convergent in the sense of Definition 4.1.*

Theorems 4.1 and 4.2 are presented in [53]. In circumstance where the derivatives of reference signal $v(t)$ are known and $v(t)$ is acceptable from application standard for output tracking, TD part can be dropped in ADRC design, and the states of TD is the derivatives of $v(t)$. The concrete design

of easy to use, in terms of Theorem 4.1 and 4.2, is linear ADRC, that is, the functions $g_i(\cdot)$ in ESO (4.39) and $\varphi(\cdot)$ in feedback control laws (4.41) and (4.48) are linear functions. The convergence of semi-global linear ADRC on stabilization, i.e., the reference signal $v(t) \equiv 0$, can also be found in [9] under the name of extended high gain approach. Linear ADRC with adaptive ESO is studied in [47]. In [38], the projected gradient estimator is used likewise linear ADRC for system (4.37) without using the nominal value of control magnifying coefficient. The nonlinear ADRC for nonlinear systems with stochastic disturbance is investigated in [21] and with actuator saturation is considered in [41]. The linear ADRC for mismatched disturbance can be found in [40]. The linear ADRC for lower triangular systems is studied in [48, 49]. The nonlinear ADRC for this kind system is considered recently in [54]. The nonlinear ADRC is extended to multi-input multi-output (MIMO) systems in [23]. The ADRC has been extended to infinite dimensional systems recently, such as wave equation [13, 15, 18], Euler-Bernoulli beam equation [14, 19], Schrodinger equation [16], coupled heat and ODE [17], and Kirchhoff plate [20], among others.

We have found many similarities of ADRC with high gain control. To have a comparison for both methods, we use a simple example for illustration. Consider

$$\begin{cases} \dot{x}_1(t) = x_2(t), \\ \dot{x}_2(t) = u(t) + w(t), \end{cases}$$

where $x_1(t) \in \mathbb{R}$ is the output, $u(t) \in \mathbb{R}$ is the control input, and $w(t) \in \mathbb{R}$ is the unknown external disturbance. The control objective is to stabilize (practically) the system. The high-gain feedback control can be designed as $u(t) = -R(\tilde{x}_1(t) + 2\tilde{x}_2(t))$, where R is the tuning gain parameter to be tuned large enough, so the high gain is used also in feedback control. The $\tilde{x}_1(t)$ and $\tilde{x}_2(t)$ are coming from the following high-gain observer:

$$\begin{cases} \dot{\tilde{x}}_1(t) = \tilde{x}_2(t) + \tilde{x}_2(t) + 2r_0(x_1(t) - \tilde{x}_1(t)), \\ \dot{\tilde{x}}_2(t) = r_0^2(x_1 - \tilde{x}_1(t)) + u(t), \end{cases}$$

where the high-gain parameter is chosen $r_0 = 200$. The ESO is designed as follows to estimate, in real time, the external disturbance $w(t)$ as well as the system state $(x_1(t), x_2(t))$:

$$\begin{cases} \dot{\hat{x}}_1(t) = \hat{x}_2(t) + 3r(t)(x_1(t) - \hat{x}_1(t)) + u(t), \\ \dot{\hat{x}}_2(t) = \hat{x}_3(t) + 3r^2(t)(x_1(t) - \hat{x}_1(t)), \\ \dot{\hat{x}}_3(t) = r^3(t)(x_1(t) - \hat{x}_1(t)), \end{cases}$$

where $r(t)$ is given as

$$(4.49) \qquad r(t) = \begin{cases} e^{at}, & 0 \leq t < \frac{1}{a} \ln r_0, \\ r_0, & t \geq \frac{1}{a} \ln r_0, \end{cases}$$

with $a = 2$ and $r_0 = 200$.

The ADRC can be designed as $u(t) = -\hat{x}_3(t) - \hat{x}_1(t) - 2\hat{x}_2(t)$, where $\hat{x}_1(t)$, $\hat{x}_2(t)$, and $\hat{x}_3(t)$ are states of ESO, $\hat{x}_1(t)$ and $\hat{x}_2(t)$ are the estimated values of $x_1(t)$ and $x_2(t)$ respectively, $\hat{x}_3(t)$ is the estimated value of external disturbance $w(t)$. Because the external disturbance is estimated and canceled by $x_3(t)$, so we do not need the high-gain in feedback loop of ADRC, which is very different from high-gain control.

In numerical simulation, we set $w(t) = 1 + \sin t$. The numerical results of high-gain control with $R = 2$ and its magnifications are plotted in Figure 9. Figure 10 plots the numerical results of high-gain control with $R = 20$ and its magnifications.

We can see from Figure 9 that the high-gain control with small gain parameter $R = 2$ cannot stabilize the system, and the system state $(x_1(t), x_2(t))$ deviates from zero obviously. As the gain parameter chosen to be as large as $R = 20$, we see from Figure 10 that the state $(x_1(t), x_2(t))$ converges to zero state very rapidly. However, the absolute value of control is very large near the initial time, and it actually exceeds 6000.

Figure 11 plots the numerical results by ADRC. Although in the feedback loop, the coefficients of estimated values of $x_1(t)$ and $x_2(t)$ are same as in high-gain control, from Figure 11 we can see that the steady performance is very well and the maximum of the absolute control value is quite small. Actually, the maximum of absolute control value is less than 10, which is sharp contrast to the maximal value 600 by classical high-gain control with $R = 2$, and the maximal value 6000 by classical high-gain control with $R = 20$.

It should also be pointed out that if the external disturbance $w(t)$ is a constant, that is, the derivative of the "total disturbance" is equal to zero, we can obtain that there exists positive constant $r_0^* > 0$ such that for any given $r_0 > r_0^*$, states of the ESO and the system driven by ADRC with the gain function

$$r(t) = \begin{cases} e^{at}, & 0 \leq t < \frac{1}{a} \ln r_0, \\ r_0, & t \geq \frac{1}{a} \ln r_0, \end{cases}$$

are convergent to zero as $t \to \infty$. On the other hand, by classical high-gain control, we can only obtain practical stability results with the error
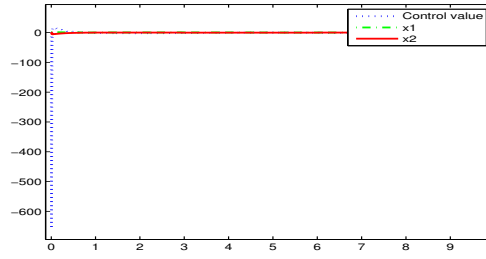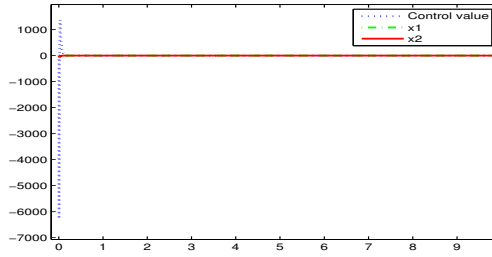
(a) State $(x_1(t), x_2(t))$ and control value u
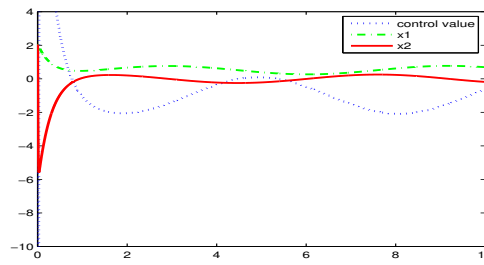
(b) Magnification of a) in y-axis
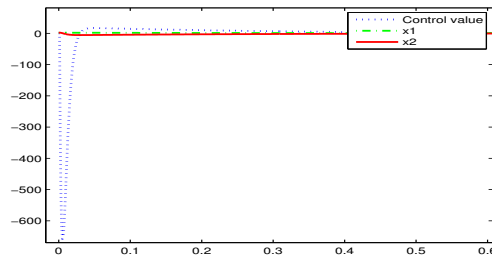
(c) Magnification of a) in y-axis

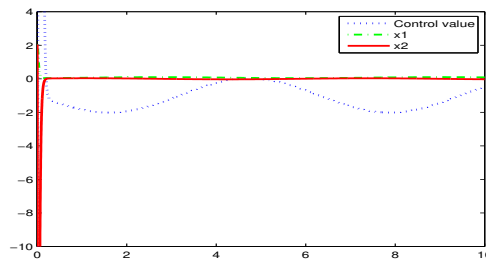Figure 9:   High-gain observer based high-gain control with $R = 2$.

estimation:

$$\|(x_1(t), x_2(t))\| \leq \frac{M}{R}, \;\; t > t_0$$

where $t_0 > 0$, and M is a w-dependent constant but is independent of R.
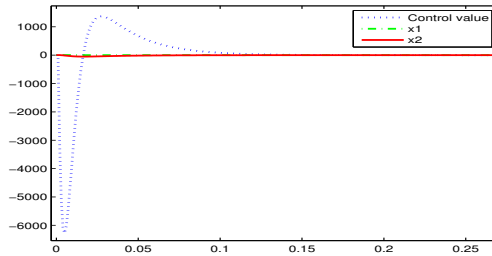
In what follows, we use a first order system to illustrate the difference of ADRC and sliding model control (SMC) numerically. The considered system

(a) State $(x_1(t), x_2(t))$ and control value u

(b) Magnification of a) in y-axis

(c) Magnification of a) in y-axis

Figure 10:   High-gain observer based high-gain control with $R = 20$.

is described by

$$(4.50) \qquad\qquad\qquad \dot{x}(t) = u(t) + w(t),$$

where $x(t) \in \mathbb{R}$ is the output, $u(t) \in \mathbb{R}$ is the control input, and $w(t) \in \mathbb{R}$ is the unknown external disturbance. The control purpose is to stabilize practically the system. The SMC can be designed as $u(t) = 2\mathrm{sign}(x(t))$. The
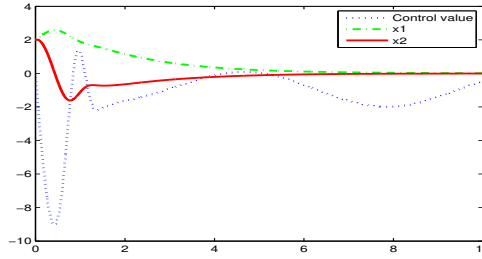
Figure 11: Extended state observer based active disturbance rejection control.

ADRC is designed as $u(t) = -\hat{w}(t) - 2x(t)$, where $\hat{w}(t)$ is coming from the following ESO to cancel the disturbance $w(t)$:

$$\begin{cases} \dot{\hat{x}}(t) = \hat{w}(t) + 2r(t)(x(t) - \hat{x}(t)) + u(t), \\ \dot{\hat{w}}(t) = r^2(t)(x(t) - \hat{x}(t)), \end{cases}$$

where $r(t)$ is given in (2.10) with $a = 2$ and $r_0 = 200$

$$\varrho(t) = \begin{cases} e^{2t}, \ 0 \le t < \frac{1}{2}\ln r, \\ r, \ t \ge \dfrac{1}{2}\ln r. \end{cases}$$

The numerical results of SMC are plotted in Figure 12. From Figure 12, we can see that the stead performance is very well. However the control $u(t)$ fills almost full the box and therefore the control energy is very large. The
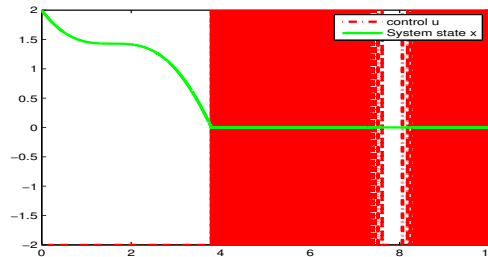

Figure 12: State of system (4.50) driven by sliding-mode control.

Figure 13 is the numerical results of ADRC. From Figure 13 we can see that

the steady performance is very well and the control energy consumption is smaller than SMC, because $-u(t)$ is almost equal to the external disturbance, while the control $u(t)$ of SMC almost equals to $+2$ or -2 along the whole process.
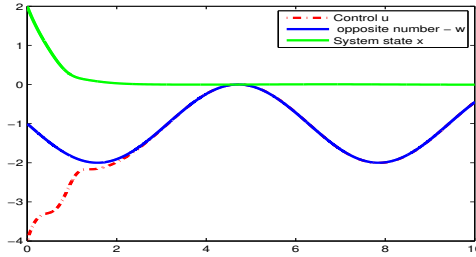

Figure 13: State of system (4.50) driven by ADRC.

To end this section, we put forward some problems that are worth investigating. In the proof of convergence for nonlinear ADRC closed-loop system, we require stronger conditions on nonlinear functions in ESO than for the closed-loop system discussed in section 3. Naturally, there are some spaces to weaken these conditions to include much larger class of nonlinear functions. Finally, the conditions for convergence of closed-loop are usually restrictive than that for convergence of ESO, which should be relaxed in further investigations. The ADRC with ESO (3.35) and (3.36) is also deserved for investigation.

Additional attention should be paid to the nonlinear systems that are not in canonical form. Systems like triangular structure as follows:

$$\begin{cases} \dot{x}_1(t) = f_1(x_1(t), x_2(t), w_1(t)), \\ \quad\vdots \\ \dot{x}_n(t) = f_n(x_1(t), \ldots, x_n(t), \xi(t), w_n(t)) + b(x(t), \xi(t), w_n(t))u(t) \\ \dot{\xi}(t) = F(x_1(t), \ldots, x_n(t), \xi(t), w_1(t), \ldots, w_n(t)), \\ y(t) = x_1(t), \end{cases}$$

where $w_i(t)(1 \leq i \leq n)$ is the external disturbance and $f_i(\cdot)(1 \leq i \leq n)$ is the system function with some uncertainty. The difficulty in dealing with the external disturbances $w_1(t), \ldots, w_{n-1}(t)$ and other uncertainties lies in that they are not in the control channel. For stabilization, the uncertainty must be control matched while for output tracking, it is not necessary.

5. Summary

In the landmark book [43], H.S. Tsien, an engineering scientist throughout his career from a Caltech professor to father of nuclear missiles of China, described control theory in his mind as "to organize the design principles used in engineering practice into a discipline that thus to exhibit the similarities between different areas of engineering practice and to emphasize the power of fundamental concepts". In his later years, Tsien had formed his Three-layers theory about science, which, relevant to control theory, is from mathematics, control theory, to finally engineering applications. As a transition layer, control theory needs the help of mathematics to build its theory on the one hand, and provides solution and principle to engineering applications on the other. However, many control practitioners when they need to solve practical control problems may have a dilemma: in applications, one needs a control method that is simple, powerful, robust, and easy to use; and on the other hand, no solutions can be found in countless of papers and books. The evidence of this awkward situation is the PID, a primitive control technology yet has dominated the control engineering practice for almost one century till today.

This paper introduces an emerging control technology, known as active disturbance rejection control (ADRC) to this day, which is generally believed, through numerous successful engineering applications in different groups in different countries, to be a viable PID alternative. This almost model free control technology, inherent nature of PID, sprouts from seed idea of Tsien [43] that the control operation "must not be influenced by internal and external disturbances". The creative idea of ADRC is that the total disturbance which may come from internal disturbance including the unknown parts or the parts of hardly to be treated by practitioner and external disturbance, is reflected entirely in the measured output and hence can be regarded as a function of time, no matter it is state dependent or time variant, and finally can be estimated. Inasmuch as the total disturbance is estimated, it can certainly be canceled in the feedback loop. This transforms a control system into a canonical form (most often it is a linear one) for which we have many ways to deal with. Although there are much selective ideas in dealing with uncertainty including organizing our daily life, the estimation/cancellatgion is among the smartest way, both logically and practically. We could certainly find the similar idea from other control methods like adaptive control and internal model principle, it is ADRC which first exploits this smart idea systematically in a very large scale not only for external disturbance but also importantly for internal disturbance as well.

This paper articulates, from a theoretical perspective, the origin, idea, possibility, applicability, as well as relations with other relevant control methods of ADRC. Three key parts, namely, tracking differentiator, extended state observer, and extended state observer based output feedback are expounded separately, with the help of numerical visualizations. The limitations and possible improvement in extended state observer design are indicated. Some unsolved theoretical problems are also put forward.

References

[1] A. Adams and John J.F. Fournier, *Sobolev Spaces.* 2rd edition, Academic press, 2003.

[2] K. Astrom and T. Hagglund, *PID Controller: Theory, Design, and Tuning, Research Triangle Park.* NC: Instrument Society of America, 1995.

[3] S. P. Bhat and D. S. Bernstein, *Finite-time stability of continuous autonomous systems.* SIAM J. Control Optim., **38** (2000), 751–766.

[4] G. Besancon, *Nonlinear Observers and Applications.* Springer-Verlag, New York, 2007.

[5] W. L. Bialkowski, *Control of the pulp and paper making process.* In: The Control Handbook (W.S. Levine, Ed.), New York: IEEE Press, 1996, 1219–1242.

[6] D. Cheng, X. Hu and T. Shen, *Analysis and Design of Nonlinear Control Systems.* Science Press Beijing and Springer-Verlag Berlin, 2010.

[7] A. M. Dabroom and H. K. Khalil, *Discrete-time implementation of high-gain observers for numerical differentiation.* Int. J. Control, **17** (1999), 1523–1537.

[8] B. Francis and W. Wonham, *The internal model principle of control theory.* Automatica, **12** (1976), 457–465.

[9] L. B. Freidovich and H. K. Khalil, *Performance recovery of feedback-linearization-based Designs.* IEEE Trans. Automat. Control, **53** (2008), 2324–2334.

[10] Z. Gao, *Scaling and bandwith-parameterization based controller tuning.* American Control Conference, IEEE, Denver, Colorado, 2003, 4989–4996.

[11] Z. Gao, *Active disturbance rejection control from an enduring idea to an emerging technology.* The 10th International Workshop on Robot Motion and Control, IEEE, Poznan University of Technology, Poznan, Poland, July 6-8, 2015, 269–282.

[12] B. Z. Guo, J. Q. Han and F. B. Xi, *Linear tracking-differentiator and application to online estimation of the frequency of a sinusoidal signal with random noise perturbation.* Internat. J. Systems Sci., **33** (2002), 351–358.

[13] B. Z. Guo and F. F. Jin, *Output feedback stabilization for one-dimensional wave equation subject to boundary disturbance.* IEEE Trans. Automat. Control, **60** (2015), 824–830.

[14] B. Z. Guo and F. F. Jin, *The active disturbance rejection and sliding mode control approach to the stabilization of Euler-Bernoulli beam equation with boundary input disturbance.* Automatica, **49** (2013), 2911–2918.

[15] B. Z. Guo and F. F. Jin, *Sliding mode and active disturbance rejection control to stabilization of one-dimensional anti-stable wave equations subject to disturbance in boundary input.* IEEE Trans. Automat. Control, **58** (2013), 1269-1274.

[16] B. Z. Guo and J. J. Liu, *Sliding mode control and active disturbance rejection control to the stabilization of one-dimensional Schrodinger equation subject to boundary control matched disturbance.* Internat. J. Robust Nonlinear Control, **24**(2014), 2194-2212.

[17] B. Z. Guo, J. J. Liu, A. S. AL-Fhaid, A. M. M. Younas and A. Asiri, *The active disturbance rejection control approach to stabilization of coupled heat and ODE system subject to boundary control matched disturbance.* Internat. J. Control, **88** (2015), 1554–1564.

[18] B. Z. Guo and H. C. Zhou, *The active disturbance rejection control to stabilization for multi-dimensional wave equation with boundary control matched disturbance.* IEEE Trans. Automat. Control, **60** (2015), 143–157.

[19] B. Z. Guo, H. C. Zhou, A. S. AL-Fhaid, A. M. M. Younas and A. Asiri, *Stabilization of Euler-Bernoulli beam equation with boundary moment control and disturbance by active disturbance rejection control and sliding model control approach.* J. Dyn. Control Syst., **20** (2014), 539–558

[20] B. Z. Guo and H. C. Zhou, *Active disturbance rejection control for rejecting boundary disturbance from multi-dimensional Kirchhoff plate via boundary control.* SIAM J. Control Optim., **52** (2014), 2800–2830.

[21] B. Z. Guo, Z. H. Wu and H. C. Zhou, *Active distubance rejection control approach to output-feedback stabilization of a class of uncertain nonlinear systems subject to stochastic disturbance.* IEEE Trans. Automat. Control, in press, `DOI:10.1109/TAC.2015.2471815`.

[22] B. Z. Guo and Z. L. Zhao, *Weak convergence of nonlinear high-gain tracking differentiator.* IEEE Trans. Automat. Control, **58** (2013), 1074–1080.

[23] B. Z. Guo and Z. L. Zhao, *On convergence of the nonlinear active disturbance rejection control for MIMO systems.* SIAM J. Control Optim., **51** (2013), 1727–1757.

[24] B. Z. Guo and Z. L. Zhao, *On the convergence of an extended state observer for nonlinear systems with uncertainty.* Systems Control Lett., **60** (2011), 420–430.

[25] B. Z. Guo and Z. L. Zhao, *On convergence of tracking differentiator.* Internat. J. Control, **84** (2011), 693–701.

[26] B. Z. Guo and Z. L. Zhao, *Weak convergence of nonlinear high-gain tracking differentiator.* IEEE Trans. Automat. Control, **58** (2013), 1074–1080.

[27] B. Z. Guo and Z. L. Zhao, *On convergence of nonlinear extended state observer for MIMO systems with uncertainty.* IET Control Theory Appl., **6** (2012), 2375–2386.

[28] B. Z. Guo and Z. L. Zhao, *On convergence of nonlinear active disturbance rejection for SISO systems.* Chinese Control and Decision Conference, 2012, 3524–3529.

[29] J. Q. Han, *Control theory: the doctrince of model or the doctrine of control?* J. Systems Sci. Math. Sci., **9**:4 (1989), 328–335, (in Chinese).

[30] J. Q. Han and W. Wang, *Nonlinear tracking-differentiator.* J. Systems Sci. Math. Sci., **14**:2 (1994), 177–183, (in Chinese).

[31] J. Q. Han, *A class of extended state observers for uncertain systems.* Control & Decision, **10**:1 (1995), 85–88 (in Chinese).

[32] J. Q. Han, *Auto-disturbance rejection control and applications.* Control & Decision, **13**:1 (1998), 19–23, (in Chinese).

[33] J. Q. Han, *Active Disturbance Rejection Control-The Technique for Extimating and Compensating Uncertainties.* National Defence Industry Press, Beijing, 2008.

[34] J. Q. Han, *From PID to active disturbance rejection control.* IEEE Trans. Ind. Electron., **56** (2009), 900–906.

[35] J. Q. Han, *A new type of controller: NLPID.* Control & Decision, **6**:6 (1994), 401–407.

[36] J. Q. Han, *The improvement of PID control law by using nonlinearity.* Information and Control, **24**:6 (1995), 356–364.

[37] J. Huang, *Nonlinear Output Regulation: Theorey and Applications.* SIAM, Philadelphia, 2004.

[38] T. T. Jiang, C. D. Huang and L. Guo, *Control of uncertain nonlinear systems based on observers and estimators.* Automatica, **59** (2015), 35–47.

[39] H. K. Khalil, *Nonlinear Systems.* Prentice Hall, New Jersey, 2002.

[40] S. H. Li, J. Yang, W. H. Chen and X. Chen, *Generalized extended state observer based control for systems with mismatched uncertainties.* IEEE Trans. Ind. Electron., **59** (2012), 4792–4802.

[41] M. P. Ran, Q. Wang and C. Y. Dong, *Stabilization of a class of nonlinear systems with actuator saturation via active disturbance rejection control.* Automatica, **63** (2016), 302–310.

[42] G. J. Silva, A. Datta and S. P. Bhattacharyya, *New Results on the synthesis of PID controllers.* IEEE Trans. Automat. Control, **47** (2002), 241–252.

[43] H. S. Tsien, *Engineering Cybernetics.* McGraw-Hill, New York, 1954.

[44] A. Teel and L. Praly, *Tools for semiglobal stabilization by partial state and output feedback.* SIAM J. Control Optim, **33** (1995), 1443–1488.

[45] M. Tomizuka, K. Chew and W. Yang, *Disturbance rejection through an external model.* ASEM J. Dynamic Systems, Measurement, and Control, **112** (1990), 559–564.

[46] V. Utkin, G. Guldner and J. X. Shi, *Sliding mode control in electro-mechanical systems.* 2nd Edition, Taylor & Francis, New York, 2009.

[47] W. C. Xue, W. Y. Bai, S. Yang, K. Song, Y. Huang and H. Xie, *ADRC with adaptive extended state observer and its application to air-fuel ratio control in gasoline engines.* IEEE Trans. Automat. Control, **62** (2015), 5847–5857.

[48] W. C. Xue and Y. Huang, *On performance analysis of ADRC for a class of MIMO lower-triangular nonlinear uncertain systems.* ISA Trans., **53** (2014), 955–962.

[49] J. Yao, Z. Jiao and D. Ma, *Extended-state-observer-based output feedback nonlinear robust control of hydraulic systems with backstepping.* IEEE Trans. Ind. Electron., **61** (2014), 6285–6293.

[50] G. Zarikian and A. Serrani, *Harmonic disturbance rejection in tracking control of Euler-Lagrange systems: a exernal model approach.* IEEE Trans. Control Systems Technology, **15** (2007), 118–129.

[51] Z. L. Zhao and B. Z. Guo, *On active disturbance rejection control for nonlinear systems using time-varying gain.* Eur. J. Control, **23** (2015), 62–70.

[52] Z. L. Zhao and B. Z. Guo, *Extended state observer for uncertain lower triangular nonlinear systems.* Systems Control Lett., **85** (2015), 100–108.

[53] Z. L. Zhao and B. Z. Guo, *On convergence of nonlinear active disturbance rejection control for a class of nonlinear systems.* J. Dyn. Control Syst., **22** (2016), 385–412.

[54] Z. L. Zhao and B. Z. Guo, *Active disturbance rejection control approach to stabilization of lower triangular systems with uncertainty.* Internat. J. Robust Nonlinear Control, in press.

[55] Z. L. Zhao and B. Z. Guo, *A nonlinear extended stated observers based on switching functions.* Preprint, Shaanxi Normal University, 2016.

[56] Q. Zheng, L. Gao and Z. Gao, *On stability analysis of active disturbance rejection control for nonlinear time-varying plants with unknow dynamics.* The 47th IEEE Conference on Decision and Control, IEEE, New Orleans, Louisiana, 2007, 3501–3506.

[57] Q. Zheng and Z. Gao, *An energy saving, factory-validated disturbance decoupling control design for extrusion processes.* The 10th World

Congress on Intelligent Control and Automation, IEEE, Piscataway, NJ, 2012, 2891–2896.

ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCE, ACADEMIA SINICA
BEIJING 100190, P.R. CHINA
AND SCHOOL OF COMPUTER SCIENCE AND APPLIED MATHEMATICS
UNIVERSITY OF THE WITWATERSRAND
WITS 2050, JOHANNESBURG, SOUTH AFRICA
E-mail address: bzguo@iss.ac.cn

SCHOOL OF MATHEMATICS AND INFORMATION SCIENCE
SHAANXI NORMAL UNIVERSITY
XI'AN 710062, SHAANXI, P.R. CHINA
E-mail address: zhiliangzhao@snnu.edu.cn