An Alternative Paradigm For Control System Design

Zhiqiang Gao [1] Yi Huang [2] Jingqing Han [2]

1) Dept. of Electrical and Computer Engineering., Cleveland State Univ., Cleveland, Ohio 44115 z.gao@csuohio.edu

2) Institute of Systems Science, AMSS, CAS, Beijing, 100080, P.R.China, yhuang@staff.iss.ac.cn

Abstract: An alternative framework for control design is presented in this paper. It compliments the existing theory in that: 1) it actively and systematically explores the use of nonlinear control mechanisms for better performance, even for linear plants; 2) it represents a control strategy that is rather independent of mathematical models of the plants, thus achieving inherent robustness and reducing design complexity. An overview of this design philosophy and associated algorithms are first introduced, followed by the summaries of preliminary analysis results and practical applications. It is evident that the proposed framework lends itself well in providing innovative solutions to practical problems while maintaining the simplicity and the intuitiveness of the existing technology, namely PID.

1. Introduction

Feedback control is a well-established engineering discipline that has enjoyed a tremendous growth in both theoretical development and applications in the last few decades. A historical perspective can be found in [1], and a comprehensive account of classical and modern control theory in [2]. The frequency domain based classical control theory provided engineers with great insight on how a single-input and single-output control (SISO) system works, as well as a set of analysis and design tools such as Bode and Nyquist plot, root-locus, etc. On the other hand, the progress made in state-space based optimal control and estimation theory, particularly the Kalman Filter theory, lead to the birth of modern control theory. Using singular value plot as a bridge, the frequency domain insight of classical control was incorporated into the framework of modern control, which lead to the well-known H_2 and H_∞ synthesis methods. This provides a modern control paradigm where both performance and robustness specifications are brought into a common mathematical framework for control synthesis of linear time invariant systems [2].

Nonlinear control is a more challenging but less mature discipline. The main design methods include feedback linearization, Lyapunov method, sliding mode, back stepping, adaptive control, etc. [3,4]. Like its linear counter part, nonlinear control law is designed based on the mathematical model of the plants. Progress has also been made in addressing the robustness issue in nonlinear control.

In the meanwhile, the use of feedback control in industry, mostly known as industrial automation, has grown into a billion dollar business. For example, it is not uncommon in manufacturing industry to see a single manufacturing cell that uses over a hundred servo loops. What was used to be a simple analog controller has now been almost exclusively converted to digital forms. With the rapid advance in computer hardware, new computational and network capabilities bring vast opportunities to implement advanced control strategies that truly bring marked benefits to the performance.

Although there are many examples of successful applications of modern control theory in industry, the core control technology used in industry has remained unchanged for several decades now. It is known as proportional-integral-derivative (PID) control, which dates back to 1922 [1,5,6], well before classical and modern control theory were born. Today, PID remains as the tool of choice in over 90% of industrial applications [7]. We believe that there is an underlining principle behind PID that makes it effective.

At the same time, in working with industry on control technology issues, it appears to us that nonlinear control means are more powerful than the linear one and, perhaps more importantly, the availability of an accurate mathematical model should not be a precondition of a good control design.

The motivation of our work was to find an alternative research direction and methodology that are more in line with how control is practiced. Philosophically, the research was carried out on the premises of Han’s vision on control theory, introduced in two seminal papers, first in 1989 [8] and then in 1999 [9]. This led to a slew of promising control techniques that are free of a few fundamental limitations, such as linearity, time invariance, accurate mathematical representation of plant, etc.

In section 2, the design philosophy and the main components of the new paradigm are introduced. The applications of the resulting new techniques are summarized in section 3. New challenges and questions to basic concepts, together with the preliminary analytical studies, are discussed in section 4. Finally, concluding remarks and comments for future research are given in section 5.

2. The New Paradigm

In his seminal paper published in 1989 [8], Han contended that the modern control theory, originated from Kalman Filter, could be alternatively viewed as “model theory” since modeling and model based analysis and synthesis are its main thrusts. While it has undoubtedly brought significant advances to the science of control, “model theory” also carried with it fundamental limitations, most notably the robustness issue (or the reliance on accurate mathematical representation of plant). These limitations could be the primary reasons that the practicing engineers have been slow to embrace the more advanced methods. Following the famous Gödel's "Incompleteness Theorem", Han suggested that the solutions to the limitations of model theory lie outside its framework [9].

In particular, even though the state space model (mostly the (A,B,C) matrices) made it possible to discover great insight in basic understanding of feedback control, such as controllability and observability, it may not be the best vehicle to carry out the control design in practice because:

1. The model is not easily available in many engineering problems;
2. Even if it is available, the resulting control law could be too dependent on the accuracy of the model parameters and suffers poor robustness;

Considering marvelous feedback control mechanisms found in Nature, Han concluded that there must exist better control laws that are beyond the restrictions of the current control theory. Taking cues from classical control theory and engineering practice, as well as the long established methodology in Guidance, Navigation and Control (GNC) in aeronautics and aerospace applications, Han proposed a new framework in seeking more effective control laws [8,10,11,12]. The key issues to be addressed are:

1. What do we have to know about the plant in order to control it? and (Q1)
2. How do we go about finding effective control mechanisms? (Q2)

The Theme Problem: To illustrate the new concepts, a GNC example is used here where the motion of an object is described as

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = a(t) + u(t) \\ y = x_1 \end{cases} \tag{1}$$

Here, x_1 and x_2 and $a(t)$ are position, velocity and "inherent acceleration", respectively, and $u(t)$ is the control signal. In many applications, $a(t)$ is usually a nonlinear function of many variables, i.e.

$$a(t) = f(t, x_1(t), x_2(t), w(t)), \tag{2}$$

where $w(t)$ represents internal and external disturbances. Han suggested that, instead of depending solely on the mathematical model of the plant, or the explicit function in (2), effective control could be achieved based on general characteristics of the physical process and certain information of the process that can be obtained in real time. In fact, engineers have long used an error-based method known as proportional-integral-derivative (PID) controller to control such processes:

$$u(t) = K_P e + K_I \int_0^t e\,dt + K_D \frac{de}{dt} \tag{3}$$

where the control action is based on the current (P), past (I) and future (D) of the error: $e(t) = y_d(t) - y(t)$. Here $y_d(t)$ is the desired output and K_p, K_I, and K_D are controller gains that can be intuitively tuned. Intuitive and effective, this design approach remains the tool of choice for engineers.

In addition to PID, classical control theory provided additional control blocks such as lead and lag compensators that further enhanced the performance of this error-based control law. The practice of control engineering demonstrates that :

1. An accurate mathematical model, such as the function $f(t,x_1(t),x_2(t),w(t))$, is not a precondition to devise an effective control law. In fact, model based design is inherently susceptible to the accuracy of the model and the dynamic change in the plant;
2. "Control" refers to a process of driving a physical variable to its destination. Model based design invariably links the control law to the *global characteristics* of the plant, represented by its model. Analysis and synthesis of the global characteristics of the open and closed loop system provide insight to such a process, but they are not the objective themselves. The engineering practice seems to suggest, that the control be respondent to the *local characteristics*, such as the behavior of the error at a particular operating point. Furthermore, a properly designed control $u(t)$ in (3) can very well overcome the unknown inherent acceleration $a(t)$.

The above discussion regarding the nature of PID brings us to the second question (Q2) and the task of systematically exploring more effective feedback mechanisms. To this end, several key developments are introduced in the following sections, including the use of nonlinear feedback; the nonlinear differentiator; the nonlinear PID; and the extended state observer. It will be shown that significant progress has been made both in developing new synthesis methods and in the analysis of the new control mechanisms. Furthermore, the new methods have made impacts in solving challenging engineering problems and contributed in new technology development [31-34].

Remarks:

1. Han's methods to control system design should be viewed as complimentary, rather than in conflict, to modern control theory in that they address the issues that are not addressed within the existing theory.

2. During the long history of control theory evolvement, there were many ideas proposed in term of modifications of control law, such as the ones in [13,14,37-39] and the references therein. While we believe the work presented here was a more systematic effort in such endeavor, which results in many unique algorithms and practical applications, a complete comparison with past results is beyond the scope of this paper.

2.1 Exploring Nonlinear Feedback

The extension of his work in canonical forms of linear system led Han to the fundamental question of linear and nonlinear systems [9,10]. The notion of linear and nonlinear system originated from the classical mechanics theory where the systems of interests have no input and output, just states. The autonomous linear and nonlinear systems with no input and output have distinct topologies, which are not interchangeable. The feedback control system, however, has an open topology that can be changed at will by employing feedback. With feedback, a linear system can become a nonlinear one, and vice versa. In other words, feedback control breaks down the boundary between linear and nonlinear system. Instead of using state feedback to place desired poles, we can think of placing desired nonlinearity into the system. Alone this line of thoughts, Han proposed a new thought: discard the traditional thinking of linear and nonlinear systems and fully explore the potentials of feedback, especially the nonlinear feedback [10, 11,12], as shown in the following examples.

Example 1: Consider a single integrator system with disturbance w_0:

$$\dot{e} = w_0 + u \tag{4}$$

the objective is to design a control law such that the closed-loop system satisfies $e(t) \to 0$, as $t \to \infty$. Comparing a standard linear proportional controller, $u = -Ke$, to a nonlinear one,

$$u = -K \cdot |e|^{\alpha} \cdot sign(e) \tag{5}$$

for the same gain K=10, when $|w_0| < 1, \forall t$, the steady state error is less than 0.1 for the linear controller and 0.01 for the nonlinear controller with $\alpha = 1/2$. This bound will be further reduced to 0.001 with $\alpha = 1/3$. In general, as $\alpha \to 0$, this bound will approach zero, which is the case in the well-known variable structure control. The reason the nonlinear controller performs better is that, when $0 < \alpha < 1$, it provides higher gain when error is small and lower gain when error is large. It completely agrees with the intuition obtained from working with practical problems. As a matter of fact, many fuzzy logic controllers and gain scheduling controllers exhibit this kind of characteristics on its error surface. Of course the fuzzy controller is much more complicated to implement.

Example 2: Once the unknown function *a(t)* in (2) is estimated and cancelled using the control signal u, the remaining double integrator control problem can be conveniently solved by using the well known time optimal control method, which yields

$$u = -a(t) - M\, sign(x_1 - v(t) + \frac{x_2 |x_2|}{2M}) \qquad (6)$$

The new closed loop system is described by

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -M sign(x_1 - v(t) + \frac{x_2 |x_2|}{2M}) \end{cases} \qquad (7)$$

where the output y = x_1 tracks the reference input v(t) in shortest time possible without overshoot. Here M represents the maximum acceleration this system can reach and is a function of the maximum actuation available in the system. This is a typical example of nonlinearity placement in Han's work where nonlinear mechanism was purposely introduced into the control law.

Note that other researchers, for example, [13,14] and the references therein, also noticed the advantages of using certain type of nonlinearities in feedback similar to (5). What's unique here is that Han proposed means to systematically seek out these nonlinear mechanisms in control design. It results in the birth of a class of new controllers with unparalleled performance. It pointed to a new direction in research for future control technologies development.

2.2 A Nonlinear Differentiator

Han proposed another use of equation (7) as a nonlinear observer, where x_1 and x_2 are the states of the observer that track the input signal v(t) and its differentiation $\dot{v}$(t), respectively. It was shown [11] that $\forall\ \varepsilon > 0$ and $T > 0$, $\exists\ M_0 > 0$, such that if $M > M_0$, $\int_0^T |x_1(t) - v(t)|\, dt < \varepsilon$. Here, the only design parameter of the filter is the gain M, which corresponds to the upper bound of acceleration, if x_1 can be viewed as a position signal. x_1 can track v(t) arbitrarily fast as long as M can be chosen arbitrarily large. x_2 is the differentiation of x_1 and therefore it approximates the generalized differentiation of v(t), when v(t) is not differentiable. This nonlinear filter is denoted as Tracking Differentiator (TD)[15].

To improve the numerical properties and avoid high frequency oscillations, a discrete time realization of TD was derived [16]

$$\begin{cases} v_1(t+h) = v_1(t) + h\ \ v_2(t) \\ v_2(t+h) = v_2(t) + h\ \ fst(v_1(t) - v(t), v_2(t), M, h) \end{cases} \qquad (8)$$

where v_1 and v_2 are the state variables, v(t) is the input signal, h is the step size and the function $fst(v_1, v_2, M, h)$ is defined as:

$$d = M\ \ h\ ;\quad d_0 = d\ h\ ;\quad y = v_1 + h\ \ v_2\ ;$$

$$a_0 = \sqrt{d^2 + 8\ \ M\ \ |y|}\ ;$$

$$a = \begin{cases} v_2 + \frac{y}{h}\ ,\ |y| < d_0 \\ v_2 + \frac{sign\ \ (y)(a_0 - d)}{2},\ |y| \ge d_0 \end{cases}\ ;$$

$$fst = \begin{cases} -M\ \ a/d\ ,\ |a| \le d \\ -M\ \ sign\ \ (a),\ |a| > d \end{cases} \qquad (9)$$

Note that this is the discrete time solution of the famous time optimal control problem from the 60s.

The impact of TD is profound. First, as a noise filter, it blocks any part of the signal with the acceleration *exceeding* M. In practice, we often know the physical boundary of a signal in terms of its acceleration rate. This knowledge can be conveniently incorporated into TD to reject noises based on the understanding of the physics of the plant. On the other hand, the traditional linear filter can only attenuate noises based on its frequency contents.

Secondly, TD has a very desirable frequency response characteristic as a filter. In particular, it has a much smaller phase shift compared to linear filters, while maintaining an extremely flat gain within the bandwidth. This research is still on-going.

Finally, perhaps the most important role of TD is its ability to obtain the derivative of a noisy signal with a good signal to noise ratio. It is well known that a pure differentiator is not physically implementable. The error is often not differentiable in practice due to the noises in the feedback and the discontinuities in the reference signal. This explains why the PID controller is used primarily as a PI controller in most applications. The use of the "D" part has been quite limited due to the extreme amplification of noises by differentiation, or it's approximations. This noise problem is resolved in TD because x_2 is obtained via integration. This idea of using integration to obtain differentiation goes back to 1920s when N. Wiener proposed the definition of "fractional differentiation" based on integration[28]. It led to the concepts of generalized function and generalized derivative, which were used widely in the theory of partial differential equations.

2.3 A Generic Nonlinear PID Control Scheme

Besides the usual problem with the differentiation, classical PID is also limited by its simple weighed sum of the current (e), past ($\int_0^t edt$), and future (de/dt) errors, as shown in (3). This very simplicity, which makes the controller attractive, also becomes a liability when it comes to performance. Since the digital control era began, various attempts have been made to enhance it with methods like gain scheduling, fuzzy logic and other nonlinear means. But the results often turned out to be quite problem dependent and not easily repeatable for different problems. Han proposed a fundament change in the way PID is designed[17]:

$$u = K_P |e|^{\alpha_P} sign(e) + K_I |e_I|^{\alpha_I} sign(e_I) + K_D |e_D|^{\alpha_D} sign(e_D)$$

$$e_I = \int_0^t e(t)dt,\ e_D = \frac{de(t)}{dt} \qquad (10)$$

The benefits of the nonlinear proportion term $K_P |e|^{\alpha_P} sign(e)$, was discussed earlier in Example 1. A similar argument can be made for the integral term as well. Note that an alternative implementation of the integral term is to bring the nonlinear term, $|e_I|^{\alpha_I}$, inside the integral, which will help with the integral windup problem often encountered in practice. By choosing α_I <1, it reduces the integral action when the error is large, which helps to prevent the integral term from being saturated. Regarding the differential term, it is advisable to make α_D >1. This is because this term plays its role of preventing output from overshooting primarily during the transient period when the error is relatively large. When the output is near steady state, the differentiation signal possesses little information and contains mostly noises. By choosing α_D >1, it makes the differential gain small when the error is small.

To enhance the numerical properties of this algorithm, Han explored various nonlinear mechanism and proposed to use the function fal(x,α,δ) to replace $|e|^{\alpha}\ sign(e)$, where

$$fal(x,\alpha,\delta)=\begin{cases}|x|^{\alpha}\ sign(x), & |x|>\delta,\\ x/\delta^{1-\alpha}, & |x|\le\delta,\end{cases}\qquad \delta>0 \qquad (11)$$

The graphical interpretation seen in Figure 1 is that the new function fal(x,α,δ) introduces a small linear region in the gain function. The purpose of it is to prevent excessive gain when error is small, which was known to cause high frequency chattering in some simulation studies.

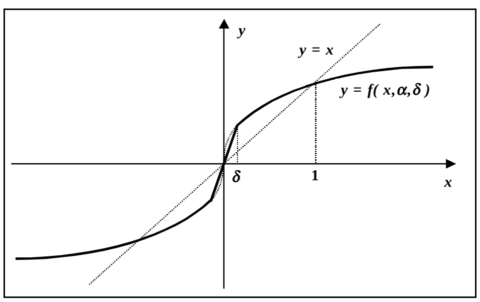

Figure 1 Comparison of linear and nonlinear gains

Finally, combining the differentiator and the nonlinear feedback discussed above, a novel nonlinear PID (NPID) control architecture is born, as shown in Figure 2.

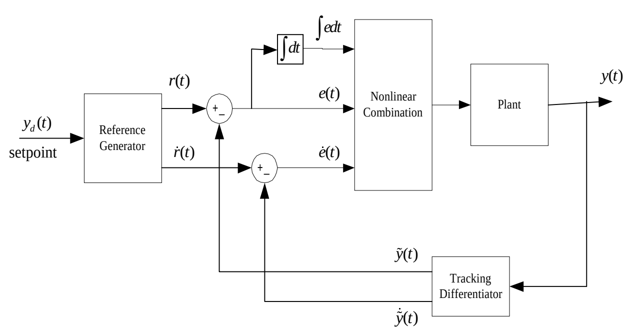

Figure 2 The Nonlinear PID Configuration

The roles of the Tracking Differentiator are two folds: 1) to filter out the noises in the measurement of y(t) and 2) to determine $\dot{y}(t)$ via integration, as discussed earlier.

The nonlinear combination (NC) is implemented as shown in (10), which can be improved using (11). r(t) and $\dot{r}(t)$ are the desired trajectories for y(t) and $\dot{y}(t)$, respectively. Instead of comparing the output to the setpoint directly to generate the error, it was proven in practice that it is more beneficial to compare the output to a predetermined desired trajectory, r(t), which is commonly known as motion profile in the motion control industry. From engineering practice, people found that this setup provides better control of not only the position, y(t), but also its first, second and third derivatives, known as velocity, acceleration, and jerk (the rate of acceleration), respectively. The selection of r(t), $\dot{r}(t)$, … etc. is an area of research by itself [18]. Note that a TD can also be used as a reference generator as well.

2.4 Extended State Observer and Active Disturbance Rejection Control

Any experienced control engineer knows the impact of the integral control on suppressing steady state error, on how fast the output enters steady state, and on disturbance rejection. On the other hand, the integral control brings inevitable lag into the system, which could even destabilize the closed loop system. The question is that is there an alternative in dealing with the steady state error and disturbances?

Recall the Theme Problem in (1). The reason the integral control is needed in practice is due to the presence of a(t), which represents both the dynamics of the physical system and possible external disturbances. Obviously the key here is the observer that tracks the value of *a(t)* closely, preferably without depending on the mathematical model of the plant. To this end, a unique nonlinear observer form was developed by Han's group and is described below.

<u>Extended State Observer</u>
Treating the unknown dynamics, *a(t)*, as an extended state and expand (1) into

$$\begin{cases}\dot{x}_1=x_2\\ \dot{x}_2=x_3+u, & x_3(t)=a(t)\underline{\underline{\Delta}}f(x_1,x_2)\\ \dot{x}_3=h(t), & h(t)=\dot{a}(t)\\ y=x_1\end{cases} \qquad (12)$$

Han's solution to this problem is to design a nonlinear observer of the form

$$\begin{cases}\dot{z}_1=z_2-g_1(z_1-y(t))\\ \dot{z}_2=z_3-g_2(z_1-y(t))+u\\ \dot{z}_3=-g_3(z_1-y(t))\end{cases} \qquad (13)$$

where $g_i(\bullet)$, i=1,2,3, are appropriate nonlinear functions such as the function fal(•) in (11), z_1, z_2 is the estimate of the state x_1, x_2, z_3 is the estimate of the extended state *a*(t).

This observer is denoted as the Extended State Observer (ESO). Interestingly, if we choose $g_i(e)=\beta_{0i}e(i=\underline{3})$, the ESO takes the form of the classical Luenberger observer. On the other hand, if $g_i(e)=\beta_{0i}(e+k_{ci}sign(e))$, then it is consistent with a form of the variable structure observer[25].

<u>The Active Disturbance Rejection Control Method</u>
For the sake of simplicity, we use the theme example (12) to illustrate a new control method. Applying the ESO given above to (12) with $g_i(\bullet)$ chosen as the fal function in (11), we have:

$$\begin{cases} \dot{z}_1 = z_2 - \beta_{01}\, fal(z_1 - y(t), \alpha_1, \delta) \\ \dot{z}_2 = z_3 - \beta_{02}\, fal(z_1 - y(t), \alpha_2, \delta) + u \\ \dot{z}_3 = -\beta_{03}\, fal(z_1 - y(t), \alpha_3, \delta) \end{cases} \tag{14}$$

where z_1, z_2, and z_3 are the estimates of x_1, x_2, and x_3=a(t), respectively. The plant can now be dynamically compensated with

$$u = u_0 - z_3 \tag{15}$$

and the difficult control problem is now simplified to a double integrator control problem of

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = u_0 \\ y = x_1 \end{cases} \tag{16}$$

where u_0 is the control input.

Since a(t) is estimated and cancelled via ESO, there is no need for integral control. Using the profile generator to generate desired state trajectory, v_1 and v_2, we proposed a nonlinear PD controller of the form:

$$u_0(t) = K_P\, fal(\varepsilon_P, \alpha_P, \delta_P) + K_D\, fal(\varepsilon_D, \alpha_D, \delta_D) \tag{17}$$

where $\varepsilon_P = v_1 - z_1$ and $\varepsilon_D = v_2 - z_2$ are error and its differentiation, respectively. K_P and K_D are obviously the controller gains. We denote this type of PD control law "Generalized PD" (GPD), because higher order differentiation can be used as needed.

Combining equation (14) to (17), together with a profile generator, the ADRC configuration is now complete, as shown in Figure 3. The name of ADRC comes from its ability to actively detect and compensate the "total disturbance" a(t), which lumps together the effects of "internal disturbance" and the "external disturbance". The former comes from system dynamics can be assumed unknown. The latter represents the external forces applied to the system that needs to be compensated by the control signal.

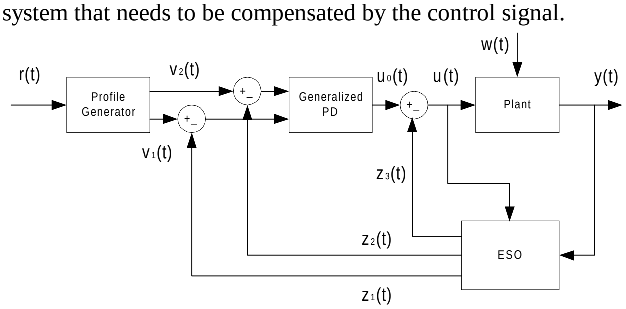

Figure 3 ADRC Configuration

Remarks:

1. The key in this design approach is ESO. It was shown in simulation and practical applications that, with appropriate selections of the nonlinear functions and gains, a single fixed ESO can tracks the states rather closely for a large class of control problems.

2. Once the ESO is properly tuned, the GPD tuning is usually straightforward since it has the same linear PD control intuition.

3. This control method proved to be very effective because it does not overly depend on mathematical model of the plant and it compensates for the internal and external disturbances dynamically. This is perhaps the most important characteristics of ADRC. It brings new meanings to the notion of robustness and disturbance rejection.

4. ADRC represents a new control concept. Its application in real world problems can take on many different forms, as shown in section 3.

Extension to MIMO Case:

The idea of ESO and ADRC can be easily generalized to multi-input and multi-output (MIMO) systems. As an extension of the Theme Problem, consider the following dynamic equation for a multi-joint manipulator:

$$\tau = M(\Theta)\ddot{\Theta} + C(\Theta, \dot{\Theta}) + G(\Theta) + W(t), \Theta \in R^n, \tag{18}$$

where $\Theta(t)$ is vector of the position of the joints, M is the $n \times n$ inertia matrix, C is the vector of Coriolis and centrifugal forces, G is the vector of gravity terms, W(t) is the vector of external disturbances and τ is the joint torque vector. Assuming that M is invertible, define

$$F(t) = -M^{-1}(\Theta)(C(\Theta, \dot{\Theta}) + G(\Theta) + W(t))$$
$$+(M^{-1}(\Theta) - M_0^{-1}(\Theta))\tau$$

where the nonsingular matrix $M_0(\Theta)$ is an estimation of $M(\Theta)$. Then the system (14) is equivalent to

$$\ddot{\Theta} = F(t) + M_0^{-1}(\Theta)\tau\ . \tag{19}$$

Viewing $F(t)$ as the "extended state" of the system, a group of n third-order ESOs can be used to estimate the state, Θ and $\dot{\Theta}$, as well as *F(t)* for each joint. These ESOs will have the same structure and similar parameter selections.

$$\begin{cases} \dot{Z}_1 = Z_2 - \beta_{01}(Z_1 - \Theta(t)) \\ \dot{Z}_2 = Z_3 - \beta_{02}\, fal(Z_1 - \Theta(t), \alpha_1, \delta) + U \\ \dot{Z}_3 = -\beta_{03}\, fal(Z_1 - \Theta(t), \alpha_2, \delta) \end{cases}$$

$$fal(Z_1 - \Theta, \alpha, \delta) = \begin{bmatrix} fal(z_{11} - \theta_1, \alpha, \delta) \\ \ldots\ldots \\ fal(z_{1n} - \theta_n, \alpha, \delta) \end{bmatrix},$$

$$Z_1 = [z_{11}, \ldots, z_{1n}]^T, \Theta = [\theta_1, \ldots, \theta_n]^T.$$

The outputs $Z_1(t)$, $Z_2(t)$ and $Z_3(t)$ will approach Θ, $\dot{\Theta}$, and *F(t)*, respectively. Once again, the plant can be reduced to a group of double integrators by applying the control law

$$\tau = M_0(\Theta)U, \qquad U = -Z_3(t) + U_0 \tag{20}$$

to the system (19), which results in $\ddot{\Theta} = U_0$ [20].

Both the *a(t)* in SISO systems, and the $F(t)$ in MIMO systems can also be regarded as a "total disturbance". Therefore, ESO can be seen as an uncertain input observer (UIO) or disturbance observer.

Disturbance rejection is an old but key problem for high performance control design. A great deal of effort has been devoted to this topic, see [37-39] and the references therein. However, these methods usually assume the knowledge of the disturbance model and/or the plant model. And usually, a higher order observer or derivatives of the measured signal are used. The breakthrough brought by ESO is that it regards all factors affecting the plant, including the nonlinear dynamics, uncertainties, the coupling effects and the external disturbances, as a “total disturbance” (extended state) to be observed. This new vision facilitates solution for a series of challenging control problems, such as disturbance rejection, dynamic linearization and decoupling control, in an ingenious way.

3. Industrial Applications

The NPID and ADRC methods have been applied as a generic solution to industrial control problems in several major industries, including the Motion Control, Tension Regulation in Web Transport System, Truck Anti-Lock Brake Systems (ABS), and Computer Numeric Control (CNC). The reasons for these problems to be selected are 1) the process exhibits significant dynamic changes during operation and/or considerable external disturbances are present; 2) the existing technology is inadequate, or limited, and improvements are badly needed; 3) the new solution will have a significant impact on industry. Much effort have been made in approaching the problem so that the successful results are not limited to one particular machine or process, but rather applicable to a large class of problems.

After preliminary simulation study for the proof of concepts, the new methods were eventually evaluated either on a realistic industrial simulator, or in hardware. Details can be found in [31-36].

<u>The Design Process</u>

1. Understanding the problem and formulating it properly. For example, in the Truck ABS problem, we had to characterize the problem as a cascade loop feedback system, before any systematic feedback control method can be applied.
2. Evaluating the problem and selecting the right tool. For example, a Nonlinear PD controller is quite adequate in handling the nonlinearity in the truck pneumatic brake system while the severe dynamic change in web tension system calls for the more capable ADRC controller.
3. Modeling and Simulation. NPID and ADRC designs do not require an explicit mathematical model of the process, but a simulation model of the process, not necessarily very accurate, helps greatly in understanding the process and determining the controller parameters. The essential information needed is how quickly the process responds to a control signal and the limitations on the sampling frequency.
4. Implementation. NPID and ADRC can be easily digitized and implemented in a high level language, such as C. The simple Euler’s method is usually good enough in digitization because of the inherent robustness of the controller.

<u>Summary of the Results</u>

Across the board, NPID and ADRC exhibit superior robustness and disturbance rejection in industrial applications unmatched by any fixed gain controllers. In motion application hardware tests, ADRC, with its parameters unchanged, overcomes significant inertia change (100%), torque disturbance (30%), friction changes, etc. and produced far much performance than the current technology. In the truck ABS application, NPID was tested against PID and a 6th order loop-shaping controller. The results showed that not only does NPID provides the better performance, it is also easy to tune and retains the same intuition of PID tuning.

In the web tension application, the process changes from a highly over-damped plant (ζ= .9) to a highly under-damped one (ζ= .2), together with a bandwidth drop of over a decade and the gain change of 40%. A fixed parameter ADRC, with the sampling rate of 1 kHz, produced an almost identical control performance during the entire operation, as though there was no change in the process.

4. New Challenges and Questions

The proposed model-independent design philosophy and nonlinear algorithms shown above pose many new challenges and interesting questions to researchers. Many traditional control concepts can now be reexamined in a new perspective. Interesting observations have been made in the process of applying the proposed methods in solving engineering problems. Some initial results in analytical study of the new control systems are also briefly introduced.

4.1 Redefining the Control Problem and Robustness

In existing control theory, the control design problem was given in terms of closed-loop characteristics, such as desired pole locations or a cost function to be minimized, that can only be captured with an accurate mathematical model. Thus it needs to be redefined in the new paradigm. For the sake of simplicity, the following assumptions were made in order to illustrate the concept.

<u>Plant</u>: Here, the class of plants of interests is described as $\ddot{x} = f(x, \dot{x}, t) + u(t)$, *y(t)=x(t).* Let *F* be the set of all possible $f(x, \dot{x}, t)$.

<u>Reference Input</u>: Let the desired output be *v(t)* which is obtained dynamically from $\ddot{v} = g(v, \dot{v})$. Let *V* be the set of all possible *v(t).*

<u>Controller</u>: A dynamic system with *y(t)* and *v(t)* as inputs and *u(t)* as output, represented by *C(p)*, where *p* is the controller parameter vector.

<u>Control problem</u>: Designing the controller, *C(p)*, such that for $\forall f(x, \dot{x}, t) \in F$ and $\forall v(t) \in V$, its output *u(t)* forces the output of the plant, *y(t)*, to follow the reference input, *v(t)*. In short, we denote {*F*, *V*} as the “control problem” and *C(p)* its solution.

Remarks

1. It seems that the above definition better describes a real control problem with a large amount of uncertainties and the lack of precise mathematical representation. The boundary of *F* can usually be characterized using the physical limitations of the plant, while *V* is determined based on the control specifications.

2. The new definition calls for a reevaluation of the robustness concept. For example, given a control problem {*F* , *V*} and its solution, $C(p_0)$, where p_0 is the controller parameter vector, define another set *P* which is the neighborhood of p_0 and satisfies that $\forall p \in P$, *C(p)* is also the solution of {*F* , *V*}, then one measure

of robustness of $C(p_0)$ is the size of P. Another measure of robustness of $C(p_0)$ is its operating range, or the size of $\{F, V\}$.

4.2 A Generic Time Constant Concept

The generic nonlinear PID control introduced in section 2.3 and the active disturbance rejection control introduced in section 2.4 are all model independent. Their inherent robustness allows them to be used for a large class of problems. But obviously there will be a limit of applicability and one controller cannot control all the processes. The questions of how to determine such limit for a controller and how to adjust controller parameters for a different class of problems prompted the research on a generic time constant concept, also known as the *time scale*[29,30]. It can also be viewed as an attempt to characterize the essential knowledge of the plant needed for control design.

Definition: Consider the problem defined in 4.1, let

$$M = \max\left\{ \max_{(x,\dot{x})\in G_1, t\geq 0, f\in F} \left| f(x,\dot{x},t) \right| ,\ \max_{(v,\dot{v})\in G_2} \left| g(v,\dot{v}) \right| \right\}$$

where G_1 and G_2 define the operating region. Then, we define $\rho = 1/\sqrt{M}$ as the *time scale* of the control problem $\{F, V\}$.

In the view of NPID or ADRC, regardless of specific forms of $f(x,\dot{x},t)$ and $g(v,\dot{v})$, ρ characterizes the problem $\{F, V\}$ as needed to determine the control parameter p. Different applications with the same time scale can now be controlled with the same controller. The exploration of effective control laws is no longer restricted by the traditional system classifications such as linear/nonlinear, time varying/time invariant, etc.

The concept of time scale also makes the controller "portable" in the sense that the control parameters can be converted easily between problems of different time scales. Using the ADRC in Section 2.4 as an example, assume $\beta_{011}, \beta_{021}, \beta_{031}, k_{P1}, k_{D1}$ are well tuned parameters for a process in (1) of the time scale ρ_1 with the sampling period of h_1. It was shown that the corresponding ADRC parameters for the system with the time scale ρ and the sampling period of h to achieve a similar performance is:

$$\beta_{01} = \frac{\rho_1}{\rho}\beta_{011},\ \beta_{02} = (\frac{\rho_1}{\rho})^2\beta_{021},\ \beta_{03} = (\frac{\rho_1}{\rho})^3\beta_{031},$$

$$k_P = (\frac{\rho_1}{\rho})^2 k_{P1}, k_D = (\frac{\rho_1}{\rho})k_{D1}, h = \frac{\rho}{\rho_1}h_1$$

More details can be found in [30] and more research is still being carried out.

4.3 Convergence and Stability Analysis of ADRC

Due to the lack of analysis tools, introducing nonlinearity into control algorithms, as shown above, brings difficulties to convergence and stability analysis. While the new concepts and methods are being developed, Han's group also has been persistent in pursuing the analysis of the new approaches. Progresses are being made in several aspects, as briefly discussed as follows.

The self-stable region (SSR) approach is a constructive continuous non-smooth synthesis method based on the improvements of the variable structure control method [22,23]. We can show the concept of SSR via an inherent property of the Theme Problem (1).

Assuming $a(t)$ in (1) is uncertain. It's clear that all convergent trajectories can't converge to the origin along any direction in the first and the third quadrants. Assuming that 1) G is the union of any two regions, which lie in the second and the forth quadrants respectively, and have only one contact point (0,0); 2) $\bar{G}$ excludes the axes except the origin, then, if there is a trajectory $(x_1(t), x_2(t))$ staying in G after certain time T, this trajectory can only converge to the origin[22,23]. This property is an inherent property, which is independent on the function $a(t)(= f(t, x_1, x_2, w))$ and the control input $u(t)$. In this sense, we call G a Self-Stable Region, or SSR, for the system in (1).

Then the purpose of control is to force all trajectories out of G entering the interior of G and forbid the trajectories inside G coming out G.

It's clear that, all stable sliding modes for the system (1) must lie in the second and the forth quadrants. Therefore the SSR concept reflects the essence of the sliding plane. Moreover, the restriction of the sliding mode being a super-plane is relaxed by replacing the sliding mode with the SSR, based on which a continuous nonlinear (non-smooth) feedback law can be constructed and the chattering in sliding mode control can be easily avoided.

Via the concept of SSR, [24] analyzed the convergence and the estimation error of ESO (13) in the second order case. [24] showed that the state of ESO does not convergence to a certain sliding mode but to a SSR, determined by $g_i(\bullet)$. The principle of choosing $g_i(\bullet)$ to guarantee the state trajectories of ESO converging to the SSR is obtained. Furthermore, the steady estimation error can be bounded by the structure of the SSR. Based on this work, a Lyapunov function is constructed for stability proof and tracking error analysis of the ADRC designed for aircraft attitude control [26].

In summary, the results presented here are merely indicative of another perspective in control research. They are by no means complete or mature. Even though there are perhaps more questions than answers at this stage, we believe that they point to a promising direction.

5. Concluding Remarks

An alternative paradigm for control system design developed on a novel design concept is introduced. A survey of various control algorithms, the initial analysis and practical applications employing the proposed framework is given. The design methods discussed here compliment the existing knowledge in that it is developed closely along the line of how control has been practiced in industry. Therefore, it provides a vehicle for improving control algorithms that are being used in practical applications.

The proposed work opens a wide array of research directions to 1) systematically explore the use of other possible nonlinear feedback mechanisms; 2) develop analysis tools for the new design; 3) expand the applications of the new methods to other industrial applications, etc. Much work is still ahead.

***Acknowledgement**:* The second author wishes to acknowledge the support by Natural Science Foundation of China (69904010) that facilitated this work. The authors would also like to thank anonymous reviewers for their suggestions and comments.

REFERENCES

[1] F.L. Lewis, *Applied Optimal Control and Estimation*, Prentice Hall, 1992.
[2] The Control Handbook, Editor W.S. Levine, pp. 198, CRC Press and IEEE Press, 1996.
[3] J.E. Slotine and W. Li, "Applied Nonlinear Control", Prentice Hall Inc, 1991.
[4] M. Krstic, I. Kanellakopoulos, and P. Kokotovic, Nonlinear and Adaptive Control Design, John Wiley and Sons Inc., 1995.
[5] N. Minorsky, "Directional Stability and Automatically Steered Bodies", J. Am. Soc. Nav. Eng., vol. 34, p. 280, 1922.
[6] J.G. Ziegler and N.B. Nichols, "Optimal Settings for Automatic Controllers", Trans. ASME, vol 64, pp759-768, 1942.
[7] K. J. Astrom and T. Hagglund, "PID Control", The Control Handbook, Editor W.S. Levine, pp. 198, CRC Press and IEEE Press, 1996.
[8] J. Han, "Control Theory: Is it a Theory of Model or Control?", Systems Science and Mathematical Sciences, Vol.9, No.4, pp.328-335, 1989.(In Chinese)
[9] J. Han, "Robustness of Control System and the Godel's Incomplete Theorem", Control Theory and Its Applications, 16(suppl.), 149-155, 1999.(In Chinese)
[10] J. Han, "The Structure of Linear Systems and The Feedback Systems Calculation", Proceedings of National Control Theory and Application Conference, Science Press, Beijing, 1981. (In Chinese)
[11] J. Han, "Linearity and Nonlinearity in Feedback Systems", Control and Decision, Vol.3, No.2, pp.27-32, 1988.(In Chinese)
[12] J. Han, "Nonlinear Design Methods for Control Systems", The Proc. Of The 14th IFAC World Congress, Beijing, 1999.
[13] V. Haimo, "Finite time controller", SIAM J. of Contr Optim, pp.760-770, 1986.
[14] S. Bhat and D. Bernstein, "Continuous finite time stabilization of the translational and rotational double integrators", IEEE Trans Automat Control, Vol. 43, No. 5, pp.678-682, 1998.
[15] J. Han and W. Wang, "Nonlinear Tracking Differentiator", Systems Science and Mathematical Sciences, Vol.14, No. 2, pp.177-183, 1994.(In Chinese)
[16] J. Han and L. Yuan, "The Discrete Form of the Tracking Differentiator", Systems Science and Mathematical Sciences, Vol.19, No.3, pp.268-273, 1999. (In Chinese)
[17] J. Han, "A New Type of Controller: NLPID", Control and Decision, Vol.9, No.6, pp.401-407, 1994. (In Chinese)
[18] R. Norton, Design of Machinery;, McGraw-Hill, Inc. 1992.
[19] J. Han, "A Class of Extended State Observers for Uncertain Systems", Control and Decision, Vol.10, No.1, pp.85-88, 1995.(In Chinese)
[20] Y. Huang, Z. Luo and S. Hosoe, Active Disturbance Rejection Control of robot systems based on Extended State Observer, the 19th Conference of Robot Society of Japan, September, 2001.
[21] J. Han, "Auto-Disturbance Rejection Control and its Applications", Control and Decision, Vol.13, No.1, pp.19-23, 1998.(In Chinese)
[22] Y. Huang and J. Han, "A new synthesis method for uncertain systems- the self-stable region approach", Int. J. Systems Sci., 1999, 30(1):33-39.
[23] Y. Huang and J. Han, "The self –stable region approach for second order nonlinear uncertain systems", Proc. 1999 IFAC World Congress, 1999, Vol. E: 135-140.
[24] Y. Huang and J. Han, "Analysis and design for nonlinear continuous Extended state Observer", Chinese Science Bulletin, 45,1373-1379,2000.
[25] V. I. Utkin, Sliding Modes in Cintrol and Optimization, Berlin, Springer-Verlag, 1992.
[26] Y. Huang, K, Xu, J. Han and J. Lam, Flight control design using Extended State Observer and non-smooth feedback, IEEE CDC conference, December, 2001.
[27] A.F.Filippov, "Differential Equations with Discontinuous Righthand Sides", Kluwer Academic Publishers, 1988.
[28] N.Wiener, The operational calculus, Math. Ann., 95, 557-584, 1926.
[29] J. Han and X. Wang, "Time Scale and Nonlinear PID Controller", Proceedings of the Chinese Control Conference, pp.314-321, Chinese Science and Technology Press, Beijing, 1994.(In Chinese)
[30] W. Zhang, Time Scale and Auot-Disturbances-Rejection-controller, Ph.D. Disstertation, Institute of System Science, AMSS, Chinese Academy of Science, 1999.
[31] Z. Gao, S. Hu, and F. Jiang, "A Novel Motion Control Design Approach Based On Active Disturbance Rejection", IEEE CDC conference, December, 2001.
[32] F. Jiang and Z. Gao, "An Adaptive Nonlinear Filter Approach to Vehicle Velocity Estimation for ABS", IEEE CCA/CACSD conference, September, 2000.
[33] F. Jiang and Z. Gao, "An Application of Nonlinear PID Control to a Class of ABS Problems", IEEE CDC conference, December, 2001.
[34] Y. Hou, Z. Gao, F. Jiang, and B. Boulter, "Active Disturbance Rejection Control for Web Tension Regulation", IEEE CDC conference, December, 2001.
[35] B. Mahawan, Z. Luo, J. Han and S. Nakajima, High-speed high precision motion control of robot using Extended State Observer, Journal of the Robotics Society of Japan, 18(2), 86-93, 2000. (In Japanese)
[36] Z. Hou, Y. Huang and J. Han, Active noise cancellation with a nonlinear mechanism, Proc. Of the IEEE Int. Conf. On Decision and Control, 1577-1578, Tampa, UAS, Dec, 1998.
[37]C.D.Johnson, A family of linear, time-invariant universal adaptive controllers for linear and non-linear plants, Int. J. Control, Vol. 49, No.4, 1217-1233, 1989.
[38]M. Corless, Tu. J, State/input estimation for a class of uncertain systems, Automatic, 34(6), 757-764, 1998.
[39]T.murakami and K. Ohnishi, Dynamic identification method of multi-degree of freedom robot based on disturbance observer, Journal of the Robotica Society of Japan, 11(1), 131-139, 1993. (In Japanese).