Cleveland State University
EngagedScholarship@CSU

Electrical Engineering & Computer Science Faculty
Publications

Electrical Engineering & Computer Science
Department

4-2002

From Linear to Nonlinear Control Means: A Practical Progression

Zhiqiang Gao
Cleveland State University, Z.GAO@csuohio.edu

Follow this and additional works at: https://engagedscholarship.csuohio.edu/enece_facpub

Part of the Controls and Control Theory Commons

How does access to this work benefit you? Let us know!

Original Citation
Gao, Z. (2002). From linear to nonlinear control means: A practical progression. ISA Transactions, 41(2), 177-189. doi:10.1016/S0019-0578(07)60077-9

Repository Citation
Gao, Zhiqiang, "From Linear to Nonlinear Control Means: A Practical Progression" (2002). *Electrical Engineering & Computer Science Faculty Publications*. 61.
https://engagedscholarship.csuohio.edu/enece_facpub/61

From linear to nonlinear control means: A practical progression

Zhiqiang Gao*

Department of Electrical and Computer Engineering, Cleveland State University, Cleveland, Ohio 44115, USA

1. Introduction

The proportional-integral-derivative (PID) based existing control technology has been in existence for close to 80 years. N. Minorsky first invented it in 1922 [1]. Ziegler and Nichols provided a set of tuning tables that has been used to this day [2]. Various control forms and tuning methods related to PID have continued to appear in literature, see, for example [3–7]. Nevertheless, the PID is still the tool of choice in over 90% of current industrial control applications [8].

Classical control theory, particularly the frequency response method, helps control engineers to get helpful insight on how and why feedback control system works, as well as how to improve it. Modern control theory, from Kalman filters to H_∞ control, represents the tremendous progress made in the last 40 years in mathematical control theory. But the level of mathematics required to understand it, the assumptions of linearity and time invariance, and its dependency on the mathematical model of the physical plant have limited its appeal for most control practitioners.

Judging from the state of control technology, it seems that the development in control theory has not been translated to a breakthrough technology. In the meanwhile, the control industry has completed the transition from analog to digital and is capable of implementing much more sophisticated algorithms than PID. But where is our next generation control technique that can tap into the power of the new digital control hardware?

This paper proposes a new vision on control technology progression. Keep in mind that the new technology

- should not be overly dependent on the mathematical model since they are often unavailable in practice

*Tel.: +1-216-687-3528; fax: +1-216-687-5405. *E-mail address*: z.gao@csuohio.edu

- needs to provide much better performance at low cost
- must be intuitive and simple to implement

It is in this background that several novel design methods are presented here. Some of these ideas, such as the use of nonlinear gains, have been utilized, in one form or the other, by practicing engineers as *ad hoc* solutions but were not systematically studied. The pioneering work of Professor Jingqing Han's in nonlinear control [9–15] provided a much needed framework for the study of such techniques.

In the following sections, several novel and practical design techniques, including nonlinear PID mechanisms and an innovative active disturbance rejection scheme, are presented. Practical applications are used to demonstrate their applicability and effectiveness. Note that this research is by no means complete. Instead, it points to a promising new direction of control technology development: break out from the constraints of the existing control practice and theory, and actively explore new control design strategies that are not overly dependent on mathematical models. Computer aided design tools, such Matlab/Simulink, provide an ideal environment for such endeavors where new ideas can be quickly tested. In fact, this is how the proposed techniques presented here were first developed.

2. Nonlinear PID control mechanisms

Engineers have long used various modifications of PID, from gain scheduling to fuzzy logic, in practice. The nonlinear variations usually provide much needed improvements of performance and flexibility of the controller. However, the design process tends to be tedious and solutions problem dependent. In this section, a systematic nonlinear PID design approach is proposed. It leads to simple, effective, and practical control solutions that apply to a wide range of industrial control problems.

2.1. Nonlinear proportional-integral control

Consider a single integrator plant

$$\dot{y} = w + u, \qquad (1)$$

with y and u as the output and input, respectively, and w as input disturbance. For convenience, the equation is written in terms of the error, $e = r - y$, where r is the desired output (constant)

$$\dot{e} = -w - u. \qquad (2)$$

The objective is to design a control law such that the closed-loop system satisfies $e(t) \rightarrow 0$, as $t \rightarrow \infty$. In practice, in the presence of the disturbance w, a proportional-integral (PI) control is usually employed:

$$u = K_p e + K_I \int e \, dt. \qquad (3)$$

It is well known that the integrator control helps to eliminate steady-state error. Unfortunately, it often causes saturation problems (known as integrator windup) and its 90° phase lag may adversely affect system stability. For these reasons, it is not uncommon to see that the integral control is gain-scheduled in practice. For example, a common practice is to zero the integral gain during the transient to prevent overshoot, saturation, and instability. This and other variations of PID are the results of ingenuities of practicing engineers. But, as control researchers, can we provide them with a more powerful set of tools that are less problematic? Let us start the investigation by examining a novel building block of a controller: a nonlinear gain.

2.1.1. Nonlinear proportional control

Intuitively, the reason integral control helps to reduce the steady state error is that it makes the controller more sensitive to small errors. This compliments the proportional control actions. An alternative is to use a nonlinear mechanism in the controller that yields higher gain for smaller errors, such as the one below:

$$u = K \cdot |e|^a \cdot \text{sgn}(e), \qquad (4)$$

which was systematically used by Han [12–14] to explore the advantage of nonlinear feedback. The sgn(e) function is defined as

$$\text{sgn}(e) = \begin{cases} 1, & e \geq 0 \\ -1, & e < 0. \end{cases}$$

Note that for $a = 0$ and 1, (4) represents the well-known bang-bang control and the linear proportional control, respectively. It is in the case of $0 < a < 1$ that an interesting phenomenon was observed. The selection of the parameter a is stud-
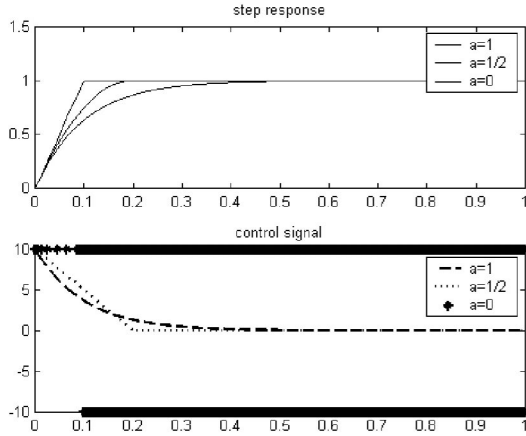
Fig. 1. Comparison of the linear and nonlinear controllers.

Fig. 2. Disturbance rejection comparisons.

ied in terms of steady-state error, transient response, disturbance rejection, and robustness, as shown below.

2.1.1.1. Steady-state error. For the same gain of $K = 10$, when $|w_0| < 1$, $\forall t$, the steady-state error is less than 0.1 for the linear controller with $a = 1$, and 0.01 for the nonlinear controller with $a = 1/2$. This bound is further reduced to 0.001 with $a = 1/3$. In general, as $a \to 0$, this bound will approach zero.

2.1.1.2. Transient response. With $K = 10$, the transient responses for $a = 0$, 1/2, 1, are shown in Fig. 1. Note that the linear control has the slowest response because it has a long tail end. This is due to its inherent nature of not being sensitive to small errors. On the other hand, the bang-bang control has the fastest response but the chattering in the control signal renders it impractical. The nonlinear control with $0 < a < 1$ seems to offer a good balance.

2.1.1.3. The choice of gain. Note that from Fig. 1, the gain K corresponds to the maximum control effort. Since the output never overshoots the target, the control signal is limited to the range of $[0, K]$, for $a > 0$. That is, the control gain can be conveniently chosen from the physical limitation of actuator.

2.1.1.4. Implication in hardware simplification. Keeping the control signal always positive is quite significant in applications where negative
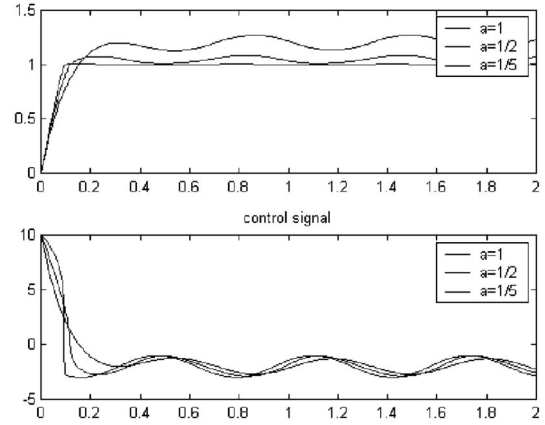
control signal requires a different set of hardware. For example, in temperature applications, the positive control signal means heating and the negative one cooling. The above results imply that, with a proper control strategy that produces fast response without overshoot, the cooling actuator may not be needed even for an aggressive control design with $0 < a < 1$. In drive applications, it may result in a motor drive that never applies the reversed torque, which not only helps to save energy, but also prolongs the life of motor and other parts in the drive train, such as coupling and transmission.

2.1.1.5. Disturbance rejection. To examine the disturbance rejection properties of linear and non-linear controllers, a disturbance of $w(t) = 2 + \sin(10t)$ is introduced. The results are shown in Fig. 2 for $a = 1$, 1/2, 1/5. Again, the linear control is the least desirable one.

2.1.1.6. Robustness. Let us examine the robustness of the control law in (4) with respect to the dynamic changes in the plant (1). Consider a first order system with a time constant τ:

$$\tau \dot{y} = -y + w + u. \tag{5}$$

With the same controller ($K = 10$, $a = 1/2$) and the disturbance, $w_0(t) = 2 + \sin(10t)$, used above, τ is varied from 0.1 to 1. The results are shown in Fig. 3. It is demonstrated that the high quality of control is maintained even though the pole of the system shifted significantly.

2.1.1.7. Remarks. (1) The reason the nonlinear controller performs better is that, with $0 < \alpha < 1$, it
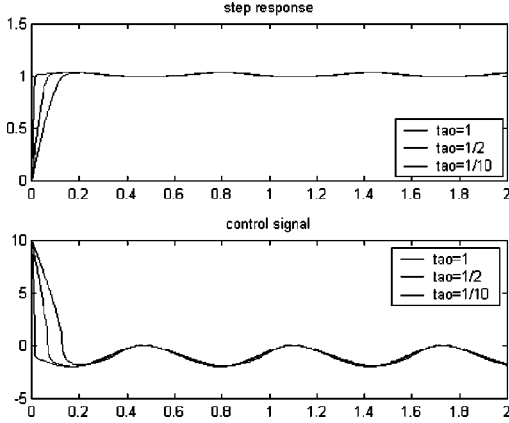
Fig. 3. Robustness of the nonlinear controller.

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 exhibit this kind of characteristics on its error surface. Of course a fuzzy controller is much more complicated to implement. (2) The above study demonstrates a desirable characteristic of nonlinear control: reducing the steady state error in the presence of disturbances does not necessarily require an integral control. This helps to enhance the stability margins of the control system by eliminating the phase delay associated with integral control. The control implementation is also simplified by removing the anti-windup mechanisms that is used to prevent the integrator from saturation. (3) In some applications, the physical plant exhibits high gain in certain operating conditions. In this case, the NPID concept should be applied so that the combined gain of the controller and the process has the shape of Fig. 4. This may require that the controller gain be actually small
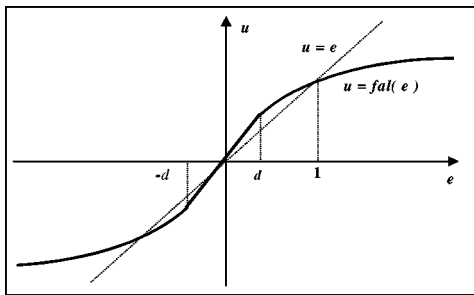

Fig. 4. Comparison of linear and nonlinear gains with a <1.

when the error is small, which is evident in the pH level regulation problems in process control.

2.1.2. Stability

Many industrial control problems, such as the drive systems and thermo processes, are approximately described by first order linear differential equation as

$$\dot{y} = -\beta y + w + bu, \qquad (6)$$

with $\beta > 0$ and $b > 0$, and w is the disturbance. But, considering many different nonlinear behaviors exhibited by physical devices, such as the dead zone, the hysteresis, and backlash, and the different forms of frictions, a better representation is

$$\dot{y} = -f(y,w) + bu, \qquad (7)$$

where $f(y,w)$ accounts for the nonlinear dynamics and disturbances in the system.

For the sake of simplicity, the disturbance w is ignored in the stability analysis. Note that although $f(y)$ is often unknown in practice, it usually shares the same sign of y, i.e., $yf(y) \geq 0$. Furthermore, the constant b generally has explicit physical meanings and is within a certain range. It represents an important relationship of how an actuator affects the system dynamically. For example, b corresponds to the torque constant of the motor in a drive system.

It was demonstrated above that the nonlinear feedback control law, such as

$$u = -K|y|^a \text{sgn}(y) \qquad (8)$$

is a very effective one. The question addressed here is if the stability of the closed-loop system can be assured with a limited knowledge of the plant.

Theorem 1. *The proposed control law in (8) guarantees the asymptotic stability of the closed-loop system if K is chosen such that*

$$bK > \sup_{y} \left(\frac{|f(y)|}{|y|^a} \right). \qquad (9)$$

Proof: Select the Lyapunov function $V(y) = y^2/2$,

$$\dot{V}(y) = y\dot{y} = -yf(y) - bKy|y|^a \text{sgn}(y)$$
$$= -yf(y) - bK|y|^{a+1}.$$

(a) If $yf(y)$ 0, and $bK>0$, then

$$\dot{V}(y)<0 \quad \text{for} \quad y\neq 0.$$

(b) If $yf(y)<0$, and $bK>0$ then for $\dot{V}(y)<0$ one needs

$$bK|y|^{a+1}>|yf(y)|,$$

or, equivalently

$$bK>\sup_{y}\left(\frac{|f(y)|}{|y|^a}\right).$$

Q.E.D.

2.1.3. The design trade-offs

The control law in (4) provides a relatively-high gain when error is small, which leads to small steady-state error, good disturbance rejection, and robustness. However, this high gain is sustained across the frequency spectrum and pushes the bandwidth of the controller upwards, which may make the controller too sensitive to noises. Since the gain of the controller in (4) increases as the error gets smaller, one possible trade-off is to limit the gain to a particular value. For example, changing (4) to

$$u=K\,\text{fal}(e). \tag{10}$$

where the function $\text{fal}(e)$, shown in Fig. 4, is defined as

$$\text{fal}(e)=\begin{cases} |e|^a \; \text{sgn}(e), & |e|>d, \\ e/d^{1-a}, & |e|\leq d, \end{cases} \quad d>0, \tag{11}$$

which limits the gain to K/d^{1-a} in the neighborhood of origin. This proves to be a very effective method in making a compromise between the effectiveness of the nonlinear control and the bandwidth and noise sensitivities of the control algorithm.

Another trade-off that can further lower the gain is adding an integral control in the neighborhood of steady-state. In particular, the integral control helps to eliminate the steady-state error without using high gain proportion control. Furthermore, the saturation and phase lag problems can be avoided, or at least lessened, by applying the integral control only in the region where the error is small. The new PI controller takes the form of

$$u=K_p\text{fal}(e)+K_I\int g_i(e)dt, \tag{12}$$

where $\text{fal}(\cdot)$ is defined in (11) and $g_i(e)$ is defined as

$$g_i(e)=\begin{cases} 0, & |e|>d_i, \\ e, & |e|\leq d_i, \end{cases} \quad d_i>0. \tag{13}$$

2.2. The insight and design of derivative control

The derivative control is used to overcome the overshoot of an underdamped second, or higher, order plant. The use of derivative control has always been a tricky part in practice. For example, it is often heard in motion control industry that the derivative control helps to reduce overshoot because it adds "damping" to the system. But it was not clear exactly why and how it happens. In addition, the differential signal is usually not measurable and is obtained by taking an approximate differentiation of a measurable signal. This makes the controller vulnerable to high frequency noises. These issues are addressed in this section.

2.2.1. The damping factor

Consider a standard motion equation of

$$J\ddot{y}=-\alpha\dot{y}+k_tu, \tag{14}$$

where y is the position of the motor shaft, u is the current in the motor armature, J is the inertia, α is the friction coefficient, and k_t is the motor torque constant. If α is very small, the plant will be highly underdamped, making it harder to control. This can be changed, however, with a derivative control. For example, the control law

$$u=-\beta\dot{y}+u_0. \tag{15}$$

will result in a dynamically compensated plant

$$J\ddot{y}=-(\alpha+\beta)\dot{y}+k_tu_0, \tag{16}$$

which is less underdamped and easier to control. Here u_0 is the new control input which can be designed, for example, using (12).

For the above reasons, one can now understand why derivative control adds "damping" to the system. This notion can also be generalized to other linear and nonlinear systems. For example, with the standard second order plant of

$$\ddot{y}=-2\xi\omega_n\dot{y}+\omega_n^2y+\omega_n^2u, \tag{17}$$

the control law of

$$u = \frac{-2}{n} \dot{y} + u_0 \qquad (18)$$

changes the damping ratio of the plant from ζ to $(\zeta + \zeta_1)$.

Based on these considerations, the following nonlinear PID control is proposed:

$$u = K_p \text{fal}(e) + K_I \int g_i(e) dt + K_D g_d(-\dot{y}), \qquad (19)$$

where the first two terms of PI control are the same as in (12), and the $g_d(x)$ is defined as

$$g_d(x) = \begin{cases} x - d_d, & |x| > d_d, \\ 0, & |x| \leq d_d, \end{cases} \quad d_d > 0, \quad (20)$$

which is known as the dead zone. It is used because once the closed-loop system enters the steady state, the differentiation signal contains mostly noises and should be discarded. That is, the derivative control is only used during the transient period to prevent overshoot and oscillations.

2.2.2. Approximate differentiation in noisy environment

Obtaining a good quality $\dot{e}$ signal is a key in any derivative control. Since the noise in the signal usually renders the pure differentiation useless, various linear approximations have been used, including

$$\frac{s}{(\tau s + 1)^m}, \quad m = 1, 2, \qquad (21)$$

and

$$\frac{1}{\tau_2 - \tau_1} \left(\frac{1}{\tau_1 s + 1} - \frac{1}{\tau_2 s + 1} \right). \qquad (22)$$

More recently, a few nonlinear observers were proposed to extract the differentiation signal from the input, including a nonlinear observer known as a tracking differentiator (TD) [10,11],

$$\dot{x}_1 = x_2,$$

$$\dot{x}_2 = -R \, \text{sgn} \left(x_1 - v(t) + \frac{x_2 |x_2|}{2R} \right), \qquad (23)$$

where $v(t)$ is the input signal, $x_1(t)$ and $x_2(t)$ are the states of the observer, and R is the filter design

parameter. This observer is based on the time optimal control theory where $x_1(t)$ tracks the input signal $v(t)$ and $x_2(t)$ converges to the generalized derivative of $v(t)$. Since $x_2(t)$ is obtained through integration, it is less sensitive to noises.

A linearized version of (18) was given in [15] of the form

$$\dot{x}_1 = x_2,$$

$$\dot{x}_2 = -mR^2 x_1 - 2\alpha R x_2 + 2mR^2 v(t), \quad (24)$$

where m and α are positive numbers. Mathematically, this is equivalent to the second order approximation in (22).

More recently, a sliding mode based differentiator known as a robust exact differentiator (RED) was introduced [16] as

$$\dot{x} = y,$$

$$y = y_1 - k|x - v(t)|^{1/2} \text{sgn}(x - v(t)), \qquad (25)$$

$$\dot{y}_1 = -a \, \text{sgn}(x - v(t)),$$

where a and k are design parameters.

The above methods in (21), (23), and (25) are compared in simulation. To make the comparison fair, the differentiators are first fed with white noises and the parameters are adjusted so that they yield the same level of noise amplification. Then the quality of the differentiation algorithms are compared based on the step response of a standard linear second order system (with $\zeta = 0.5$ and $\omega_n = 200$) where the exact derivative of the output is known and the output measurement is corrupted by about 10% white noises. The results are shown in Fig. 5. Four different differentiators are compared against the true derivative. The TD has the best performance, followed closely by the second order linear approximation in (22). The first order linear approximation and the RED performed poorly.

2.3. Soft start

Engineers have long realized that a gradual setpoint change, such as a ramp function, has certain advantages over the sudden change, such as a step or pulse function. The extensive use of the motion profiles, such as the trapezoidal and the S curve, are prime examples. In general, a "profile," instead of a step function, offers the following advantages:
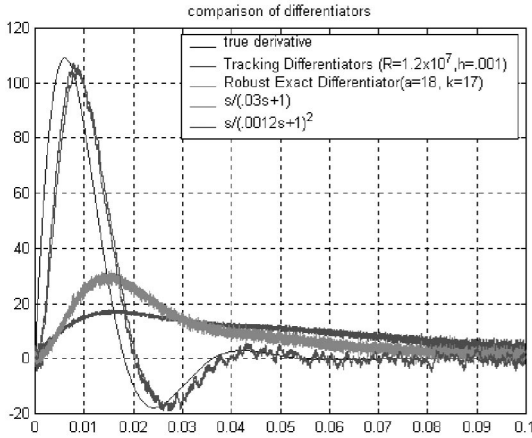
Fig. 5. Comparison of differentiators.

Fig. 6. Profiles generated from the TD.

(1) It provides a more reasonable trajectory for the plant to follow, resulting in a transient process where the error stays relatively small. This allows a more aggressive control design (such as higher gains).

(2) It allows the designer to specify the desired 1st, 2nd, or even 3rd derivatives of the output. This is especially important in motion control since the maximum speed, acceleration and its derivative (known as the Jerk) are all key design parameters.

(3) It makes it easier to avoid actuator saturation and the problems associated with it.

From our experience, the TD is an excellent profile generator with both desired output and its derivative. Instead of storing a fixed lookup table in the memory, TD is a simple equation that can be implemented together with the controller in software. The speed of the profile can be easily adjusted, even on the fly, with one parameter [R, in Eq. (23)]. For example, for $R = 10$, 50, and 100, the output of the TD and its derivative are shown in Fig. 6.

2.4. Design example 1

The proposed nonlinear PID scheme has been used in a number of applications, most notably in anti-lock brake systems (ABS) design for heavy commercial truck trailers [17]. Here, a relatively simple application of a digitally controlled dc-dc power converter is used to illustrate the design concept and process.
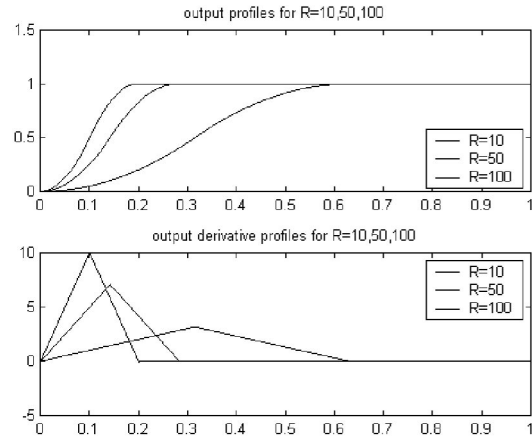
The transfer function, obtained through hardware tests [18], and the simulation model of the converter are shown in Fig. 7. The DAC and the ADC operate at the sampling frequency of 20 kHz, with an 8-bit resolution. The ranges of the input and output are 0–240 (PWM count) and 0–28 V (output dc voltage), respectively. The simulation model is normalized by two scaling factors of 240 and 1/28 so that both the input and output are in the range of [10]. All of these details are reflected in the simulation model, which yields a response that closely matches the hardware data.

The linear PI control gains were first tuned for the best results which yielded $K_p = 0.4$ and $K_I = 700$. The nonlinear PID gains were then similarly obtained in experiment: the proportional gain is 2 with $a = 0.4$ and $d = 0.2$; the integral gain is 5000 with $d_i = 0.2$ (integral control kicks in when error is less than 20% of the final value); the derivative gain is set at -0.0006, with $d_d = 0.1$. A TD with $R = 10^7$ is used to obtain the derivative of the output. The step response comparison of PI and NPID is shown in Fig. 8. A load disturbance of 15% is added at $t = 0.01$ sec. The numerical data indicate that the output voltage of PI dips 1.14 V while that of NPID dips only 0.2 V, i.e., the disturbance rejection for the NPID is more than five times better than that of the PI.
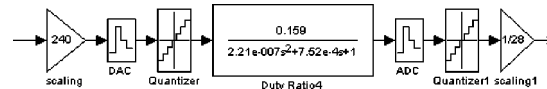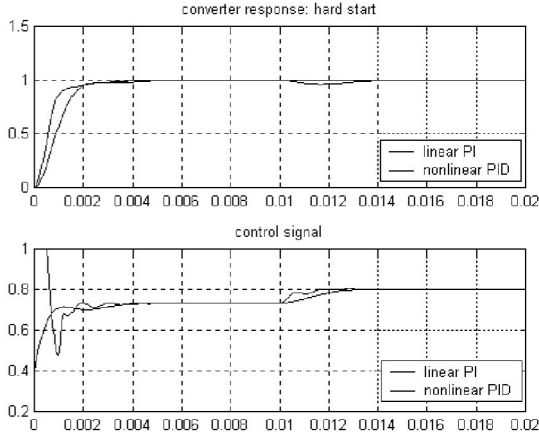

Fig. 7. Simulink model of the dc/dc converter.

Fig. 8. PI and NPID comparison for a dc/dc converter.

3. Active disturbance rejection

The primary reason for using the feedback control is to deal with the variations and uncertainties of the plant dynamics (internal) and unknown forces from the outside that exert influences on the behavior of the plant (external). Here, a generic design methodology is proposed to deal with the combination of both quantities, denoted as disturbance.

Consider a second order plant

$$\dot{x}_1 = x_2,$$

$$\dot{x}_2 = f(t, x_1, x_2, w) + bu, \qquad (26)$$

$$y = x_1,$$

where $f(t, x_1, x_2, w)$, primarily unknown, represents the combination of the dynamics of the plant and the external disturbance w. Traditional control theory relies on the knowledge of the function $f(t, x_1, x_2, w)$, which is usually not available in practice. The ensuing modeling and/or identification efforts are usually tedious. It also tends to make the control design highly dependent on the mathematical model of the plant, which creates the so-called robustness problem. Here, an alternative approach is proposed based on Han's work [12–14].

Definition. Active disturbance rejection control (ADRC) is defined as the control method where the value of $f(t, x_1, x_2, w)$ is estimated in real time and compensated by the use of the control signal, u.

The implication of this concept, in the sense of digital control, is that one does not have to know the analytical function of $f(t, x_1, x_2, w)$ in order to control the plant in (26), as long as he or she knows its value at each sampling instant, or $a(t) = f(t, x_1, x_2, w)$, $t = kt_s$, $k = 1, 2, \ldots$. With this knowledge, for example, a simple control law of

$$u = -a(t)/b + u_0 \qquad (27)$$

reduces the plant to a double integrator form:

$$\dot{x}_1 = x_2,$$

$$\dot{x}_2 = bu_0, \qquad (28)$$

$$y = x_1,$$

where u_0 is the new control input that can be easily designed using a PD controller. This control law can also be viewed as an *integrator-less PID*, since the dynamically compensated plant provides the integral action.

3.1. A disturbance observer

A simple estimation of $a(t) = f(t, x_1, x_2, w)$ is

$$a(t) = \ddot{x}(t) - b_0 u(t), \qquad (29)$$

where $\ddot{x}(t)$ is the estimated double differentiation of the output $x(t)$, and b_0 is the approximated value of b. Obviously, this method can very easily run into noise problems because of the double differentiation. A more systematic method, proposed by Han [12–14], is based on the well known state space observer design concept, as shown below.

We first expand the state space model of (26) to

$$\dot{x}_1 = x_2,$$

$$\dot{x}_2 = x_3 + bu, \quad x_3(t) = a(t) \triangleq f(t, x_1, x_2, w), \qquad (30)$$

$$\dot{x}_3 = h(t), \quad h(t) = \dot{a}(t),$$

$$y = x_1,$$

so that the unknown function $a(t)$ is now an extended state, x_3. Then, a state observer can be constructed to estimate all three states. With $a(t)$ and $h(t)$ unknown, a linear state space model of the plant can be expressed as

$$\dot{x} = Ax + Bu,$$

$$y = Cz, \qquad (31)$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ b \\ 0 \end{bmatrix}, \quad C = [1 \ \ 0 \ \ 0].$$

Now the state space observer, also known as the Luenberger observer, of (31) can be constructed as

$$\dot{z} = Az + Bu + L(y - \hat{y}),$$

$$\hat{y} = Cz, \qquad (32)$$

and L is the observer gain vector, which can be obtained using various known methods such as pole placement, or LQ:

$$L = -[\beta_{01}\beta_{02}\beta_{03}]. \qquad (33)$$

This observer is denoted as the extended state observer (ESO) of (26).

Intuitively, the closed-loop observer, or the correction term, $L(y - \hat{y})$ in particular, is used to accommodate the unknown initial states, the uncertainties in parameters, and disturbances. Whether such an observer can meet the control requirements is largely dependent on the observer design and $f(t, x_1, x_2, w)$.

In order to overcome the unknown function $f(t, x_1, x_2, w)$ and make $z \to x$ faster, the linear gains are now replaced with the nonlinear ones in the ESO, which is denoted as the nonlinear ESO (NESO) as

$$\dot{z}_1 = z_2 - \beta_{01}\mathrm{fal}(e, a_1, d_1),$$

$$\dot{z}_2 = z_3 - \beta_{02}\mathrm{fal}(e, a_2, d_2) + b_0 u, \qquad (34)$$

$$\dot{z}_3 = -\beta_{03}\mathrm{fal}(e, a_3, d_3),$$

where the function $\mathrm{fal}(e, a, d)$ is defined in (11) and b_0 is a rough estimate of b. This ingenious nonlinear observer form was originally proposed by Han [12,13].

3.1.1. Remarks

(1) Without the knowledge of $f(t, x_1, x_2, w)$, the linear observer will be hard pressed in providing reasonably accurate estimations in real time. The use of nonlinear gains are intended to make the observer converge faster, similar to the use of non-

linear feedback gains discussed earlier to make the output of the plant converge to the setpoint faster.

(2) If part or all of $f(t, x_1, x_2, w)$, say $f_1(t, x_1, x_2, w)$, is known, then it should be incorporated into the observer as

$$\dot{z}_1 = z_2 - \beta_{01}\mathrm{fal}(e, a_1, d_1),$$

$$\dot{z}_2 = f_1(t, z_1, z_2, w) + z_3 - \beta_{02}\mathrm{fal}(e, a_2, d_2) + b_0 u, \qquad (35)$$

$$\dot{z}_3 = h(t, z_1, z_2, w) - \beta_{03}\mathrm{fal}(e, a_3, d_3).$$

This will make the observer more efficient because z_3 will now track $a(t) = f(t, x_1, x_2, w) - f_1(t, x_1, x_2, w)$. Furthermore, if b is unknown, then a rough estimate, b_0, should be used in its place.

(3) The invention of the nonlinear extended state observer is a truly revolutionary concept that could change the landscape of control theory and practice. It helps to break down the boundary between control theory and practice and makes it possible to develop a truly model-independent design approach. This concept was tested and verified time and again in key industrial control problems [19–23]. For this reason and from this point on, this observer (ESO and NESO) will be referred to as Han's observer.

3.2. The control configuration

With the observer properly designed and $z_i \to x_i$, $i = 1,2,3$, the control design becomes straightforward. In particular,

$$u(t) = u_0(t) - z_3(t)/b_0 \qquad (36)$$

reduces the plant to (28) and a NPD control law such as

$$u_0(t) = -K_P\mathrm{fal}(e) - K_D\mathrm{fal}(\dot{e}) \qquad (37)$$

can then be applied. Because it actively estimates and compensates for the disturbance (including the internal and the external ones), this control law is denoted as active disturbance rejection controller (ADRC) [12 13].

The closed-loop control configuration is shown in Fig. 9. Here, $v(t)$ is the setpoint, $r(t)$ and $\dot{r}(t)$ are the desired trajectories for $y(t)$ and $\dot{y}(t)$, respectively, which are commonly known as a motion profile in the motion control industry.
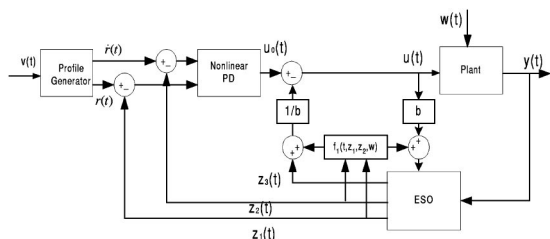
Fig. 9. ADRC configuration.

3.2.1. Observations

How do the new ideas proposed here fit into everyday practice? The following observations offer some insight.

(1) The existing PI control is still the simplest controller that can provide adequate performance in most applications.

(2) Currently, when the PI control fails to meet the requirements, engineers will try a variety of fixes, ranging from gain scheduling, to adaptive and fuzzy logic controls. This is where the nonlinear PID concepts shown above come in.

(3) In the spirit of keeping the control as simple as possible, one should start from the nonlinear proportional control, and proceed to nonlinear PD or ADRC only when necessary.

(4) Comparing ADRC to other more complex control methods, such as adaptive control, one notes that ADRC represents a completely different thinking on how a control problem is solved systematically. The basic idea of using nonlinear gains and Han's observer can be very well incorporated into other designs for better performance. In this sense, it does not compete but rather complements the existing techniques.

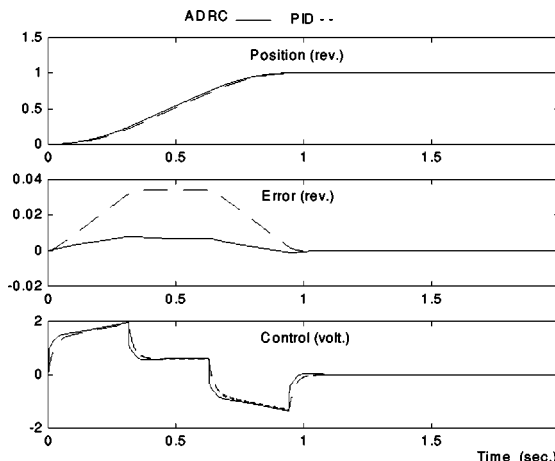(5) Similar to any observer based design, ADRC generally requires higher sampling rate, com-

Fig. 11. Nominal reponses.

pared to PID, to ensure that the observer provides reasonable estimates. From the limited simulation and experimentation studies, it appears that the ADRC requires 100–200 controller and observer updates during the transient period in a typical step response. That is, the sampling frequency is generally 25–50 times faster than that of the closed-loop response in a typical application.

(6) By tuning ADRC parameters carefully, the sampling rate can be reduced, but the process can be quite time consuming. In one application, it was discovered that a sampling rate less than five times that of the closed-loop bandwidth works reasonably well [20], although the performance can be much improved at a higher rate. Obviously, this is an important design trade off in industrial applications.
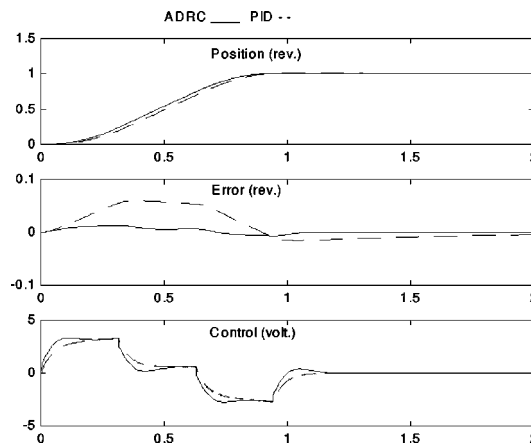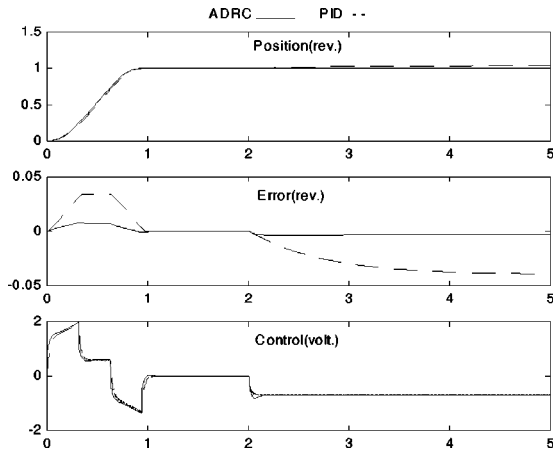

Fig. 10. The motion control test setup.

Fig. 12. Reponses to a 100% inertia change.

Fig. 13. Reponses to a 20% torque disturbance at $t = 2$ sec.

Fig. 15. Frequency response from T_d to y.

(7) Although a mathematical model of the plant is not required by the proposed methods, it should be incorporated into the design if such a model is available. This will help the observer to perform better and may help to lower the sampling rate.

3.3. Design example 2

An industrial motion control test bed, shown in Fig. 10, made by ECP is used to validate the ADRC design. The requirements associated with motion control dictate that the steady-state error be zero, the performance is robust in the face of inertia change, friction change, and torque disturbance. For simulation purposes, the mathematical model of the motion system is derived as
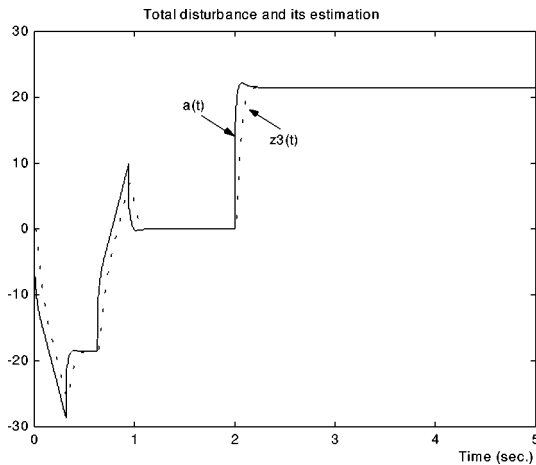
$$\ddot{y} = (-1.41\dot{y} + 23.2T_d) + 23.2u, \qquad (38)$$
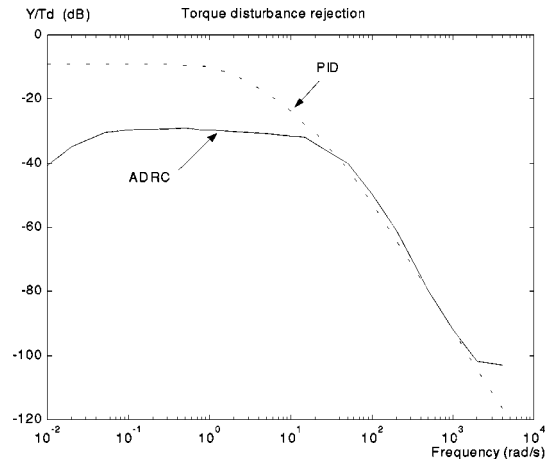
where u is the control signal and T_d is the torque disturbance. To see if ADRC can deal with a totally unknown system, this model is not used in the NESO. To make the test more realistic, the parameter b in NESO is chosen about 30% larger than the real one ($b = 30.71$).

3.3.1. Simulation

The ADRC algorithms shown in (34), (36), and (37) were digitized and compared to a PID controller. To make the comparison fair, both PID and ADRC are well tuned for the nominal plant, with a fixed inertia, minimal friction, and zero torque disturbance, as shown in Fig. 11. The robustness and
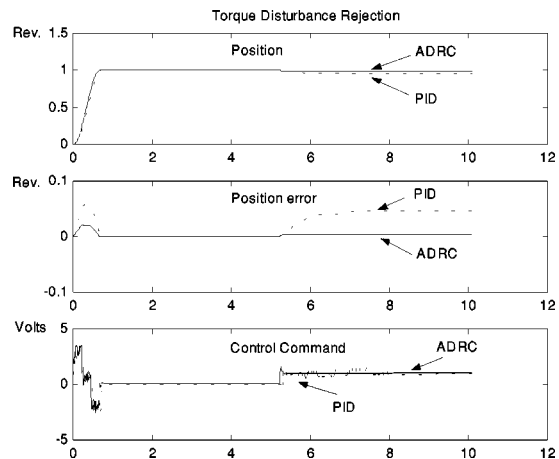

Fig. 14. Convergence of $x_3(t)$.

Fig. 16. Hardware disturbance rejection tests.

Table 1
Experimental results with ADRC and PID schemes.

		Over-shoot (%)	Settling time (s)	Steady-state error (rev)	RMS error (rad)
ADRC	Nominal case	0.00	0.615	6.87×10^{-4}	0.029
	Setpoint change	0.00	2.45	3.80×10^{-3}	0.116
	Total inertia doubed	0.30	0.858	1.25×10^{-4}	0.024
	With extra friction	0.00	0.625	4.13×10^{-3}	0.037
	30% of T_{max} (0.129 N.m) disturbance	N/A	0.19	4.13×10^{-3}	0.030
PID	Nominal case	0.20	0.639	1.87×10^{-4}	0.066
	Setpoint change	0.34	2.43	1.86×10^{-2}	0.216
	Total inertia doubled	0.98	0.872	5.31×10^{-3}	0.089
	With extra friction	0.00	0.669	1.61×10^{-2}	0.128
	30% of T_{max} (0.129 N.m) disturbance	N/A	2.11	4.69×10^{-2}	0.196

disturbance rejection are compared in Figs. 12 and 13, respectively. The ADRC shows a much better performance.

3.3.2. Validity of the ESO and ADRC concepts

The good performance of ADRC is not coincidental. For this example, the uncertainty function to be estimated is

$$a(t) = -1.41\dot{y} + 23.2T_d - 7.51u. \qquad (39)$$

Fig. 14 shows that $x_3(t)$ closely tracks it throughout the simulation.

The good disturbance property can be traced back to the frequency response between the disturbance input, T_d, and the output y, as shown in Fig. 15. The ADRC has a much better disturbance attenuation than the PID.

3.3.3. Hardware test

The PID and ADRC are programmed in C using a simple Euler's method. The controller ran on a 133 MHz Pentium computer with a decoder for the position signal. The sampling rate is 1 kHz. Overall, the hardware test results closely resemble that of the simulation. For example, the disturbance rejection test shown in Fig. 16 is quite similar to Fig. 13 except that the hardware disturbance has an opposite sign and the signals are noisier.

To quantify the difference between the PID and ADRC, overshoot, settling time (2%), steady-state error and the RMS error are used as criteria for comparison, as shown in Table 1. Overall, the improvements made by ADRC have been quite significant.

4. Concluding remarks

A practical nonlinear PID control framework and design techniques are introduced. An active disturbance rejection methodology, based on Han's observer, is presented. The new methods can be viewed as a natural progression of the existing PID technology that can tap into the new hardware capabilities and produce much more powerful control means. The new methods are still simple and intuitive for practicing engineers. They were tested successfully in software simulation and hardware verification of two practical problems.

References

[1] Minorsky, N., Directional stability and automatically steered bodies. J. Am. Soc. Nav. Eng. **34**, 280 (1922).

[2] Ziegler, J. G. and Nichols, N. B., Optimal settings for automatic controllers. Trans. ASME **64**, 759–768 (1942).

[3] Eucker, R. J. and Gao, Z., A novel self-tuning control design approach for industrial applications, in: IEEE Industrial Application Society 2000 Annual Meeting and World Conference on Industrial Applications of Electrical Energy, 8–12 October 2000.

[4] Hamdan, M. (Marv) and Gao, Z., A novel PID controller for pneumatic proportional valves with hysteresis, in: IEEE Industrial Application Society 2000 Annual Meeting and World Conference on Industrial Applications of Electrical Energy, 8–12 October 2000.

[5] Gao, Z. and Trautzsch, T. A., A stable self-tuning fuzzy logic control system for industrial temperature control problems, in: IEEE Industrial Application Society 2000 Annual Meeting and World Conference on Industrial Applications of Electrical Energy, 8–12 October 2000.

[6] Boulter, B. and Gao, Z., A novel approach for efficient self-tuning web tension regulation, in: Proceedings of the 4th IEEE Conference on Control Applications; Real-Time Systems (Special Issue). The Engineering of Complex Real-Time Computer Control Systems. The International Journal of Time-Critical Computing Systems, **11** (3) (November 1996).

[7] Dhayagude, N., Gao, Z., and Mrad, F., Fuzzy logic control of automated screw fastening. J. Robotics Comput. Aided Manufacturing **12** (3), 235–242 (1996).

[8] Astrom, K. J. and Hagglund, T., PID Control, The Control Handbook, edited by W. S. Levine. CRC Press, New York, 1996, p. 198.

[9] Han, J., Control theory: is it a theory of model or control?. Syst. Sci. Math. Sci. **9** (4), 328–335 (1989) (In Chinese).

[10] Han, J. and Wang, W., Nonlinear tracking differentiator, Syst. Sci. Math. Sci. **14** (2), 177–183 (1994) (In Chinese).

[11] Han, J. and Yuan, L., The discrete form of the tracking differentiator. Syst. Sci. Math. Sci. **19** (3), 268–273 (1999).

[12] Han, J., Nonlinear state error feedback control. Control and Decision **10** (3), 221–225 (1994) (In Chinese).

[13] Han, J., Nonlinear Design Methods for Control Systems, The Proc. Of The 14th IFAC World Congress, Beijing, 1999.

[14] Han, J., Robustness of control system and the Godel's incomplete theorem. Control theory and its applications, **16** (suppl.) 149–155 (1999) (In Chinese).

[15] Guo, B. and Han, J., A linear differentiator and application to the online estimation of the frequency of a sinusoidal signal, in: Proceedings of the 2000 IEEE International Conference on Control Applications, Anchorage, Alaska, 25–27 September 2000.

[16] Levant, A., Robust exact differentiation via sliding mode technique. Automatica **34** (3), 379–384 (1998).

[17] Jiang, F., A novel feedback control approach to a class of anti-lock brake problems, Ph.D. dissertation, Department of Electrical and Computer Engineering, Cleveland State University, 1 May 2000.

[18] Stimac, T., Digital Control of a DC-DC Converter for Space Applications, Department of Electrical and Computer Engineering, Cleveland State University, December 2000.

[19] Hou, Y., Novel control approaches for web tension regulation, Ph.D dissertation, Department of Electrical and Computer Engineering, Cleveland State University, May 2001.

[20] Hu, S., On high performance servo control solutions, Ph.D. Dissertation, Department of Electrical and Computer Engineering, Cleveland State University, May 2001.

[21] Hou, Y., Gao, Z. Jiang, F., and Boulter, B. Active disturbance rejection control for web tension regulation, submitted for publication.

[22] Gao, Z., Hu, S., and Jiang, F., A novel motion control design approach based on active disturbance rejection, submitted for publication.

[23] Feng, G. and Huang, L. H., A robust nonlinear controller for induction motor, IFAC'99 G, pp. 103–108.

Dr. Zhiqiang Gao is an Associate Professor of Electrical Engineering at Cleveland State University. He received a Ph.D. in Electrical Engineering from the University of Notre Dame in 1990. He is currently the director of the Applied Control Research Laboratory at CSU and has collaborated extensively with industry. He and his group have developed a set of innovative solutions including nonlinear control, fuzzy logic, self-tuning, a CAD software package, etc. Applications of these techniques can be found in many industry sectors such as automotive, paper and pulp, manufacturing, medical instrumentation, and space electronics. Dr. Gao also lectures broadly in industry on advanced control topics.