# ADAPTIVE SYSTEMS THEORY: SOME BASIC CONCEPTS, METHODS AND RESULTS*

GUO Lei
(*Institute of Systems Science, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing* 100080, *China*)

**Abstract.** The adaptive systems theory to be presented in this paper consists of two closely related parts: adaptive estimation (or filtering, prediction) and adaptive control of dynamical systems. Both adaptive estimation and control are nonlinear mappings of the on-line observed signals of dynamical systems, where the main features are the uncertainties in both the system's structure and external disturbances, and the non-stationarity and dependency of the system signals. Thus, a key difficulty in establishing a mathematical theory of adaptive systems lies in how to deal with complicated nonlinear stochastic dynamical systems which describe the adaptation processes.

In this paper, we will illustrate some of the basic concepts, methods and results through some simple examples. The following fundamental questions will be discussed: How much information is needed for estimation? How to deal with uncertainty by adaptation? How to analyze an adaptive system? What are the convergence or tracking performances of adaptation? How to find the proper rate of adaptation in some sense? We will also explore the following more fundamental questions: How much uncertainty can be dealt with by adaptation ? What are the limitations of adaptation ? How does the performance of adaptation depend on the prior information ? We will partially answer these questions by finding some "critical values" and establishing some "Impossibility Theorems" for the capability of adaptation, for several basic classes of nonlinear dynamical control systems with either parametric or nonparametric uncertainties.

**Key words.** Adaptive systems, estimation, control, uncertainty, stochastic systems, stability.

## 1 Introduction

Since the modelling, analysis, intervention or control of complex systems play an important role in the development of the modern science and technology, it is widely recognized that the research on complex systems is a frontier of science in the twenty first century. As is well-known, the Santa Fe Institute (SFI) has conducted extensive research activities on complex adaptive systems over the past two decades, while the Institute of Systems Science (ISS) of the Chinese Academy of Sciences has been one of the research centers on systems and control in China since its establishment in 1979. Investigations on complex systems and control systems are closely related in many cases. Therefore, it is desirable to provide a forum for experts in both complex systems and control to get together for a series of technical exchanges to explore the theory and methodology of intervention and adaptation of complex systems. As an initial step towards this, the SFI and ISS has jointly organized an International Symposium on "Intervention and Adaptation in Complex Systems" in Beijing in October 2002.

*This work is supported by the National Natural Science Foundation of China and the National Key Project of China. This paper is based on the presentation at the International Symposium on "Intervention and Adaptation in Complex Systems" held in Beijing from October 21–25, 2002.

The investigation of complex adaptive systems has been one of the focus of the above mentioned symposium. For example, Tom Kepler gave a general introduction to complex adaptive systems, John Holland talked about guiding complex adaptive systems, and Simon Levin talked about resiliency in complex adaptive systems. In this paper, we will not discuss complex adaptive systems directly, instead, we would like to take a different approach from different perspectives to answer such questions as what can be said about the quantitative theory for simple adaptive systems? We hope that this can offer some implications for a quantitative theory of complex adaptive systems.

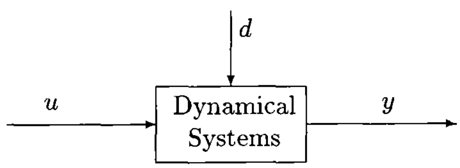
We will first explain the concept of adaptation from the perspective of systems and control. Then we will briefly present some standard methods used in adaptation, and indicate some theoretical difficulties even for analyzing some simple adaptive systems. Finally, we will present some quantitative results about the capability and limitations of adaptation.

## 2 What is Adaptation?

If we open a dictionary, we will probably find that "to adapt" means to change oneself to conform to a new or changed circumstance. If we examine this explanation carefully, we will find that there are two ingredients in it: one is the knowledge from the new information, and the other is the corresponding responses.

In technical terms, it means that adaptation consists of two steps. The first step is estimation or adaptive estimation (or identification, learning), and the second step is the control or decision (or intervention). These two steps are closely related, but can also be separated depending on applications. Obviously, adaptation is the use of system information based on the observed signals of systems. The information contained in the signals of dynamic systems can reduce the uncertainty in the system structure. Also, uncertainties in the system structure always exist in the modelling of practical dynamic systems. Basically, uncertainty can be classified into two types. One is called system uncertainty, meaning the internal or the structure uncertainty of the system, and the other is environmental uncertainty treated normally as a disturbance. But these two kinds of uncertainties are closely related: the environmental uncertainty may influence the structure uncertainty, and vice versa, so this classification is simply for the ease of study. The classification may also be situation dependent, and quite artificial sometimes. Furthermore, there are three cases of structure uncertainties typically. The first case is that a system contains structure uncertainties but the uncertainties can be represented by a certain unknown parameter (say $\theta$). In the second case, the system structure uncertainty is represented by a sequence of signals or time-varying unknown parameters (say $\{\theta(t)\}$). The last case is the functional uncertainty. For example, we know that there is a nonlinear $f(\cdot)$ between two things, but we don't know exactly what the relationship is. We will briefly discuss all the three cases in this paper.

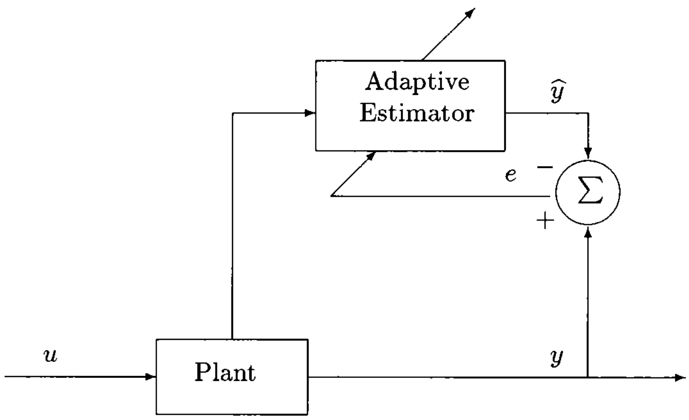
Now, let's have a look at adaptation in dynamical systems. Here is a diagram for dynamical systems:




where $u$ is the input sequence, $y$ is the output sequence, and $d$ is the disturbance in the dynamic system. Here the model for a system contains two parts actually. The first part is the prior knowledge about the system, and the second part is represented by some uncertainties. This understanding of model is different from what we are familiar with in, for example, the usual study of differential equations. Also, the key point here for a dynamic system is the so-called posterior information, meaning a collection of the input-output process at any time $t$,

$$\{y_0, y_1, \cdots, y_t; u_0, u_1, \cdots, u_t\},$$

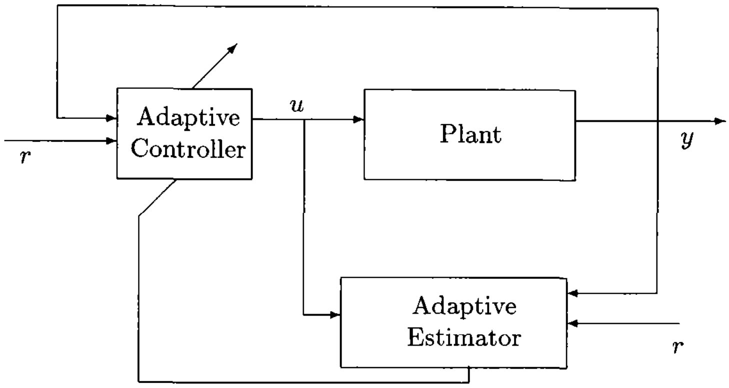
which is the information that we know during the operation of the system. It is the posterior information that makes it possible to reduce uncertainties or to respond to the changes through the process of adaptation. That is actually the process of adaptive estimation and/ or adaptive control. Intuitively, adaptive estimation can be explained as a parameter or structure estimator, which can be updated on-line based on the observed information of an uncertain system. The following diagram is an example




In the above, we have a plant with input sequence $u$, and output sequence $y$. We have an estimated model on the top with its simulated output $\hat{y}$ compared with the actual output $y$, resulting in a difference $e$ called the prediction error. Then one can use this difference to update the estimated model. This is a typical way. But, how to update the model? Usually, in the parametric case the on-line parameter estimator is derived from minimization of certain prediction errors, i.e.,

$$\theta_t = \underset{\theta}{\operatorname{argmin}}\{\text{Certain Prediction Errors}\}.$$

Next, what is adaptive control? Adaptive control is usually regarded as a controller with adjustable parameters (or structures) together with a mechanism for adjusting them(see, e.g. [1]–[3]). The following diagram gives an illustration.




Here, again, we have a plant which contains uncertainties in the structure. We don't know the precise structure of the system, so we use the online input and output information to get an estimate for the structure. Based on the estimated model at each time, we can then make a decision, which is usually called adaptive control. This online decision is then used to influence the plant dynamics, and so the model as well as the decision may change from time to time according to some objectives. Mathematically, it can be seen that both adaptive estimation and adaptive control are actually nonlinear mapping of the online observed signals of uncertain systems.

## 3 Some Standard Methods of Adaptation

In this section, we will outline some standard methods of adaptation together with some standard results. First, we consider adaptive estimation.

### 3.1 Adaptive Estimation

**(a) Parameter estimation**

There are many standard ways of parameter estimation.

Now, let's consider a typical example, i.e., the widely used linear regression model

$$y_{t+1} = \theta^\tau \phi_t + w_{t+1},$$

where $\theta$ is an uncertain parameter vector, $w_t$ is the noise signal, and $\phi(t)$ is called the regression vector. Normally, $\phi(t)$ is linear or nonlinear functions of the observed system signals. For example, $\phi(t)$ may be an (arbitrary) combination of the input and output data up to time $t$. A typical and widely used way to estimate the unknown parameter $\theta$ is the so-called Least-Squares(LS) method, which is mean square optimal in the Gaussian case. The recursive form of LS is as follows:

$$\theta_{t+1} = \theta_t + P_{t+1}\phi_t(y_{t+1} - \theta_t^\tau \phi_t),$$
$$P_{t+1} = P_t - P_t\phi_t\phi_t^\tau P_t(1 + \phi_t^\tau P_t\phi_t)^{-1},$$

where $P_t = (\sum_{i=0}^{t-1} \phi_i\phi_i^\tau)^{-1}$ is the estimation "covariance" matrix (or $P_t^{-1}$ is the information matrix). Roughly speaking, the current model updated by using the innovation gives a new model. In general, the LS minimizes the following prediction error:

$$J_t(\theta) = \sum_{i=1}^{t}(y_{i+1} - \varphi_i^\tau \theta)^2.$$

A basic question in parameter estimation is: how much information is needed for the convergence or strong consistency of the LS estimator, i.e.,

$$\theta_t \quad \longrightarrow \quad \theta \quad ?$$

A basic result in this direction is as follows: the minimum information (in a certain sense) required for strong consistency of LS is the following Lai-Wei[4] condition

$$\frac{\log \lambda_{\max}(P_t^{-1})}{\lambda_{\min}(P_t^{-1})} \longrightarrow 0, \qquad \text{as } t \longrightarrow \infty,$$

where $P_t^{-1}$ is the information matrix mentioned above. This condition is for the consistence of the LS parameter estimator. For practical systems, we usually also have something else unknown in the model. For example, the order of a linear system. The above condition can also be used to estimate the dimension or order of linear control systems without using any additional prior knowledge about the system structure (see e.g., [5]).

**(b) Parameter/Signal tracking**

Consider again the basic linear time-varying signal model

$$y_{t+1} = \theta_t^\tau \phi_t + w_{t+1},$$

where the only difference with the previous model is that $\theta_t$ is a time-varying process rather than a constant parameter. This model is important because in many cases, a nonlinear system can be satisfactorily approximated by a time-varying model. Our objective is to track $\theta_t$ based on the on-line observed information. The adaptive tracking algorithm usually takes the following general form,

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \mu L_t(y_{t+1} - \varphi_t^\tau \hat{\theta}_t),$$

the new estimated model is formed from the old estimated model updated by using the innovation or the new information of the system. Here $\mu$ stands for the rate of adaptation. Obviously, small $\mu$ gives slow adaptation because the update is not so large, while a large $\mu$ gives a large modification meaning fast adaptation. The gain matrix $L_t$ reflects the direction of adaptation, its main feature is that

$$L_t \nrightarrow 0, \qquad \text{as } t \to \infty.$$

Otherwise, the algorithm will not be able to track non-trivial time-varying parameters.

Let us briefly show some standard methods.

The first case is the Least Mean Squares(LMS), which is widely used in adaptive signal processing and adaptive control. The LMS is formed by simply taking $L_t$ in the above algorithms as (or $L_t$ in normalized case):

$$L_t = \varphi_t \qquad \left(\text{or} \qquad L_t = \frac{\phi_t}{1 + \|\phi_t\|^2}\right).$$

Actually, this corresponds to a gradient algorithm aiming at minimizing the expected prediction error,

$$e_t(\theta) = E(y_{t+1} - \phi_t^\tau \theta)^2.$$

The second standard case is the Forgetting Factor (FF) algorithm with $L_t$ given by

$$L_t = P_t \varphi_t,$$

$$P_t = \frac{1}{1-\mu}\left\{P_{t-1} - \mu\frac{P_{t-1}\varphi_t\varphi_t^\tau P_{t-1}}{1-\mu+\mu\varphi_t^\tau P_{t-1}\varphi_t}\right\},$$

which gives an estimate $\hat{\theta}_t$ that minimizes

$$\sum_{i=1}^{t}(1-\mu)^{t-i}(y_{i+1} - \varphi_i^\tau\theta)^2,$$

where $(1-\mu)$ is the "forgetting factor".

A remarkable feature of this algorithm is that the "old data" are discounted exponentially fast, in order to capture the most recent parameter changes.

The third standard case is the well-known Kalman Filtering (KF)-based algorithm, where the gain matrix $L_t$ has the following form:

$$L_t = \frac{P_{t-1}\varphi_t}{R+\varphi_t^\tau P_{t-1}\varphi_t},$$
$$P_t = P_{t-1} - \frac{P_{t-1}\varphi_t\varphi_t^\tau P_{t-1}}{R+\varphi_t^\tau P_{t-1}\varphi_t} + Q, \quad (R, Q > 0).$$

A main feature of this algorithm is that it is optimal in the mean square sense, if the system noise and the parameter variation process are Gaussian white noises with covariance matrices $R$ and $Q$. In fact, in this case, $\widehat{\theta}_t$ is the conditional expectation of $\theta_t$ given the observations:

$$\widehat{\theta}_t = E[\theta_t|\sigma\{\varphi_i, i < t\}].$$

This result is slightly different from the traditional ones, since $\phi_t$ is an adapted process.

Now, one may ask some basic questions:

(i) How much information on the system signals is needed in order to guarantee the stability of the tracking algorithms?

(ii) How to analyze / calculate the tracking performances measured by the covariance matrix of the tracking error:

$$\Pi_t = E[\widetilde{\theta}_t\widetilde{\theta}_t^\tau], \qquad \widetilde{\theta}_t = \theta_t - \widehat{\theta}_t \quad ?$$

(iii) Can we get the "best rate of adaptation" in a certain sense ?

To answer the first question, let us introduce the following condition called "conditional excitation" to describe the information needed

$$E\Big[\sum_{i=k+1}^{k+h}\frac{\phi_i\phi_i^\tau}{1+\|\phi_i\|^2}\,|\,\mathcal{F}_k\Big] \geq \delta I > 0, \quad \forall k$$

for some $h > 0$ and $\delta > 0$, where $\mathcal{F}_k \triangleq \sigma\{\phi_i, i \leq k\}$.

This condition is imposed on the regression vector which contains the information about input and output signals of a system. We remark that this condition is the minimum information needed for stability in some standard cases, and is a sufficient condition for stability in the general case[6].

As for the second question, we have a rather general formula for the tracking errors under some reasonable conditions[7]

$$E[\widetilde{\theta}_k\widetilde{\theta}_k^\tau] \quad \sim \quad \Pi = \mu\overline{R}_v + \frac{\gamma^2}{\mu}\overline{Q}_w,$$

where $\mu$, as mentioned before, is the rate of adaptation, $\gamma$ represents the speed of parameter variation, $R_v$ is a quantity reflecting the noise variance, and $Q_w$ is a quantity reflecting the variance of parameter variations. From this, we can see that the "best adaptation rate" is a tradeoff between tracking ability and noise sensitivity, e.g., for the RLS algorithm[7]:

$$\mu = \frac{\gamma}{R_v}\sqrt{\frac{\mathrm{Tr}(Q_w)}{\mathrm{Tr}(S^{-1})}},$$

which means that the "best" adaptation rate is proportional to the speed of parameter variation.

**(c) Functional estimation**

Let us consider a typical simple nonparametric model as follows:

$$y_{t+1} = f(y_t) + u_t + w_{t+1},$$

which is a stochastic dynamical system with unknown function $f$. There are many ways to estimate $f$ by using techniques from nonparametric estimation in statistics. One of the method is the so-called Kernel estimation described as follows.

We first define a Kernel function $K(x)$ satisfying the following three conditions:

(i) $K(x)$ is bounded and has compact support;

(ii) $K(0) > 0$, $K(-x) = K(x)$;

(iii) $\int K(x)dx = 1$, $\int [K^2(x) + |x|K(x)]dx < \infty$.

Then using this Kernel function to define a new, shifted function

$$\delta_j(x, y) = K(j^a(x - y)), \quad j \geq 1,$$

where $a > 0$, and $\delta_0 = 0$.

By using this, we can then get a kernel estimation $f(y)$ in the following way.

For any $y \in R$, the kernel estimate for $f(y)$ is defined as

$$\hat{f}_t(y) = \begin{cases} [N_{t-1}(y)]^{-1} \sum_{i=1}^{t} \delta_{i-1}(y_{i-1}, y)(y_i - u_{i-1}), & \text{if} \quad N(\cdot) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$N_{t-1}(y) = \sum_{i=1}^{t} \delta_{i-1}(y_{i-1}, y).$$

This nonparametric estimate can be used in adaptive control to get asymptotically optimal control systems in some cases (see [8,9]).

Another typical method is the nearest neighbor estimator. We first define $i_t$ as

$$i_t = \underset{0 \leq i \leq t-1}{\operatorname{argmin}} |y_t - y_{i-1}|.$$

In other words, $i_t$ corresponds to the point in the observations which is closest to the current observation value. Then for the above nonparametric model, the nearest neighbor estimator for $f(y_t)$ can be written as follows[10]

$$\hat{f}(y_t) = y_{i_t+1} - u_{i_t},$$

which is a simple, intuitive and powerful method in the design of adaptation laws[10].

### 3.2 Adaptive Control

**(a) "Certainty Equivalence" control**

In adaptive control, a basic concept is the so-called "certainty equivalence" principle, which normally means that the adaptive controller is designed by replacing an unknown parameter $\theta_0$ in a non-adaptive controller by its online estimate $\widehat{\theta}_t$, where the bias in the estimation is neglected

$$u_t = u(\theta_0, \phi_t)|_{\theta_0=\widehat{\theta}_t}.$$

Let's look at a simple (probably the simplest nontrivial) example — the first order linear control system:

$$y_{t+1} = ay_t + bu_t + w_{t+1}, \quad b \neq 0, \quad t = 0, 1, 2, \cdots ,$$

where $u_t$ is an input sequence, $w_t$ is the white noise sequence (for simplicity), $y_t$ is the output sequence. The parameters $a$ and $b$ are unknown, which can be any parameter in the plane except the line $b = 0$.

Our objective is as follows: given a desired reference signal $\{y_t^*\}$, to design a feedback control $u_t \in \mathcal{F}_t \triangleq \sigma\{y_i, i \leq t\}$, s.t. the following averaged tracking errors is minimized:

$$\lim_{T\longrightarrow\infty} \frac{1}{T} \sum_{t=1}^{T} (y_t - y_t^*)^2 = \min, \quad \text{a.s.}$$

Note that

$$E[y_{t+1} - y_{t+1}^*]^2 = \sigma_w^2 + E(ay_t + bu_t - y_{t+1}^*)^2.$$

Letting the last term to be zero, then we find that the minimum tracking error is $\sigma_w^2$, the variance of the noise. Also, the optimal control in the case where both $a$ and $b$ were known is

$$u_t = -\frac{1}{b}(ay_t - y_{t+1}^*).$$

Since $a$ and $b$ are unknown to us, this controller is not available. Now, let $(a_t, b_t)$ be any parameter estimate at time $t$ (for example, the LS estimate). Then, the certainty equivalence adaptive control is

$$u_t = -\frac{1}{b_t}(a_t y_t - y_{t+1}^*).$$

**(b) Cautious/Dual control**

There are also other ways to design an adaptive controller. Let $\theta = (a, b)^\tau$ and $\{w_t\}$ be independent and Gaussian distributed, and let $\{u_t\}$ be any feedback sequence. Then by the (generalized) Kalman filtering theory, the LS estimate $\theta_t$ for $\theta$, can be expressed as

$$\theta_t = E[\theta|\mathcal{F}_t^y], \quad P_t = E[\widetilde{\theta}_t \widetilde{\theta}_t^\tau|\mathcal{F}_t^y], \qquad \mathcal{F}_t^y = \sigma\{y_i, i \leq t\}.$$

By solving the following one-step-ahead minimization problem,

$$\min_{u_t \in \mathcal{F}_t^y} E[(y_{t+1} - y_{t+1}^*)^2|\mathcal{F}_t^y],$$

we can get the so-called cautious control represented by

$$u_t = -\frac{1}{b_t^2 + p_{22}(t)}(a_t b_t y_t - p_{12}(t) - b_t y_{t+1}^*),$$

which is different from the certainty equivalence control, since the estimation uncertainty measured by $P_t = (p_{ij}(t))$ has been taken into account. (In the certainty equivalence control, the terms $p_{12}(t)$ and $p_{22}(t)$ do not appear).

It can be seen that when the variance $p_{22}(t)$ of the estimate for $b$ is large, the feedback gain will be small. In this sense the above controller is cautious. Cautious control is only the result of a one-step-ahead minimization. The general case gives the so-called Dual Control, which is obtained from the following multistep minimization problem

$$V(t+1) \triangleq \min_{u_t,\cdots,u_N} E\left\{\sum_{i=t}^{N}(y_{i+1}-y_{i+1}^*)^2|\mathcal{F}_t^y\right\}.$$

From this, the following Bellman equation can be deduced:

$$V(t+1) = \min_{u_t} E\left\{(y_{t+1}-y_{t+1}^*)^2 + V(t+2)|\mathcal{F}_t^y\right\}.$$

Unfortunately, this equation is hard to solve in general, even in the case where $\theta$ is a constant parameter vector. Anyway, from the above equation, one may find that the optimal control has dual effects[11]: it should be a compromise between the control action and probing (or learning) action. This gives us some very valuable guidelines in designing better adaptive controllers (see, e.g., [1]).

## 4 Theoretical Obstacles

Now, let's discuss the theoretical difficulties in analyzing even some very simple adaptive systems.

Theoretically, we are interested in many properties of a closed-loop control system, e.g.,

a) Stability

$$\limsup_{T\longrightarrow\infty} \frac{1}{T}\sum_{t=1}^{T}(y_t^2+u_t^2) < \infty, \quad \text{a.s.}$$

b) Optimality

$$\lim_{T\longrightarrow\infty} \frac{1}{T}\sum_{t=1}^{T}(y_t-y_t^*)^2 = \min, \quad \text{a.s.}$$

c) Rate of convergence

$$\sum_{t=1}^{T}(y_t-y_t^*-w_t)^2 = O(?), \quad \text{a.s.}$$

d) Consistency, robustness $\cdots$

Now, let's go back to the simple LS-based-adaptive control system as mentioned in Section 3.2. We put all the related equations together to get the following closed-loop equation:

$$\begin{cases} y_{t+1} = ay_t + bu_t + w_{t+1}, \\ u_t = -\frac{1}{b_t}(a_t y_t - y^*_{t+1}), \\ \begin{pmatrix} a_t \\ b_t \end{pmatrix} = \begin{pmatrix} \sum\limits_{i=0}^{t-1} y_i^2, & \sum\limits_{i=0}^{t-1} u_i y_i \\ \sum\limits_{i=0}^{t-1} u_i y_i, & \sum\limits_{i=0}^{t-1} u_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum\limits_{i=0}^{t-1} y_i y_{i+1} \\ \sum\limits_{i=0}^{t-1} u_i y_{i+1} \end{pmatrix}. \end{cases}$$

The first is the system equation with unknown parameters $a$ and $b$. Then $a$ and $b$ are estimated by LS in the third equation using the online input and output data, giving the estimates $a_t$ and $b_t$ at each step $t$. It is a very complicated nonlinear function from data to $a_t$ and $b_t$. Then use $a_t$ and $b_t$ in the second equation to form a certainty equivalence controller, which then influence the system dynamics through the input $u_t$. One can see that even for this simple case, the output $y_t$ is generated from a very complicated nonlinear stochastic dynamical system, which can be written in the following general form:

$$y_{t+1} = f_t(y_0, y_1, \cdots, y_t) + w_{t+1}.$$

Now, one may ask: why the analysis of the closed-loop equation is complicated? Some of the main reasons are described as follows:

(i) The closed-loop system is a very complicated nonlinear stochastic dynamical system (even for the simple case), which defines the data to be used in the estimation and control.

(ii) No useful statistical properties, like stationarity or independency of the system signals are available or can be directly used without proof.

(iii) Also, because of (ii), no properties of $(a_t, b_t)$ are known *a priori.*

Actually, from the above closed-loop equation, one is likely to have a circular argument as follows: If we require $(a_t, b_t)$ to have good properties, then that implies that the signals used to estimate $(a_t, b_t)$ should be good in a certain sense. However, why the system signals are good? Note that the system signals are affected by a controller which is formed by using $(a_t, b_t)$, and so $(a_t, b_t)$ should have good properties in order for $(u_t, y_t)$ to have good properties. This is certainly a circular argument! In fact, it has been a long standing issue in adaptive control theory how to avoid this kind of arguments[12,13].

From the above analysis we can see that there is a key difference between adaptive systems theory and the standard statistical theory. We would like to remark that almost all adaptive estimation and control problems of dynamical systems have the similar theoretical difficulties.

## 5 Capability and Limitations of Adaptation

Now, let's mention some recent results on capability and limitations of adaptation for some typical nonlinear dynamical systems.

Let's first formulate the problem in a mathematical way.

Since it is the posterior information that enables one to reduce the uncertainty in a system, let us denote $I_t$ as the posterior information available at any time $t$,

$$I_t = \{y_0, y_1, \cdots, y_t; u_0, u_1, \cdots, u_{t-1}\}.$$

Then the estimation (in the abstract sense) at any time $t$ is actually a mapping from the information space to the parameter space:

$$\widehat{\theta}_t = \theta_t(I_t), \qquad \theta_t(\cdot) : \mathbb{R}^{\dim(I_t)} \longrightarrow \mathbb{R}^{\dim(\theta)}.$$

The feedback at $t$ is then a mapping from the information-parameter space to the control space:

$$u_t = u_t(\widehat{\theta}_t, I_t), \qquad u_t(\cdot) : I\!R^{\dim(I_t)+\dim(\theta)} \longrightarrow I\!R^{\dim(u)}.$$

Now, substituting $\widehat{\theta}_t$ into this leads to

$$u_t = g_t(I_t), \qquad g_t(\cdot) : I\!R^{\dim(I_t)} \longrightarrow I\!R^{\dim(u)}$$

which says that the feedback control is essentially a nonlinear mapping from the information space to the control space.

Adaptation law is defined as a collection of such $u_t$,

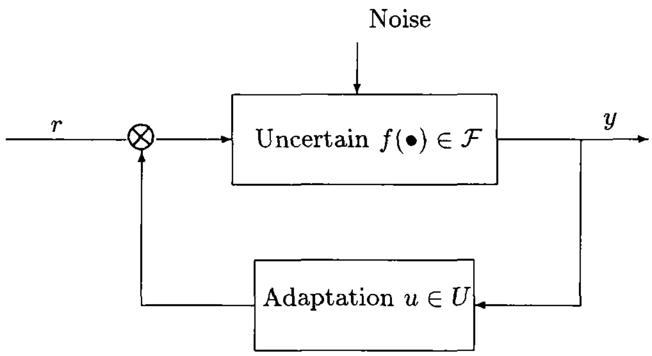
$$u = \{u_t, t = 0, 1, 2, \cdots\}.$$

The adaptation mechanism is then defined as the collection of all adaptation laws:

$$U = \{u| \quad u \text{ is any adaptation law } \}.$$

Obviously, the adaptation mechanism thus defined contains all the possible feedback controls or adaptation laws that can be constructed by using the posterior information. So the capability of adaptation means the capability of all the adaptation laws in $U$, not of a specific class of adaptation laws. Essentially, it is the capability of all possible feedbacks laws.

Below is a diagram explaining our problem formulation




Here we have an uncertain system with its model represented by some function $f$. This function is not known to us. What we would like to do is to find an adaptation law such that this system behaves well under some noise disturbances.

Since the function $f$ is uncertain, it is can be treated as any point in a set of functions. So the uncertainty can be measured by a set of functions. The adaptation law should be able to deal with any $f$ in the function set $F$. This is certainly a kind of robustness problems.

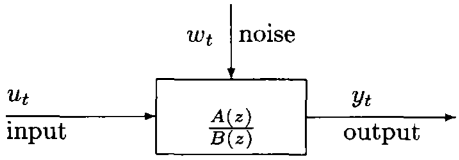
However, since we are interested in the capability of the whole class of adaptation laws (not only a particular adaptation law), our problem is: how large the size of the uncertainty set $\mathcal{F}$ can be dealt with by the set of adaptation laws $U$? This is a basic problem setting. Mathematically, it is a complicated optimization problem:

$$\sup_{u \in U} \left\{ Size(\mathcal{F}) : \sup_t \|y_t(f, u)\| < \infty, \quad \forall f \in \mathcal{F} \right\}.$$

To solve this problem is very hard in general. Here, we just show some cases where the solutions have been found.

**a) Linear time-invariant stochastic systems**

Here we have a transfer function model:




where $A(z)$ and $B(z)$ are coprime polynomials of finite order. If the noise is zero, then by solving a set of linear algebraic equations, one can get the precise value of the transfer function, as the transfer function only contains a finite number of parameters of the linear system. Hence, it appears to be a trivial adaptive control problem to not consider the noise effects, at least theoretically. In the noise case, however, there does indeed exist some essential difficulties in establishing a rigorous adaptive theory, as delineated in the last section. Despite of this, much progress has been made over the past several decades. The basic understanding now is that without requiring further prior information, adaptation is capable of achieving (asymptotically) the same control performance as the case where the system transfer function is known, for either tracking problems or LQ /pole-placement problems(see, e.g.[1,2],[12–15]).

We would like to remark that similar general assertions may not be true when either the "linear" assumption or the "time-invariant" assumption is removed. This will be explained later.

**b) Nonlinear parametric systems**

Consider the following nonlinear system with unknown parameters entering into the system in a linear way:

$$\begin{cases} y_{t+1} = \theta^\tau f(\phi_t) + w_{t+1}, \\ \phi_t = [y_t, y_{t-1} \cdots y_{t-p+1} u_t \cdots u_{t-q+1}]^\tau, \end{cases}$$

where the nonlinear function $f$ is assumed to have a linear growth rate:

$$\|f(x)\| \le c_1 + c_2\|x\|, \quad x \in R^{p+q}.$$

In this case, it can be shown that adaptive control results can also be established based on the existing results for linear systems.

The nontrivial case is the nonlinear growth rate. We now consider the following simple system

$$y_{t+1} = \theta f(y_t) + u_t + w_{t+1}$$

where the uncertainty is represented by an unknown parameter $\theta$, and $f(\cdot)$ is a known function with growth rate

$$f(x) = O(x^b), \quad \text{as} \quad x \to \infty.$$

Obviously, if $b$ is less than or equal to 1, then the growth rate is dominated by linear growth, which can be dealt with easily. Our question is: does adaptation have the capability of dealing with any nonlinear growth rate $b > 1$?

It turns out that $b = 4$ is a critical case for adaptive stabilizability[16]. To be precise, if $b < 4$, then one can design an adaptation law (for example, the LS-based adaptive control law) to globally stabilize the uncertain system.

However, if $b \geq 4$, it can be shown that there is no stabilizing adaptation law for the uncertain system. In other words, for any adaptation law $u$ in the adaptation mechanism $U$, it is always true that

$$E|y_t|^2 \to \infty, \text{ as } t \longrightarrow \infty.$$

**c) Linear time-varying systems**

We consider the following first order system

$$x_{t+1} = a(\theta_t)x_t + u_t + w_{t+1},$$

where $\theta_t \in \{1, 2\}$, is a Markov chain with transition probability matrix $\{p_{ij}\}$, $a(1) \triangleq a_1$, $a(2) \triangleq a_2$, and $p_{12} = p_{21}$.

It is easy to see that in the non-adaptive case where $\theta_t$ is available, one can cancel the first term in the system equation by choosing $u_t = -a(\theta_t)x_t$. If the noise is bounded then we have bounded output(stability).

In the adaptive case, however, there is a limitation on the capability of adaptation. In fact, it can be shown that the system is adaptively stabilizable if and only if the following inequality is satisfied (see [17])

$$(a_1 - a_2)^2(1 - p_{12})p_{12} < 1.$$

From this, we know that the capability of adaptation is inversely proportional to the uncertainty $Q = p_{12}(1 - p_{12})$, and the commonly used "rate of parameter changes" is not a proper measure in characterizing the capability of adaptation.

**d) Nonlinear and nonparametric systems**

Consider the following basic system:

$$y_{t+1} = f(y_t) + u_t + w_{t+1}, \quad y, w, u \in R^1,$$

where we assume that $f$ is unknown, but satisfies a Lipschitz condition:

$$|f(x) - f(y)| \leq L|x - y|, \quad L > 0, \quad x, y \in R^d.$$

It is known that an asymptotically optimal adaptive tracking control can be designed based on kernel estimation whenever $L \in (0, 1)$([8, 9]). Our question is: how large the Lipschitz condition $L$ is allowed by adaptation? Now we denote the class of uncertain Lipschitz functions as

$$\mathcal{F}(L) = \{f \mid |f(x) - f(y)| \leq L|x - y|, \quad \forall x, y \in R^1\}.$$

The size of this set can also be regarded as a measure of uncertainty of the above control systems.

It has been found that the Lipschitz constant $L = \frac{3}{2} + \sqrt{2}$ is a critical value for adaptive stabilizability[10]. In other words, if $L < \frac{3}{2} + \sqrt{2}$, then one can construct an adaptation law $\{u_t\}$, such that the closed-loop system is globally stable for all $f \in \mathcal{F}(L)$; while if $L \geq \frac{3}{2} + \sqrt{2}$, then there is no stabilizing adaptation law $\{u_t\}$ for the class of uncertain systems described by $\mathcal{F}(L)$.

## 6 Concluding Remarks

Adaptive techniques have been applied successfully in many engineering systems over the past several decades. In this paper, we have only presented some of the basic concepts, methods and results developed in adaptive systems theory. Various extensions of the existing theoretical results are possible, but many are hard and still open.

Finally, a natural question is: what can be done towards a mathematical theory of complex adaptive systems based on the existing results in adaptive systems? It is open for investigation.

## References


[1] K. J. Aström and B. Wittenmark, *Adaptive Control*, Addison-Wesley, Reading, MA, 2nd ed., 1995.

[2] H. Chen and L. Guo, *Identification and Stochastic Adaptive Control*, Boston, MA: Birkhäuser, 1991.

[3] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction and Control*, Prentice-Hall, Englewood Cliffs NJ, 1984.

[4] T. L. Lai and C. Z. Wei, Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, *Ann, Statist.,* 1982, **10**: 154–166.

[5] D. Huang and L. Guo, Estimation of nonstationary ARMAX models based on Hannan-Rissanen method, *The Annals of Statistics*, 1990, **18**(4): 1729–1756.

[6] L. Guo, Stability of recursive stochastic tracking algorithms, *SIAM J. on Control and Optimization,* 1994, **32**(5): 1195–1225.

[7] L. Guo and L. Ljung, Performance analysis of general tracking algorithms, *IEEE Trans.on Automatic Control,* 1995, **40**(8): 1388–1402.

[8] M. Duflo, *Random Iterative Models*, Berlin: Springer-Verlag, 1997.

[9] L. Guo and L. L. Xie, An asymptotically optimal non-parametric adaptive controller, *Science in China(Series E)*, 2000, **30**(3): 250–257.

[10] L. L. Xie and L. Guo, How much uncertainty can be dealt by feedback? *IEEE Trans. on Automatic Control,* 2000, **45**(12): 2203–2217.

[11] A. A. Feldbaum, *Optimal Control Theory*, New York: Academic Press, 1965.

[12] L. Guo and H. F. Chen, The Aström-Wittenmark self-tuning regulator revisited and ELS-based adaptive trackers, *IEEE Trans. Automat. Contr.*, 1991, **36**(7): 802–812.

[13] L. Guo, Convergence and logarithm laws of self-tuning regulations, *Automatica,* 1995, **31**(3): 435–450.

[14] L. Guo, Self-convergence of weighted least-squares with applications to stochastic adaptive control, *IEEE Trans. Automat. Contr.*, 1996, **41**(1): 79–89.

[15] T. L. Duncan, L. Guo and B. Pasik-Duncan, Continuous-time linear-quadratic Gaussian adaptive control, *IEEE Trans. Automatic Control*, 1999, **44**(9): 1653–1662.

[16] L. Guo, On critical stability of discrete-time adaptive nonlinear control, *IEEE Trans. Automat. Contr.*, 1997, **42**(11): 1488–1499.

[17] F. Xue, L. Guo and M. Y. Huang, Towards understanding the capability of adaptation for time-varying systems, *Automatica*, 2001, **37**: 1551–1560.