2

Adaptive Estimation and Control for Systems with Parametric and Nonparametric Uncertainties

Hongbin Ma* and Kai-Yew Lum†
Temasek Laboratories, National University of Singapore
*tslmh@nus.edu.sg**
tsllumky@nus.edu.sg†

Abstract

Adaptive control has been developed for decades, and now it has become a rigorous and mature discipline which mainly focuses on dealing parametric uncertainties in control systems, especially linear parametric systems. Nonparametric uncertainties were seldom studied or addressed in the literature of adaptive control until new areas on exploring limitations and capability of feedback control emerged in recent years. Comparing with the approach of robust control to deal with parametric or nonparametric uncertainties, the approach of adaptive control can deal with relatively larger uncertainties and gain more flexibility to fit the unknown plant because adaptive control usually involves adaptive estimation algorithms which play role of "learning" in some sense.

This chapter will introduce a new challenging topic on dealing with both parametric and nonparametric internal uncertainties in the same system. The existence of both two kinds of uncertainties makes it very difficult or even impossible to apply the traditional recursive identification algorithms which are designed for parametric systems. We will discuss by examples why conventional adaptive estimation and hence conventional adaptive control cannot be applied directly to deal with combination of parametric and nonparametric uncertainties. And we will also introduce basic ideas to handle the difficulties involved in the adaptive estimation problem for the system with combination of parametric and nonparametric uncertainties. Especially, we will propose and discuss a novel class of adaptive estimators, i.e. *information-concentration (IC) estimators*. This area is still in its infant stage, and more efforts are expected in the future for gainning comprehensive understanding to resolve challenging difficulties.

Furthermore, we will give two concrete examples of semi-parametric adaptive control to demonstrate the ideas and the principles to deal with both parametric and nonparametric uncertainties in the plant. (1) In the first example, a simple first-order discrete-time nonlinear system with both kinds of internal uncertainties is investigated, where the uncertainty of non-parametric part is characterized by a Lipschitz constant L, and the nonlinearity of parametric part is characterized by an exponent index b. In this example, based on the idea of the IC estimator, we construct a unified adaptive controller in both cases of $b = 1$ and

$b > 1$, and its closed-loop stability is established under some conditions. When the parametric part is bilinear ($b = 1$), the conditions given reveal the magic number $\frac{3}{2}+\sqrt{2}$ which appeared in previous study on capability and limitations of the feedback mechanism. (2) In the second example with both parametric uncertainties and non-parametric uncertainties, the controller gain is also supposed to be unknown besides the unknown parameter in the parametric part, and we only consider the noise-free case. For this model, according to some *a priori* knowledge on the non-parametric part and the unknown controller gain, we design another type of adaptive controller based on a gradient-like adaptation law with time-varying deadzone so as to deal with both kinds of uncertainties. And in this example we can establish the asymptotic convergence of tracking error under some mild conditions, althouth these conditions required are not as perfect as in the first example in sense that $L < 0.5$ is far away from the best possible bound $\frac{3}{2}+\sqrt{2}$.

These two examples illustrate different methods of designing adaptive estimation and control algorithms. However, their essential ideas and principles are all based on the *a priori* knowledge on the system model, especially on the parametric part and the non-parametric part. From these examples, we can see that the closed-loop stability analysis is rather nontrivial. These examples demonstrate new adaptive control ideas to deal with two kinds of internal uncertainties simultaneously and illustrates our elementary theoretical attempts in establishing closed-loop stability.

1. Introduction

This chapter will focus on a special topic on adaptive estimation and control for systems with parametric and nonparametric uncertainties. Our discussion on this topic starts with a very brief introduction to adaptive control.

1.1 Adaptive Control

As stated in [SB89], "Research in adaptive control has a long and vigorous history" since the initial study in 1950s on adaptive control which was motivated by the problem of designing autopilots for air-craft operating at a wide range of speeds and altitudes. With decades of efforts, adaptive control has become a rigorous and mature discipline which mainly focuses on dealing parametric uncertainties in control systems, especially linear parametric systems.

From the initial stage of adaptive control, this area has been aiming at study how to deal with *large uncertainties* in control systems. This goal of adaptive control essentially means that one adaptive control law cannot be a *fixed* controller with fixed structure and fixed parameters because any fixed controller usually can only deal with *small uncertainties* in control systems. The fact that most fixed controllers with certain structure (e.g. linear feedback control) designed for an exact system model (called *nominal model*) can also work for a small range of changes in the system parameter is often referred to as *robustness*, which is the kernel concept of another area, *robust control*. While robust control focuses on studying the stability margin of fixed controllers (mainly linear feedback controller), whose design essentially relies on priori knowledge on exact nominal system model and bounds of uncertain parameters, adaptive control generally does not need a priori information about the bounds on the uncertain or (slow) time-varying parameters. Briefly speaking, comparing with the approach of robust control to deal with parametric or nonparametric uncertainties, the approach of adaptive control can deal with relatively larger uncertainties and gain more flexibility to fit the unknown plant because adaptive control usually involves adaptive estimation algorithms which play role of "learning" in some sense.

The advantages of adaptive control come from the fact that adaptive controllers can adapt themselves to modify the control law based on estimation of unknown parameters by recursive identification algorithms. Hence the area of adaptive control has close connections with system identification, which is an area aiming at providing and investigating mathematical tools and algorithms that build dynamical models from measured data. Typically, in system identification, a certain model structure is chosen by the user which contains unknown parameters and then some recursive algorithms are put forward based on the structural features of the model and statistical properties of the data or noise. The methods or algorithms developed in system identification are borrowed in adaptive control in order to estimate the unknown parameters in the closed loop. For convenience, the parameter estimation methods or algorithms adopted in adaptive control are often referred to as *adaptive estimation* methods. Adaptive estimation and system identification share many similar characteristics, for example, both of them originate and benefit from the development of statistics. One typical example is the frequently used least-squares (LS) algorithm, which gives parameter estimation by minimizing the sum of squared errors (or residuals), and we know that LS algorithm plays important role in many areas including statistics, system identification and adaptive control. We shall also remark that, in spite of the significant similarities and the same origin, adaptive estimation is different from system identification in sense that adaptive estimation serves for adaptive control and deals with dynamic data generated in the closed loop of adaptive controller, which means that statistical properties generally cannot be guaranteed or verified in the analysis of adaptive estimation. This unique feature of adaptive estimation and control brings many difficulties in mathematical analysis, and we will show such difficulties in later examples given in this paper.

1.2 Linear Regression Model and Least Square Algorithm

Major parts in existing study on regression analysis (a branch of statistics) [DS98, Ber04, Wik08j], time series analysis [BJR08, Tsa05], system identification [Lju98, VV07] and adaptive control [GS84, AW89, SB89, CG91, FL99] center on the following linear regression model

$$z_k = \theta^\tau \phi_k + v_k \tag{1}$$

where $\{z_k\}$, ϕ_k, v_k represent observation data, regression vector and noise disturbance (or external uncertainties), respectively. Here θ is the unknown parameter to be estimated. Linear regression models have many applications in many disciplines of science and engineering [Wik08g, web08, DS98, Hel63, Wei05, MPV07, Fox97, BDB95]. For example, as stated in [web08], *Linear regression is probably the most widely used, and useful, statistical technique for solving environmental problems. Linear regression models are extremely powerful, and have the power to empirically tease out very complicated relationships between variables.* Due to the importance of model (1.1), we list several simple examples for illustration:

- Assume that a series of (stationary) data (x_k, y_k) $(k = 1, 2, \cdots, N)$ are generated from the following model

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where β_0, β_1 are unknown parameters, $\{x_k\}$ are i. i. d. taken from a certain probability distribution, and $\varepsilon_k \approx N(0, \sigma^2)$ is random noise independent of X. For this model, let $\theta = [\beta_0, \beta_1]^\tau$, $\varphi_k = [1, x_k]^\tau$, then we have $y_k = \theta^\tau \phi_k + \varepsilon_k$. This example is a classic topic in statistics to study the statistical properties of parameter estimates $\hat{\theta}_N$ as the data size N grows to infinity. The statistical properties of interests may include $\mathrm{E}(\hat{\theta} - \theta), \mathrm{Var}(\hat{\theta})$, and so on.

- Unlike the above example, in this example we assume that x_k and x_{k+1} have close relationship modeled by

$$x_{k+1} = \beta_0 + \beta_1 x_k + \varepsilon_k$$

where β_0, β_1 are unknown parameters, and $\varepsilon_k \approx N(0, \sigma^2)$ are i. i. d. random noise independent of $\{x_1, x_2, \cdots, x_k\}$.

This model is an example of linear time series analysis, which aims to study asymptotic statistical properties of parameter estimates $\hat{\theta}_N$ under certain assumptions on statistical properties of ε_k. Note that for this example, it is possible to deduce an explicit expression of x_k in terms of ε_j $(j = 0, 1, \cdots, k-1)$.

- In this example, we consider a simple control system

$$x_{k+1} = \beta_0 + \beta_1 x_k + bu_k + \varepsilon_k$$

where $b \neq 0$ is the controller gain, ε_k is the noise disturbance at time step k. For this model, in case where b is known *a priori*, we can take; $\theta = [\beta_0, \beta_1]^\tau$, $\phi_k = [1, x_{k-1}]^\tau$, $z_k = x_k - bu_{k-1}$; otherwise, we can take $\theta = [\beta_0, \beta_1, b]^\tau$, $\phi_k = [1, x_{k-1}]^\tau$, $z_k = x_k - bu_{k-1}$. In both cases, the system can be rewritten as

$$z_k = \theta^\tau \phi_k + \varepsilon_k$$

which implies that intuitively, θ can be estimated by using the identification algorithm since both data z_k and ϕ_k are available at time step k. Let $\hat{\theta}_k$ denote the parameter estimates at time step $\hat{\theta}_k$, then we can design the control signal u_k by regarding as the real parameter θ:

$$u_k = \frac{1}{\hat{b}}[r_{k+1} - \hat{\beta}_0 - \hat{\beta}_1 x_k]$$

where $\{r_k\}$ is the known reference signal to be tracked, and $\hat{b}$, $\hat{\beta}_0$, $\hat{\beta}_1$ are estimates of b, β_0, β_1, respectively. Note that for this example, the closed-loop system will be very complex because the data generated in the closed loop essentially depend on all history signals. In the closed-loop system of an adaptive controller, generally it is difficult to analyze or verify statistical properties of signals, and this fact makes that adaptive estimation and control cannot directly employ techniques or results from system identification. Now we briefly introduce the frequently-used LS algorithm for model (1.1) due to its importance and wide applications [LH74, Gio85, Wik08e, Wik08f, Wik08d]. The idea of LS algorithm is simply to minimize the sum of squared errors, that is to say,

$$\hat{\theta}_n^{LS} \triangleq \arg\min_{\zeta} \sum_{k=1}^{n} [z_k - \zeta^\tau \phi_k]^\tau [z_k - \zeta^\tau \phi_k] \tag{1.2}$$

This idea has a long history rooted from great mathematician Carl Friedrich Gauss in 1795 and published first by Legendre in 1805. In 1809, Gauss published this method in volume two of his classical work on celestial mechanics, *heoria Motus Corporum Coelestium in sectionibus conicis solem ambientium*[Gau09], and later in 1829, Gauss was able to state that the LS estimator is optimal in the sense that in a linear model where the errors have a mean of zero, are uncorrelated, and have equal variances, the best linear unbiased estimators of the coefficients is the least-squares estimators. This result is known as the Gauss-Markov theorem [Wik08a].

By Eq. (1.2), at every time step, we need to minimize the sum of squared errors, which requires much computation cost. To improve the computational efficiency, in practice we often use the recursive form of LS algorithm, often referred to as *recursive LS algorithm*, which will be derived in the following. First, introducing the following notations

$$Z_n = \begin{bmatrix} z_1^\tau \\ \vdots \\ z_n^\tau \end{bmatrix}, \quad \Phi_n = \begin{bmatrix} \phi_1^\tau \\ \vdots \\ \phi_n^\tau \end{bmatrix}, \quad V_n = \begin{bmatrix} v_1^\tau \\ \vdots \\ v_n^\tau \end{bmatrix}, \tag{1.3}$$

and using Eq. (1.1), we obtain that

$$Z_n = \Phi_n \theta + V_n.$$

Noting that

$$\begin{aligned}
& \sum_{k=1}^{n}[z_k - \zeta^\tau \phi_k]^\tau [z_k - \zeta^\tau \phi_k] \\
= & [Z_n - \Phi_n \zeta]^\tau [Z_n - \Phi_n \zeta] \\
= & Z_n \tau Z_n - 2\Phi_n \zeta + \zeta^\tau \Phi_n^\tau \Phi_n \zeta \\
= & [\zeta - (\Phi_n^\tau \Phi_n)^+ \Phi_n^\tau Z_n]^\tau \Phi_n^\tau \Phi_n [\zeta - (\Phi_n^\tau \Phi_n)^+ \Phi_n^\tau Z_n] \\
& + Z_n^\tau [I - \Phi_n (\Phi_n^\tau \Phi_n)^+ \Phi_n^\tau] Z_n
\end{aligned}$$

where the last equation is derived from properties of Moore-Penrose pseudoinverse [Wik08h]

$$\Phi_n^\tau = \Phi_n^\tau \Phi_n \Phi_n^+ = \Phi_n^\tau \Phi_n (\Phi_n^\tau \Phi_n)^+ \Phi_n^\tau$$

we know that the minimum of $[Z_n - \Phi_n \varsigma]^\tau [Z_n - \Phi_n \varsigma]$ can be achieved at

$$\hat{\theta}_n^{LS} = (\Phi_n^\tau \Phi_n)^+ \Phi_n^\tau Z_n \tag{1.4}$$

which is the LS estimate of θ. Let

$$P_n \triangleq (\Phi_n^\tau \Phi_n)^+$$

and then, by Eq. (1.3), with the help of matrix inverse identity

$$[A^{-1} + A^{-1} B^\tau C^{-1} B A^{-1}]^{-1} = A - B^\tau (C + B A^{-1} B^\tau)^{-1} B$$

we can obtain that

$$\begin{aligned}
P_n &= (P_{n-1}{}^{-1} + \phi_n \phi_n^\tau)^{-1} \\
&= [A^{-1} + A^{-1} B^\tau C^{-1} B A^{-1}]^{-1} \\
&= P_{n-1} - (P_{n-1}\phi_n)[1 + (\phi_n^\tau P_{n-1}) P_{n-1}^{-1} (P_{n-1}\phi_n)]^{-1} (\phi_n^\tau P_{n-1}) \\
&= P_{n-1} - a_n P_{n-1} \phi_n \phi_n^\tau P_{n-1}
\end{aligned}$$

where

$$a_n = (1 + \phi_n^\tau P_{n-1} \phi_n)^{-1}$$

Further,

$$
\begin{aligned}
\hat{\theta}_n &= P_n \sum_{k=1}^{n} \phi_k z_k^{\tau} \\
&= [P_{n-1} - a_n P_{n-1}\phi_n\phi_n^{\tau} P_{n-1}][\sum_{k=1}^{n-1} \phi_k z_k^{\tau} + \phi_n z_n^{\tau}] \\
&= \hat{\theta}_{n-1} - a_n P_{n-1}\phi_n\phi_n^{\tau} P_{n-1}\hat{\theta}_{n-1} + P_{n-1}\phi_n z_n^{\tau} - a_n P_{n-1}\phi_n\phi_n^{\tau} P \\
&= \hat{\theta}_{n-1} - a_n P_{n-1}\phi_n\phi_n^{\tau} P_{n-1}\hat{\theta}_{n-1} + P_{n-1}\phi_n[1 - a_n\phi_n^{\tau} P_{n-1}\phi_n] z_n^{\tau} \\
&= \hat{\theta}_{n-1} - a_n P_{n-1}\phi_n\phi_n^{\tau} P_{n-1}\hat{\theta}_{n-1} + a_n P_{n-1}\phi_n z_n^{\tau} \\
&= \hat{\theta}_{n-1} + a_n P_{n-1}\phi_n(z_n^{\tau} - \phi_n^{\tau}\theta_{n-1})
\end{aligned}
$$

Thus, we can obtain the following recursive LS algorithm

$$\hat{\theta}_n = \hat{\theta}_{n-1} + a_n P_{n-1}\phi_n(z_n^{\tau} - \phi_n^{\tau}\theta_{n-1})$$

where P_{n-1} and θ_{n-1} reflect only information up to step $n-1$, while a_n, ϕ_n and $z_n^{\tau} - \phi_n^{\tau}\theta_{n-1}$ reflect information up to step n.

In statistics, besides linear parametric regression, there also exist generalized linear models [Wik08b] and non-parametric regression methods [Wik08i], such as kernel regression [Wik08c]. Interested readers can refer to the wiki pages mentioned above and the references therein.

1.3 Uncertainties and Feedback Mechanism

By the discussions above, we shall emphasize that, in a certain sense, linear regression models are kernel of classical (discrete-time) adaptive control theory, which focuses to cope with the parametric uncertainties in linear plants. In recent years, parametric uncertainties in nonlinear plants have also gained much attention in the literature[MT95, Bos95, Guo97, ASL98, GHZ99, LQF03]. Reviewing the development of adaptive control, we find that parametric uncertainties were of primary interests in the study of adaptive control, no matter whether the considered plants are linear or nonlinear. Nonparametric uncertainties were seldom studied or addressed in the literature of adaptive control until some new areas on understanding limitations and capability of feedback control emerged in recent years. Here we mainly introduce the work initiated by Guo, who also motivated the authors' exploration in the direction which will be discussed in later parts.

Guo's work started from trying to understand fundamental relationship between the uncertainties and the feedback control. Unlike traditional adaptive theory, which focuses on investigating closed-loop stability of certain types of adaptive controllers, Guo began to think over a general set of adaptive controllers, called *feedback mechanism*, i.e., all possible feedback control laws. Here the feedback control laws need not be restricted in a certain class of controllers, and any series of mappings from the space of history data to the space of control signals is regarded as a feedback control law. With this concept in mind, since the most fundamental concept in automatic control, *feedback*, aims to reduce the effects of the plant uncertainty on the desired control performance, by introducing the set *F* of internal uncertainties in the plant and the whole feedback mechanism *U*, we wonder the following basic problems:

1. Given an uncertainty set *F*, does there exist any feedback control law in *U* which can stabilize the plant? This question leads to the problem of how to characterize the maximum capability of feedback mechanism.
2. If the uncertainty set *F* is too large, is it possible that any feedback control law in *U* cannot stabilize the plant? This question leads to the problem of how to characterize the limitations of feedback mechanism.

The philosophical thoughts to these problems result in fruitful study [Guo97, XG00, ZG02, XG01, LX06, Ma08a, Ma08b].

The first step towards this direction was made in [Guo97], where Guo attempted to answer the following question for a nontrivial example of discrete-time nonlinear polynomial plant model with parametric uncertainty: What is the largest nonlinearity that can be dealt with by feedback? More specifically, in [Guo97], for the following nonlinear uncertain system

$$y_{t+1} = \theta\phi_t + u_t + w_{t+1}, \quad \phi_t = O(y_t^b), \qquad b > 0 \tag{1.5}$$

where θ is the unknown parameter, *b* characterizes the nonlinear growth rate of the system, and $\{w_t\}$ is the Gaussian noise sequence, a critical stability result is found — system (1.5) is not a.s. globally stabilizable if and only if $b \geq 4$. This result indicates that there exist limitations of the feedback mechanism in controlling the discrete-time nonlinear adaptive systems, which is not seen in the corresponding continuous-time nonlinear systems (see [Guo97, Kan94]). The "impossibility" result has been extended to some classes of uncertain nonlinear systems with unknown vector parameters in [XG99, Ma08a] and a similar result for system (1.5) with bounded noise is obtained in [LX06].

Stimulated by the pioneering work in [Guo97], a series of efforts ([XG00, ZG02, XG01, MG05]) have been made to explore the maximum capability and limitations of feedback mechanism. Among these work, a breakthrough for non-parametric uncertain systems was made by Xie and Guo in [XG00], where a class of first-order discrete-time dynamical control systems

$$y_{t+1} = f(y_t) + u_t + w_{t+1}, \quad f(\cdot) \in \mathcal{F}(L) \tag{1.6}$$

is studied and another interesting critical stability phenomenon is proved by using new techniques which are totally different from those in [Guo97]. More specifically, in [XG00], *F*(*L*) is a class of nonlinear functions satisfying Lipschitz condition, hence the Lipschitz constant *L* can characterize the size of the uncertainty set *F*(*L*). Xie and Guo obtained the following results: if $L \geq \frac{3}{2} + \sqrt{2}$, then there exists a feedback control law such that for any $f \in F(L)$, the corresponding closed-loop control system is globally stable; and if $L < \frac{3}{2} + \sqrt{2}$, then for any feedback control law and any $y_0 \in R^1$, there always exists some $f \in F(L)$ such that the corresponding closed-loop system is unstable. So for system (1.6), the "magic" number $\frac{3}{2}+\sqrt{2}$ characterizes the capability and limits of the whole feedback mechanism. The impossibility part of the above results has been generalized to similar high-order discrete-time nonlinear systems with single Lipschitz constant [ZG02] and multiple Lipschitz constants [Ma08a]. From the work mentioned above, we can see two different threads: one is focused on parametric nonlinear systems and the other one is focused on non-parametric nonlinear systems. By examining the techniques in these threads, we find that different difficulties exist in the two threads, different controllers are designed to deal with the uncertainties and completely different methods are used to explore the capability and limitations of the feedback mechanism.

1.4 Motivation of Our Work

From the above introduction, we know that only parametric uncertainties were considered in traditional adaptive control and non-parametric uncertainties were only addressed in recent study on the whole feedback mechanism. This motivates us to explore the following problems: When both parametric and non-parametric uncertainties are present in the system, what is the maximum capability of feedback mechanism in dealing with these uncertainties? And how to design feedback control laws to deal with both kinds of internal uncertainties? Obviously, in most practical systems, there exist parametric uncertainties (unknown model parameters) as well as non-parametric uncertainties (e.g. unmodeled dynamics). Hence, it is valuable to explore answers to these fundamental yet novel problems. Noting that parametric uncertainties and non-parametric uncertainties essentially have different nature and require completely different techniques to deal with, generally it is difficult to deal with them in the same loop. Therefore, adaptive estimation and control in systems with parametric and non-parametric uncertainties is a new challenging direction. In this chapter, as a preliminary study, we shall discuss some basic ideas and principles of adaptive estimation in systems with both parametric and non-parametric uncertainties; as to the most difficult adaptive control problem in systems with both parametric and non-parametric uncertainties, we shall discuss two concrete examples involving both kinds of uncertainties, which will illustrate some proposed ideas of adaptive estimation and special techniques to overcome the difficulties in the analysis closed-loop system. Because of significant difficulties in this new direction, it is not possible to give systematic and comprehensive discussions here for this topic, however, our study may shed light on the aforementioned problems, which deserve further investigation.

The remainder of this chapter is organized as follows. In Section 2, a simple semi-parametric model with parametric part and non-parametric part will be introduced first and then we will discuss some basic ideas and principles of adaptive estimation for this model. Later in Section 3 and Section 4, we will apply the proposed ideas of adaptive estimation and investigate two concrete examples of discrete-time adaptive control: in the first example, a discrete-time first-order nonlinear semi-parametric model with bounded external noise disturbance is discussed with an adaptive controller based on information-contraction estimator, and we give rigorous proof of closed-loop stability in case where the uncertain parametric part is of linear growth rate, and our results reveal again the magic number

$\frac{3}{2}+\sqrt{2}$; in the second example, another noise-free semi-parametric model with parametric uncertainties and non-parametric uncertainties is discussed, where a new adaptive controller based on a novel type of update law with deadzone will be adopted to stabilize the system, which provides yet another view point for the adaptive estimation and control problem for the semi-parametric model. Finally, we give some concluding remarks in Section 5.

2. Semi-parametric Adaptive Estimation: Principles and Examples

2.1 One Semi-parametric System Model

Consider the following semi-parametric model

$$z_k = \theta^\tau \phi_k + f(\phi_k) + \varepsilon_k \tag{2.1}$$

where $\theta \in \Theta$ denotes unknown parameter vector, $f(\cdot) \in F$ denotes unknown function and $\varepsilon_k \in \Delta_k$ denote external noise disturbance. Here Θ, F and Δk represent *a priori* knowledge on possible θ, $f(\phi_k)$ and ε_k, respectively. In this model, let

$$v_k = f(\phi_k) + \epsilon_k$$

then Eq. (2.1) becomes Eq. (1.1). Because each term of right hand side of Eq. (2.1) involves uncertainty, it is difficult to estimate θ, $f(\phi_k)$ and ε_k simultaneously.

Adaptive estimation problem can be formulated as follows: Given *a priori* knowledge on θ, $f(\cdot)$ and ε_k, how to estimate θ and $f(\cdot)$ according to a series of data $\{\phi_k, z_k; k = 1,2,\cdots,n\}$ Or in other words, given *a priori* knowledge on θ and v_k, how to estimate θ and v_k according to a series of data $\{\phi_k, z_k; k = 1,2,\cdots,n\}$.

Now we list some examples of *a priori* knowledge to show various forms of adaptive estimation problem.

Example 2.1 *As to the unknown parameter θ, here are some commonly-seen examples of* a priori *knowledge:*

- *There is no any a priori knowledge on θ except for its dimension. This means that θ can be arbitrary and we do not know its upper bound or lower bound.*
- *The upper and lower bounds of θ are known, i.e. $\underline{\theta} \le \theta \le \overline{\theta}$, where $\underline{\theta}$ and $\overline{\theta}$ are constant vector and the relationship "≤" means element-wise "less or equal".*
- *The distance between θ and a nominal θ_0 is bounded by a known constant, i.e. $\|\theta - \theta_0\| \le r_\theta$, where $r_\theta \ge 0$ is a known constant and θ_0 is the center of set Θ.*
- *The unknown parameter lies in a known countable or finite set of values, that is to say, $\theta \in \{\theta_1, \theta_2, \theta_3, \cdots\}$.*

Example 2.2 *As to the unknown function $f(\cdot)$, here are some possible examples of* a priori *knowledge:*

- *$f(x) = 0$ for all x. This case means that there is no unmodeled dynamics.*

• *Function f is bounded by other known functions, that is to say,* $\underline{f}(x) \le f(x) \le \overline{f}(x)$ *for any x.*

• *The distance between f and a nominal* f_0 *is bounded by a known constant, i.e.* $||f - f_0|| \le r_f$, *where* $r_f \ge 0$ *is a known constant and* f_0 *can be regarded as the center of a ball F in a metric functional space with norm* $||\cdot||$.

• *The unknown function lies in a known countable or finite set of functions, that is to say,* $f \in \{f_1, f_2, f_3, \cdots\}$.

• *Function f is Lipschitz, i.e.* $f(x_1) - f(x_2) \le L|x_1 - x_2|$ *for some constant* $L > 0$.

• *Function f is monotone (increasing or decreasing) with respect to its arguments.*

• *Function f is convex (or concave).*

• *Function f is even (or odd).*

Example 2.3 *As to the unknown noise term* ε_k, *here are some possible examples of* a priori *knowledge:*

• *Sequence* $\varepsilon_k = 0$. *This case means that no noise/disturbance exists.*

• *Sequence* ε_k *is bounded in a known range, that is to say,* $\underline{\varepsilon} \le \varepsilon_k \le \overline{\varepsilon}$ *for any k. One special case is* $\underline{\varepsilon} = -\overline{\varepsilon}$.

• *Sequence* ε_k *is bounded by a diminishing sequence, e.g,* $|\varepsilon_k| \le \frac{1}{k}$ *for any* k. *This case means that the noise disturbance converges to zero with a certain rate. Other typical rate sequences include* $\{\frac{1}{k^2}\}$, $\{\delta^k\}$ $(0 < \delta < 1)$, *and so on.*

• *Sequence* ε_k *is bounded by other known sequences, that is to say,* $\underline{\epsilon}_k \le \epsilon_k \le \bar{\epsilon}_k$ *for any k. This case generalizes the above cases.*

• *Sequence* ε_k *is in a known finite set of values, that is to say,* $\varepsilon_k \in \{e_1, e_2, \cdots, e_N\}$. *This case may happen in digital systems where all signals can only take values in a finite set.*

• *Sequence* ε_k *is oscillatory with specific patterns, e.g.* $\varepsilon_k > 0$ *if k is even and* $\varepsilon_k < 0$ *if k is odd.*

• *Sequence* ε_k *has some statistical properties, for example,* $Ee_k = 0$, $Ee_k^2 = \sigma^2$ *;; for another example, sequence {* ε_k *} is i.i.d. taken from a probability distribution e.g.* $\varepsilon_k \approx U(0,1)$.

Parameter estimation problems (without non-parametric part) involving statistical properties of noise disturbance are studied extensively in statistics, system identification and traditional adaptive control. However, we shall remark that other non-statistic descriptions on *a priori* knowledge is more useful in practice yet seldom addressed in existing literature. In fact, in practical problems, usually the probability distribution of the noise/disturbance (if any) is not known and many cases cannot be described by any probability distribution since noise/disturbance in practical systems may come from many different types of sources. Without any *a priori* knowledge in mind, one frequently-used way to handle the noise is to simply assume the noise is Gaussian white noise, which is reasonable in a certain sense. But in practice, from the point of view of engineering, we can usually conclude the noise/disturbance is bounded in a certain range. This chapter will focus on uncertainties with non-statistical *a priori* knowledge. Without loss of generality, in this section we often regard $v_k = f(\phi_k) + \varepsilon_k$ as a whole part, and correspondingly, *a priori* knowledge on v_k, (e.g. $\underline{v}_k \leq v_k \leq \overline{v}_k$), should be provided for the study.

2.2 An Example Problem

Now we take a simple example to show that it may not be appropriate to apply traditional identification algorithms blindly so as to get the estimate of unknown parameter.
Consider the following system

$$z_k = \theta\phi_k + f(\phi_k, k) + \varepsilon_k \tag{2.2}$$

where θ, $f(\cdot)$ and ε_k are unknown parameter, unknown function and unmeasurable noise, respectively. For this model, suppose that we have the following *a priori* knowledge on the system:

- No *a priori* knowledge on θ is known.
- At any step k, the term $f(\phi_k, k)$ is of form $f(\phi_k, k) = \exp(\xi_k \phi_k)$. Here $\{\xi_k\}$ is an unknown sequence satisfying $0 \leq \xi_k \leq 1$.
- Noise ε_k is diminishing with $|\epsilon_k| \leq \frac{1}{k}$.

And in this example, our problem is how to use the data generated from model (2.2) so as to get a good estimate of true value of parameter θ. In our experiment, the data is generated by the following settings (k = 1, 2, $\cdots$, 50):

$$\theta = 5,\ \phi_k = \frac{k}{10},\ f(\phi_k, k) = \exp(|\sin k|\, \phi_k),\ \varepsilon_k = \frac{1}{k}(\alpha_k - 0.5)$$

where $\{\alpha_k\}$ are i.i.d. taken from uniform distribution $U(0, 1)$. Here we have N = 50 groups of data (ϕ_k, z_k).
Since model (2.2) involves various uncertainties, we rewrite it into the following form of linear regression

$$z_k = \theta\phi_k + v_k \tag{2.3}$$

by letting

$$v_k = f(\phi_k, k) + \varepsilon_k .$$

From the *a priori* knowledge for model (2.2), we can obtain the following *a priori* knowledge for the term v_k

$$\underline{v}_k \leq v_k \leq \bar{v}_k$$

where

$$\bar{v}_k = \begin{cases} \exp(\phi_k) + \frac{1}{k} & \text{if } \phi_k \geq 0 \\ 1 + \frac{1}{k} & \text{if } \phi_k < 0 \end{cases}$$
$$\underline{v}_k = \begin{cases} 1 + \frac{1}{k} & \text{if } \phi_k \geq 0 \\ \exp(\phi_k) + \frac{1}{k} & \text{if } \phi_k < 0 \end{cases}$$

Since model (2.3) has the form of linear regression, we can use try traditional identification algorithms to estimate θ. Fig. 1 illustrates the parameter estimates for this problem by using standard LS algorithm, which clearly show that LS algorithm cannot give good parameter estimate in this example because the final parameter estimation error $\tilde{\theta}_k = \hat{\theta} - \theta \approx 5.68284$ is very large.

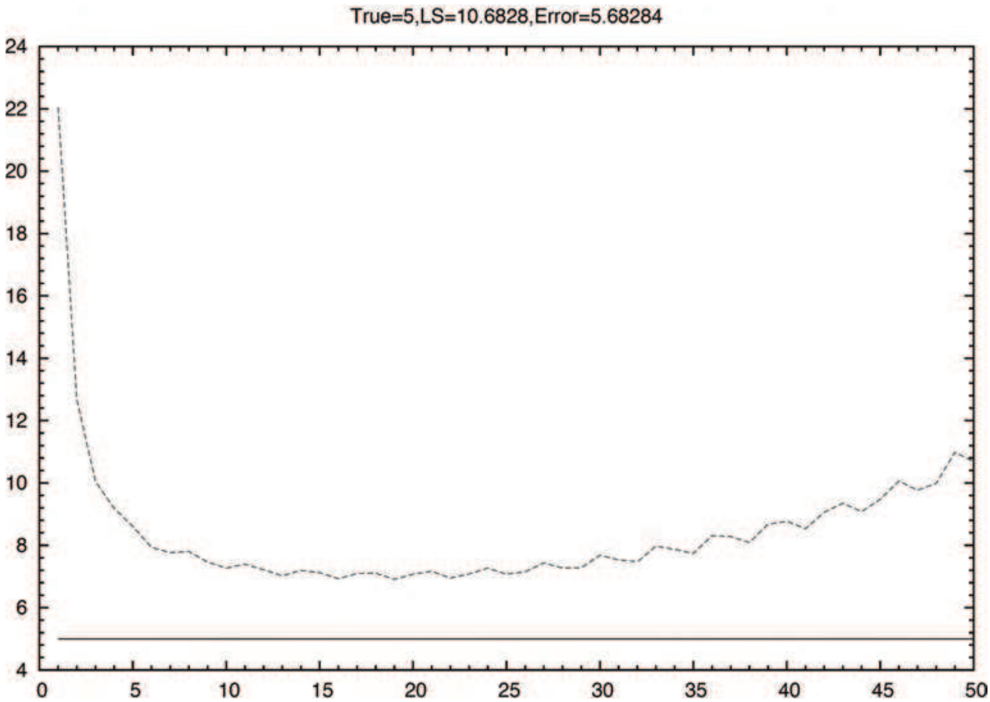

Fig. 1. The dotted line illustrates the parameter estimates obtained by standard least-squares algorithm. The straight line denotes the true parameter.

One may then argue that why LS algorithm fails here is just because the term v_k is in fact biased and we indeed do not utilize the *a priori* knowledge on v_k. Therefore, we may try a modified LS algorithm for this problem: let

$$
\begin{aligned}
c_k &= \frac{1}{2}(\underline{v}_k + \bar{v}_k) \\
d_k &= \frac{1}{2}(\bar{v}_k - \underline{v}_k) \\
w_k &= v_k - c_k \\
y_k &= z_k - c_k
\end{aligned}
$$

then we can conclude that $y_k = \theta^{\tau}\phi_k + w_k$ and $w_k \in [-d_k, d_k]$, where $[-d_k, d_k]$ is a symmetric interval for every k. Then, intuitively, we can apply LS algorithm to data $\{(\phi_k, z_k)$, $k = 1, 2, \cdots, N\}$. The curve of parameter estimates obtained by this modified LS algorithm is plotted in Fig. 2. Since the modified LS algorithm has removed the bias in the *a priori* knowledge, one may expect the modified LS algorithm may give better parameter estimates, which can be verified from Fig. 2 since the final parameter estimation error $\tilde{\theta}_N = \hat{\theta}_N - \theta \approx -1.83314$. In this example, although the modified LS algorithm can work better than the standard LS algorithm, the modified LS algorithm in fact does not help much in solving our problem since the estimation error is still very large comparing with the true value of the unknown parameter.

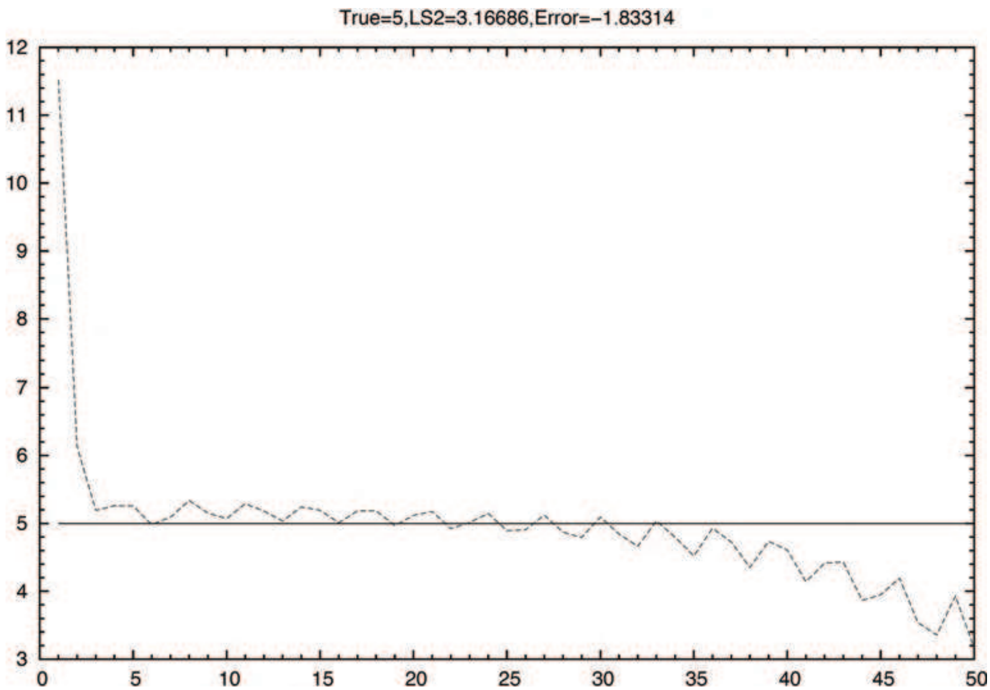

Fig. 2. The dotted line illustrates the parameter estimates obtained by modified least-squares algorithm. The straight line denotes the true parameter.

From this example, we do not aim to conclude that traditional identification algorithms developed in linear regression are not good, however, we want to emphasize the following particular point: *Although traditional identification algorithms (such as LS algorithm) are very powerful and useful in practice, generally it is not wise to apply them blindly when the matching conditions, which guarantee the convergence of those algorithms, cannot be verified or asserted a priori.* This particular point is in fact one main reason why the so-called *minimum-variance self tuning regulator*, developed in the area of adaptive control based on the LS algorithm, attracted several leading scholars to analyze its closed-loop stability throughout past decades from the early stage of adaptive control.

To solve this example and many similar examples with *a priori* knowledge, we will propose new ideas to estimate the parametric uncertainties and the non-parametric uncertainties.

2.3 Information-Concentration Estimator

We have seen that there exist various forms of *a priori* knowledge on system model. With the *a priori* knowledge, how can we estimate the parametric part and the non-parametric part? Now we introduce the so-called information-concentration estimator. The basic idea of this estimator is, the *a priori* knowledge at each time step can be regarded as some constraints of the unknown parameter or function, hence the growing data can provide more and more information (constraints) on the true parameter or function, which enable us to reduce the uncertainties step by step. We explain this general idea by the simple model

$$z_k = \theta^\tau \phi_k + v_k \tag{2.4}$$

with *a priori* knowledge that $\theta \in \Theta \subseteq R^d, \upsilon_k \in V_k$. Then, at k-th step ($k \geq 1$), with the current data k, ϕ_k, z_k we can define the so-called *information set* I_k at step k:

$$I_k \triangleq \{\theta \in \Theta : z_k - \theta^\tau \phi_k \in V_k\}. \tag{2.5}$$

For convenience, let $I_0 = \Theta$. Then we can define the so-called *concentrated information set* C_k at step k as follows

$$C_k = \bigcap_{j=0}^{k} I_k \tag{2.6}$$

which can be recursively written as

$$C_k = C_{k-1} \cap I_k \tag{2.7}$$

with initial set $C_0 = \Theta$. Eq. (2.7) with Eq. (2.5) is called *information-concentration estimator* (short for *IC estimator*) throughout this chapter, and any value in the set C_k can be taken as one possible estimate of unknown parameter θ at time step k. The IC estimator differs from existing parameter identification in the sense that the IC estimator is in fact a setvalued estimator rather than a real-valued estimator. In practical applications, generally C_k is a domain in R^d , and naturally we can take the center point of C_k as $\hat{\theta}_k$.

Remark 2.1 *The definition of* information set *varies with system model. In general cases, it can be extended to the set of possible instances* of θ (and/or f) *which do not contradict with the data at step k. We will see an example involving unknown f in next section.*

From the definition of the IC estimator, the following proposition can be obtained without difficulty:

Proposition 2.1 *Information-concentration estimator has the following properties:*

(i) Monotonicity: $C_0 \supseteq C_1 \supseteq C_2 \supseteq \cdots$

(ii) Convergence: Sequence {C_k} has a limit set $C_\infty = \bigcap_{k=1}^{\infty} C_k$ *;*

(iii) If the system model and the a priori *knowledge are correct, then* C_∞ *must be a non-empty set with property* $\theta \in C_\infty$ *and any element of* C_∞ *can match the data and the model;*

(iv) If $C_\infty = \varnothing$ *, then the data* $\{\phi_k, z_k\}$ *cannot be generated by the system model used by the IC estimator under the specified* a priori *knowledge.*

Proposition 2.1 tells us the following particular points of the IC estimator: property (i) implies that the IC estimator will provide more and more exact estimation; property (ii) means that the there exists a limitation in the accuracy of estimation; property (iii) means that true parameter lies in every C_k if the system model and *a priori* knowledge are correct; and property (iv) means that the IC estimator provides also a method to validate the system model and the *a priori* knowledge. Now we discuss the IC estimator for model (2.4) in more details. In the following discussions, we only consider a typical *a priori* knowledge on $\underline{v}_k \le v_k \le \bar{v}_k$ are two known sequences of vectors (or scalars).

2.3.1 Scalar case: *d* = 1

By Eq. (2.5), we have

$$I_k = \{\theta \in \Theta : \underline{v}_k \le z_k - \theta\phi_k \le \bar{v}_k\}$$

Solving the inequality in I_k, we obtain that

$$\theta\phi_k \in [z_k - \underline{v}_k, z_k - \bar{v}_k]$$

and consequently, if $\phi_k \neq 0$, then we have

$$\theta \in [\underline{b}_k, \bar{b}_k]$$

where

$$\begin{aligned} \underline{b}_k &= \frac{\min(\text{sign}(\phi_k)(z_k - \underline{v}_k), \text{sign}(\phi_k)(z_k - \bar{v}_k))}{|\phi_k|}, \\ \bar{b}_k &= \frac{\max(\text{sign}(\phi_k)(z_k - \underline{v}_k), \text{sign}(\phi_k)(z_k - \bar{v}_k))}{|\phi_k|}, \end{aligned}$$

Here sign(x) denotes the sign of x: sign(x) = 1, 0,−1 for positive number, zero, and negative number, respectively. Then, by Eq. (2.7), we can explicitly obtain that

$$C_k = [\underline{\beta}_k, \bar{\beta}_k]$$

where $\underline{\beta}_k$ and $\bar{\beta}_k$ can be recursively obtained by

$$\begin{aligned} \underline{\beta}_k &= \max(\underline{\beta}_{k-1}, \underline{b}_k) \\ \bar{\beta}_k &= \min(\bar{\beta}_{k-1}, \bar{b}_k) \end{aligned}$$

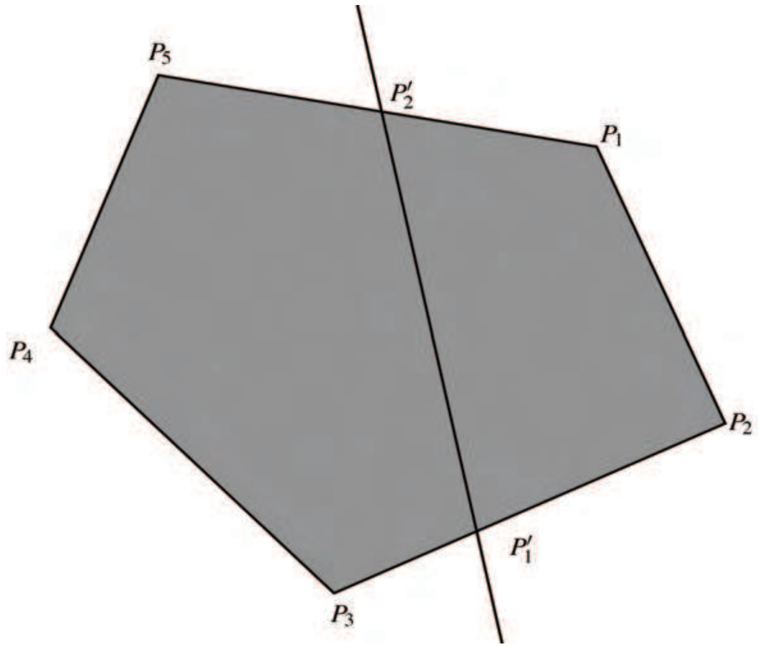

Fig. 3. The straight line may intersect the polygon V and split it into two sub-polygons, one of which will become new polygon V'. The polygon V' can be efficiently calculated from the polygon V.