Part I

Introduction and preliminaries

1

Introduction

Consider a living organism or an artificial mechanism, which we shall refer to for the moment as a system, aiming to perform optimally in an uncertain environment. Despite the fact that the environment may be uncertain, we will suppose that we know the structure of the physical laws of the environment determining plausible motions of the system. Suppose that we even know what the system's action might be and assume that criteria of optimality according to which the system must determine its actions are available. Would we be able to decide a priori which particular action a system must execute or how it should adjust itself in order to maintain its behavior at the optimum?

Depending on the language describing the system's behavior, environment, and uncertainties a number of theoretical frameworks can be employed to find an answer to this non-trivial question. If the available information about the system is limited to a statistical description of the events and their likelihoods are known, then a good methodological candidate is the theory of statistical decision making. On the other hand, if the more sophisticated and involved apparatus of stochastic calculus is used to formalize the behavior of a system in an uncertain environment then a reasonable way to approach the analysis of such an object is to employ the theory of stochastic control and regulation. Despite these differences in how the behavior of a system may be described in various settings, there is a fundamental similarity in the corresponding theoretical frameworks. This similarity, if expressed informally, is that every framework should contain a description of the system's *actions*, *mechanisms for maintaining* and *adjusting* its behavior, and *criteria of optimality* or *goals*. These are in essence components of what we usually understand when calling a system adaptive.

In biology, according to the *Encyclopedia Britannica*, adaptation is described as a "process by which an animal or plant species becomes fitted to its environment; it is the result of natural selection's acting upon heritable variation. Even the simpler organisms must be adapted in a great variety of ways: in their structure, physiology, and genetics, in their locomotion or dispersal, in their means of defense and attack, in their reproduction and development, and in other respects." Actions, regulation and adjustments, criteria of optimality (fitness) are all present in this definition.

In systems theory there is less consensus on what the term "an adaptive system" describes. According to Evleigh (1967) a system is called adaptive if it "is a system which is provided with a means of continuously monitoring its own performance in relation to a given figure of merit or optimal condition and a means of modifying its own parameters by a closed-loop action so as to approach this optimum." Other definitions of an adaptive system have been provided by e.g. L. Zadeh, R. Bellman and R. Kalaba, J. G. Truxall, and V. A. Yakubovich, which we will consider in detail in Chapter 3. Yet they all share the very same ingredients such as actions, adjustments, and criteria of optimality. In this book we will also use the same general understanding of what an adaptive system means, though we will allow some technical deviations from these classical definitions.

Because the phenomenon of adaptation is generally understood as a special regulatory process in which a system maintains its performance at the optimum by adjusting itself and its actions, a natural language to analyze the phenomenon of adaptation is the language of systems and control theories. There are many inspiring and excellent monographs covering the general topic of adaptation. A non-exhaustive list of influential texts includes Tsypkin (1968), Tsypkin (1970), Narendra and Annaswamy (1989), Sastry and Bodson (1989), Krstić *et al.* (1995), and Fradkov *et al.* (1999). Hence it is reasonable to ask whether anything new can be added to this wealth of intellectual resources by one more text. As is often the case in science, novelty is a frequent consequence of a new formulation of a known problem, or it emerges as a result of answering new questions about familiar objects.

The purpose of this monograph is to contribute to the theory of adaptive systems by presenting a list of challenging questions and providing a unified theory that would allow one to find answers to these questions in a rigorous and systematic way. There are numerous examples illustrating the benefits of mathematical analysis of the phenomenon of adaptation: they range from solving the problems of crisis predictions (Gorban *et al.* 2010) to explaining plausible mechanisms of cell functioning in biology (Moreau and Sontag 2003), understanding the evolution of species (Gorban 2007), and motor learning (Smith *et al.* 2006). It is the author's hope that the methods developed here will also be useful for addressing open questions in science.

Below we present several examples of these questions emerging across the disciplines ranging from brain modeling to the issues of precise perturbation compensation in engineering and the problems of signal classification and pattern analysis in artificial intelligence. These examples are split into two major groups related to the problems of observation and regulation. For each of these groups we provide informal statements of the corresponding adaptation problems. These statements are not to be considered final and we will reshape them later on in the text. The function of these statements is to emphasize different facets of the problem of adaptation. There was no specific reason for choosing particular subject areas from which the examples are taken except probably the author's personal interests and bias.

1.1 Observation problems

The problem of state and parameter reconstruction of dynamical systems from the values of just few variables is a common task in the domain of mathematical modeling. Despite the fact that this problem received substantial attention in the past (see e.g. Bastin and Dochain (1990) and Ljung (1999)), there are gray spots in the literature for which finding a computationally plausible and theoretically rigorous solution remains a non-trivial task. The usual sources of difficulties are the presence of nonlinear parametrization, and the fact that we are not allowed to influence the system's behavior by varying its inputs in a reasonably broad class of functions.

There are numerous observation problems of this kind in physics. We start by presenting two examples from the domains of biophysics and neuroscience.

1.1.1 Example: quantitative modeling in biophysics and neuroscience

Let us consider the problem of simultaneous state and parameter reconstruction of models describing the dynamics of neural cells. Most of the available models of individual biological neurons are systems of ordinary differential equations describing the cell's response to stimulation; their parameters characterize variables such as time constants, conductances, and response thresholds, which are important for relating the model responses to the behavior of biological cells. Even the simplest models in this class, such as the Morris–Lecar model (Morris and Lecar 1981), are a great source of inspiration from the modeler's perspective (see Figure 1.1). This model is defined by the following system of equations:

$$\begin{aligned} \dot{V} &= \frac{1}{C}(-\bar{g}_{\mathrm{Ca}} m_\infty(V)(V - E_{\mathrm{Ca}}) - \bar{g}_{\mathrm{K}} w (V - E_{\mathrm{K}}) - \bar{g}_{\mathrm{L}}(V - E_{\mathrm{L}})) + I, \\ \dot{w} &= -\frac{1}{\tau(V)} w + \frac{w_\infty(V)}{\tau(V)}, \end{aligned} \tag{1.1}$$

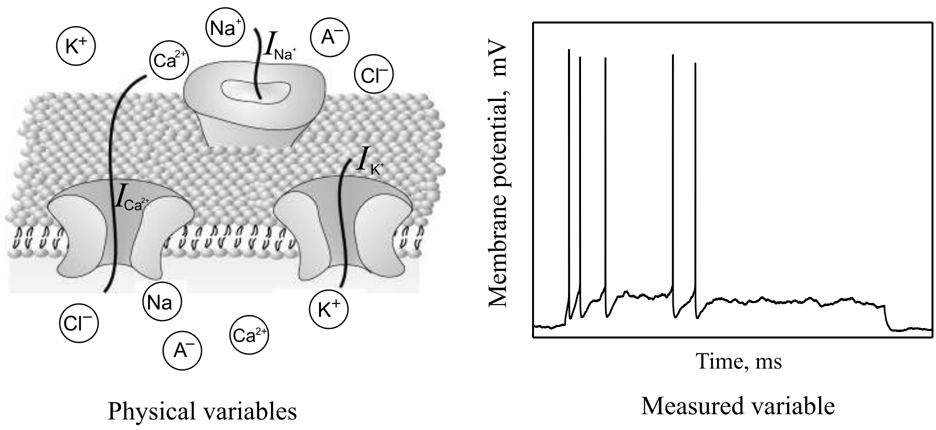

Figure 1.1 Incompleteness of information in quantitative modeling of a cell's behavior. The diagram on the left shows a basic phenomenological description of how currents propagate through a patch of the cell's membrane. There is a number of voltage-dependent channels, such as for Ca, Na, and K depicted in the figure. These channels pump ions through the membrane, and each of these channels has its own dynamics. The problem is that recording currents through every single channel in the membrane simultaneously is not always possible. Thus they must be estimated from available measurements, such as the membrane potentials depicted in the right diagram.

where

$$m_\infty(V) = 0.5\left(1 + \tanh\left(\frac{V - V_1}{V_2}\right)\right),$$

$$w_\infty(V) = 0.5\left(1 + \tanh\left(\frac{V - V_3}{V_4}\right)\right),$$

$$\tau(V) = T_0 \frac{1}{\cosh\left((V - V_3)/2V_4\right)}.$$

The variable V in (1.1) corresponds to the measured membrane potential, and I models an external stimulation current. The parameters $\bar{g}_{\mathrm{Ca}}$, $\bar{g}_{\mathrm{K}}$, and $\bar{g}_{\mathrm{L}}$ stand for the maximal conductances of the calcium, potassium, and leakage currents, respectively; C is the membrane capacitance; V_1, V_2, V_3, and V_4 are the parameters of the gating variables; T_0 is the parameter regulating the time scale of ionic currents; E_{Ca} and E_{K} are the Nernst potentials of the calcium and potassium currents; and E_{L} is the rest potential.

The total number of parameters in system (1.1) is 12, excluding the stimulation current I. Some of these parameters can be considered typical. For example the values of the Nernst potentials for calcium and potassium channels, E_{Ca} and E_{K}, are known and usually are set as $E_{\mathrm{Ca}} = 100\,\mathrm{mV}$ and $E_{\mathrm{K}} = -70\,\mathrm{mV}$ (Koch 2002).

The value of the rest potential, E_{L}, can be measured explicitly. The values of the parameters, $\bar{g}_{\mathrm{Ca}}$, $\bar{g}_{\mathrm{K}}$, $\bar{g}_{\mathrm{L}}$, and T_0, however, may vary substantially from one cell to another, and in general they are dependent on the conditions of the experiment. For example, the values of $\bar{g}_{\mathrm{Ca}}$, $\bar{g}_{\mathrm{K}}$, and $\bar{g}_{\mathrm{L}}$ depend on the density of ion channels in a patch of the membrane; and the value of T_0 is dependent on temperature. Hence, to be able to model the dynamics of individual cells, we have to recover these values from data.

Another example of the same nature is a model predicting the force generated by rat skeletal muscles during brief isometric contractions (Wexler *et al.* 1997). The model consists of three coupled nonlinear differential equations,

$$\begin{aligned}
\dot{F} &= aT\left(1 - \frac{F}{F_{\mathrm{m}}}\right) - \frac{F}{\tau_1 + \tau_2 T/T_0},\\
\dot{T} &= k_1 T_0 C^2 - (k_1 C^2 + k_2)T,\\
\dot{C} &= 2(k_1 C^2 + k_2)T - 2k_1 T_0 C^2 + kC_0 - (k + k_0)C,
\end{aligned} \tag{1.2}$$

where F is the force generated by the muscles, T is the concentration of Ca^{2+}–troponin complex, and C is the concentration of Ca^{2+} in the sarcoplasmic reticulum. The parameters τ_1, C_0, and k are fixed, while the parameters k_0, k_1, k_2, τ_2, F_{m}, a, and T_0 are free. The values of T and C are not available for direct observation, and the values of F over time can be measured. The question is whether it is possible to reconstruct the free parameters of the model together with the values of the concentrations T and C from the measurements of F. As in the previous example, we are dealing with an uncertain system in which the unknown parameters enter the right-hand side of the corresponding differential equations nonlinearly.

1.1.2 Example: adaptive classification in neural networks

The problem of estimating parameters of ordinary differential equations is not limited to the domain of modeling. It has an important relative in the field of artificial intelligence, namely the problem of adaptive classification of signals. An example of this problem is provided below.

Consider a set of signals defined as

$$\begin{aligned}
&\mathcal{F} = \{f_i(\xi(t), \theta_i)\},\ i \in \{1, \ldots, N_f\},\\
&f_i : \mathbb{R} \times \mathbb{R} \to \mathbb{R},\ f_i(\cdot, \cdot) \in \mathcal{C}^0,\\
&\xi : \mathbb{R}_{\geq 0} \to \mathbb{R},\ \xi(\cdot) \in \mathcal{C}^1 \cap L_\infty[0, \infty],
\end{aligned} \tag{1.3}$$

where $\theta_i \in \Omega_\theta \subset \mathbb{R}$ are parameters of which the values are unknown a priori, $\Omega_\theta = [\theta_{\min}, \theta_{\max}]$ is a bounded interval, and $\xi(t)$ is a known and bounded function.

Signals $f_i(\xi(t), \theta_i)$ constitute the set of variables chosen to represent the state of an object.

Let $s \in \mathcal{F}$ be an element of class $\mathcal{F}$. The values of $s(t, \theta)$ are fed into the following system of differential equations:

$$\begin{aligned}
\dot{x}_j &= \sum_{m=1}^{N} c_{j,m}\sigma(\mathbf{w}_{j,m}^{\mathrm{T}}\mathbf{x} + w_{s,j,m}s(t) + w_{\xi,j,m}\xi + b_{j,m}), \\
j &\in \{1, \ldots, N_x\}, \\
\mathbf{x} &= \mathrm{col}(x_1, \ldots, x_{N_x}), \ \mathbf{x}(t_0) = \mathbf{x}_0. \qquad (1.4)
\end{aligned}$$

System (1.4) is often referred to as the recurrent neural network with standard multilayer perceptron structure. Here $c_{j,m}$, $\mathbf{w}_{j,m}$, $w_{s,j,m}$, $w_{\xi,j.m}$, and $b_{j,m}$ are parameters of which the values are fixed, and the function $\sigma : \mathbb{R} \to \mathbb{R}$ is sigmoidal:

$$\sigma(p) = \frac{1}{1 + e^{-p}}.$$

The problem of classification can now be stated as follows: is there a network of type (1.4) that is able to recover uncertain parameters i and θ_i from the input $s(t)$ (see Figure 1.2)? Informally, this means that there exist two sets of functions of the network state $\mathbf{x}$ and input $s(t)$:

$$\begin{aligned}
&\{h_{f,j}(\mathbf{x}(t), s(t))\}, \ \{h_{\theta,j}(\mathbf{x}(t), s(t))\}, \\
&h_{f,j} : \mathbb{R}^{N_x} \times \mathbb{R} \to \mathbb{R}, \ h_{\theta,j} : \mathbb{R}^{N_x} \times \mathbb{R} \to \mathbb{R}, \ j \in \{1, \ldots, N_f\},
\end{aligned}$$

such that the values of i and θ_i can be inferred from $\{h_{f,j}(\mathbf{x}(t), s(t))\}$ and $\{h_{\theta,j}(\mathbf{x}(t), s(t))\}$, respectively, within a given finite interval of time.

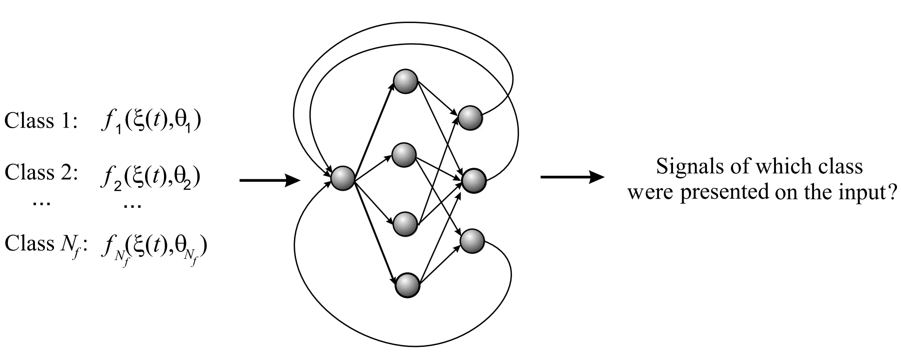

Figure 1.2 Adaptive classification of temporal signals in recurrent neural networks with fixed weights.

Networks (1.4) form an important class of computational structures of which the practical utility and capabilities are widely acknowledged in the literature (Haykin 1999). This class has been shown to be successful in dealing with a wide range of classification problems, including that of classifying signals from (1.3), provided that the values of θ_i in (1.3) are known. Empirical studies suggest that recurrent neural networks of this type are able to solve the classification problem (Feldkamp and Puskorius 1997; Prokhorov *et al.* 2002a) even if θ_i are unknown. The question, however, is how to show that this is indeed the case.

The problem of adaptive classification may look different from the previous examples in the domain of modeling. Indeed, here we have an existence question, whereas in the examples before we asked for a specific estimation algorithm. Despite these differences, there is substantial similarity in these problems. To be able to see this similarity, we would like to state the observation problem in a more general context below.

1.1.3 Preliminary statement of the problem

Let us generalize model (1.1) to the following class of dynamical systems:

$$\begin{aligned} &\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) + \mathbf{g}(\mathbf{x}, \boldsymbol{\theta}) u(t), \quad \mathbf{x}(t_0) \in \Omega_x \subset \mathbb{R}^n, \\ &y = h(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n, \quad \boldsymbol{\theta} \in \Omega_\theta, \ \Omega_\theta \subset \mathbb{R}^d, \quad y \in \mathbb{R}, \end{aligned} \tag{1.5}$$

where $\mathbf{f}, \mathbf{g} : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}$, and $u : \mathbb{R} \to \mathbb{R}$ are continuous and differentiable functions. The variable x stands for the state vector, $u \in \mathcal{U} \subset \mathcal{C}^1[t_0, \infty)$ is the known input, $\boldsymbol{\theta}$ is the vector of unknown parameters, and y is the output of (1.5). Given that the right-hand side of (1.5) is differentiable, for any $\mathbf{x}' \in \Omega_x$, $u \in \mathcal{C}^1[t_0, \infty)$ there exists a time interval $\mathcal{T} = [t_0, t_1]$, $t_1 > t_0$ such that a solution $\mathbf{x}(t, \mathbf{x}')$ of (1.5) passing through $\mathbf{x}'$ at t_0 exists for all $t \in \mathcal{T}$. Hence, $y(t) = h(\mathbf{x}(t))$ is defined for all $t \in \mathcal{T}$. For the sake of convenience we will assume that the interval $\mathcal{T}$ of the solutions is large enough or even coincides with $[t_0, \infty)$ when necessary.

Taking these notations into account, we can now state the observation problem as follows: suppose that we are able to measure the values of $y(t)$ precisely; can the values of $\mathbf{x}'$ and the parameter vector $\boldsymbol{\theta}$ be recovered from the observations of $y(t)$, and, if so, how? In particular, we are interested in finding a computational algorithm

$$\dot{\boldsymbol{\xi}} = \mathbf{p}(\boldsymbol{\xi}, t, u(t), y(t)), \qquad \boldsymbol{\xi}_0 = \boldsymbol{\xi}(t_0) \in \Omega_\xi, \tag{1.6}$$

such that for some known functions $\mathbf{h}_x(\boldsymbol{\xi})$ and $\mathbf{h}_\theta(\boldsymbol{\xi})$ and given number $\varepsilon > 0$ the following property holds:

$$\begin{aligned} \limsup_{t\to\infty} \|\mathbf{h}_x(\boldsymbol{\xi}(t,\boldsymbol{\xi}_0)) - \mathbf{x}(t)\| \leq \varepsilon, \\ \limsup_{t\to\infty} \|\mathbf{h}_\theta(\boldsymbol{\xi}(t,\boldsymbol{\xi}_0)) - \boldsymbol{\theta}\| \leq \varepsilon \ \forall\, \boldsymbol{\xi}_0 \in \Omega_\xi. \end{aligned} \tag{1.7}$$

In order to see how this statement is related to the adaptive classification problem in neural networks it is sufficient to notice that (1) the right-hand side of (1.4) can approximate an arbitrary continuous function in a bounded domain (hence it can model the right-hand side of (1.6)), and (2) the function s in (1.4) may be modeled as an output of system (1.5).

System (1.5) can be viewed as an external object or environment, and computational algorithm (1.6) and the functions $\mathbf{h}_x(\boldsymbol{\xi})$ and $\mathbf{h}_\theta(\boldsymbol{\xi})$ constitute the adapting system. The system responds to changes in the environment so that its performance (defined here by (1.7)) reaches an acceptable level and is maintained at this level indefinitely. If (1.5) were linearly parametrized, i.e. the functions $\mathbf{f}(\mathbf{x},\boldsymbol{\theta})$ and $\mathbf{g}(\mathbf{x},\boldsymbol{\theta})$ were linear in $\boldsymbol{\theta}$, then in order to answer this question we could employ the well-developed machinery of standard adaptive observers design (Marino and Tomei 1995b). Yet, as model (1.1) illustrates, the assumption of linear parametrization does not always hold. Hence alternative methods are needed.

This question (as well as other related issues of parameter estimation of nonlinear ordinary equations) is discussed in detail in Chapter 5. In addition to presenting sufficient conditions stipulating the mere existence of solutions to the observation problem, we provide specific computational algorithms (1.6) satisfying the required asymptotic properties (1.7). Special attention is paid to the analysis of the convergence rates of these algorithms. One may expect that the rates of convergence are likely to depend on the classes of nonlinearities in the models. This is indeed the case, as we illustrate in Chapter 5.

1.2 Regulation problems

Suppose now that we are not interested in reconstructing the values of the state and parameters of system (1.5). We do, however, require that the system's state is regulated to a given set in the system's state space for all $\boldsymbol{\theta} \in \Omega_\theta$. Consider for example the following system:

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= -x_1 - x_2 + g(x_1, x_2, \boldsymbol{\theta}) + u, \end{aligned} \tag{1.8}$$

where x_1 and x_2 are the state variables, $\boldsymbol{\theta} \in \Omega_\theta$, $\Omega_\theta \subset \mathbb{R}^d$ is the vector of unknown parameters, $g : \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is a continuous function, and $u : \mathbb{R} \to \mathbb{R}$ is an input. Equations (1.8) describe a large class of mechanical and chemical systems. If we accept a simplified interpretation in which x_1 is the position of an object in space and x_2 is its velocity then $g(x_1, x_2, \boldsymbol{\theta})$ could stand for the friction terms (Canudas de Wit and Tsiotras 1999). If (1.8) is a model of a bio-reactor then x_1 and x_2 are the substrate concentrations and $g(x_1, x_2, \boldsymbol{\theta})$ could stand for the standard Michaelis–Menten nonlinearity (Bastin and Dochain 1990). In all these cases the function $g(x_1, x_2, \boldsymbol{\theta})$ is nonlinear in $\boldsymbol{\theta}$. The question is whether there is a function $u(x_1, x_2, \hat{\boldsymbol{\theta}})$ such that the solutions of (1.8) converge to the origin for all $\boldsymbol{\theta} \in \Omega_\theta$.

1.2.1 Example: non-dominating adaptive regulation

If no additional constraints are imposed then the above problem can be easily solved within the framework of dominating functions (Lin and Qian 2002b; Putov 1993). In this framework the original nonlinearly parametrized uncertainty $g(x_1, x_2, \boldsymbol{\theta})$ is replaced by a dominating linearly parametrized one $|g(x_1, x_2, \boldsymbol{\theta})| \leq \bar{g}(x_1, x_2)^{\mathrm{T}} \boldsymbol{\eta}$ and the problem is then solved using the standard method of Lyapunov functions (see Lin and Qian (2002b) for details). Although practical, this approach is not necessarily optimal for systems with limited resources. If the system is a living organism then using resources excessively may be an important limiting factor. The same argument applies for artificial yet autonomous systems. For these classes of systems a reasonable assumption is that the system is penalized for excessive use of domination terms in control.

One of the simplest examples of such non-dominating control schemes is the compensatory control $u = -g(x_1, x_2, \boldsymbol{\theta})$. If the value of $\boldsymbol{\theta}$ were known then this feedback would be able to steer the system to the origin. The problem, however, is that the values of $\boldsymbol{\theta}$ are unknown and the function $g(x_1, x_2, \boldsymbol{\theta})$ is nonlinearly parametrized. A possible strategy would be to make an initial guess at $\boldsymbol{\theta}$ and then adjust its value over time. The question, however, is how should one do this? This is a typical example of the non-dominating adaptation problem, of which a more formal statement is provided at the end of this section.

1.2.2 Example: adaptive tuning to bifurcations

In the previous case the set to which the system solutions are to converge was a priori known. There are systems for which information of this kind is not explicitly available. Their goal is not to reach a given state in the system's state space but rather to maintain adaptively a certain functional property of the system. An interesting example is the problem of adaptive self-tuning of a hearing nerve cell (Moreau and