

1

Introduction

We consider a physical system where variables and parameters interact in a domain of interest. Variables are physical quantities that are observable or measurable, and their values change with position and time to form signals. We may express system structures and properties as parameters, which may also change with position and time. For a given system, we understand its dynamics based on underlying physical principles describing the interactions among the variables and parameters. We adopt mathematical tools to express the interactions in a manageable way.

A physical excitation to the system is an input and its response is an output. The response is always accompanied by some form of energy transfer. The input can be applied to the system externally or internally. For the internal input, we may also use the term “source”. Observations or measurements can be done at the outside, on the boundary and also on the inside of the system. For the simplicity of descriptions, we will consider boundary measurements as external measurements. Using the concept of the generalized system, we will introduce the forward and inverse problems of a physical system.

1.1 Forward Problem

The generalized system H in Figure 1.1 has a system parameter p , input x and output y . We first need to understand how they are entangled in the system by understanding underlying physical principles. A mathematical representation of the system dynamics is the forward problem formulation. We formulate a forward problem of the system in Figure 1.1 as

$$y = H(p, x), \quad (1.1)$$

where H is a nonlinear or linear function of p and x . We should note that the expression in (1.1) may not be feasible in some cases where the relation between the input and output can only be described implicitly.

To treat the problem in a computationally manageable way, we should choose core variables and parameters of most useful and meaningful information to quantify their interrelations. The expression in (1.1), therefore, could be an approximation of complicated

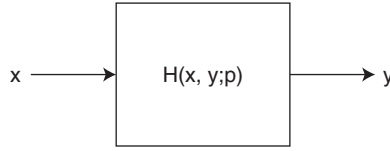


Figure 1.1 Forward problem for a system with parameter p , input x and output y

interactions among variables and parameters. In practice, we may not be able to control the input precisely because of technical limitations, and the measured output will always be contaminated by noise. Solving a forward problem is to find the output from a given input and system parameter. Evaluation of (1.1) suffices for its solution.

A simple example is a sound recording and reproduction system including a microphone, amplifier and speaker. An input sound wave enters the system through the microphone, goes through the amplifier and exits the system as an output sound wave through the speaker. The system characteristics are determined by the electrical and mechanical properties of the system components, including the gain or amplitude amplification factor, phase change, frequency bandwidth, power and so on.

Example 1.1.1 *The output signal $y(t)$ at time t of a system H is found to be $y(t) = Kx(t - \tau)$, where $x(t)$ is the input signal, K is a fixed gain and τ is a fixed time delay. Taking the Fourier transform of both sides, we can find the frequency transfer function $H(i\omega)$ of the system to be*

$$H(i\omega) = \frac{\mathcal{F}\{y(t)\}}{\mathcal{F}\{x(t)\}} = \frac{K e^{-i\omega\tau} X(i\omega)}{X(i\omega)} = K e^{-i\omega\tau}, \quad (1.2)$$

where $i = \sqrt{-1}$ and ω is the angular frequency. This means that, for all ω ,

$$|H(i\omega)| = K \quad (\text{flat magnitude response}), \quad (1.3)$$

$$\theta(i\omega) = \angle H(i\omega) = -\tau\omega \quad (\text{linear phase response}). \quad (1.4)$$

When (1.3) and (1.4) are satisfied within a frequency range of the input signal, the output is a time-delayed amplified version of the input signal without any distortion. For a sinusoidal input signal

$$x(t) = A \cos(\omega t + \theta),$$

the corresponding output signal is

$$y(t) = K A \cos(\omega t + \theta - \tau\omega) = K A \cos\{\omega(t - \tau) + \theta\}.$$

When the input signal is a sum of many sinusoids with different frequencies, that is,

$$x(t) = \sum_{j=1}^n A_j \cos(\omega_j t + \theta_j),$$

the corresponding output signal is

$$y(t) = K \sum_{j=1}^n A_j \cos(\omega_j t + \theta_j - \tau \omega_j) = K \sum_{j=1}^n A_j \cos\{\omega_j(t - \tau) + \theta_j\}.$$

This system with parameters K and τ is a linear system with no distortion. Given the forward problem expressed as $y(t) = Kx(t - \tau)$ with known values of K and τ , we can find the output y for any given input x .

In most physical systems, inputs are mixed within the system to produce outputs. The mixing process is accompanied by smearing of information embedded in the inputs. Distinct features of the inputs may disappear in the outputs, and the effects of the system parameters may spread out in the observed outputs.

Exercise 1.1.2 For a continuous-time linear time-invariant system with impulse response $h(t)$, where t is time, find the expression for the output $y(t)$ corresponding to the input $x(t)$.

Exercise 1.1.3 For a discrete-time linear time-invariant system with impulse response $h[n]$, where n is time, find the expression for the output $y[n]$ corresponding to the input $x[n]$.

1.2 Inverse Problem

For a given forward problem, we may consider two types of related inverse problems as in Figure 1.2. The first type is to find the input from a measured output and identified system parameter. The second is to find the system parameter from a designed input and measured output. We symbolically express these two cases as follows:

$$x = H_1^+(p, y) \quad (1.5)$$

and

$$p = H_2^+(x, y), \quad (1.6)$$

where H_1^+ and H_2^+ are nonlinear or linear functions. We may need to design multiple inputs carefully to get multiple input–output pairs with enough information to solve the inverse problems.

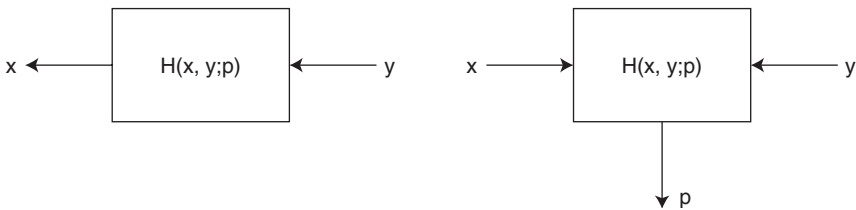


Figure 1.2 Two different inverse problems for a system with parameter p , input x and output y

Example 1.2.1 We consider the linear system in Example 1.1.1 and assume that we have measured the output $y(t)$ subject to the sinusoidal input $x(t) = A \cos(\omega t + \theta)$ with unknown A and θ . We can find the amplitude and phase of the sinusoidal input, A and θ , respectively, by performing the following phase-sensitive demodulation process. For the in-phase channel,

$$Y_I = \frac{1}{T} \int_{t_0}^{t_0+T} y(t) \cos \omega t \, dt = \frac{KA}{2} \cos(\theta - \tau\omega), \quad (1.7)$$

where $T = 2\pi/\omega$ is the period of the sinusoid and t_0 is an arbitrary time. For the quadrature channel,

$$Y_Q = \frac{1}{T} \int_{t_0}^{t_0+T} y(t) \sin \omega t \, dt = \frac{KA}{2} \sin(\theta - \tau\omega). \quad (1.8)$$

We recover A and θ as

$$A = \frac{\sqrt{4Y_I^2 + 4Y_Q^2}}{K} \quad \text{and} \quad \theta = \tan^{-1} \frac{Y_Q}{Y_I} + \tau\omega, \quad (1.9)$$

assuming that we know the system parameters K and τ .

Exercise 1.2.2 Assume that we have measured the output $y(t)$ of the linear system in Example 1.1.1 for the known sinusoidal input $x(t)$. Find the system parameters: the gain K and the delay τ .

Exercise 1.2.3 Consider a discrete-time linear time-invariant system with impulse response $h[n]$, where n is time. We have measured its output $y[n]$ subject to the known input $x[n]$. Discuss how to find $h[n]$.

In general, most inverse problems are complicated, since the dynamics among inputs, outputs and system parameters are attributed to complex, possibly nonlinear, physical phenomena. Within a given measurement condition, multiple inputs may result in the same output for given system parameters. Similarly, different system parameters may produce the same input–output relation. The inversion process, therefore, suffers from the uncertainty that originates from the mixing process of the corresponding forward problem.

To seek a solution of an inverse problem, we first need to understand how those factors are entangled in the system by understanding the underlying physical principles. Extracting core variables of most useful information, we should properly formulate a forward problem to quantify their interrelations. This is the reason why we should investigate the associated forward problem before trying to solve the inverse problem.

1.3 Issues in Inverse Problem Solving

In solving an inverse problem, we should consider several factors. First, we have to make sure that there exists at least one solution. This is the issue of the existence of a solution, which must be checked in the formulation of the inverse problem. In practice, it may

not be a serious question, since the existence is obvious as long as the system deals with physically existing or observable quantities. Second is the uniqueness of a solution. This is a more serious issue in both theoretical and practical aspects, and finding a unique solution of an inverse problem requires careful analyses of the corresponding forward and inverse problems. If a solution is not unique, we must check its optimality in terms of its physical meaning and practical usefulness.

To formulate a manageable problem dealing with key information, we often go through a simplification process and sacrifice some physical details. Mathematical formulations of the forward and inverse problems, therefore, suffer from modeling errors. In practice, measured data always include noise and artifacts. To acquire a quantitative numerical solution of the inverse problem, we deal with discretized versions of the forward and inverse problems. The discretization process may add noise and artifacts. We must carefully investigate the effects of these practical restrictions in the context of the existence and uniqueness of a solution.

We introduce the concept of well-posedness as proposed by Hadamard (1902). When we construct a mathematical model of a system to transform the associated physical phenomena into a collection of mathematical expressions and data, we should consider the following three properties.

1. Existence: at least one solution exists.
2. Uniqueness: only one solution exists.
3. Continuity: a solution depends continuously on the data.

In the sense of Hadamard, a problem is well-posed when it meets the above requirements of existence, uniqueness and continuity. If these requirements are not met, the problem is ill-posed.

If we can properly formulate the forward problem of a physical system and also its inverse problem, we can safely assume that a solution exists. Non-uniqueness often becomes a practically important issue, since it is closely related with the inherent mixing process of the forward problem. Once the inputs are mixed, uniquely sorting out some inputs and system parameters may not be feasible. The mixing process may also cause sensitivity problems. When the sensitivity of a certain output to the inputs and/or system parameters is low, small changes in the inputs or system parameters may result in small and possibly discontinuous changes in the output, with measurement errors. The inversion process in general includes a step where the measured output values are divided by sensitivity factors. If we divide small measured values, including errors, by a small sensitivity factor, we may amplify the errors in the results. The effects of the amplified errors may easily dominate the inversion process and result in useless solutions, which do not comply with the continuity requirement.

Considering that mixing processes are embedded in most forward problems and that the related inverse problems are ill-posed in many cases, we need to devise effective methods to deal with such difficulties. One may incorporate as much *a priori* information as possible in the inversion process. Preprocessing methods such as denoising and feature extraction can be employed. One may also need to implement some regularization techniques to find a compromise between the robustness of an inversion method and the accuracy or sharpness of its solution.

1.4 Linear, Nonlinear and Linearized Problems

Linearity is one of the most desirable features in solving an inverse problem. We should carefully check whether the forward and inverse problems are linear or not. If not, we may try to approximately linearize the problems in some cases. We first define the linearity of the forward problem in Figure 1.1 as follows.

1. Homogeneity: if $y_1 = H(x)$, then $y_2 = H(Kx) = KH(x) = Ky_1$ for any constant K .
2. Additivity: if $y_1 = H(x_1)$ and $y_2 = H(x_2)$, then $y_3 = H(x_1 + x_2) = H(x_1) + H(x_2) = y_1 + y_2$.

For a linear system, we can, therefore, apply the following principle of superposition: if $y_1 = H(x_1)$ and $y_2 = H(x_2)$, then

$$y_3 = H(K_1x_1 + K_2x_2) = K_1y_1 + K_2y_2 \quad (1.10)$$

for any constants K_1 and K_2 .

For the inverse problems in Figure 1.2, we may similarly define the linearity for two functions H_1^+ and H_2^+ . Note that we should separately check the linearity of the three functions H , H_1^+ and H_2^+ . Any problem, either forward or inverse, is nonlinear if it does not satisfy both of the homogeneity and additivity requirements.

Example 1.4.1 *Examples of nonlinear systems are*

1. $y = Kx^2$,
2. $y = K_1e^{K_2x}$.

For a nonlinear problem $y = H(x)$, we may fix $x = x_0$ and consider a small change Δx around x_0 . As illustrated in Figure 1.3, we can approximate the corresponding change Δy as

$$\Delta y = H(x_0 + \Delta x) - H(x_0) \approx \partial_x H(x)|_{x_0} \Delta x = S_{x_0} \Delta x, \quad (1.11)$$

where S_{x_0} is the sensitivity of Δy to Δx at $x = x_0$, which can be found from the analysis of the problem. The approximation in (1.11) is called the linearization to find $y_1 = H(x_0 + \Delta x) \approx y_0 + \Delta y$ where $y_0 = H(x_0)$. The approximation is accurate only for a small Δx .

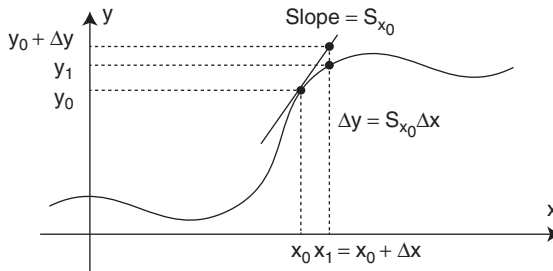


Figure 1.3 Illustration of a linearization process

Problems either forward or inverse can be linear or nonlinear depending on the underlying physical principles. Proper formulations of forward and inverse problems using mathematical tools are essential steps before any attempt to seek solution methods. In the early chapters of the book, we study mathematical backgrounds to deal with linear and nonlinear problems. In the later chapters, we will introduce several imaging modalities.

Reference

Hadamard J 1902 Sur les problèmes aux dérivées partielles et leur signification physique. *Bull. Univ. Princeton*, **13**, 49–52.

