NONLINEAR INVERSE PROBLEMS IN IMAGING
NONLINEAR INVERSE PROBLEMS IN IMAGING

Jin Keun Seo  
*Yonsei University, Korea*

Eung Je Woo  
*Kyung Hee University, Korea*
Contents

Preface xi

List of Abbreviations xiii

1 Introduction 1
1.1 Forward Problem 1
1.2 Inverse Problem 3
1.3 Issues in Inverse Problem Solving 4
1.4 Linear, Nonlinear and Linearized Problems 6
References 7

2 Signal and System as Vectors 9
2.1 Vector Spaces 9
2.1.1 Vector Space and Subspace 9
2.1.2 Basis, Norm and Inner Product 11
2.1.3 Hilbert Space 13
2.2 Vector Calculus 16
2.2.1 Gradient 16
2.2.2 Divergence 17
2.2.3 Curl 17
2.2.4 Curve 18
2.2.5 Curvature 19
2.3 Taylor’s Expansion 21
2.4 Linear System of Equations 23
2.4.1 Linear System and Transform 23
2.4.2 Vector Space of Matrix 24
2.4.3 Least-Squares Solution 27
2.4.4 Singular Value Decomposition (SVD) 28
2.4.5 Pseudo-inverse 29
2.5 Fourier Transform 30
2.5.1 Series Expansion 30
2.5.2 Fourier Transform 32
2.5.3 Discrete Fourier Transform (DFT) 37
2.5.4 Fast Fourier Transform (FFT) 40
5 Numerical Methods 129
5.1 Iterative Method for Nonlinear Problem 129
5.2 Numerical Computation of One-Dimensional Heat Equation 130
  5.2.1 Explicit Scheme 132
  5.2.2 Implicit Scheme 135
  5.2.3 Crank–Nicolson Method 136
5.3 Numerical Solution of Linear System of Equations 136
  5.3.1 Direct Method using LU Factorization 136
  5.3.2 Iterative Method using Matrix Splitting 138
  5.3.3 Iterative Method using Steepest Descent Minimization 140
  5.3.4 Conjugate Gradient (CG) Method 143
5.4 Finite Difference Method (FDM) 145
  5.4.1 Poisson Equation 145
  5.4.2 Elliptic Equation 146
5.5 Finite Element Method (FEM) 147
  5.5.1 One-Dimensional Model 147
  5.5.2 Two-Dimensional Model 149
  5.5.3 Numerical Examples 154
References 157
Further Reading 158

6 CT, MRI and Image Processing Problems 159
6.1 X-ray Computed Tomography 159
  6.1.1 Inverse Problem 160
  6.1.2 Basic Principle and Nonlinear Effects 160
  6.1.3 Inverse Radon Transform 163
  6.1.4 Artifacts in CT 166
6.2 Magnetic Resonance Imaging 167
  6.2.1 Basic Principle 167
  6.2.2 k-Space Data 168
  6.2.3 Image Reconstruction 169
6.3 Image Restoration 171
  6.3.1 Role of $p$ in (6.35) 173
  6.3.2 Total Variation Restoration 175
  6.3.3 Anisotropic Edge-Preserving Diffusion 180
  6.3.4 Sparse Sensing 181
6.4 Segmentation 184
  6.4.1 Active Contour Method 185
  6.4.2 Level Set Method 187
  6.4.3 Motion Tracking for Echocardiography 189
References 192
Further Reading 194

7 Electrical Impedance Tomography 195
7.1 Introduction 195
7.2 Measurement Method and Data 196
## Contents

7.2.1 Conductivity and Resistance 196  
7.2.2 Permittivity and Capacitance 197  
7.2.3 Phasor and Impedance 198  
7.2.4 Admittivity and Trans-Impedance 199  
7.2.5 Electrode Contact Impedance 200  
7.2.6 EIT System 201  
7.2.7 Data Collection Protocol and Data Set 202  
7.2.8 Linearity between Current and Voltage 204

7.3 Representation of Physical Phenomena 205  
7.3.1 Derivation of Elliptic PDE 205  
7.3.2 Elliptic PDE for Four-Electrode Method 206  
7.3.3 Elliptic PDE for Two-Electrode Method 209  
7.3.4 Min–Max Property of Complex Potential 210

7.4 Forward Problem and Model 210  
7.4.1 Continuous Neumann-to-Dirichlet Data 211  
7.4.2 Discrete Neumann-to-Dirichlet Data 212  
7.4.3 Nonlinearity between Admittivity and Voltage 214

7.5 Uniqueness Theory and Direct Reconstruction Method 216  
7.5.1 Calderón’s Approach 216  
7.5.2 Uniqueness and Three-Dimensional Reconstruction: Infinite Measurements 218  
7.5.3 Nachmann’s D-bar Method in Two Dimensions 221

7.6 Back-Projection Algorithm 223

7.7 Sensitivity and Sensitivity Matrix 226  
7.7.1 Perturbation and Sensitivity 226  
7.7.2 Sensitivity Matrix 227  
7.7.3 Linearization 227  
7.7.4 Quality of Sensitivity Matrix 229

7.8 Inverse Problem of EIT 229  
7.8.1 Inverse Problem of RC Circuit 229  
7.8.2 Formulation of EIT Inverse Problem 231  
7.8.3 Ill-Posedness of EIT Inverse Problem 231

7.9 Static Imaging 232  
7.9.1 Iterative Data Fitting Method 232  
7.9.2 Static Imaging using Four-Channel EIT System 233  
7.9.3 Regularization 237  
7.9.4 Technical Difficulty of Static Imaging 237

7.10 Time-Difference Imaging 239  
7.10.1 Data Sets for Time-Difference Imaging 239  
7.10.2 Equivalent Homogeneous Admittivity 240  
7.10.3 Linear Time-Difference Algorithm using Sensitivity Matrix 241  
7.10.4 Interpretation of Time-Difference Image 242

7.11 Frequency-Difference Imaging 243  
7.11.1 Data Sets for Frequency-Difference Imaging 243  
7.11.2 Simple Difference $F_{t,ω2} - F_{t,ω1}$ 244  
7.11.3 Weighted Difference $F_{t,ω2} - αF_{t,ω1}$ 244
7.11.4 Linear Frequency-Difference Algorithm using Sensitivity Matrix
7.11.5 Interpretation of Frequency-Difference Image

References 247

8 Anomaly Estimation and Layer Potential Techniques 251
8.1 Harmonic Analysis and Potential Theory 252
  8.1.1 Layer Potentials and Boundary Value Problems for Laplace Equation 252
  8.1.2 Regularity for Solution of Elliptic Equation along Boundary of Inhomogeneity 259
8.2 Anomaly Estimation using EIT 266
  8.2.1 Size Estimation Method 268
  8.2.2 Location Search Method 274
8.3 Anomaly Estimation using Planar Probe 281
  8.3.1 Mathematical Formulation 282
  8.3.2 Representation Formula 287
References 290
Further Reading 291

9 Magnetic Resonance Electrical Impedance Tomography 295
9.1 Data Collection using MRI 296
  9.1.1 Measurement of \( B_z \) 297
  9.1.2 Noise in Measured \( B_z \) Data 299
  9.1.3 Measurement of \( \mathbf{B} = (B_x, B_y, B_z) \) 301
9.2 Forward Problem and Model Construction 301
  9.2.1 Relation between \( \mathbf{J}, B_z \) and \( \sigma \) 302
  9.2.2 Three Key Observations 303
  9.2.3 Data \( B_z \) Traces \( \sigma \nabla u \times \mathbf{e}_z \) Directional Change of \( \sigma \) 304
  9.2.4 Mathematical Analysis toward MREIT Model 305
9.3 Inverse Problem Formulation using \( \mathbf{B} \) or \( \mathbf{J} \) 308
  9.4 Inverse Problem Formulation using \( B_z \) 309
  9.4.1 Model with Two Linearly Independent Currents 309
  9.4.2 Uniqueness 310
  9.4.3 Defected \( B_z \) Data in a Local Region 314
9.5 Image Reconstruction Algorithm 315
  9.5.1 \( J \)-substitution Algorithm 315
  9.5.2 Harmonic \( B_z \) Algorithm 317
  9.5.3 Gradient \( B_z \) Decomposition and Variational \( B_z \) Algorithm 319
  9.5.4 Local Harmonic \( B_z \) Algorithm 320
  9.5.5 Sensitivity Matrix-based Algorithm 322
  9.5.6 Anisotropic Conductivity Reconstruction Algorithm 323
  9.5.7 Other Algorithms 324
9.6 Validation and Interpretation 325
  9.6.1 Image Reconstruction Procedure using Harmonic \( B_z \) Algorithm 325
  9.6.2 Conductivity Phantom Imaging 326
Preface

Imaging techniques in science, engineering and medicine have evolved to expand our ability to visualize the internal information in an object such as the human body. Examples may include X-ray computed tomography (CT), magnetic resonance imaging (MRI), ultrasound imaging and positron emission tomography (PET). They provide cross-sectional images of the human body, which are solutions of corresponding inverse problems. Information embedded in such an image depends on the underlying physical principle, which is described in its forward problem. Since each imaging modality has limited viewing capability, there have been numerous research efforts to develop new techniques producing additional contrast information not available from existing methods.

There are such imaging techniques of practical significance, which can be formulated as nonlinear inverse problems. Electrical impedance tomography (EIT), magnetic induction tomography (MIT), diffuse optical tomography (DOT), magnetic resonance electrical impedance tomography (MREIT), magnetic resonance electrical property tomography (MREPT), magnetic resonance elastography (MRE), electrical source imaging and others have been developed and adopted in application areas where new contrast information is in demand. Unlike X-ray CT, MRI and PET, they manifest some nonlinearity, which result in their image reconstruction processes being represented by nonlinear inverse problems.

Visualizing new contrast information on the electrical, optical and mechanical properties of materials inside an object will widen the applications of imaging methods in medicine, biotechnology, non-destructive testing, geophysical exploration, monitoring of industrial processes and other areas. Some are advantageous in terms of non-invasiveness, portability, convenience of use, high temporal resolution, choice of dimensional scale and total cost. Others may offer a higher spatial resolution, sacrificing some of these merits.

Owing primarily to nonlinearity and low sensitivity, in addition to the lack of sufficient information to solve an inverse problem in general, these nonlinear inverse problems share the technical difficulties of ill-posedness, which may result in images with a low spatial resolution. Deep understanding of the underlying physical phenomena as well as the implementation details of image reconstruction algorithms are prerequisites for finding solutions with practical significance and value.

Research outcomes during the past three decades have accumulated enough knowledge and experience that we can deal with these topics in graduate programs of applied mathematics and engineering. This book covers nonlinear inverse problems associated with some of these imaging modalities. It focuses on methods rather than applications.
The methods mainly comprise mathematical and numerical tools to solve the problems. Instrumentation will be treated only in enough detail to describe practical limitations imposed by measurement methods.

Readers will acquire the diverse knowledge and skills needed to deal effectively with nonlinear inverse problems in imaging by following the steps below:

1. Understand the underlying physical phenomena and the constraints imposed on the problem, which may enable solutions of nonlinear inverse problems to be improved. Physics, chemistry and also biology play crucial roles here. No attempt is made to be comprehensive in terms of physics, chemistry and biology.
2. Understand forward problems, which usually are the processes of information loss. They provide strategic insights into seeking solutions of nonlinear inverse problems. The underlying principles are described here so that readers can understand their mathematical formulations.
3. Formulate forward problems in such a way that they can be dealt with systematically and quantitatively.
4. Understand how to probe the imaging object and what is measurable using available engineering techniques. Practical limitations associated with the measurement sensitivity and specificity, such as noise, artifacts, interface between target object and instrument, data acquisition time and so on, must be properly understood and analyzed.
5. Understand what is feasible in a specific nonlinear inverse problem.
6. Formulate proper nonlinear inverse problems by defining the image contrast associated with physical quantities. Mathematical formulations should include any interrelation between those qualities and measurable data.
7. Construct inversion methods to produce images of contrast information.
8. Develop computer programs and properly address critical issues of numerical analysis.
9. Customize the inversion process by including a priori information.
10. Validate results by simulations and experiments.

This book is for advanced graduate courses in applied mathematics and engineering. Prerequisites for students with a mathematical background are vector calculus, linear algebra, partial differential equations and numerical analysis. For students with an engineering background, we recommend taking linear algebra, numerical analysis, electromagnetism, signal and system and also preferably instrumentation.

Lecture notes, sample codes, experimental data and other teaching material are available at http://mimaging.yonsei.ac.kr/NIPI.
List of Abbreviations

General Notation

$\mathbb{R}$ : the set of real numbers
$\mathbb{R}^n$ : $n$-dimensional Euclidean space
$\mathbb{C}$ : the set of complex numbers
$i = \sqrt{-1}$
$\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ : the set of non-negative integers
$\mathbb{N} = \{1, 2, \ldots\}$ : the set of positive integers
$r = (x, y, z)$ : position
$B_r(a)$ : the ball with radius $r$ at the center $a$ and $B_r = B_r(0)$
$e_j$ : $j$th unit vector in $\mathbb{R}^n$, for example, $e_2 = (0, 1, 0, \ldots, 0)$
$\forall$ : for all
$\exists$ : there exist(s)
$\because$ : because
$\therefore$ : therefore

Electromagnetism

$E$ : electric field intensity
$B$ : magnetic flux density
$J$ : current density
$D$ : electric flux density
$\sigma$ : electrical conductivity
$\varepsilon$ : electrical permittivity
$\gamma = \sigma + i\omega\varepsilon$ : admittivity
$u$ : voltage (electrical potential)
Notations for Domains and Vector Spaces

$\Omega$ : a domain in $\mathbb{R}^n$
$\partial \Omega$ : the boundary of the domain $\Omega$
$n$ : the unit outward normal vector to the boundary
$C(\Omega)$ : the set of all continuous functions in $\Omega$
$C^k(\Omega)$ : the set of continuously $k$th differentiable functions defined in the domain $\Omega$
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

\[
y = H(p, x),
\]

(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...
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 $y(t) = K x(t - \tau)$, where $x(t)$ is the input signal, $K$ is a fixed gain and $\tau$ 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

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

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

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

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

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

$$y(t) = KA \cos(\omega(t - \tau) + \theta) = KA \cos(\omega t - \tau \omega).$$

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 \( \tau \) is a linear system with no distortion. Given the forward problem expressed as \( y(t) = K x(t - \tau) \) with known values of \( K \) and \( \tau \), 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:

\[ x = H_1^+(p, y) \]  \hspace{1cm} (1.5)

and

\[ p = H_2^+(x, y), \]  \hspace{1cm} (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](image-url) 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 $\theta$. We can find the amplitude and phase of the sinusoidal input, $A$ and $\theta$, respectively, by performing the following phase-sensitive demodulation process. For the in-phase channel,

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

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

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

We recover $A$ and $\theta$ as

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

assuming that we know the system parameters $K$ and $\tau$.

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 $\tau$.

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

\[
y_3 = H(K_1 x_1 + K_2 x_2) = K_1 y_1 + K_2 y_2 \quad (1.10)
\]

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 \( H_1^+, H_2^+, H_3^+ \). 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. \( y = K x^2 \),
2. \( y = K_1 e^{K_2 x} \).

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

![Figure 1.3 Illustration of a linearization process](image-url)
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

Signal and System as Vectors

To solve forward and inverse problems, we need mathematical tools to formulate them. Since it is convenient to use vectors to represent multiple system parameters, variables, inputs (or excitations) and outputs (or measurements), we first study vector spaces. Then, we introduce vector calculus to express interrelations among them based on the underlying physical principles. To solve a nonlinear inverse problem, we often linearize an associated forward problem to utilize the numerous mathematical tools of the linear system. After introducing such an approximation method, we review mathematical techniques to deal with a linear system of equations and linear transformations.

2.1 Vector Spaces

We denote a signal, variable or system parameter as \( f(\mathbf{r}, t) \), which is a function of position \( \mathbf{r} = (x, y, z) \) and time \( t \). To deal with a number \( n \) of signals, we adopt the vector notation \((f_1(\mathbf{r}, t), \ldots, f_n(\mathbf{r}, t))\). We may also set vectors as \((f(\mathbf{r}_1, t), \ldots, f(\mathbf{r}_n, t))\), \((f(\mathbf{r}, t_1), \ldots, f(\mathbf{r}, t_n))\) and so on. We consider a set of all possible such vectors as a subset of a vector space. In the vector space framework, we can add and subtract vectors and multiply vectors by numbers. Establishing the concept of a subspace, we can project a vector into a subspace to extract core information or to eliminate unnecessary information. To analyze a vector, we may decompose it as a linear combination of basic elements, which we handle as a basis or coordinate of a subspace.

2.1.1 Vector Space and Subspace

Definition 2.1.1 A non-empty set \( V \) is a vector space over a field \( F = \mathbb{R} \) or \( \mathbb{C} \) if there are operations of vector addition and scalar multiplication with the following properties.

Vector addition

1. \( \mathbf{u} + \mathbf{v} \in V \) for every \( \mathbf{u}, \mathbf{v} \in V \) (closure).
2. \( \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u} \) for every \( \mathbf{u}, \mathbf{v} \in V \) (commutative law).
3. \((u + v) + w = u + (v + w)\) for every \(u, v, w \in V\) (associative law).
4. There exist \(0 \in V\) such that \(u + 0 = u\) for every \(u \in V\) (additive identity).
5. For all \(u \in V\) there exists \(-u \in V\) such that \(u + (-u) = 0\) and \(-u\) is unique (additive inverse).

Scalar multiplication
1. For \(a \in \mathbb{F}\) and \(u \in V\), \(au \in V\) (closure).
2. For \(a, b \in \mathbb{F}\) and \(u, v \in V\), \(a(bu) = (ab)u\) (associative law).
3. For \(a \in \mathbb{F}\) and \(u, v \in V\), \(a(u + v) = au + av\) (first distributive law).
4. For \(a, b \in \mathbb{F}\) and \(u \in V\), \((a + b)u = au + bu\) (second distributive law).
5. \(1u = u\) for every \(u \in V\) (multiplicative identity).

A subset \(W\) of a vector space \(V\) over \(F\) is a subspace of \(V\) if and only if \(au + v \in W\) for all \(a \in \mathbb{F}\) and for all \(u, v \in W\). The subspace \(W\) itself is a vector space.

**Example 2.1.2** The following are examples of vector spaces.

- \(\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_1, \ldots, x_n \in \mathbb{R}\}\), \(n\)-dimensional Euclidean space.
- \(\mathbb{C}^n = \{(x_1, x_2, \ldots, x_n) : x_1, \ldots, x_n \in \mathbb{C}\}\).
- \(C([a,b])\), the set of all complex-valued functions that are continuous on the interval \([a,b]\).
- \(C^1([a,b]) := \{f \in C([a,b]) : f' \in C[a,b]\}\), the set of all functions in \(C([a,b])\) with continuous derivative on the interval \([a,b]\).
- \(L^2((a,b)) := \{f : (a,b) \to \mathbb{R} : \int_a^b |f(x)|^2 \, dx < \infty\}\), the set of all square-integrable functions on the open interval \((a,b)\).
- \(H^1((a,b)) := \{f \in L^2(a,b) : \int_a^b |f'(x)|^2 \, dx < \infty\}\).

**Definition 2.1.3** Let \(G = \{u_1, \ldots, u_n\}\) be a subset of a vector space \(V\) over a field \(\mathbb{F} = \mathbb{R}\) or \(\mathbb{C}\). The set of all linear combinations of elements of \(G\) is denoted by \(\text{span}\, G\):

\[
\text{span}\, G := \left\{ \sum_{j=1}^n a_j u_j : a_j \in \mathbb{F} \right\}.
\]

The \(\text{span}\, G\) is the smallest subspace of \(V\) containing \(G\). For example, if \(V = \mathbb{R}^3\), then \(\text{span}\{(1, 2, 3), (1, -2, 0)\}\) is the plane \(\{a(1, 2, 3) + b(1, -2, 0) : a, b \in \mathbb{R}\}\).

**Definition 2.1.4** The elements \(u_1, \ldots, u_n\) of a vector space \(V\) are said to be linearly independent if

\[
\sum_{j=1}^n a_j u_j = 0 \quad \text{holds only for } a_1 = a_2 = \cdots = a_n = 0.
\]

Otherwise, \(u_1, \ldots, u_n\) are linearly dependent.

If \((u_j)_{j=1}^n\) is linearly independent, no vector \(u_j\) can be expressed as a linear combination of other vectors in the set. If \(u_1\) can be expressed as \(u_1 = a_2 u_2 + \cdots + a_n u_n\), then \(a_1 u_1 + a_2 u_2 + a_n u_n = 0\) with \(a_1 = -1\), so they are not linearly independent. For
example, \{(4, 1, 5), (2, 1, 3), (1, 0, 1)\} is linearly dependent since 

\[-(4, 1, 5) + (2, 1, 3) + 2(1, 0, 1) = (0, 0, 0)\].

The elements \((2, 1, 3)\) and \((1, 0, 1)\) are linearly independent because 

\[a_1(2, 1, 3) + a_2(1, 0, 1) = (0, 0, 0)\] implies \(a_1 = a_2 = 0\).

Example 2.1.5 In Figure 2.1, the two vectors \{\(u, v\)\} are linearly independent whereas the four vectors \{\(u, v, p, q\)\} are linearly dependent. Note that \(p\) and \(q\) are linearly independent.

### 2.1.2 Basis, Norm and Inner Product

#### Definition 2.1.6

Let \(W\) be a subspace of a vector space \(V\). If \(\text{span}\{u_1, \ldots, u_n\} = W\) and \(\{u_1, \ldots, u_n\}\) is linearly independent, then \(\{u_1, \ldots, u_n\}\) is said to be a basis for \(W\).

If \(\{u_j\}_{j=1}^n\) is a basis for \(W\), then any vector \(v \in W\) can be expressed uniquely as \(v = \sum_{j=1}^n a_j u_j \in W\). If \(G'\) is another basis of \(W\), then \(G'\) contains exactly the same number \(n\) of elements.

#### Definition 2.1.7

Let \(W\) be a subspace of a vector space \(V\). Then \(W\) is \(n\)-dimensional if the number of elements of the basis of \(W\) is \(n\); \(W\) is finite-dimensional if \(\text{dim} W < \infty\); otherwise \(W\) is infinite-dimensional.

To quantify a measure of similarity or dissimilarity among vectors, we need to define the magnitude of a vector and the distance between vectors. We use the norm \(\|u\|\) of a vector \(u\) to define such a magnitude. In the area of topology, the metric is also used for defining a distance. To distinguish different vectors in a vector space, we define a measure of distance or metric between two vectors \(u\) and \(v\) as the norm \(\|u - v\|\). The norm must satisfy the following three rules.

#### Definition 2.1.8

A normed vector space \(V\) is a vector space equipped with a norm \(\| \cdot \|\) that satisfies the following:

1. \(0 \leq \|u\| < \infty, \forall u \in V\) and \(\|u\| = 0\) iff \(u = 0\).
2. \(\|au\| = |a| \|u\|, \forall u \in V\) and \(\forall a \in \mathbb{F}\).
3. \(\|u + v\| \leq \|u\| + \|v\|, \forall u, v \in V\) (triangle inequality).

Here, the notation \(\forall\) stands for “for all” and iff stands for “if and only if”.
Example 2.1.9 Consider the vector space $\mathbb{C}^n$. For $\mathbf{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{C}^n$ and $1 \leq p < \infty$, the $p$-norm of $\mathbf{u}$ is

$$
\|\mathbf{u}\|_p = \begin{cases}
    \left( \sum_{j=1}^{n} |u_j|^p \right)^{1/p} & \text{for } 1 \leq p < \infty, \\
    \max_{1 \leq j \leq n} |u_j| & \text{for } p = \infty.
\end{cases}
$$

(2.1)

In particular, $\|\mathbf{u} - \mathbf{v}\|_2 = \sqrt{(u_1 - v_1)^2 + \cdots + (u_n - v_n)^2}$ is the standard distance between $\mathbf{u}$ and $\mathbf{v}$. We should note that, when $0 < p < 1$, $\|\mathbf{u}\|_p$ is not a norm because it does not satisfy the triangle inequality.

Example 2.1.10 Consider the vector space $V = C([0, 1])$. For $f, g \in V$, the distance between $f$ and $g$ can be defined by

$$
\|f - g\| = \sqrt{\int_0^1 |f(x) - g(x)|^2 \, dx}.
$$

In addition to the distance between vectors, it is desirable to establish the concept of an angle between them. This requires the definition of an inner product.

Definition 2.1.11 Let $V$ be a vector space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. We denote the complex conjugate of $a \in \mathbb{C}$ by $\bar{a}$. A vector space $V$ with a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ is an inner product space if:

1. $0 \leq \langle \mathbf{u}, \mathbf{u} \rangle < \infty$, $\forall \mathbf{u} \in V$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if $\mathbf{u} = \mathbf{0}$;
2. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$, $\forall \mathbf{u}, \mathbf{v} \in V$;
3. $\langle a\mathbf{u} + b\mathbf{v}, \mathbf{w} \rangle = a\langle \mathbf{u}, \mathbf{w} \rangle + b\langle \mathbf{v}, \mathbf{w} \rangle$, $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $\forall a, b \in \mathbb{F}$.

In general, $\langle \mathbf{u}, \mathbf{v} \rangle$ is a complex number, but $\langle \mathbf{u}, \mathbf{u} \rangle$ is real. Note that $\langle \mathbf{w}, a\mathbf{u} + b\mathbf{v} \rangle = \bar{a}\langle \mathbf{w}, \mathbf{u} \rangle + \bar{b}\langle \mathbf{w}, \mathbf{v} \rangle$ and $\langle \mathbf{u}, \mathbf{0} \rangle = 0$. If $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in V$, then $\mathbf{u} = \mathbf{0}$. Given any inner product, $\sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = \|\mathbf{u}\|$ is a norm on $V$.

For a real inner product space $V$, the inner product provides angle information between two vectors $\mathbf{u}$ and $\mathbf{v}$. We denote the angle $\theta$ between $\mathbf{u}$ and $\mathbf{v}$ as

$$
\theta = \angle(\mathbf{u}, \mathbf{v}) = \cos^{-1} \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|} \quad \text{and} \quad \langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta.
$$

We interpret the angle as follows.

1. If $\theta = 0$, then $\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\| \|\mathbf{v}\|$ and $\mathbf{v} = a\mathbf{u}$ for some $a > 0$. The two vectors $\mathbf{u}$ and $\mathbf{v}$ are in the same direction.
2. If $\theta = \pi$, then $\langle \mathbf{u}, \mathbf{v} \rangle = -\|\mathbf{u}\| \|\mathbf{v}\|$ and $\mathbf{v} = a\mathbf{u}$ for some $a < 0$. The two vectors $\mathbf{u}$ and $\mathbf{v}$ are in opposite directions.
3. If $\theta = \pm \pi/2$, then $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. The two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

Definition 2.1.12 The set $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ in an inner product space $V$ is said to be an orthonormal set if $\langle \mathbf{u}_j, \mathbf{u}_k \rangle = 0$ for $j \neq k$ and $\|\mathbf{u}_j\| = 1$. A basis is an orthonormal basis if it is an orthonormal set.
2.1.3 Hilbert Space

When we analyze a vector \( f \) in a vector space \( V \) having a basis \( \{ u_j \}_{j=1}^{\infty} \), we wish to represent \( f \) as

\[
 f = \sum_{j=1}^{\infty} a_j u_j.
\]

Computation of the coefficients \( a_j \) could be very laborious when the vector space \( V \) is not equipped with an inner product and \( \{ u_j \}_{j=1}^{\infty} \) is not an orthonomal set. A Hilbert space is a closed vector space equipped with an inner product.

**Definition 2.1.13** A vector space \( H \) over \( \mathbb{F} = \mathbb{R} \) or \( \mathbb{C} \) is a Hilbert space if:

1. \( H \) is an inner product space;
2. \( H = \overline{H} \) (\( H \) is a closed vector space), that is, whenever \( \lim_{n \to \infty} \| u_n - u \| = 0 \) for some sequence \( \{ u_n \} \subset H \), \( u \) belongs to \( H \).

For \( u, v, w \in H \), a Hilbert space, we have the following properties.

- Cauchy–Schwarz inequality: \( |\langle u, v \rangle| \leq \| u \| \| v \| \).
- Triangle inequality: \( \| u + v \| \leq \| u \| + \| v \| \).
- Parallelogram law: \( \| u + v \|^2 + \| u - v \|^2 = 2\| u \|^2 + 2\| v \|^2 \).
- Polarization identity: \( 4\langle u, v \rangle = \| u + v \|^2 - \| u - v \|^2 + i\| u + iv \|^2 - i\| u - iv \|^2 \).
- Pythagorean theorem: \( \| u + v \|^2 = \| u \|^2 + \| v \|^2 \) if \( \langle u, v \rangle = 0 \).

**Exercise 2.1.14 (Gram–Schmidt process)** Let \( H \) be a Hilbert space with a basis \( \{ v_1, v_2, \ldots \} \). Assume that \( \{ u_1, u_2, \ldots \} \) is obtained from the following procedure depicted in Figure 2.2:

1. Set \( w_1 = v_1 \) and \( u_1 = w_1/\| w_1 \| \);
2. Set \( w_2 = v_2 - \langle v_2, u_1 \rangle u_1 \) and \( u_2 = w_2/\| w_2 \| \);
3. For \( n = 2, 3, \ldots \),

\[
 w_{n+1} = v_{n+1} - \sum_{j=1}^{n} (v_{j+1} u_j) u_j \quad \text{and} \quad u_{n+1} = \frac{w_{n+1}}{\| w_{n+1} \|}.
\]
Prove that \( \{u_1, u_2, \ldots \} \) is an orthonormal basis of \( H \), that is,

\[
\text{span}\{u_1, \ldots, u_n\} = \text{span}\{v_1, \ldots, v_n\}.
\]

**Theorem 2.1.15 (Projection theorem)** Let \( G \) be a closed convex subset of a Hilbert space \( H \). For every \( u \in H \), there exists a unique \( u_\ast \in G \) such that \( \|u - u_\ast\| \leq \|u - v\| \) for all \( v \in G \).

For the proof of the above theorem, see Rudin (1970). Let \( S \) be a closed subspace of a Hilbert space \( H \). We define the orthogonal complement \( S^\perp \) of \( S \) as

\[
S^\perp := \{v \in H \mid \langle u, v \rangle = 0 \text{ for all } u \in S\}.
\]

According to the projection theorem, we can define a projection map \( P_S : H \rightarrow S \) such that the value \( P_S(u) \) satisfies

\[
\|u - P_S(u)\| \leq \|u - (P_S(u) + tv)\| \quad \text{for all } v \in S \text{ and } t \in \mathbb{R}.
\]

This means that \( f(t) = \|u - P_S(u) + tv\|^2 \) has its minimum at \( t = 0 \) for any \( v \in S \) and, therefore,

\[
0 = f'(0) = \langle u - P_S(u), v \rangle \quad \text{for all } v \in S
\]

or

\[
u - P_S(u) \in S^\perp.
\]

Hence, the projection theorem states that every \( u \in H \) can be uniquely decomposed as \( u = v + w \) with \( v \in S \) and \( w \in S^\perp \) and we can express the Hilbert space \( H \) as

\[
H = S \oplus S^\perp.
\]

From the Pythagorean theorem,

\[
\|u\|^2 = \|P_S(u)\|^2 + \|u - P_S(u)\|^2.
\]

Figure 2.3 illustrate the projection of a vector \( u \) onto a subspace \( S \).

![Figure 2.3](image-url)