Kalman-like Inversion with ODE/SDE Formulations and Adaptive Algorithms

Yuchen Yang, student ID: 4264270

January 29, 2020

Contents

Chapter 1

Introduction

This thesis focuses on inverse problems. Similar concepts are known as, e.g., data assimilation in historical matching and numerical prediction, parameter estimation (optimization) in estimation theory (optimum control theory), and model training in machine learning and deep learning. The class of inverse problems that we studied are aimed to infer unknown parameters as inputs of forward models when the outputs are given/observed.

This chapter introduces inverse problems and discusses how they arise in applications. Then, we present an abbreviation of objectives, contributions, and layouts of this thesis.

1.1 Overview of inverse problems

In this section, inverse problems will be discussed as follows. Firstly, we will intuitively describe what inverse problems are. Secondly, we will list real applications of inverse problems. Thirdly, we will consider some infinite-dimensional inverse problems governed by three basic types of second order (hyperbolic, parabolic, and elliptic) partial differential equations (PDEs).

1.1.1 Descriptions of inverse problems

Inverse problem is a general concept, that means, for a system, to calculate causal factors (inputs of the system) from given observations (outputs of the system). This process is the 'inverse' of forward problem which starts with inputs and calculates outputs.

There exist some simple examples that forward problems and inverse problems are equivalent: consider a finite-dimensional full-rank square matrix $\mathbf{A}$, so that, $\mathbf{y} = \mathbf{A}\mathbf{x}$ if and only if $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$; consider the exponential and logarithm functions on the real line, so that, $y = \exp(x)$ if and only if $x = \log(y)$. However, these kinds of trivial examples are less interesting. In most real applications, the direction of 'forward' and 'inverse' cannot be inverted, because of two facts: 1) physically, there exist consequences from causes to effects e.g. the second law of thermodynamics, and 2) mathematically, inverse problems are usually ill-posed even though forward models are well-posed.

The concept of well-posedness was proposed by Jacques Hadamard [47]. A problem is well-posed if it has a unique solution continuously depending on conditions. A problem is ill-posed if it is not well-posed. In order to explain the idea more specifically, consider a continuous map $\mathcal{G} : \mathcal{X} \to \mathcal{Y}$ from a separable Banach space $\mathcal{X}$ to another separable Banach space $\mathcal{Y}$. Then, the forward problem is to compute $y = \mathcal{G}(x)$ given $x \in \mathcal{X}$. Clearly, the forward problem is well-posed, as there exists a unique y given x, and y is continuous in x. On the other hand, the inverse problem is to find $x \in \mathcal{X}$, such that, the equality $y = \mathcal{G}(x)$ holds given $y \in \mathrm{Ran}(\mathcal{G})$. However, the inverse problem could be ill-posed because of two facts: 1) there may be multiple solutions of x, and 2) the solution x may not be continuously depending on data y. In the following, we use two examples to show the ill-posedness of inverse problems.

Example 1.1.1 (under-determined system of linear equations)**.** *This example shows that an under-determined linear system has multiple solutions. Consider a system of linear equations* $\mathbf{y} = \mathbf{A}\mathbf{x}$*, where* $\mathbf{A}$ *is an* $n \times m$ *real-valued matrix with* $n < m$ *(under-determined). Assume* $\mathbf{A}$ *is full-rank. Then, given any* $\mathbf{y} \in \mathbb{R}^n$*, the solution* $\mathbf{x} \in \mathbb{R}^m$ *exists but not unique. Thus, the inverse problem is ill-posed.*

Example 1.1.2 (Fredholm integral equation)**.** *This example shows that the solution of Fredholm integral equation [60] is not continuous. Let* $K \in L^2([a,b] \times [a,b]; \mathbb{R})$ *be a symmetric positive-semi-definite kernel. Given a data* $g \in L^2([a,b]; \mathbb{R})$*, we aim to find the solution* $f \in L^2([a,b]; \mathbb{R})$*, such that, the Fredholm integral equation holds, for all* $s \in [a,b]$*,*

$$g(s) = \int_a^b K(s,t) f(t) \, \mathrm{d}t \tag{1.1.1}$$

More clearly, let λ_i be the eigenvalues (sorted from the largest value to the lowest value) and let φ_i be the corresponding (orthonormal) eigenfunctions, such that, the following equation holds for all $s \in [a, b]$,

$$\lambda_i \varphi_i(s) = \int_a^b K(s,t)\varphi_i(t)\,\mathrm{d}t \tag{1.1.2}$$

Thus, $\{\varphi_i : i \in \mathbb{N}_1\}$ form an orthonormal basis in the separable Hilbert space $L^2([a,b];\mathbb{R})$. Furthermore, for any $i = 1, 2, ...,$ we have

$$\int_a^b g(s)\varphi_i(s)\,\mathrm{d}s = \int_a^b \left(\int_a^b K(s,t) f(t)\,\mathrm{d}t \right) \varphi_i(s)\,\mathrm{d}s \tag{1.1.3}$$

$$= \int_a^b \left(\int_a^b K(s,t)\varphi_i(s)\,\mathrm{d}s \right) f(t)\,\mathrm{d}t \tag{1.1.4}$$

$$= \lambda_i \int_a^b \varphi_i(t) f(t)\,\mathrm{d}t \tag{1.1.5}$$

We will show that the solution f is not continuous depending on g. Let $g_1, g_2 \in L^2([a,b];\mathbb{R})$ be two different data, and let $f_1, f_2 \in L^2([a,b];\mathbb{R})$ be the corresponding solutions. Equation (1.1.5) *tells that the difference of the solutions satisfies*

$$\langle f_1 - f_2, \varphi_i \rangle_{L^2} = \langle g_1 - g_2, \varphi_i \rangle_{L^2} / \lambda_i \tag{1.1.6}$$

Therefore, we have

$$\|f_1 - f_2\|_{L^2}^2 = \sum_{i=1}^{\infty} \langle f_1 - f_2, \varphi_i \rangle_{L^2}^2 = \sum_{i=1}^{\infty} \langle g_1 - g_2, \varphi_i \rangle_{L^2}^2 / \lambda_i^2 \tag{1.1.7}$$

Notice the fact that, $\langle g_1 - g_2, \varphi_i \rangle_{L^2}$ is non-zero, but $\lim_{i \to +\infty} \lambda_i = 0$. Thus, $f_1 - f_2$ is unbounded. Although a discrete subspace with finite dimensions is adopted numerically, the condition number of the discrete matrix can be very large.

Ill-posed inverse problems need to be regularized. We should mention two pioneers who made contributions in this area. They are two Russian mathematicians, Tikhonov and Morozov. Nowadays, Tikhonov regularization is regarded as the fundamental theory in inverse problems. He suggested that, an ill-posed inverse problem can be regularized by adding a penalty functional, such that the regularized problems is well-posed. In addition, Morozov' discrepancy principle suggests how to determine the penalty functional in Tikhonov regularization. Their main works were published in journals 1960s and 1970s

(see Tikhonov's collection book [10] and references therein). After Tikhonov, there appear many regularization methods using different penalty functionals, e.g. total variation regularization, low-rank regularization, etc. (see reference [67] for a collection of modern regularization methods). In conclusion, these approaches form a branch of inverse problems, called variational inversion, which aims to minimize an objective functional (cost functional plus penalty functional).

On the other hand, the Bayesian approach [8] adopts prior distributions (from experience) to regularize ill-posed inverse problems, and interprets posterior distributions (via the Bayes' formula) to infer unknown parameters. Bayesian inversion becomes more popular nowadays, since there is an explosive growth of computer science after 1990s, which enables Markov chain Monte Carlo (MCMC) methods [24] numerically solving the Bayesian estimation. Bayesian inversion accounts for an entire distribution of all possible estimates rather than a point estimate, which requires much more computational work. According to the history of MCMC [19], although Hasting proposed the rejection sampling algorithm in 1970 [103], MCMC methods only came to mainstream statisticians after the realization by Gelfand and Smith in 1990 [5]. However, the Bayesian approach has advantages in information update and uncertainty quantification, since Bayesian inference naturally determines how to learn information from data, and also determines the implications of probability distributions.

1.1.2 Applications of inverse problems

Inverse problems are very important and widely appear in science and engineering. There are numerous applications in optics, acoustics, quantum chemistry, astronomy, geophysics, hydrology, oceanography, atmospheric sciences, systems biology, medical imaging, nondestructive testing, signal processing, artificial intelligence, and many other fields. This subsection tries to list and discuss academic literature for each of the applications mentioned above, in order to help reader achieve intuitive senses of inverse problems in the real world.

Inverse problems in fluid mechanics

consider imcompressible non-thermal fluids. The physical behavior of fluid flow is generally governed by the Navier-Stokes equations [43, 45], which describes states (e.g. velocity and pressure) of fluid flow. For convenience, we classify two types of fluid flow: flow in free space, such as winds and rivers, and flow in porous media, such as water flow through soils and sands.

In free space This case mainly results in an inverse boundary problem, as fluid flow passes through a solid which is a part of boundaries of the fluid flow domain. This is the so-called fluid-structure interaction problem [52]. For this problem, one is likely to know the flow states on the boundary by collecting experimental data. For example, [34] investigates how to obtain the inflow velocity field from the knowledge of wind loads around bridge sections. This kind of inverse problem is valuable in civil engineering, because the interaction of winds and buildings is important in construction design.

In porous media This case mainly results in an inverse coefficient problem, as the viscous resisting forces of fluids depend on the properties of the porous medium (the dependency is usually described by the Darcy's law in pracice). In this inverse problem, one can thus estimate the medium properties by measuring the flow states. For example, [62, 64, 66, 63, 61] investigate how to obtain subsurface permeabilities by measuring pressures of underground flow. This kind of inverse problems is meaningful in reservoir engineering, as it is helpful to improve the characterization of the geophysical properties of the subsurface.

Inverse problems in acoustic waves

Consider non-fluid non-thermal acoustic waves. In continuum mechanics, acoustic waves occur due to vibration of sources. Since stress-strain responses of media are different, the speeds of propagation of acoustic waves in these media are also different, and this makes acoustic waves scatter at the interfaces of media. Thus, an acoustic wave field in

a domain implies information of the media and sources inside. One can infer properties of the media or locations of the sources from data of acoustic waves. Generally speaking, the forward problems in this area are about solving the wave equation in different cases, and the inverse scattering problems consist of object identification or source detectation.

Acoustic waves are widely used in geophysics, engineering, and medical sciences. For example, determination of hypocenters in earthquakes [89], seismic tomography of the earth [51, 107], sonar used for several purposes in oceanography [105, 84], ultrasounds used for medical imaging [90] and nondestructive testing [53]. However, in these applications, the sizes of objects are different, so frequencies of applied acoustic waves are also different. More specifically, sonar applies frequencies higher than seismic waves but lower than ultrasounds.

Inverse problems in electromagnetic waves

Electric currents, radio waves, microwaves, lights, and X-ray are about electromagnet fields, but there exist differences. Electric fields in electrical circuits are assumed to be conservative fields (a special case of the Maxwell's equations), since the frequencies are very low or zero. Radio waves and microwaves used in telecommunication engineering are governed by the Maxwell's equations. Lights (infrared, visible, ultraviolet) studied in optics are also governed by the Maxwell's equations, but sometimes the photoelectric effect (beyond the Maxwell's equations) should be considered, especially for ultraviolet. For X-ray, the wave-particle duality (beyond the Maxwell's equations) has to be accounted due to the very high frequencies of these kinds of electromagnetic fields such that X-ray photons carry enough energy penetrating objects (e.g. human body).

Electromagnetic waves are widely applied in the scenarios of inverse problems. Since attenuation and scattering effects of electromagnetic waves in media are different, objects can be identified by applying electromagnetic waves and collecting the response data. As discussed in the last paragraph, the physical behaviors of electromagnetic fields vary with respect to different frequencies. Therefore, various inverse problems arise under different regimes: 1) in low frequency, the interaction is dominated by the Ampère's circuital law (electrostatics); 2) medium frequency, in the interaction is dominated by the Maxwell's equations (electrodynamics); 3) in high frequency, the interaction is dominated by the

photoelectric effect (wave-particle duality); and 4) in intrinsic frequency of nucleus, the interaction is dominated by the Bloch equations (nuclear magnetic resonance).

Low frequency This is a special case governed by Ampère's circuital law (with Maxwell's addition), since this case is (quasi)-electrostatic. The inverse problem aims to investigate electrical properties of materials by applying electrical currents into media and measuring the resulting voltages. This is similar to the inverse coefficient problem of Darcy's flow, but the coefficient here is an electrical property. Two typical examples are electrical impedance tomography (EIT) used in medical imaging [69, 57] and electrical resistivity tomography (ERT) used in geophysical imaging [102, 7]. These applications focus on different electrical properties due to different objects, but mathematically they are the same.

Medium frequency This is a more general case governed by the Maxwell's equations and leading to the wave equation. Thus, it is similar to the inverse scattering problem of acoustic waves, but the wave here is an electromagnetic wave. Theis has several applications in atmospheric optics [58, 56, 4], radar [12], and astronomy [14]. The main idea is to monitor or identify objects by analyzing signals of electromagnetic waves.

High frequency This case results in the inverse problems that uses the penetrating property of high-energy photons, in order to recover inside images of objects. For example, X-ray photography is used in medical imaging [18] and nondestructive testing of mateirals [86].

Magnetic resonance The previous three cases discuss the interaction between media and electromagnetic fields. An additional case studies the effects of nuclear magnetic resonance [99, 40], which involves the interation between nuclei and magnetic fields (not electric fields). Its applications are well-known as the techinique: magnetic resonance imaging (MRI), which is more and more popular in clinical science nowadays

[95, 55, 72].

Inverse problems in probability waves

In quantum mechanics, a quantum state is described by a wave function $\Phi(x,t)$, whose amplitude is the probability that the particle appears in location x at time t, so it is also called a probability wave. Generally, the wave function $\Phi(x,t)$ is governed by the Schrödinger equation with a Hamiltonian operator [48]. Specially, the Schrödinger equation of a single nonrelativistic particle describes the interaction between the potential energy and the wave function of the particle.

In inverse problems, the most natural examples include finding the potential energy curves from the knowledge of the ro-vibrational spectra, or determining Hamiltonian matrix elements from the knowledge of experimental energy levels. These inverse problems are discussed in [48].

Inverse problems in gravitational waves

The Einstein field equations are widely accepted to describe gravity. In astronomy, researchers use the Laser Interferometer Gravitational-wave Observatory (LIGO) to detect cosmic gravitational waves and to develop gravitational-wave observations as an astronomical tool [13]. The inverse problem for a network of laser interferometer gravitational wave detectors is discussed in [27] as the analysis of cosmic signals.

Inverse problems in complex systems and artificial models

For some complex systems (e.g. weather forecasting [109], system biology [41, 68]), there is lack of perfect theoretical characteration, or it is impposible to excatly solve the physical governing euqtaions because of the complexity of PDEs and the chaotic behaviour of the systems, so some empirical models with charaterizing parameters are adopted. On the other hand, for some modern applications especially in computer science and machine learning, researchers construct artificial models (e.g. neuron networks [76]) to deal with some challenging problems such as computer vision [21] and natural language processing [87, 25]. The inverse problems consist of training parameters involved in artificial models.

1.1.3 PDE-constrained inverse problems

Inverse problems constitute a branch of applied mathematics, and in practice, there are many implementations of the general theory of inverse problems (as discussed in the last subsection). This subsection classifies some commonly used inverse problems governed by linear second-order PDEs, which helps reader obtain further understanding of inverse problems with more mathematical views. Since these PDE-constrained inverse problems are in function spaces (infinite-dimensional), there may be some theoretical difficulties for readers who are not familiar with this area, but the theory of inverse problems has been well established even for infinite-dimensional cases [8].

Inverse problems with hyperbolic equations

Consider two kinds of waves: 1) (non-fluid non-thermal) acoustic waves, and 2) electromagnetic waves (governed by the Maxwell's equations). Mathematically, they have the same structure as follows. The (source-free) wave equation of a scalar potential v in an isotropic homogeneous and linear medium is a hyperbolic partial differential equation,

$$\frac{\partial^2 v}{\partial t^2} = c^2 \nabla^2 v \tag{1.1.8}$$

where v is the acoustic pressure in the context of acoustic waves [82] or the electric potential (under Lorenz gauge) in the context of electromagnetic waves [15], t is time, $c > 0$ is the speed of wave propagation in the medium, and ∇^2 is the Laplace operator on spatial domain. By considering time-harmonic wave or applying Fourier transform from time domain to frequency domain, the wave equation (1.1.8) can be rewritten as Helmholtz equation,

$$\nabla^2 \widehat{v} + k^2 \widehat{v} = 0 \tag{1.1.9}$$

where $\widehat{v}$ is the spectrum of v, $k = 2\pi\xi/c$ is the wavenumber, and ξ is the frequency.

Inverse scattering problems are widely studied for the purpose of target identification in many applications of radar, optics, sonar, and ultrasound. Mathematically, given the Sommerfeld's radiation condition at infinity [11, 92] and the (Lipschitz) boundary ∂D of an object in $\mathbb{R}^3$, the Helmholtz equation (1.1.9) has a solution $\widehat{v}$. The inverse problem is following: given the (partially and inaccurately) observed data of $\widehat{v}$, to recover the boundary ∂D of the object [23].

However, in general, it is not easy to analytically solve the forward problems as well as the inverse problems. Practically, the solution $\hat{v}$ of the Helmholtz equation (1.1.9) can be represented by an integral form involving the Green's function, and then is estimated by some approximation methods, such as Born approximation and Rytov approximation [100, 39]. The simplest but widely used method is the eikonal approximation [26], which only considers the leading component of the Helmholtz equation, that means only the direct wave is captured and any other scattered waves are ignored. Then the distance between the target and the observer can be calculated from the travel time of wave propagation.

Inverse problems with parabolic equations

The (source-free) heat equation of a scalar potential v in an isotropic homogeneous and linear medium is the simplest example of parabolic partial differential equations,

$$\frac{\partial v}{\partial t} = \alpha \nabla^2 v \tag{1.1.10}$$

where v is the temperature, t is the time, $\alpha > 0$ is thermal diffusivity of the medium, and ∇^2 is the Laplace operator on spatial domain. To solve the heat equation (1.1.10), one needs the boundary condition of the temperature v on ∂D and the initial condition $(t = 0)$ of the temperature v on D, where D is a Lipschitz domain. The inverse initial condition problem is the following: given the boundary condition and the final condition $(t = T)$ of the temperature v, to calculate the unknown initial condition.

Solving the inverse initial condition problem is much harder than solving the forward problem of the heat equation (1.1.10). In physics, according the second law of thermodynamics, it is well-known that heat transfer is irreversible. Mathematically, the inverse initial condition of the heat equation is ill-posed [101, 70, 8]. For this reason, finding the initial condition of the heat equation is a typical example showing the ill-posedness in inverse problems.

Beyond the simple heat equation (1.1.10), consider the more general parabolic equation [35, 36, 37],

$$\frac{\partial v}{\partial t} + \frac{1}{2} \sum_{i,j=1}^{d} a^{ij}(x,t) \frac{\partial^2 v}{\partial x^i \partial x^j} + \sum_{i=1}^{d} b^i(x,t) \frac{\partial v}{\partial x^i} + c(x,t)u + f(x,t) = 0 \tag{1.1.11}$$

where v is the state, d is the number of dimensions of the spatial domain, and the coefficient a^{ij} must strictly satisfy the elliptic condition (positive-definiteness). Note: this is a backward parabolic equation, that means to solve this equation, the final condition ($t = T$) rather than the initial condition ($t = 0$) should be given. The parabolic approach (represent the sate v in equation (1.1.11) as an expectation of underlying diffusion processes) is related to solving the Navier-Stokes equations in fluid mechanics, also related to pricing derivatives in quantitative finance. The inverse problem, e.g. for option pricing, is parameter calibration of a^{ij}, b^i, and c with respect to market data.

Inverse problems with elliptic equations

Consider two scenarios: 1) the Darcy's law for fluid flow in porous media, and 2) the Ampère's circuital law (with Maxwell's addition) for electromagnetic fields in electrical media. Mathematically, they have the same structure as follows. The static equilibrium in an isotropic heterogeneous and linear medium is described by an elliptic partial differential equation,

$$-\nabla \cdot (\kappa \nabla v) = f \tag{1.1.12}$$

where v is the scalar potential (e.g. v is the pressure for Darcy flow or the voltage for electrical circuits), $\kappa > 0$ is the property of the isotropic heterogeneous medium (e.g. v is the permeability for Darcy flow or the impedance for electrical circuits), ∇ is the nabla operator on spatial domain, and f is the source recharge.

The inverse problem is to recover the coefficient κ with observed data of v. Since the number of observation is finite and κ is a infinite-dimensional (a function on the spatial domain), the inverse problem is highly underdetermined [70, 8]. Since the elliptic equation is coercive and it is easy to be solved with finite element methods, so elliptic equation is also commonly used as a benchmark for testing inverse algorithms.

1.2 Abbreviation of this thesis

This section introduce motivations, contributions and layouts of this thesis.

There exist computational challenges in real applications of inverse problems, as inverse estimation usually requires multiple (from tens to millions) forward simulations,

but forward simulation of complex models (e.g. PDE-constrained problems) is computationally expensive. It is needed to design efficient and robust inverse algorithms for both academic research and real applications.

Mathematicians and statisticians consider inverse problems with two different approaches, the varitional approach and the Bayesian approach. The former is related to mathematical optimization and the latter is related to statistical inference. These two approaches sometimes can be connected, since in some cases, the varitional approach is equivalent to the maximum a posteriori (MAP) estimation in the Bayesian approach, e.g. Tikhonov regularization method and the MAP estimation with Gaussian error and Gaussian prior. Both of the variational approach and the Bayesian approach will be introduced in this thesis.

Another question is how to find the minimum point of an objective functional, or how to calculate the posterior distribution from the Bayes' formula. For linear inverse problems, there exists the closed form of solutions (details will be discussed later in this thesis). For nonlinear inverse problems, numerical algorithms (details will be discussed later in this thesis) are needed:

- Firstly, Tikhonov regularization is a method using objective functionals formulated in L^2 norm. Optimization in L^2 norm is well-known as the least-square method (minimization of sum of squares). The standard numerical methods solving nonlinear least-square problems are the Gaussian-Newton algorithm (GNA) [2, 22] and the Levenberg-Marquardt algorithm (LMA) [22]. LMA is a modification of GNA using a trust region approach with damping factors. Conversely, GNA can be regarded as a special case of LMA with the damping factors equaling to zeros. Generally, LMA with suitable damping factors is more robust than GNA.

- Secondly, the objective of Bayesian inversion is to draw samples from the posterior distribution. The most popular sampling algorithms are known as the MCMC methods [24] or some variants e.g. sequential Monte Carlo methods with MCMC mutations [3, 1]. For infinite-dimensional sampling, it should be very careful about the proposal transitions in the Metropolis-Hastings rejection sampling, in order to avoid singularity of probability measures. This thesis adopts the **p**reconditioned

Crank-**N**icolson (PCN)-MCMC method [94], which is a mesh-invariant sampler regardless of number of dimensions.

However, there are some drawbacks of the standard algorithms: 1) LMA searches for a local optimum, which may be far from the global optimum for a highly nonlinear problem, and there lacks uncertainty quantification; 2) MCMC methods are accurate and can capture the entire distribution of multiple estimates, but these sampling algorithms are very inefficient in practice. In order to solve these issues, we aim to apply some approximate methods. These heuristic algorithms are related to the Kalman filter [38, 33]. How to come up with the Kalman-like methods for inverse problems is briefly introduced as follows:

1. Inverse problems can be rewritten in a tempering setting (see formula (3.1.20) and formula (3.1.21)). This technique is widely applied in simulated annealing [104], annealed importance sampling [88], sequential Monte Carlo method [3, 1], etc. The terminology 'tempering/annealing' originally comes from metallurgy. Statisticians borrow the similar idea to design mathematical algorithms. The motivation is that: it could be inefficient to directly draw samples from a distribution with multiple sharp peaks, because numerically the samples may be sticky around one peak and other peaks cannot be captured. However, it is easier to gradually sample from a sequence of distributions from the prior to the posterior (from the flat to the sharp) as the 'temperature' goes down. The tempering setting here is a mathematical concept rather than a physical fact, i.e. it is 'simulated' annealing rather than 'real' annealing. The terminology, simulated annealing, usually indicates a special type of mathematical optimization algorithms. To void confusion, this thesis uses a different terminology, tempering setting, that indicates the general mathematical formulation.

2. Inverse problems formulated via the tempering setting (will be simply called *tempered inverse problems*) can be equivalently regarded as filtering problems. Thus filtering algorithms, e.g. Kalman filter and its variants like extended Kalman filter (EKF) and ensemble Kalman filter (EnKF), can be applied. Kalman filter is the linear filtering algorithm with a closed form; EKF and EnKF are nonlin-

ear (approximate) filtering algorithms. Then, the Kalman-like filters can be used to (approximately) solve the tempered inverse problems, which is very efficient in practice. We should mention that, using the Kalman-like filters on the tempered inverse problems is different from the common sense of 'filter', since the 'filtering' here is not for a real process but for the tempering setting. Therefore, this thesis prefers to use different names as *Kalman inversion*, *extended Kalman inversion* (EKI), and *ensemble Kalman inversion* (EnKI), in order to specify the fact that Kalman-like approaches are used to solve tempered inverse problems.

This thesis mainly focuses on the two approximate algorithms: EKI and EnKI. The former is more like a deterministic optimization algorithm; and the latter is more like a statistical sampling algorithm. On one hand, EKI is a point estimation method searching for a sub-optimum. The performance of EKI is similar like LMA, but EKI can further (approximately) quantifies the uncertainty via covariance update. EKI works for problems with continuously differentiable forward models. On the other hand, EnKI is an approximate sampling algorithm. EnKI only works for problems whose parameters and observations have strong linear dependence, but this method is derivative-free and much more efficient than MCMC sampling.

The layouts of this thesis are arranged as follows. Chapter 1 is an introduction of inverse problems including descriptions and applications. Chapter 2 discusses two standard approaches dealing with inverse problems, including variational approach and Bayesian approach. The fundamental theories and standard algorithms can be fond in this chapter. Chapter 3 is the main part of this thesis, where we propose a new framework for computational inverse problems. The tempering setting and the approximate algorithms (e.g. EKI and EnKI) can be found in this chapter. Chapter 4 gathers all theoretical discussions and proofs which support the results in previous chapters. Chapter 5 conducts numerical tests of the proposed algorithms with a PDE-constrained inverse problem. Chapter 6 is a summary which highlights the main contents and contributions of this thesis.

Chapter 2

Inverse Problems on Hilbert Spaces

This chapter introduces inverse problems placed in infinite-dimensional spaces. Finite-dimensional parameter estimation is well-known in many fields including engineering, physics, econometrics. Although, after discretization, the number of parameters is always finite, it is benifitial to keep the mathematical structure in infinite-dimensional spaces and leave the discretization in the last step [8]. There are two main approaches dealing with inverse problems, variational inversion and Bayesian inversion. Both will be discussed and compared in detail.

This chapter can be regarded as a structured literature review. Section 2.1 represents preliminaries about functional analysis and probability theory. Section 2.2 introduces the two standard approaches (variational and Bayesian) for inverse problems. Section 2.3 discusses Karhunen-Loève expansion that is used to represent prior random fields. For posterior estimation, sections 2.4 and 2.5 introduce the Levenberg-Marquardt algorithm (LMA) and Markov chain Monte Carlo (MCMC) methods as the standard algorithms for the variational inversion and the Bayesian inversion, respectively. Section 2.6 is a toy example showing how to apply LMA and MCMC to solve infinite-dimensional nonlinear inverse problems.

2.1 Preliminaries

We consider infinite-dimensional inverse problems placed in function spaces. This section presents a list of relevant definitions and theorems.

The required theories can be found in any books or well-structure lecture notes related to functional analysis and probability/measure theory. In functional analysis, we need to know compact operators in separable Hilbert spaces, and theorems like Riesz representation theorem, spectral theorem, and Mercer's theorem (we will cite results from references [29, 50, 106, 46, 44] about functional analysis). In probability/measure theory, we need to know probability/measure space in separable Hilbert spaces, and theorems like Radon-Nikodym theorem, Fernique's theorem, and Cameron-Martin theorem (we will cite results from [8, 32] about probability/measure theory).

2.1.1 Functional analysis

Let $\mathcal{H}$ be a Hilbert space over $\mathbb{R}$, equipped with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. The inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is a symmetry bilinear operator such that, for all $u \in \mathcal{H}$,

$$\langle u, u \rangle_{\mathcal{H}} \geq 0 \tag{2.1.1}$$

where the equality holds if and only if $u = 0$. In addition, the norm $\| \cdot \|_{\mathcal{H}} : \mathcal{H} \to \mathbb{R}$ is induced from the inner product such that, for all $u \in \mathcal{H}$,

$$\|u\|_{\mathcal{H}} := \sqrt{\langle u, u \rangle_{\mathcal{H}}} \tag{2.1.2}$$

Let $\mathcal{H}$ be a Hilbert space over $\mathbb{R}$. The dual space of $\mathcal{H}$ is denoted by $\mathcal{H}^*$, which is a set of all bounded linear maps from $\mathcal{H}$ to $\mathbb{R}$. For any $u \in \mathcal{H}$, the *dual* of u is denoted by $u^* \in \mathcal{H}^*$ such that, for all $v \in \mathcal{H}$,

$$u^*(v) = \langle u, v \rangle_{\mathcal{H}} \tag{2.1.3}$$

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces. For any bounded linear operator $\mathcal{A} : \mathcal{H}_1 \to \mathcal{H}_2$, the *Hermitian adjoint* of $\mathcal{A}$ is the bounded linear operator $\mathcal{A}^* : \mathcal{H}_2 \to \mathcal{H}_1$ such that, for all $u_1 \in \mathcal{H}_1$ and $u_2 \in \mathcal{H}_2$,

$$\langle \mathcal{A} u_1, u_2 \rangle_{\mathcal{H}_2} = \langle u_1, \mathcal{A}^* u_2 \rangle_{\mathcal{H}_1} \tag{2.1.4}$$

Existence and uniqueness of $(\cdot)^*$ follows from the Riesz representation theorem. In fact, the Riesz representation theorem shows that a real-(complex-)valued Hilbert space and its dual space are isometric (anti-)isomorphism.

Theorem 2.1.1 (Riesz representation theorem [29]). *Let $\mathcal{H}$ be a Hilbert space. For any $g \in \mathcal{H}^*$, there exists a unique $u \in \mathcal{H}$, such that for all $x \in \mathcal{H}$, $g(x) = \langle u, x \rangle_{\mathcal{H}}$. Moreover, $\|u\|_{\mathcal{H}} = \|g\|_{\mathcal{H}^*}$.*

Let $\mathcal{H}$ be a Hilbert space. A bounded linear operator $\mathcal{A} : \mathcal{H} \to \mathcal{H}$ is said to be *self-adjoint* if

$$\mathcal{A}^* = \mathcal{A} \tag{2.1.5}$$

A self-adjoint operator $\mathcal{A} : \mathcal{H} \to \mathcal{H}$ is said to be *positive-semi-definite* if for all $u \in \mathcal{H}$

$$\langle \mathcal{A}u, u \rangle_{\mathcal{H}} \geq 0 \tag{2.1.6}$$

and is said to be *positive-definite* if the equality in above formula only holds on the condition $u = 0$.

Let $\mathcal{H}$ be a separable Hilbert space. Let $\{e_i\}$ be an orthonormal basis of $\mathcal{H}$. A linear operator $\mathcal{A} : \mathcal{H} \to \mathcal{H}$ is said to be *trace-class* if

$$\mathrm{Tr}(\mathcal{A}) := \sum_{i=1}^{\infty} \left\langle (\mathcal{A}^*\mathcal{A})^{1/2} e_i, e_i \right\rangle_{\mathcal{H}} < \infty \tag{2.1.7}$$

The sum is independent on the choice of orthonormal bases. A linear operator $\mathcal{A} : \mathcal{H} \to \mathcal{H}$ is said to be *Hilbert-Schmidt* if

$$\mathrm{Tr}(\mathcal{A}^*\mathcal{A}) = \sum_{i=1}^{\infty} \|\mathcal{A}e_i\|_{\mathcal{H}}^2 < \infty \tag{2.1.8}$$

The sum is independent on the choice of orthonormal bases. The concepts of trace-class and Hilbert-Schmidt are closely related.

Proposition 2.1.2. *[44] A bounded linear operator is trace-class if and only if it is a multiplication of two Hilbert-Schmidt operators.*

Proposition 2.1.3. *[44] A trace-class operator must be a Hilbert-Schmidt operator, a Hilbert-Schmidt operator must be a compact operator.*

For self-adjoint compact operators in Hilbert spaces, a fundamental result is the spectral theorem.

Theorem 2.1.4 (Spectral theorem for compact operators in Hilbert spaces (more general discussions are in [106])). *A self-adjoint compact operator $\mathcal{A}$ in a Hilbert space $\mathcal{H}$ is unitarily diagonalizable. Namely, there exists a (countable) sequence of eigenvalues $\{\lambda_i : i \in \mathbb{N}_1\}$ ($|\lambda_i|$ is sorted from largest to smallest) and corresponding (orthonormal) eigenvectors $\{\varphi_i : i \in \mathbb{N}_1\}$,*

$$\mathcal{A}\varphi_i = \lambda_i \varphi_i \tag{2.1.9}$$

such that the eigenvalues vanish to zero, i.e.

$$\lim_{i \to +\infty} \lambda_i = 0 \tag{2.1.10}$$

and the eigenvectors form an orthonormal basis of the range of operator $\mathcal{A}$, i.e. for any bounded $x \in \mathcal{H}$,

$$\mathcal{A}x = \sum_{i=1}^{\infty} \lambda_i \langle x, \varphi_i \rangle_{\mathcal{H}} \varphi_i \tag{2.1.11}$$

Closely related to the spectral theorem, another important theoretical tool is Mercer's theorem, which is used to characterize symmetric positive-semi-definite kernels.

Theorem 2.1.5 (Mercer's theorem [50]). *Let $K \in C(\overline{D} \times \overline{D}; \mathbb{R})$ be a continuous symmetric positive-semi-definite kernel on a compact set $\overline{D} \subset \mathbb{R}^d$. Then K can be represented by,*

$$K(s,t) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(s) \varphi_i(t) \tag{2.1.12}$$

with absolute and uniform convergence in $\overline{D}$, where $\{\lambda_i\}$ and $\{\varphi_i\}$ are the eigenvalues and (orthonormal) eigenfunctions,

$$\lambda_i \varphi_i(s) = \int_D K(s,t) \varphi_i(t) \, \mathrm{d}t \tag{2.1.13}$$

In order to regularize an ill-posed inverse problem on a separable Hilbert space $\mathcal{H}$, usually a self-adjoint positive-semi-definite trace-class operator $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ is applied, which restricts the ill-posed problem from the original Hilbert space $\mathcal{H}$ to the reproducing kernel Hilbert space/the Cameron-Martin space[1], $E = \mathrm{Ran}\left(\mathcal{C}^{1/2}\right)$. The essential property here is trace-class, which means that the eigenvalues of $\mathcal{C}$ are summable. Also, it

[1]the terminology *reproducing kernel Hilbert space* [46] is usually used in functional analysis when discussing integral operators, and the terminology *Cameron-Martin space* [8] is usually used in probability theory when discussing Gaussian measures.

implies $\mathcal{C}^{1/2}$ is Hilbert-Schmidt, so $\mathcal{C}^{1/2}$ is a compact operator. Thus the Cameron-Martin space/reproducing kernel Hilbert space $E = \mathrm{Ran}\left(\mathcal{C}^{1/2}\right)$ is a relatively compact subspace of $\mathcal{H}$, equipped with the inner product

$$\langle \cdot, \cdot \rangle_E \equiv \langle \cdot, \cdot \rangle_{\mathcal{C}} := \left\langle \mathcal{C}^{-1/2}(\cdot), \mathcal{C}^{-1/2}(\cdot) \right\rangle_{\mathcal{H}} \tag{2.1.14}$$

In the next, we make a typical example of a trace-class operator, and the associated reproducing kernel Hilbert space.

Example 2.1.6. *Let Δ be the Laplace operator over the interval $[-\pi, \pi]$ with the homogeneous Dirichlet boundary condition. The eigenfunctions of the Laplace operator are $\sin(nx)$, $x \in [-\pi, \pi]$, for all $n \in \mathbb{N}_1$, and the corresponding eigenvalues are $-n^2$ for all $n \in \mathbb{N}_1$. Now, let $\mathcal{C} = -\Delta^{-1}$ be an operator on the Hilbert space $L_0^2([-\pi, \pi]; \mathbb{R})$. Then, the eigenvalues of $\mathcal{C}$ are $1/n^2$ for all $n \in \mathbb{N}_1$, which are summable $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$. Thus, $\mathcal{C}$ is a trace-class operator. Furthermore, the associated reproducing kernel Hilbert space is*

$$\begin{aligned}
E = \mathrm{Ran}\left(\mathcal{C}^{1/2}\right) &= \left\{ u \in L_0^2([-\pi, \pi]; \mathbb{R}) : \left\| \mathcal{C}^{-1/2} u \right\|_{L_0^2([-\pi,\pi];\mathbb{R})} < +\infty \right\} && (2.1.15)\\
&= \left\{ u \in L_0^2([-\pi, \pi]; \mathbb{R}) : \left\langle u, \mathcal{C}^{-1} u \right\rangle_{L_0^2([-\pi,\pi];\mathbb{R})} < +\infty \right\} && (2.1.16)\\
&= \left\{ u \in L_0^2([-\pi, \pi]; \mathbb{R}) : -\left\langle u, \Delta u \right\rangle_{L_0^2([-\pi,\pi];\mathbb{R})} < +\infty \right\} && (2.1.17)\\
&= \left\{ u \in L_0^2([-\pi, \pi]; \mathbb{R}) : -\int_{-\pi}^{\pi} u(x) \Delta u(x) \, \mathrm{d}x < +\infty \right\} && (2.1.18)\\
&= \left\{ u \in L_0^2([-\pi, \pi]; \mathbb{R}) : \int_{-\pi}^{\pi} \nabla u(x) \cdot \nabla u(x) \, \mathrm{d}x < +\infty \right\} && (2.1.19)
\end{aligned}$$

which is exactly the Sobolev space $W_0^{1,2}([-\pi, \pi]; \mathbb{R})$.

2.1.2 Probability/measure theory

Definition 2.1.7 (Bochner space [32])**.** *Let (X, Σ, μ) be a measure space, that is X is a set, Σ is a σ-field over X, and μ is a measure on (X, Σ). Let E be a real-valued separable Banach space equipped with a norm $\| \cdot \|_E$. For any $p \geq 1$, the notation $\mathcal{L}^p(X, \mu; E)$ is used to denote the set of all measurable functions $f : X \to E$ such that*

1. *either $p \in [1, \infty)$,*

$$\|f\|_p := \left(\int_X \|f(x)\|_E^p \, \mu(\mathrm{d}x) \right)^{1/p} < \infty \tag{2.1.20}$$

2. or $p = \infty$,

$$\|f\|_\infty := \inf\{b \in \mathbb{R} : \mu(\{x \in X : \|f(x)\|_E > b\}) = 0\} < \infty \tag{2.1.21}$$

$\mathcal{L}^p(X, \mu; E)$ is a seminorm space equipped with the seminorm $\|\cdot\|_p$. In order to obtain a norm space, let $L^p(X, \mu; E)$ be the Bochner space of equivalence class up to μ-null set such that, the equivalence relation $\sim$ is defined as $f \sim g \iff \mu(\{x \in E : f(x) \neq g(x)\}) = 0$.

Definition 2.1.8 (Gaussian measures on Euclidean spaces)**.** *A measure μ on an Euclidean space $\mathbb{R}^m$ is called a (non-degenerate) Gaussian measure, if the map $\mu : \mathcal{B}(\mathbb{R}^m) \to [0, 1]$ is determined via, for all $\Omega \in \mathcal{B}(\mathbb{R}^m)$,*

$$\mu(\Omega) = \frac{1}{\sqrt{(2\pi)^m \det(\mathbf{C})}} \int_\Omega \exp\left(-\frac{1}{2}\left\|\mathbf{C}^{-1/2}(\mathbf{x} - \mathbf{m})\right\|_{\mathbb{R}^m}^2\right) \mathrm{d}\mathbf{x} \tag{2.1.22}$$

where $\mathbf{m} \in \mathbb{R}^m$ is mean, and $\mathbf{C} : \mathbb{R}^m \to \mathbb{R}^m$ is the covariance matrix which is an m-dimensional symmetric positive-(semi-)definite matrix. A Gaussian measure on an Euclidean spaces is noted by $\mathcal{N}(\mathbf{m}, \mathbf{C})$.

Definition 2.1.9 (Gaussian measures on separable Banach spaces)**.** *A measure μ on a real-valued separable Banach space X is called a (non-degenerate) Gaussian measure, if its pushforward measure $\varphi_*(\mu)$ is a (non-degenerate) Gaussian measure on $\mathbb{R}$ for all bounded (non-zero) linear functional $\varphi \in X^*$, where the pushforward measure is defined as, for all $U \in \mathcal{B}(\mathbb{R})$,*

$$[\varphi_*(\mu)](U) = \mu(\varphi^{-1}(U)) \tag{2.1.23}$$

In the next, we will show an example about how to apply the abstract definition of Gaussian measures on separable Banach spaces in the special case when the separable Banach spaces are Euclidean spaces.

Example 2.1.10. *Let $\mu = \mathcal{N}(\mathbf{m}, \mathbf{C})$ be a (non-degenerate) Gaussian measure on $\mathbb{R}^m$, i.e., the map $\mu : \mathcal{B}(\mathbb{R}^m) \to [0, 1]$ is determined via formula* (2.1.22)*. Let $\mathbf{g} \in \mathbb{R}^m$ be a (non-zero) element, and let $\mathbf{g}^* : \mathbb{R}^n \to \mathbb{R}$ be the dual of $\mathbf{g}$, i.e., $\mathbf{g}^*$ is a linear functional, such that, $\mathbf{g}^*(\mathbf{x}) = \langle \mathbf{g}, \mathbf{x} \rangle_{\mathbb{R}^m}$ for all $\mathbf{x} \in \mathbb{R}^m$. Let $\mathbb{P} : \mathcal{B}(\mathbb{R}) \to [0, 1]$ be the pushforward measure of μ corresponding to $\mathbf{g}^*$. By using the definition of pushforward measure* (2.1.23)*, $\mathbb{P}$ is represented as the measure on $\mathbb{R}$, such that, for all $U \in \mathcal{B}(\mathbb{R})$,*

$$\mathbb{P}(U) = \frac{1}{\sqrt{(2\pi)^m \det(\mathbf{C})}} \int_{A(U)} \exp\left(-\frac{1}{2}\left\|\mathbf{C}^{-1/2}(\mathbf{x} - \mathbf{m})\right\|_{\mathbb{R}^m}^2\right) \mathrm{d}\mathbf{x} \tag{2.1.24}$$

with the integral domain

$$A(U) = \{\mathbf{x} \in \mathbb{R}^m | \mathbf{g}^* \mathbf{x} \in U\}) \tag{2.1.25}$$

Furthermore, by using the coordinate transformation $x = \mathbf{g}^* \mathbf{x}$*, the pushforward measure* $\mathbb{P}$ *can be rewritten as*

$$\mathbb{P}(U) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_U \exp\left(-\frac{(x-\lambda)^2}{2\sigma^2}\right) \mathrm{d}x \tag{2.1.26}$$

where

$$\lambda = \mathbf{g}^* \mathbf{m} \qquad \sigma^2 = \mathbf{g}^* \mathbf{C} \mathbf{g} \tag{2.1.27}$$

Thus, $\mathbb{P}$ *is exactly the (non-degenerate) Gaussian measure on* $\mathbb{R}$ *with mean* λ *and variance* σ^2*. This result holds for all bounded (non-zero) linear functional* $\mathbf{g}^*$*.*

Theorem 2.1.11 (Radon-Nikodym theorem [8])**.** *For two* σ*-finite measures* ν *and* μ *on a measurable space* (X, Σ)*, if* ν *is absolutely continuous with respect to* μ*, then there exits a measurable function* $f : X \to [0, +\infty)$ *such that, for all* $S \in \Sigma$*,*

$$\nu(S) = \int_S f(u) \, \mathrm{d}\mu \tag{2.1.28}$$

where f *is unique up to a* μ*-null set, and called the Radon-Nikodym derivative of the two measures, noted by* $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$*. For convenience, this thesis will sometimes use the notation* $\frac{\nu(\mathrm{d}x)}{\mu(\mathrm{d}x)} := \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)$ *to denote a point value at* $x \in X$*.*

An application of Radon-Nikodym theorem is to determine probability density functions with respect to Lebesgue measures in Euclidean spaces. In the next, we will show an example of Gaussian densities.

Example 2.1.12. *Consider the Gaussian measure* $\mu = \mathcal{N}(\mathbf{m}, \mathbf{C})$ *defined in formula* (2.1.22)*. Then, the integrand in the right hand side of formula* (2.1.22) *is the Radon-Nikodym derivative of the Gaussian measure* μ *with respect to the Lebesgue measure on* $\mathbb{R}^m$*. This integrand is thus defined as the probability density function* $f : \mathbb{R}^m \to [0, +\infty)$ *of the normal distribution, i.e., for all* $\mathbf{x} \in \mathbb{R}^m$*,*

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m \det(\mathbf{C})}} \exp\left(-\frac{1}{2} \left\| \mathbf{C}^{-1/2}(\mathbf{x} - \mathbf{m}) \right\|_{\mathbb{R}^m}^2\right) \tag{2.1.29}$$

Theorem 2.1.13 (Fernique's theorem [8]). *For a zero-mean Gaussian measure μ on a separable Banach space X, there exists $\alpha > 0$ such that*

$$\int_X \exp\left(\alpha \|x\|_X^2\right) \mu(\mathrm{d}x) < \infty \tag{2.1.30}$$

Corollary 2.1.14. *A natural corollary follows the Fernique's theorem is that: a Gaussian measure μ has any finite moment for $k \geq 0$,*

$$\int_X \|x\|_X^k \, \mu(\mathrm{d}x) < \infty \tag{2.1.31}$$

Proposition 2.1.15. *[8] The covairance operator of a (non-degenerate) Gaussian measure on a real-valued separable Hilbert space is a self-adjoint positive-(semi-)definite trace-class operator. Conversely, a self-adjoint positive-(semi-)definite trace-class operator forms the covairance operator of a (non-degenerate) Gaussian measure on a real-valued separable Hilbert.*

Above proposition tells that, a Gaussian measure on a real-valued separable Hilbert $\mathcal{H}$ is uniquely determined via the mean and covariance. Thus, a Gaussian measure on $\mathcal{H}$ is usually noted by its mean and covariance as $\mathcal{N}(m, \mathcal{C})$, where $m \in \mathcal{H}$ is the mean, and $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ is the covariance operator. An example of Gaussian measures on real-valued separable Hilbert spaces can be explicitly shown in the special case when the separable Hilbert spaces are Euclidean spaces. In this special case, the Gaussian measure has an explicit form only relying on the mean and covariance, shown in formula (2.1.22).

Theorem 2.1.16 (Cameron-Martin theorem [8]). *Two Gaussian probability measures $\mu_i = \mathcal{N}(m_i, \mathcal{C}_i)$, $i = 1, 2$, on a Hilbert space $\mathcal{H}$ are either singular or equivalent. They are equivalent if and only if the following three conditions hold:*

1. $\mathrm{Ran}\left(\mathcal{C}_1^{1/2}\right) = \mathrm{Ran}\left(\mathcal{C}_2^{1/2}\right) := E$,
2. $m_1 - m_2 \in E$,
3. *The operator* $T := \left(\mathcal{C}_1^{-1/2}\mathcal{C}_2^{1/2}\right)\left(\mathcal{C}_1^{-1/2}\mathcal{C}_2^{1/2}\right)^* - \mathcal{I}$ *is Hilbert-Schmidt in* $\overline{E}$.

For regularization of an inverse problem on a separable Hilbert space $\mathcal{H}$, a prior covariance operator $\mathcal{C}_0 : \mathcal{H} \to \mathcal{H}$ is considered, which leads to the reproducing kernel

Hilbert space/the Cameron-Martin space $E = \mathrm{Ran}\left(\mathcal{C}_0^{1/2}\right)$. Moreover, a prior mean $m_0 \in \mathcal{H}$ sometimes is also used for translation of space $E \to m_0 + E$. If $m_0 \in E$, then the space keeps the same after the translation, i.e. $E = m_0 + E$. Furthermore, if $m_0 \in E$, the Gaussian measures also keep the equivalence under the translation, as shown in the Cameron-Martin theorem.

2.2 Well-posed inverse problems

This section introduces how to define an inverse problem with well-posed mathematical structures via the two standard approaches, variational approach and Bayesian approach.

2.2.1 The formulation

We consider inverse problems with finite observations and countable parameters. Namely, let $\mathbb{R}^n$ be the observation space, where n is the number of observations, and let $\mathcal{H}$ be the parameter space, where $\mathcal{H}$ is a real-valued separable Hilbert space. The simplest way connecting observations and parameters is the additive noise model,

$$y = \mathcal{G}(x) + e \tag{2.2.1}$$

where $y \in \mathbb{R}^n$ is the observation, $x \in \mathcal{H}$ is the hidden parameter, $\mathcal{G} : \mathcal{H} \to \mathbb{R}^n$ is the forward map for mathematical prediction, and $e \in \mathbb{R}^n$ is the error between the real observation y and the mathematical prediction $\mathcal{G}(x)$.

'Inversion' means to infer the unknown parameter $x \in \mathcal{H}$ given the observation $y \in \mathbb{R}^n$. We assume that y is in the range of operator $\mathcal{G}$, namely $y \in \mathrm{Ran}(\mathcal{G})$. This assumption makes sure that there exists at least a solution $u \in \mathcal{H}$ such that $y = \mathcal{G}(u)$. However, in general cases, the direct approach finding the solution u as an estimate of the truth x is ill-posed, because of the following two facts: 1) for under-determined problems, there could be multiple solutions; 2) even though there exists a unique solution, the solution may be unstable, that means a little change in the data may lead to a big change in the solution, so that, the estimation is not reliable for noisy data.

An ill-posed inverse problem needs regularization. There are two standard approaches for prior regularization, i.e. variational approach and Bayesian approach. Variational ap-

proach determines a point estimate of the truth by minimizing an objective functional. The objective functional is usually the sum of a cost functional and a penalty functional, where the cost functional quantifies the difference between real observations and mathematical predictions, and the penalty functional is required for additional regularization if minimization of the cost functional is ill-posed. On the other hand, Bayesian approach characterizes the probability distribution of all possible estimates. The initial guess of estimates is characterized by a prior distribution, which also provides the prior regularization of ill-posed problems. The conditional distribution (from the Bayes' rule) given observations is thus regarded as the posterior distribution of estimates. More precisely, these two approaches are formulated as follows:

1. The variational method regards inversion as deterministic optimization, which aims to find the minimum point $\widehat{x}(y) \in \mathcal{H}$ of an objective functional,

$$\widehat{x}(y) = \arg\min_{u \in \mathcal{H}} \{\Phi(u|y) + R(u)\} \tag{2.2.2}$$

   where $\Phi(\cdot|y) : \mathcal{H} \to [0, +\infty)$ is the cost functional given observation $y \in \mathbb{R}^n$, and $R : \mathcal{H} \to [0, +\infty)$ is the penalty functional. For the additive noise model (2.2.1), the cost functional can be expressed as $\Phi(\cdot|y) = \rho(y - \mathcal{G}(\cdot))$ with a non-negative function $\rho : \mathbb{R}^n \to [0, +\infty)$ which quantifies the utility of the error $\epsilon \equiv y - \mathcal{G}(u)$ for any $u \in \mathcal{H}$. Moreover, the penalty functional R is user-specified. A typical example is Tikhonov regularization, which assumes the penalty functional is determined by using L^2 norm.

2. The Bayesian method regards inversion as statistical inference, which aims to interpret the conditional probability measure $\mathbb{P}(\cdot|y) : \mathcal{B}(\mathcal{H}) \to [0, 1]$ via the Bayes' formula,

$$\mathbb{P}(\mathrm{d}u|y) \propto \mathcal{L}(u|y)\mathbb{P}(\mathrm{d}u) \tag{2.2.3}$$

   where $\mathcal{L}(\cdot|y) : \mathcal{H} \to [0, +\infty)$ is the likelihood function given observation $y \in \mathbb{R}^n$, and $\mathbb{P} : \mathcal{B}(\mathcal{H}) \to [0, 1]$ is the prior probability measure of estimates. For the additive noise model (2.2.1), the likelihood function can be expressed as $\mathcal{L}(\cdot|y) = \pi(y - \mathcal{G}(\cdot))$ with a probability density function $\pi : \mathbb{R}^n \to [0, +\infty)$ which characterizes the distribution of the error $\epsilon \equiv y - \mathcal{G}(u)$ for all $u \in \mathcal{H}$. Moreover, the prior probability

$\mathbb{P}$ is user-specified. A common example is Gaussian prior, which assumes the prior probability is a Gaussian measure on $\mathcal{H}$.

Now, we consider a specific form of the varitional approach (2.2.2) and the Bayesian approach (2.2.3). Namely, we consider 1) Tikhonov regularization for variational approach, and 2) Gaussian error and Gaussian prior for Bayesian approach. Tikhonov regularization considers minimization in L^2 norm, that means the cost functional and the penalty functional are in the form of sum of squares[2]. On the other hand, for Bayesian inversion, Gaussian distributions are exactly corresponding to the L^2 norm used for the variational inversion. Thus, Gaussian distributions build a bridge from variational inversion to Bayesian inversion. More precisely, the variational inversion (2.2.2) and the Bayesian inversion (2.2.3) can be specified in the quadratic form:

1. Tikhonov regularization for the variational approach,

$$\widehat{x}(y) = \arg\min_{u\in\mathcal{H}} \left\{ \frac{1}{2} \left\| \gamma^{-1/2}(y - \mathcal{G}(u)) \right\|_{\mathbb{R}^n}^2 + \frac{1}{2} \left\| \mathcal{C}_0^{-1/2} (u - m_0) \right\|_{\mathcal{H}}^2 \right\} \tag{2.2.4}$$

2. Gaussian error and Gaussian prior for the Bayesian approach,

$$\mathbb{P}(\mathrm{d}u|y) \propto \exp\left(-\frac{1}{2} \left\| \gamma^{-1/2}(y - \mathcal{G}(u)) \right\|_{\mathbb{R}^n}^2 \right) \mathbb{P}(\mathrm{d}u) \quad \text{with} \quad \mathbb{P} = \mathcal{N}(m_0, \mathcal{C}_0) \tag{2.2.5}$$

where $\gamma : \mathbb{R}^n \to \mathbb{R}^n$ is a symmetric positive-definite bounded matrix (the covariance matrix of error), $m_0 \in \mathcal{H}$ is a bounded element (the prior mean), and $\mathcal{C}_0 : \mathcal{H} \to \mathcal{H}$ is a self-adjoint positive-semi-definite trace-class operator (the prior covariance operator). The Tikhonov regularization (2.2.4) and the Bayesian inference with Gaussian error and Gaussian prior (2.2.5) are connected as the minimum point of (2.2.4) equals to the MAP (maximum a posteriori) estimate of (2.2.5) [8].

In conclusion, inversion is a general concept about inference of inputs of a system provided with observed outputs. Mathematically, an inverse problem is usually treated

[2]Minimization in L^2 norm is also known as the least squares method, but the least squares method typically has more special meaning, i.e. to regress data in a finite-dimensional over-determined system (the number of observations is finite and more than the number of parameters). In order to avoid confusion, the terminology 'minimization in L^2 norm' used in this thesis indicates more general meaning regardless of its applications in over-determined or under-determined, finite-dimensional or infinite-dimensional, linear or nonlinear problems.

as a mathematical optimization problem (variational inversion) or a statistical inference problem (Bayesian inversion). For the variational inversion (2.2.2), our purpose is to find an optimum $\widehat{x}(y)$ given the observation $y \in \mathbb{R}^n$. The optimum $\widehat{x}(y)$ is regarded as a point estimate of the unknown parameter $x \in \mathcal{H}$. For the Bayesian inversion (2.2.3), our purpose is to interpret the entire posterior distribution $\mathbb{P}(\mathrm{d}u|y)$ rather than a point estimate. The posterior distribution characterizes all possible estimates of the unknown parameter $x \in \mathcal{H}$ given the observation $y \in \mathbb{R}^n$. Practically, we need to conduct mathematical optimization algorithms (e.g. gradient descent, Gauss-Newton, Levenberg-Marquardt) for variational inversion, and conduct numerical sampling algorithms (e.g. importance sampling, MCMC) for Bayesian inversion.

2.2.2 The well-posedness

This subsection aims to show the well-posedness of method (2.2.4) and method (2.2.5). First of all, we consider linear problems whose solutions have a closed form, so the well-posedness can be shown explicitly. After that, we consider more general cases of nonlinear problems, where the well-posedness is described in a more abstract way.

For linear problems

Now, we consider the simplest case that the forward map $\mathcal{G}$ is an affine operator. Namely, there exists a bounded linear operator $\mathcal{A} : \mathcal{H} \to \mathbb{R}^n$ and a bounded element $b \in \mathcal{H}$, such that, the forward map $\mathcal{G} : \mathcal{H} \to \mathbb{R}^n$ can be represented by

$$\mathcal{G}(\cdot) = b - \mathcal{A}(\cdot) \tag{2.2.6}$$

(Note: a negative sign is used of the operator $\mathcal{A}$, because this is more convenient for us to keep the sign consistently in this thesis. Mathematically, it does not matter to use negative sign or positive sign. The most important thing is to keep the consistency of the signs.)

Theorem 2.2.1. *If the forward map $\mathcal{G} : \mathcal{H} \to \mathbb{R}^n$ can be represented by formula* (2.2.6)*, then the posterior probability measure $\mathbb{P}(\cdot|y)$ in formula* (2.2.5) *is well-defined and it is a Gaussian measure $\mathbb{P}(\cdot|y) = \mathcal{N}(m_1(y), \mathcal{C}_1)$, where the mean $m_1(y) \in \mathcal{H}$ and covariance*

$\mathcal{C}_1 : \mathcal{H} \to \mathcal{H}$ *are determined by*

$$m_1(y) = m_0 - \mathcal{C}_0\mathcal{A}^* \left(\gamma + \mathcal{A}\mathcal{C}_0\mathcal{A}^*\right)^{-1} \left(\mathcal{A}m_0 - b + y\right) \tag{2.2.7}$$

$$\mathcal{C}_1 = \mathcal{C}_0 - \mathcal{C}_0\mathcal{A}^* \left(\gamma + \mathcal{A}\mathcal{C}_0\mathcal{A}^*\right)^{-1} \mathcal{A}\mathcal{C}_0 \tag{2.2.8}$$

Proof. See Example 6.23 in [8]. □

Remark 2.2.2. *Since the posterior distribution* $\mathbb{P}(\cdot|y) = \mathcal{N}(m_1(y), \mathcal{C}_1)$ *is Gaussian, the posterior mean* $m_1(y)$ *is also the MAP estimator. Furthermore, the MAP estimator exactly equals to the minimum point of the Tikhonov regularization. Thus, formula* (2.2.7) *also provides the unique solution* $\widehat{x}(y) = m_1(y)$ *of the Tikhonov regularization* (2.2.4) *as long as the forward map* $\mathcal{G}$ *has a form of* (2.2.6)*.*

According to theorem 2.2.1, the solution has a closed-form for linear problems. The well-posedness of the solution is clear. Since the covariance matrix γ is positive-definite, the matrix inversion $\left(\gamma + \mathcal{A}\mathcal{C}_0\mathcal{A}^*\right)^{-1}$ exists and it is bounded. Thus the posterior mean $m_1(y)$ and covariance $\mathcal{C}_1$ exist and they are bounded. Furthermore, formula (2.2.7) shows that $m_1(y)$ is continuously depending on data y. As the result, for linear problems, both the variational inversion and the Bayesian inversion are always well-posed.

For nonlinear problems

For nonlinear problems, the situation is more challenging. In order to ensure the well-posedness of the inverse problems, some regularity properties of the forward map $\mathcal{G}$ are needed. See the following statement of the assumptions:

Assumption 2.2.3 (Assumption 2.7 in [8])**.** *Assume that the forward map* $\mathcal{G} : \mathcal{H} \to \mathbb{R}^n$ *satisfies the following two conditions:*

1. *For every* $\epsilon > 0$ *there is an* $M = M(\epsilon) \in \mathbb{R}$ *such that, for all* $u \in \mathcal{H}$*,*

$$\left\|\gamma^{-1/2}\mathcal{G}(u)\right\|_{\mathbb{R}^n} \leq \exp(\epsilon\|u\|_{\mathcal{H}}^2 + M) \tag{2.2.9}$$

2. *For every* $r > 0$ *there is a* $K = K(r) > 0$ *such that, for all* $u_1, u_2 \in \mathcal{H}$ *with* $\max\{\|u_1\|_{\mathcal{H}}, \|u_2\|_{\mathcal{H}}\} < r$*,*

$$\left\|\gamma^{-1/2}(\mathcal{G}(u_1) - \mathcal{G}(u_2))\right\|_{\mathbb{R}^n} \leq K\|u_1 - u_2\|_{\mathcal{H}} \tag{2.2.10}$$

Condition 1 in assumption 2.2.3 ensures that the cost functional $\left\|\gamma^{-1/2}(y-\mathcal{G}(\cdot))\right\|_{\mathbb{R}^n}$ has an exponential tail, so it is integrable with respect to Gaussian measures according to the Fernique's theorem. Condition 2 in assumption 2.2.3 ensures the Lipschitz continuity of the cost functional constrained on any bounded subsets.

With the two conditions in assumption 2.2.3, it is sufficient to show that the Bayesian inversion (2.2.5) is well-posed. See the following theorem.

Theorem 2.2.4. *If the forward map $\mathcal{G}:\mathcal{H}\to\mathbb{R}^n$ satisfies condition 1 and condition 2 in assumption 2.2.3, then the posterior probability measure $\mathbb{P}(\cdot|y)$ in formula* (2.2.5) *is well-defined. Furthermore, the posterior distribution is Lipschitz continuous in data y with respect to the Hellinger distance: for all $r>0$ there is a $C=C(r)>0$ such that, for all $y_1,y_2\in\mathbb{R}^n$ with $\max\left\{\left\|\gamma^{-1/2}y_1\right\|_{\mathbb{R}^n},\left\|\gamma^{-1/2}y_2\right\|_{\mathbb{R}^n}\right\}<r$,*

$$d_{Hell}\left(\mathbb{P}(\cdot|y_1),\mathbb{P}(\cdot|y_2)\right)\leq C\left\|\gamma^{-1/2}(y_1-y_2)\right\|_{\mathbb{R}^n} \tag{2.2.11}$$

Proof. See Theorem 4.1, Theorem 4.2, and Corollary 4.4 in [8]. □

The Hellinger distance mentioned in theorem 2.2.4 is defined as follows.

Definition 2.2.5 (Hellinger distance)**.** *The Hellinger distance of two probability measures ν and μ on a measurable space (X,Σ) is defined as,*

$$d_{Hell}(\nu,\mu)=\sqrt{\frac{1}{2}\int_X\left(\sqrt{\frac{\mathrm{d}\nu}{\mathrm{d}\lambda}}-\sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}}\right)^2\mathrm{d}\lambda} \tag{2.2.12}$$

where ν and μ are absolutely continuous with respect to a third probability measure λ. The definition is independent on choice of λ.

For the variational inversion (2.2.4), it is easy to show the existence of a minimum point, but it is much more difficult to show the uniqueness and Lipschitz continuity in the data. Nevertheless, a weaker statement holds. See the following theorem.

Theorem 2.2.6. *If the forward map $\mathcal{G}:\mathcal{H}\to\mathbb{R}^n$ satisfies condition 2 in assumption 2.2.3, then there exists a minimum point $\widehat{x}(y)$ in formula* (2.2.4) *and every minimum must belong to a compact subset, i.e.*

$$\left\|\mathcal{C}_0^{-1/2}(\widehat{x}(y)-m_0)\right\|_{\mathcal{H}}\leq\left\|\gamma^{-1/2}(y-\mathcal{G}(m_0))\right\|_{\mathbb{R}^n} \tag{2.2.13}$$

Furthermore, let $\{y_k\}$ *and* $\{x_k\}$ *be sequences where* $y_k \to y$ *and* $x_k \equiv \widehat{x}(y_k)$ *is a minimizer of* (2.2.4) *with* y *replaced by* y_k*. Then there exists a convergent subsequence of* $\{x_k\}$ *and the limit of every convergent subsequence is a minimizer of* (2.2.4)*.*

Proof. This proof consists of two parts. In the first part, we need to prove that, for any (bounded) data $y \in \mathbb{R}^n$, there exists a minimum $\widehat{x}(y)$ and the inequality (2.2.13) holds. For this part, please see theorem 4.1.4 in section 4.1. In the second part, we need to prove that, there exists a convergent subsequence to the minimizer. For this part, we use the result of the first part that, the sequence of minimums $\{x_k\}$ exists corresponding to the sequence of data $\{y_k\}$, and for any k, we have

$$\left\|\mathcal{C}_0^{-1/2}(x_k - m_0)\right\|_{\mathcal{H}} \leq \left\|\gamma^{-1/2}(y_k - \mathcal{G}(m_0))\right\|_{\mathbb{R}^n} \tag{2.2.14}$$

Since every data y_k is bounded, then we can define a constant c as

$$c := \sup_k \left\{\left\|\gamma^{-1/2}(y_k - \mathcal{G}(m_0))\right\|_{\mathbb{R}^n}\right\} \tag{2.2.15}$$

Therefore, every element in the sequence $\{x_k\}$ satisfies

$$\left\|\mathcal{C}_0^{-1/2}(x_k - m_0)\right\|_{\mathcal{H}} \leq c \tag{2.2.16}$$

that means, for any k, x_k is in the compact set $U := \left\{u \in \mathcal{H} : \left\|\mathcal{C}_0^{-1/2}(u - m_0)\right\|_{\mathcal{H}} \leq c\right\}$. Thus, the forward map $\mathcal{G} : \mathcal{H} \to \mathbb{R}^n$ can be constrained from the entire Hilbert space $\mathcal{H}$ to the compact subset U, which leads to $\mathcal{G} : U \to \mathbb{R}^n$. Since $\mathcal{G}$ is a continuous operator on the compact set, we can apply Theorem 1, Theorem 2 in [97] and Theorem 2.1 in [42] to show the existence of convergent subsequence. (Note: [97] considers forward maps between Banach spaces, and [42] considers forward maps between Hilbert spaces. We consider a special case with the forward map $\mathcal{G} : U \to \mathbb{R}^n$ from the compact set to the Euclidean space. In fact, the compact subset U is used instead of the entire Hilbert space $\mathcal{H}$ in order to satisfy the 'coresive' property discussed in [97].) □

2.3 Random fields for prior construction

To deal with an inverse problems, a prior probability measure needs to be predetermined. A typical example is using a Gaussian measure specified with mean and covariance. How-

ever, there exists some mathematical difficulties when consider infinite-dimensional random variables. It is not difficult to generalize the concept of random variables from the real line to an Euclidean space, but it seems not intuitive with an infinite-dimensional random variable. On a Euclidean space, probability distributions can be deduced by using the Lebesgue measure which leads to probability density functions, but this is not possible on infinite-dimensional spaces since there are not analogues of the Lebesgue measure. Nevertheless, Gaussian measures on infinite-dimensional spaces still exists. Moreover, given an infinite-dimensional random variable, it is desired to explicitly represent the random variable in a simple way. A techinique is know as the Karhunen-Loéve (KL) expansion [8, 85]. The core idea is using eigen-decomposition and basis transformation, such that, an infinite-dimensional random variable can be represented by a sum of uncorrelated one-dimensional random variables. In this section, we introduce random fields.

2.3.1 Square-integrable random fields on $\overline{D} \subset \mathbb{R}^d$

There are two equivalent ways describing random fields. The first way regards an $\mathbb{R}$-valued random field as a collection of $\mathbb{R}$-valued random variables, and another way regards a random field as an infinite-dimensional random variable in a function space.

In applied mathematics, the terminology 'random field' usually indicates the first way. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, where Ω is a set, $\mathcal{F}$ is a sigma-algebra over Ω, and $\mathbf{P} : \mathcal{F} \to [0, 1]$ is a probability measure on the measurable space $(\Omega, \mathcal{F})$. Let $D \subset \mathbb{R}^d$ be a bounded domain and $\overline{D}$ be the closure of D. Then a random field $u : \overline{D} \times \Omega \to \mathbb{R}$ is a collection of $\mathbb{R}$-valued random variables indexed over the set $\overline{D}$, such that, for each $r \in \overline{D}$, $u(r, \cdot) : \Omega \to \mathbb{R}$ is an $\mathbb{R}$-valued random variable. The real-valued random field u is called square-integrable if

$$\int_\Omega \int_D u(r, \omega)^2 \, \mathrm{d}r \, \mathbf{P}(\mathrm{d}\omega) < \infty \tag{2.3.1}$$

If u is square-integrable, then its mean $m : \overline{D} \to \mathbb{R}$ and autocovariance $C : \overline{D} \times \overline{D} \to \mathbb{R}$ are defined as the functions, such that, for almost every $r, s \in \overline{D}$,

$$m(r) = \int_\Omega u(r, \omega) \, \mathbf{P}(\mathrm{d}\omega) \tag{2.3.2}$$

$$C(r, s) = \int_\Omega (u(r, \omega) - m(r))(u(s, \omega) - m(s)) \, \mathbf{P}(\mathrm{d}\omega) \tag{2.3.3}$$

In probability theory, a random field can be described as an infinite-dimensional random variable with a probability measure over a function space (this thesis considers Hilbert spaces). Let $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ be a measurable space, where $\mathcal{H}$ is a separable Hilbert space, and $\mathcal{B}(\mathcal{H})$ is the Borel sigma-algebra over $\mathcal{H}$. Let $u : \Omega \to \mathcal{H}$ denote an $\mathcal{H}$-valued random variable (measurable function). Then, the pushforward measure of $\mathbf{P}$ is the probability measure $\mu : \mathcal{B}(\mathcal{H}) \to [0, 1]$ over the Hilbert space, such that, $\mu(S) = \mathbf{P}(\{\omega \in \Omega | u(\omega) \in S\})$ for any $S \in \mathcal{B}(\mathcal{H})$. The $\mathcal{H}$-valued random variable u is called square-integrable if

$$\int_{\mathcal{H}} \|x\|_{\mathcal{H}}^2 \, \mu(\mathrm{d}x) < \infty \tag{2.3.4}$$

If u is square-integrable, then the mean $m \in \mathcal{H}$ and the covariance operator $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ are defined as, for all bounded $v, w \in \mathcal{H}$,

$$\langle m, v \rangle_{\mathcal{H}} = \int_{\mathcal{H}} \langle x, v \rangle_{\mathcal{H}} \, \mu(\mathrm{d}x) \tag{2.3.5}$$

$$\langle v, \mathcal{C}w \rangle_{\mathcal{H}} = \int_{\mathcal{H}} \langle v, x - m \rangle_{\mathcal{H}} \langle x - m, w \rangle_{\mathcal{H}} \, \mu(\mathrm{d}x) \tag{2.3.6}$$

The two ways describing random fields are equivalent, since the random field $u : \overline{D} \times \Omega \to \mathbb{R}$ over $\overline{D}$ can be also treated as an infinite-dimensional random variable $u : \Omega \to L^2(\overline{D}; \mathbb{R})$ in the separable Hilbert space $L^2(\overline{D}; \mathbb{R})$, and the covariance operator $\mathcal{C} : L^2(\overline{D}; \mathbb{R}) \to L^2(\overline{D}; \mathbb{R})$ is determined by the autocovariance function $C : \overline{D} \times \overline{D} \to \mathbb{R}$ via integral transform, such that, for all bounded $w \in L^2(\overline{D}; \mathbb{R})$ and for almost every $r \in \overline{D}$,

$$[\mathcal{C}w](r) = \int_D C(r, s) w(s) \mathrm{d}s \tag{2.3.7}$$

Explicitly, given the mean m and the autocovariance C, the square-integrable random field u can be expressed by the KL expansion with mean-square convergence,

$$u = m + \sum_{j=1}^{\infty} \sqrt{\lambda_j} \xi_j \varphi_j \tag{2.3.8}$$

$$\int_D C(r, s) \varphi_j(s) \, \mathrm{d}s = \lambda_j \varphi_j(r) \tag{2.3.9}$$

$$\int_D \varphi_i(s) \varphi_j(s) \, \mathrm{d}s = \delta_{ij} \tag{2.3.10}$$

where $\{\lambda_i : i = 1, 2, ...\}$ and $\{\varphi_i : i = 1, 2, ...\}$ are the eigenvalues and eigenfunctions, and $\{\xi_i : i = 1, 2, ...\}$ are uncorrelated $\mathbb{R}$-valued standard random variables. If the random

filed u is Gaussian, then ξ_i are i.i.d. from $\mathcal{N}(0,1)$. If the autocovariance function C is continuous on $\overline{D}\times\overline{D}$, then the infinite sum (2.3.8) uniformly converges on $\overline{D}$, ensured by the Mercer's theorem for continuous symmetric positive-semi-definite kernels on compact sets. The property of uniform convergence does not hold for discontinuous autocovariance functions.

2.3.2 Gaussian random fields on $\overline{D} \subset \mathbb{R}^d$

The regularity properties (Hölder continuity and Sobolev embedding) of random fields should be considered. This thesis particularly focuses on Gaussian fields.

To show the regularity properties, we need make some assumptions: according to [8], a linear operator $\mathcal{L}$ is called 'Laplacian-like' on $L^2(\overline{D};\mathbb{R})$ if the following assumptions hold:

Assumption 2.3.1 (Assumption 2.9. in [8]). *$\mathcal{L}$ is a linear operator, densely defined on a Hilbert space $\mathcal{H} \subset L^2(\overline{D};\mathbb{R})$, that satisfies the following properties.*

1. *$\mathcal{L}$ is self-adjoint, positive-definite and invertible.*
2. *The eigenfunctions/eigenvalues $\{\varphi_k, \kappa_k\}$ of $\mathcal{L}$, indexed by $k \in \mathbb{K} \subset \mathbb{Z}^d \setminus \{0\}$, form an orthonormal basis for $\mathcal{H}$.*
3. *There exist $C^{\pm} > 0$ such that the eigenvalues satisfy, for all $k \in \mathbb{K}$,*
$$C^- \leq \frac{\kappa_k}{|k|^2} \leq C^+ \tag{2.3.11}$$
4. *There exists $C > 0$ such that*
$$\sup_{k\in\mathbb{K}} \left\{ \|\varphi_k\|_\infty + \frac{1}{|k|}\|\mathrm{D}\varphi_k\|_\infty \right\} \leq C \tag{2.3.12}$$

Proposition 2.3.2 (Lemma 6.25 [8]). *Let the operator $\mathcal{L}$ satisfies assumptions 1-4 in 2.3.1. Consider a Gaussian measure $\mu = \mathcal{N}(0,\mathcal{C})$ with $\mathcal{C} = \mathcal{L}^{-\alpha}$ with $\alpha > d/2$. Then a draw $u \sim \mu$ is almost surely s-Hölder continuous for all $0 < s < \min\{1, \alpha - d/2\}$, where the s-Hölder continuity means that there exists a constant C such that, for all $x, y \in \overline{D}$, the following condition holds almost surely,*
$$|u(x) - u(y)| < C\|x - y\|^s \tag{2.3.13}$$

Proposition 2.3.3 (Lemma 6.27 [8]). *Let the operator $\mathcal{L}$ satisfies assumptions 1-3 in 2.3.1. Consider a Gaussian measure $\mu = \mathcal{N}(0, \mathcal{C})$ with $\mathcal{C} = \mathcal{L}^{-\alpha}$ with $\alpha > d/2$. Then a draw $u \sim \mu$ is in $\mathcal{H}^s$ almost surely for all $s \in [0, \alpha - d/2)$, where $\mathcal{H}^s$ is the separable Hilbert space associated with the operator $\mathcal{L}$,*

$$\mathcal{H}^s = \{u \in \mathcal{H} : u^* \mathcal{L}^s u < \infty\} \tag{2.3.14}$$

Consider a centered Gaussian field on $\overline{D}$ with the covariance operator $\mathcal{C} = \mathcal{L}^{-\alpha}$, where $\mathcal{L}$ is the operator satisfying assumption 2.3.1. Then, for any $\alpha > d/2$, the draws from the Gaussian random filed are almost surely in $L^2(\overline{D}; \mathbb{R})$ (proposition 2.3.3), and the draws are almost surely Hölder continuous (proposition 2.3.2) so that they are almost surely in $C(\overline{D}; \mathbb{R})$. For any non-centered Gaussian field, if the mean m is in the Cameron-Martin space $E = \text{Ran}\left(\mathcal{C}^{1/2}\right)$, then the draws are also almost surely in $C(\overline{D}; \mathbb{R}) \subset L^2(\overline{D}; \mathbb{R})$ for any $\alpha > d/2$, because the non-centered Gaussian measure $\mathcal{N}(m, \mathcal{C})$ with $m \in E$ is equivalent to the centered Gaussian measure $\mathcal{N}(0, \mathcal{C})$, according to the Cameron-Martin theorem.

2.3.3 Whittle-Matérn random fields on $\overline{D} \subset \mathbb{R}^d$

A weak-stationary random field is a square-integrable random field with translation-invariant mean and autocovariance. More specifically, let $\mu \in \mathbb{R}$ and $\sigma > 0$ be the mean and standard deviation of the steady state, and let $\text{ACF} : \mathbb{R}^d \to \mathbb{R}$ be the autocorrelation function of the weak-stationary field. Then the mean $m : \overline{D} \to \mathbb{R}$ and autocovariance $C : \overline{D} \times \overline{D} \to \mathbb{R}$ of the random field is determined by, for all $r, s \in \overline{D}$,

$$m(r) = \mu \tag{2.3.15}$$

$$C(r, s) = \sigma^2 \text{ACF}(r - s) \tag{2.3.16}$$

Moreover, its covariance operator $\mathcal{C} : L^2(\overline{D}; \mathbb{R}) \to L^2(\overline{D}; \mathbb{R})$ is determined by, for all $w \in L^2(\overline{D}; \mathbb{R})$ and for almost every $r \in \overline{D}$,

$$[\mathcal{C}w](r) := \sigma^2 \int_D \text{ACF}(r - s) w(s) \mathrm{d}s \tag{2.3.17}$$

For a weak-stationary field, the most important thing is the autocorrelation function, which determines the behavior of the random field. A function $\text{ACF} : \mathbb{R}^d \to \mathbb{R}$ is an

autocorrelation function of a weak-stationary field if ACF can be represented by convolution $\text{ACF} = g * g$, where $g : \mathbb{R}^d \to \mathbb{R}$ is an even function with $\|g\|_2 = 1$ and $*$ is the convolution on $\mathbb{R}^d$. g is even which ensures that ACF is even, $\|g\|_2 = 1$ ensures the normalizing condition $\text{ACF}(0) = 1$, and the convolution property $\text{ACF} = g * g$ ensures the positive-semi-definiteness of the kernel ACF since its Fourier transform is squared and thus never negative.

There are many choices of autocorrelation functions in applications of machine learning and signal processing [16]. One of them, Whittle-Matérn fields form an important class of stationary Gaussian fields with specified autocorrelation functions,

$$\text{ACF}(\cdot) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \left\|L^{-1}(\cdot)\right\|\right)^{\nu} K_\nu \left(\sqrt{2\nu} \left\|L^{-1}(\cdot)\right\|\right) \tag{2.3.18}$$

where $\nu \in (0, +\infty]$ is the smoothness parameter, $L : \mathbb{R}^d \to \mathbb{R}^d$ is the length-scale parameter which is a lower triangular matrix with positive and bounded diagonal entries, Γ is the gamma function, and K_ν is the modified Bessel function of the second kind of order ν. The smoothness parameter determines different types of autocorrelation functions, for example $\nu = 0.5$ leads to the exponential kernel,

$$\text{ACF}(\cdot) = \exp\left(-\left\|L^{-1}(\cdot)\right\|\right) \tag{2.3.19}$$

and $\nu = +\infty$ leads to the Gaussian kernel,

$$\text{ACF}(\cdot) = \exp\left(-\frac{1}{2}\left\|L^{-1}(\cdot)\right\|^2\right) \tag{2.3.20}$$

The Fourier transform of the Whittle-Matérn autocorrelation function (2.3.18) has an analytic form,

$$\widehat{\text{ACF}}(\cdot) = \frac{\det(L)(2\pi/\nu)^{d/2}\Gamma(\nu + \frac{d}{2})}{\Gamma(\nu)} \left(1 + \frac{\|2\pi L^{\mathrm{T}}(\cdot)\|^2}{2\nu}\right)^{-(\nu+d/2)} \tag{2.3.21}$$

Formula (2.3.21) represents the spectrum, where the base $\kappa_k = 1 + \frac{\|2\pi L^{\mathrm{T}} k\|^2}{2\nu}$ with $k \in \mathbb{K} \subset \mathbb{Z}^d \setminus \{0\}$ is positive-definite and grows in square rate, and the power $\alpha = \nu + d/2$ with $\nu > 0$ is greater than $d/2$. Thus, the draws of the Whittle-Matérn field are almost surely continuous, as the consequence of proposition 2.3.2 and proposition 2.3.3.

Some examples of Whittle-Matérn fields are shown in figure 2.1, figure 2.2 and figure 2.3. Figure 2.1 presents the autocorrelation functions with respect to different smoothness

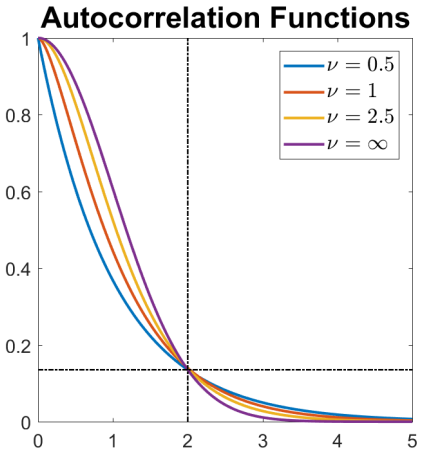


Figure 2.1: The Whittle-Matérn autocorrelation functions with different smoothness parameters.

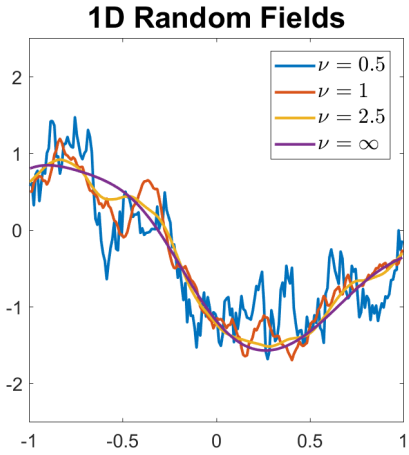


Figure 2.2: 1D examples of Whittle-Matérn fields on $[-1, 1]$ with the fixed length-scale parameter $L = 0.4$.

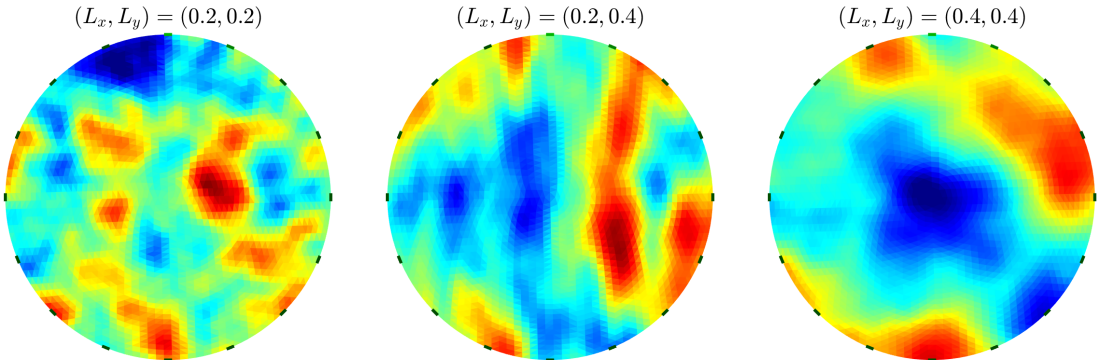


Figure 2.3: 2D examples of Whittle-Matérn fields on $\{s \in \mathbb{R}^2 : \|s\| < 1\}$ with the fixed smoothness parameter $\nu = 2$.

parameters. Figure 2.2 presents some draws from 1D Whittle-Matérn fields on $[-1, 1]$ with a fixed length-scale parameter and different smoothness parameters. Figure 2.3 presents some draws from 2D Whittle-Matérn fields on $\{s \in \mathbb{R}^2 : \|s\| < 1\}$ with a fixed smoothness parameter and different length-scale parameters.

2.3.4 Discretization of random fields on $\overline{D} \subset \mathbb{R}^d$

For numerical implementation, random fields should be discretized. We consider to use the KL expansion to numerically construct prior random fields. KL expansion of a random filed are shown in formulas (2.3.8) (2.3.9) and (2.3.10). We consider to use piece-wise

constant approximation and derive the matrix form after discretization.

Consider a discrete mesh on a Lipshitz domain $D \subset \mathbb{R}^d$. We use the following notation to describe the discrete mesh.

- Let M be the number of discrete elements.
- Let $D_i \subset \mathbb{R}^d$ be the domain of the ith element such that,
  1. for any $i = 1, ..., M$, D_i is a non-empty convex set;
  2. for any $i, j = 1, ..., M$, $i \neq j \implies D_i \cap D_j = \emptyset$;
  3. $\cup_{i=1}^{M} D_i = D$.
- Let $s_i \in \mathbb{R}^d$ be a point in the ith element such that,
  1. for any $i = 1, ..., M$, $s_i \in \overline{D}_i$, where $\overline{D}_i$ is the closure of D_i;
  2. for any $i, j = 1, ..., M$, $i \neq j \implies s_i \neq s_j$.
- Let $h_i = \|D_i\|$ be the volume of the ith element, where $\|\cdot\|$ denotes the Lebesgue measure on $\mathbb{R}^d$.
- For all $s \in D$, let $\mathbf{1}(s)$ be an M-dimensional column vector whose ith component is the indicator function $\mathbf{1}_i(s)$, where $\mathbf{1}_i(s)$ is defined as

$$\mathbf{1}_i(s) = \begin{cases} 1 & \text{if } s \in D_i \\ 0 & \text{if } s \in D \setminus D_i \end{cases} \tag{2.3.22}$$

With above notation, the functions in formulas (2.3.8) (2.3.9) and (2.3.10) are discretized and then represented by the matrix system,

- The random field u: $\forall s \in D$,

$$u(s) \approx \mathbf{1}^{\mathrm{T}}(s) \cdot \mathbf{u} \tag{2.3.23}$$

  where $\mathbf{u}$ is the discrete random vector, defined as an M-dimensional column vector such that, for any $i = 1, ..., M$, the entry is a random variable given by

$$\mathbf{u}_i = u(s_i) \tag{2.3.24}$$

- The mean m: $\forall s \in D$,

$$m(s) \approx \mathbf{1}^{\mathrm{T}}(s) \cdot \mathbf{m} \tag{2.3.25}$$

where $\mathbf{m}$ is the discrete mean vector, defined as an M-dimensional column vector such that, for any $i = 1, ..., M$, the entry is given by

$$\mathbf{m}_i = m(s_i) \tag{2.3.26}$$

- The autocovariance function C: $\forall r, s \in D$,

$$C(r, s) \approx \mathbf{1}^{\mathrm{T}}(r) \cdot \mathbf{C} \cdot \mathbf{1}(s) \tag{2.3.27}$$

where $\mathbf{C}$ is the discrete covariance matrix, defined as an $M \times M$ matrix such that, for any $i, j = 1, ..., M$, the entry is given by

$$\mathbf{C}_{ij} = C(s_i, s_j) \tag{2.3.28}$$

- The eigenfunction φ_j for any $j = 1, ..., M$: $\forall s \in D$,

$$\varphi_j(s) \approx \mathbf{1}^{\mathrm{T}}(s) \cdot \mathbf{\Phi}_{\cdot j} \tag{2.3.29}$$

where $\mathbf{\Phi}_{\cdot j}$ is the jth column of the discrete eigenfunction matrix $\mathbf{\Phi}$, and $\mathbf{\Phi}$ is defined as an $M \times M$ matrix such that, for any $i, j = 1, ..., M$, the entry is given by

$$\mathbf{\Phi}_{ij} = \varphi_j(s_i) \tag{2.3.30}$$

As the results, substitute the piece-wise constant approximations (2.3.23) (2.3.25) (2.3.27) and (2.3.29) into formulas (2.3.8) (2.3.9) and (2.3.10), which leads to

$$\mathbf{1}^{\mathrm{T}} \cdot \mathbf{u} = \mathbf{1}^{\mathrm{T}} \cdot \mathbf{m} + \sum_{j=1}^{M} \sqrt{\lambda_j} \xi_j \mathbf{1}^{\mathrm{T}} \cdot \mathbf{\Phi}_{\cdot j} \tag{2.3.31}$$

$$\int_D \mathbf{1}^{\mathrm{T}}(r) \cdot \mathbf{C} \cdot \mathbf{1}(s) \cdot \mathbf{1}^{\mathrm{T}}(s) \cdot \mathbf{\Phi}_{\cdot j} \, \mathrm{d}s = \lambda_i \mathbf{1}^{\mathrm{T}}(r) \cdot \mathbf{\Phi}_{\cdot j} \tag{2.3.32}$$

$$\int_D \mathbf{\Phi}_{\cdot i}^{\mathrm{T}} \cdot \mathbf{1}(s) \cdot \mathbf{1}^{\mathrm{T}}(s) \cdot \mathbf{\Phi}_{\cdot j} \, \mathrm{d}s = \delta_{ij} \tag{2.3.33}$$

Calculating and rearranging above formulas results in the following eigenvalue problem in matrix system (with $\mathbf{m}$, $\mathbf{C}$, $\mathbf{H}$, ξ as inputs and $\mathbf{\Lambda}$, $\mathbf{\Phi}$, $\mathbf{u}$ as outputs),

$$\mathbf{u} = \mathbf{m} + \mathbf{\Phi} \cdot \mathbf{\Lambda}^{1/2} \cdot \xi \tag{2.3.34}$$

$$\mathbf{C} \cdot \mathbf{H} \cdot \mathbf{\Phi} = \mathbf{\Phi} \cdot \mathbf{\Lambda} \tag{2.3.35}$$

$$\mathbf{\Phi}^{\mathrm{T}} \cdot \mathbf{H} \cdot \mathbf{\Phi} = \mathbf{I} \tag{2.3.36}$$

where ξ is an M-dimensional column vector whose entries are uncorrelated standard random variables, $\mathbf{\Lambda}$ is the discrete eigenvalue matrix that is an $M \times M$ diagonal matrix with $\{\lambda_i : i = 1, ..., M\}$ as the diagonal entries, and $\mathbf{H}$ is the discrete mesh size matrix that is an $M \times M$ diagonal matrix with $\{h_i : i = 1, ..., M\}$ as the diagonal entries.

2.4 Levenberg-Marquardt algorithm

Levenberg-Marquardt algorithm (LMA) is a modification of Gauss-Newton algorithm (GNA). LMA is also known as damped Gauss-Newton algorithm or damped least-squares method, as LMA is the Gauss-Newton using trust region approach with damping factors. Same as GNA, LMA aims to find a local optimum minimizing a sum of squares, though LMA is more robust than GNA because of the damping factors.

This section is a short introduction about LMA. First of all, we introduce the general form of LMA for minimization in L^2 norm. After that, we classify LMA by two types. The first type deals with regularized problems (over-determined, well-posed), and the second type deals with unregularized problems (under-determined, ill-posed). The essential difference between the two types are the stropping criteria: 1) for regularized problems, the LMA keeps iterations until the solution converges to a stationary point; 2) for unregularized problems, the solution (minimum) is not stable, so iterations must be stopped by a discrepancy principle before the algorithm loses stability.

2.4.1 LMA as an iterative method

This subsection introduces LMA as a standard and robust algorithm for minimization of sum of squares. Without loosing generality, consider the form of minimization in L^2 norm below,

$$\min_{x \in \mathcal{X}} \left\{ \frac{1}{2} \|H(x)\|_{\mathcal{Y}}^2 \right\} \tag{2.4.1}$$

where $H : \mathcal{X} \to \mathcal{Y}$ is a function from a separable Hilbert space $\mathcal{X}$ to another separable Hilbert space $\mathcal{Y}$. Assume that H is continuously differentiable. Let $\mathrm{D}H(x) : \mathcal{X} \to \mathcal{Y}$

denote the Fréchet derivative of H at $x \in \mathcal{X}$. Let $\widetilde{H}(x'; x)$ denote the first order expansion of H at $x \in \mathcal{X}$, such that for all x' in a neighborhood around x,

$$H(x') \approx \widetilde{H}(x'; x) := [\mathrm{D}H(u)](x' - x) + H(x) \tag{2.4.2}$$

GNA and LMA are iterative algorithms minimizing (2.4.1): pick an initial value $x_0 \in \mathcal{X}$, and for $k > 0$, x_k is determined by the following iterative algorithms:

- GNA: the basic method without damping,

$$x_k = \arg\min_{x \in \mathcal{X}} \left\{ \frac{1}{2} \left\| \widetilde{H}(x; x_{k-1}) \right\|_{\mathcal{Y}}^2 \right\} \tag{2.4.3}$$

  where $\widetilde{H}$ is the linear expansion (2.4.2) of H. Explicitly, the above formula has a closed form,

$$x_k = x_{k-1} - \left(\mathrm{D}H(x_{k-1})^* \mathrm{D}H(x_{k-1})\right)^{-1} \mathrm{D}H(x_{k-1})^* H(x_{k-1}) \tag{2.4.4}$$

  In general cases, GNA is not robust because of two facts. First of all, the matrix/operator inversion in the above formula could be almost singular. If $\mathrm{D}H(x_{k-1})^* \mathrm{D}H(x_{k-1})$ is singular or has a very large condition number, then it additionally requires singular value decomposition (SVD) [81]. In fact, by using SVD, the entire of expression $\left(\mathrm{D}H(x_{k-1})^* \mathrm{D}H(x_{k-1})\right)^{-1} \mathrm{D}H(x_{k-1})^*$ is replaced by the pseudoinverse of $\mathrm{D}H(x_{k-1})$. Moreover, if the objective function H is highly nonlinear, the GNA iteration $x_{k-1} \to x_k$ is unstable and sometimes performs badly, because the next point x_k may be too far from the last point x_{k-1} missing sensitive areas between the two points. If so, the GNA points $x_0, x_1, x_2, x_3...$ would be fluctuating and/or jagged, but not convergent.

- LMA: modified GNA with damping factors $\lambda_k > 0$,

$$x_k = \arg\min_{x \in \mathcal{X}} \left\{ \frac{1}{2} \left\| \widetilde{H}(x; x_{k-1}) \right\|_{\mathcal{Y}}^2 + \frac{\lambda_k}{2} \left\| x - x_{k-1} \right\|_{\mathcal{X}}^2 \right\} \tag{2.4.5}$$

  where $\widetilde{H}$ is the linear expansion (2.4.2) of H. Explicitly, the above formula has a closed form,

$$x_k = x_{k-1} - \left(\mathrm{D}H(x_{k-1})^* \mathrm{D}H(x_{k-1}) + \lambda_k \mathcal{I}\right)^{-1} \mathrm{D}H(x_{k-1})^* H(x_{k-1}) \tag{2.4.6}$$

If λ_k is large, LMA is like gradient descent with small step size; if λ_k is small, LMA is like Gauss-Newton. The dampping factor is initially chosen as a relatively large number (starts as gradient descent), and finally becomes to relatively small number (ends in Gauss-Newton). LMA predicts the next point x_k within a quadratic surface depending on the damping factor λ_k. The proposal x_k is accepted or rejected by comparing the actual reduction $ared_k = \frac{1}{2}\|H(x_{k-1})\|_{\mathcal{Y}}^2 - \frac{1}{2}\|H(x_k)\|_{\mathcal{Y}}^2$ relative to the predicted reduction $pred_k = \frac{1}{2}\|H(x_{k-1})\|_{\mathcal{Y}}^2 - \frac{1}{2}\|\widetilde{H}(x_k; x_{k-1})\|_{\mathcal{Y}}^2$ [22]. If $ared_k/pred_k < c$ ($c \geq 0$ is user-specified), then x_k is rejected and the damping factor should be increased for another trial.

The iterations must stop with a stop criterion. As we mentioned before, the stopping rules for regularized problems and unregularized problems are different. In the next subsection, we will explain the two kinds of situations associated with the corresponding stopping rules.

2.4.2 Apply LMA on regularized/unregularized problems

In the last subsection, we generally discussed an objective function $H : \mathcal{X} \to \mathcal{Y}$ between two separable Hilbert spaces. In this subsection, we make a slight restriction, such that, the observation space $\mathcal{Y}$ is finite-dimensional. Then, we consider a function $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ from a separable Hilbert space $\mathcal{H}$ to an Euclidean space $\mathbb{R}^n$, where $\mathcal{Z}$ is called the data-misfit function. Assume $\mathcal{Z}$ is continuously differentiable. Let $\mathrm{D}\mathcal{Z}(u) : \mathcal{H} \to \mathbb{R}^n$ denote the Fréchet derivative of $\mathcal{Z}$ at $u \in \mathcal{H}$. Let $\widetilde{\mathcal{Z}}(u'; u)$ denote the first order expansion of $\mathcal{Z}$ at $u \in \mathcal{H}$, such that for all u' in a neighborhood around u,

$$\mathcal{Z}(u') \approx \widetilde{\mathcal{Z}}(u'; u) := [\mathrm{D}\mathcal{Z}(u)](u' - u) + \mathcal{Z}(u) \tag{2.4.7}$$

Moreover, let $(m_0, \mathcal{C}_0)$ be a pair of parameters (for prior regularization), where $m_0 \in \mathcal{H}$ is a bounded element, and $\mathcal{C}_0 : \mathcal{H} \to \mathcal{H}$ is a self-adjoint positive-semi-definite trace-class operator.

For regularized problems

With Tikhonov regularization, the objective is to minimize

$$\min_{u\in\mathcal{H}}\left\{\frac{1}{2}\|\mathcal{Z}(u)\|_{\mathbb{R}^n}^2+\frac{1}{2}\|u-m_0\|_{\mathcal{C}_0}^2\right\} \tag{2.4.8}$$

The regularized minimization (2.4.8) is well-posed. Thus, we aim to search for a stationary point $\widetilde{x}$ such that the derivative of the objective functional is nearly zero,

$$\left\|\mathcal{C}_0^{-1/2}\left(\widetilde{x}-m_0\right)+\mathcal{C}_0^{1/2}\mathrm{D}\mathcal{Z}\left(\widetilde{x}\right)^*\mathcal{Z}\left(\widetilde{x}\right)\right\|_{\mathcal{H}}^2\leq\epsilon_0 \tag{2.4.9}$$

where $\epsilon_0>0$ is a user-specified parameter for accuracy control, e.g. $\epsilon_0=0.01$.

Apply the LMA to minimize formula (2.4.8): for $k=0$, let $v_0=m_0$ be the initial point; for $k>0$, let $\lambda_k>0$ be the kth damping factor, and then v_k is determined via the iteration below,

$$v_k=\arg\min_{u\in\mathcal{H}}\left\{\frac{1}{2}\left\|\widetilde{\mathcal{Z}}(u;v_{k-1})\right\|_{\mathbb{R}^n}^2+\frac{1}{2}\left\|\mathcal{C}_0^{-1/2}\left(u-m_0\right)\right\|_{\mathcal{H}}^2+\frac{\lambda_k}{2}\left\|\mathcal{C}_0^{-1/2}\left(u-v_{k-1}\right)\right\|_{\mathcal{H}}^2\right\} \tag{2.4.10}$$

where $\widetilde{\mathcal{Z}}$ is the linear expansion (2.4.7) of $\mathcal{Z}$. The above minimization has the unique solution v_k satisfying,

$$v_k=v_{k-1}-\alpha_k\left(\mathcal{C}_0\mathrm{D}\mathcal{Z}(v_{k-1})^*\widetilde{z}_k+v_{k-1}-m_0\right)\qquad\widetilde{z}_k=\widetilde{\mathcal{Z}}(v_k;v_{k-1}) \tag{2.4.11}$$

where $\alpha_k\equiv 1/(1+\lambda_k)$. The above implicit equations can be explicitly solved, and $\widetilde{z}_k$ equals to

$$\widetilde{z}_k=\left(\mathbf{I}+\alpha_k\mathrm{D}\mathcal{Z}(v_{k-1})\mathcal{C}_0\mathrm{D}\mathcal{Z}(v_{k-1})^*\right)^{-1}\left(\mathcal{Z}(v_{k-1})-\alpha_k\mathrm{D}\mathcal{Z}(v_{k-1})(v_{k-1}-m_0)\right) \tag{2.4.12}$$

The LMA iteration (2.4.11) should stop at the first time for some $K\geq 0$ when the value of derivative satisfies condition (2.4.9), i.e.,

$$\left\|\mathcal{C}_0^{-1/2}\left(v_K-m_0\right)+\mathcal{C}_0^{1/2}\mathrm{D}\mathcal{Z}\left(v_K\right)^*\mathcal{Z}\left(v_K\right)\right\|_{\mathcal{H}}^2\leq\epsilon_0 \tag{2.4.13}$$

Then, v_K is the final estimate for the regularized problem.

For unregularized problems

Without any prior regularization, the direct approach is to minimize the L^2 norm of the data-misfit,

$$\min_{u\in\mathcal{H}}\left\{\frac{1}{2}\|\mathcal{Z}(u)\|_{\mathbb{R}^n}^2\right\} \tag{2.4.14}$$

However, the unregularized minimization (2.4.14) is ill-posed. For this issue, we have to use other estimator instead of the minimum since the minimum estimator is unstable. The Morozov's discrepancy principle [98, 80, 96] suggests to find an estimator $\widetilde{x}$ such that,

$$\|\mathcal{Z}(\widetilde{x})\|_{\mathbb{R}^n} \leq \tau\delta \tag{2.4.15}$$

where $\tau > 1$ is an accuracy control parameter, and $\delta > 0$ is the given noise level satisfying

$$\|\mathcal{Z}(x)\|_{\mathbb{R}^n} \leq \delta \tag{2.4.16}$$

where $x \in \mathcal{H}$ is the true value of the unknown parameter.

The LMA provides regularization within iterations for the unregularized problem (2.4.14) such that: for $k = 0$, let $v_0 = m_0$ be the initial point; for $k > 0$, let $\lambda_k > 0$ be the kth damping factor, and then v_k is determined via the iteration below,

$$v_k = \arg\min_{u\in\mathcal{H}} \left\{ \frac{1}{2} \left\| \widetilde{\mathcal{Z}}(u; v_{k-1}) \right\|_{\mathbb{R}^n}^2 + \frac{\lambda_k}{2} \left\| \mathcal{C}_0^{-1/2} (u - v_{k-1}) \right\|_{\mathcal{H}}^2 \right\} \tag{2.4.17}$$

where $\widetilde{\mathcal{Z}}$ is the linear expansion (2.4.7) of $\mathcal{Z}$. The above minimization has the unique solution v_k satisfying,

$$v_k = v_{k-1} - \beta_k \mathcal{C}_0 \mathrm{D}\mathcal{Z}\left(v_{k-1}\right)^* \widetilde{z}_k \qquad \widetilde{z}_k = \widetilde{\mathcal{Z}}(v_k; v_{k-1}) \tag{2.4.18}$$

where $\beta_k \equiv 1/\lambda_k$. The above implicit equations can be explicitly solved, and $\widetilde{z}_k$ equals to

$$\widetilde{z}_k = \left(\mathbf{I} + \beta_k \mathrm{D}\mathcal{Z}(v_{k-1})\mathcal{C}_0 \mathrm{D}\mathcal{Z}(v_{k-1})^*\right)^{-1} \mathcal{Z}(v_{k-1}) \tag{2.4.19}$$

The LMA iteration (2.4.18) should stop at the first time for some $K \geq 0$ when the discrepancy principle (2.4.15) holds, i.e.,

$$\|\mathcal{Z}(v_K)\|_{\mathbb{R}^n} \leq \tau\delta \tag{2.4.20}$$

Then, v_K is the final estimate for the unregularized problem.

2.4.3 Determine the damping factors

By now, we have had the formulation of the LMA for regularized/unregularized problems. Furthermore, another question is how to determine the damping factors λ_k. In practice,

the damping factors are usually determined by trials (acceptance or rejection of proposals). Sometimes, there are theoretical approaches in particular cases. In the following two subsections, we will introduce two approaches. One is the trust region approach, which is a general procedure about acceptance or rejection of proposals. Another is the regularizing Levenberg-Marquardt scheme (RLMS), which requires some good properties of the data-misfit function.

The trust region approach

LMA is usually regarded as a trust region approach using quadratic approximation in each iteration. The radius of the trust region is equivalently characterized by the damping factor. In practice, the damping factor is usually determined by trials. Several damping strategies are suggested in references [28, 22, 73]. The main idea can be summarized as follows:

1. Propose an initial value (sufficiently large) of the damping factor $\lambda \leftarrow \lambda_0$, where $\lambda_0 > 0$ is user-specified.

2. Calculate the actual reduction $ared_k$ and the predicted reduction $pred_k$ in each iteration from x_{k-1} to x_k, where $k = 1, 2, ..., N$ is the number of iterations, and

$$ared_k = \frac{1}{2}\|H(x_{k-1})\|_{\mathcal{Y}}^2 - \frac{1}{2}\|H(x_k)\|_{\mathcal{Y}}^2 \tag{2.4.21}$$

$$pred_k = \frac{1}{2}\|H(x_{k-1})\|_{\mathcal{Y}}^2 - \frac{1}{2}\|\widetilde{H}(x_k; x_{k-1})\|_{\mathcal{Y}}^2 \tag{2.4.22}$$

3. Decide whether accept or reject the proposal, depending on the value of the ratio $\rho = ared_k/pred_k$. If $\rho \geq c$, the proposal is accepted, otherwise rejected, where $c \geq 0$ is user-specified.

4. Adjust the value of the damping factor for the next iteration, depending on the value of the ratio $\rho = ared_k/pred_k$ obtained in the current iteration. If $\rho \geq c_1$, $\lambda \leftarrow \omega_{down}\lambda$ is decreased for the next iteration; if $c_1 > \rho \geq c_0$, $\lambda \leftarrow \lambda$ remains the same for the next iteration; if $c_0 > \rho$, $\lambda \leftarrow \omega_{up}\lambda$ is increased for the next iteration, where $0 \leq c_0 \leq c_1$ and $0 < \omega_{down} < 1 < \omega_{up}$ are user-specified.

However, sometimes, it is not convenient to choose these algorithmic parameters, since the choices are highly depending on users' experience or numerical tuning. Even though the tuning parameters are suitable for one model, it does not mean they are suitable for other models. Nevertheless, there are some suggestions from references. The initial damping factor is usually large, which can be selected between $\sqrt{J(x_0)/n}$ and $J(x_0)/n$, applied in [63], where n is the number of observations and $J(\cdot) \equiv \frac{1}{2}\|H(\cdot)\|^2_{\mathcal{Y}}$ is the objective functional. The parameter $c = 10^{-4}$ or $c = 0.25$ is suggested in [22]. The multipliers $\omega_{up} = 10$, $\omega_{down} = 1/10$ was originally suggested by Marquardt [28]. Other literature [73] suggests $\omega_{up} = 2$, $\omega_{down} = 1/3$ for moderate size problems, and $\omega_{up} = 1.5$, $\omega_{down} = 1/5$ for larger problems. There is no a benchmark. How to choose the damping factors depends on practical trials.

The regularizing Levenberg-Marquardt scheme

Relevantly, there is an adaptive scheme determing damping factors for LMA. This scheme is called the regularizing Levenberg-Marquardt scheme (RLMS), proposed by Hanke [71]. The main idea of Hanke's method is that: under some assumptions, the proposal v_k determined via the LMA iteration (2.4.18) with damping factor λ_k determined via the RLMS is always a better estimate than v_{k-1}.

However, we should mention that, it is not easy to check the priority assumption of RLMS in practice, and sometimes the assumption does not hold. In this situation, RLMS may fail due to two facts: 1) there is no solution of the damping factor, and 2) even though the solution exists, RLMS may lose the property of better and convergent estimation. If the priority assumption of RLMS does not hold, we have to use the trust region approach (by trials). Nevertheless, once the priority assumption of RLMS holds, then RLMS works well. In the following is a brief introduction of the RLMS, on the condition that the priority assumption holds.

The RLMS has a priority assumption (which is formula (2.1) in Hanke's paper [71]) such that, the estimates $\{v_j\}$ produced by the LMA in all iterations satisfy, for all $j = 0, 1, 2, ..., K$, where K is the number of iterations,

$$\left\| \widetilde{\mathcal{Z}}\left(x; v_j\right) \right\|_{\mathbb{R}^n} \leq \frac{\rho}{\tau} \left\| \mathcal{Z}(v_j) \right\|_{\mathbb{R}^n} \tag{2.4.23}$$

where $\widetilde{\mathcal{Z}}$ is the linearization of $\mathcal{Z}$ (2.4.7), x is the true value of the unknown parameter, $\{v_j : j = 0, 1, ..., K\}$ are the LMA points determined via the iterative formula (2.4.18), and $0 < \rho < 1 < \tau$ are fixed parameters. Then, RLMS determines the damping factor λ_k in the kth iteration by making the following equation hold

$$\|\widetilde{z}_k(\lambda_k)\|_{\mathbb{R}^n} = \rho \|\mathcal{Z}(v_{k-1})\|_{\mathbb{R}^n} \tag{2.4.24}$$

where $\widetilde{z}_k$ relying on λ_k is determined in formula (2.4.19). Hanke has proved that v_k in formula (2.4.18) with the dampping factor λ_k determined by the RLMS (2.4.24) is a better estimate of x than v_{k-1} (proposition 2.1 in [71]) based on the priority assumption (2.4.23).

Furthermore, a suitable stop criterion has to be supplied. As discussed by Hanke, "*for the present version of the Levenberg-Marquardt iteration the discrepancy principle is an appropriate stopping rule for this purpose*". Namely, the stopping rule (2.4.20) is adopted. In order to ensure convergence and stability of the algorithm (theorem 2.3 in [71]), the following conditions should hold:

1. The parameter τ in the stop rule (2.4.20) should satisfy $\tau > 1/\rho$;

2. The function $\mathcal{Z}$ should be locally bounded;

3. There exists a constant $C > 0$ such that, the Taylor remainder of $\mathcal{Z}$ is bounded by

$$\left\| \mathcal{Z}(u') - \widetilde{\mathcal{Z}}(u'; u) \right\|_{\mathbb{R}^n} \leq C \|u' - u\|_{\mathcal{C}_0} \|\mathcal{Z}(u') - \mathcal{Z}(u)\|_{\mathbb{R}^n} \tag{2.4.25}$$

Some numerical success of RLMS has been shown by Hanke [71], where the Darcy flow model (an elliptic PDE) is used as the testing model. This is a very special application, since the forward model governed by the elliptic PDE has good convex properties. For more general use, there are some concerns of the RLMS, because the priority assumption (2.4.23) may not hold. From our experience, (2.4.23) practically holds if the forward model has good convex properties, and if the LMA initial point is chosen carefully. However, the priority assumption seems not hold for highly nonlinear functions.

2.5 Markov chain Monte Carlo methods

Markov chain Monte Carlo (MCMC) methods are the most popular numerical sampling algorithms for Bayesian inference. In this thesis, we will firstly introduce the ergodic theorem which shows that the time average equals to the spatial average. This property hence leads to the construction of ergodic Markov chains with an invariant measure, such that, the Markov chains convergent to the steady state. After that, we will introduce the fundamental concepts of the Metropolis-Hastings algorithm, which provides the core idea of MCMC methods about rejection sampling of Markov transitions. Finally, we will introduce a class of Langevin MCMC methods, and PCN-MCMC is the simplest example. It will be emphasized that the Langevin MCMC methods are suitable for functional settings (infinite dimensions) whereas the vallina random-walk MCMC methods collapse as the number of dimensions goes to infinity [94].

2.5.1 The ergodic theorem

Let (V, Σ, μ, T) be a measure-preserving dynamical system, i.e.

1. V is a non-empty set.
2. Σ is a σ-algebra over V.
3. $\mu : \Sigma \to [0, 1]$ is a probability measure on the measurable space (V, Σ).
4. $T : V \to V$ is a measurable transformation preserving the measure μ, i.e. $\mu(T^{-1}(S)) = \mu(S)$ for all $S \in \Sigma$.

Theorem 2.5.1 (Ergodic theorem [83]). *If T is ergodic, i.e. given $S \in \Sigma$, $T^{-1}(S) = S$ implies either $\mu(S) = 0$ or $\mu(S) = 1$, then for all $f \in L^1(V, \mu; \mathbb{R})$ and for almost every $v \in V$, the following equation holds*

$$\lim_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} f\left(T^k v\right) = \int_V f(\xi)\, \mu(\mathrm{d}\xi) \tag{2.5.1}$$

Ergodic theorem tells that the expected value of a random variable under a probability measure can be equivalently calculated by a chain of samples under any measure-preserving transformation. This is the core idea to develop Markov chain Monte Carlo methods.

2.5.2 Markov chain with invariant measure

Let $K : V \times \Sigma \to [0, 1]$ be a Markov kernel, i.e.

1. The map $v \mapsto K(v, S)$ is measurable for all $S \in \Sigma$.

2. The map $S \mapsto K(v, S)$ is a probability measure for all $v \in V$.

The ergodic theorem tells that in order to draw samples from measure μ, it is possible to apply the indirect way that constructs an ergodic Markov chain with an invariant transition kernel K preserving measure μ, i.e.

$$\mu(\mathrm{d}w) = \int_V K(v, \mathrm{d}w)\, \mu(\mathrm{d}v) \tag{2.5.2}$$

Above formula is called the *balance equation* in [49]. Sometimes, a stronger but easier condition is adopted instead of formula (2.5.2), that is

$$K(w, \mathrm{d}v)\mu(\mathrm{d}w) = K(v, \mathrm{d}w)\mu(\mathrm{d}v) \tag{2.5.3}$$

Notice that formula (2.5.2) is implied by integrating both sides of formula (2.5.3) with respect to the dummy variable v, but the converse is not true. Formula (2.5.3) is called the *detailed balance equation* in [49].

2.5.3 The Metropolis-Hastings algorithm

As discussed in the last subsection, the key point of MCMC methods is to construct a measure-preserving transition kernel K of a Markov chain with respect to the target measure μ. The Metropolis-Hastings algorithm [49] is a reject sampling method that constructs the transition kernel by

$$K(v, \mathrm{d}w) = \alpha(v, w)q(v, \mathrm{d}w) \tag{2.5.4}$$

where $q : V \times \Sigma \to [0, 1]$ is a user-proposed transition kernel, and α is the acceptance rate of the proposal,

$$\alpha(v, w) := \min\left\{1, \frac{q(w, \mathrm{d}v)\mu(\mathrm{d}w)}{q(v, \mathrm{d}w)\mu(\mathrm{d}v)}\right\} \tag{2.5.5}$$

Let η denote the measure $\eta(\mathrm{d}v, \mathrm{d}w) = q(w, \mathrm{d}v)\mu(\mathrm{d}w)$, and let $\eta^{\perp}$ denote the measure by reversing the roles of v and w, i.e. $\eta^{\perp}(\mathrm{d}v, \mathrm{d}w) = q(v, \mathrm{d}w)\mu(\mathrm{d}v)$. If η and $\eta^{\perp}$ are

equivalent, then the Radon-Nikodym theorem leads to a well-defined α. Otherwise, the acceptance rate is only $\alpha = 0$, that means all proposals are rejected and the Markov chain is draped in sticky points, so the chain is not ergodic and the algorithm fails.

If α is well-posed (the Radon-Nikodym derivatives exist), then it is easy to check that equation (2.5.4) associated with equation (2.5.5) satisfy the detailed balance equation (2.5.3), shown below.

1. Assume that $\frac{q(w,\mathrm{d}v)\mu(\mathrm{d}w)}{q(v,\mathrm{d}w)\mu(\mathrm{d}v)} \geq 1$. Then $\alpha(v,w)$ and $\alpha(w,v)$ are given by

$$\alpha(v,w) := \min\left\{1, \frac{q(w,\mathrm{d}v)\mu(\mathrm{d}w)}{q(v,\mathrm{d}w)\mu(\mathrm{d}v)}\right\} = 1 \tag{2.5.6}$$

$$\alpha(w,v) := \min\left\{1, \frac{q(v,\mathrm{d}w)\mu(\mathrm{d}v)}{q(w,\mathrm{d}v)\mu(\mathrm{d}w)}\right\} = \frac{q(v,\mathrm{d}w)\mu(\mathrm{d}v)}{q(w,\mathrm{d}v)\mu(\mathrm{d}w)} \tag{2.5.7}$$

Consequently, $K(v,\mathrm{d}w)\mu(\mathrm{d}v)$ and $K(w,\mathrm{d}v)\mu(\mathrm{d}w)$ are given by

$$K(v,\mathrm{d}w)\mu(\mathrm{d}v) = \alpha(v,w)q(v,\mathrm{d}w)\mu(\mathrm{d}v) = q(v,\mathrm{d}w)\mu(\mathrm{d}v) \tag{2.5.8}$$

$$K(w,\mathrm{d}v)\mu(\mathrm{d}w) = \alpha(w,v)q(w,\mathrm{d}v)\mu(\mathrm{d}w) = \frac{q(v,\mathrm{d}w)\mu(\mathrm{d}v)}{q(w,\mathrm{d}v)\mu(\mathrm{d}w)}q(w,\mathrm{d}v)\mu(\mathrm{d}w) = q(v,\mathrm{d}w)\mu(\mathrm{d}v) \tag{2.5.9}$$

Thus, we have

$$K(v,\mathrm{d}w)\mu(\mathrm{d}v) = K(w,\mathrm{d}v)\mu(\mathrm{d}w) \tag{2.5.10}$$

2. Assume that $\frac{q(w,\mathrm{d}v)\mu(\mathrm{d}w)}{q(v,\mathrm{d}w)\mu(\mathrm{d}v)} < 1$. Then $\alpha(v,w)$ and $\alpha(w,v)$ are given by

$$\alpha(v,w) := \min\left\{1, \frac{q(w,\mathrm{d}v)\mu(\mathrm{d}w)}{q(v,\mathrm{d}w)\mu(\mathrm{d}v)}\right\} = \frac{q(w,\mathrm{d}v)\mu(\mathrm{d}w)}{q(v,\mathrm{d}w)\mu(\mathrm{d}v)} \tag{2.5.11}$$

$$\alpha(w,v) := \min\left\{1, \frac{q(v,\mathrm{d}w)\mu(\mathrm{d}v)}{q(w,\mathrm{d}v)\mu(\mathrm{d}w)}\right\} = 1 \tag{2.5.12}$$

Consequently, $K(v,\mathrm{d}w)\mu(\mathrm{d}v)$ and $K(w,\mathrm{d}v)\mu(\mathrm{d}w)$ are given by

$$K(v,\mathrm{d}w)\mu(\mathrm{d}v) = \alpha(v,w)q(v,\mathrm{d}w)\mu(\mathrm{d}v) = \frac{q(w,\mathrm{d}v)\mu(\mathrm{d}w)}{q(v,\mathrm{d}w)\mu(\mathrm{d}v)}q(v,\mathrm{d}w)\mu(\mathrm{d}v) = q(w,\mathrm{d}v)\mu(\mathrm{d}w) \tag{2.5.13}$$

$$K(w,\mathrm{d}v)\mu(\mathrm{d}w) = \alpha(w,v)q(w,\mathrm{d}v)\mu(\mathrm{d}w) = q(w,\mathrm{d}v)\mu(\mathrm{d}w) \tag{2.5.14}$$

Thus, we have

$$K(v,\mathrm{d}w)\mu(\mathrm{d}v) = K(w,\mathrm{d}v)\mu(\mathrm{d}w) \tag{2.5.15}$$

2.5.4 The Metropolis-adjusted Langevin algorithm

In MCMC methods, there is a particular class of algorithms using the Langevin diffusion as proposals [9, 79]. Langevin MCMC methods are placed in function spaces so they are suitable for infinite-dimensional problems, i.e. the Langevin proposals ensure the well-posedness of formula (2.5.5). We will introduce a Langevin proposal conditioned to Gaussian prior. The clue of this subsection is: Itô diffusion $\rightarrow$ Langevin diffusion $\rightarrow$ Langevin diffusion with invariant Gaussian measure.

Consider an Itô diffusion $\{X_t : t \geq 0\}$ on V (for simplicity, we only formulate the diffusion in finite-dimensional case $V = \mathbb{R}^M$),

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t \tag{2.5.16}$$

where $b : V \rightarrow V$ and $\sigma : V \rightarrow \{V \rightarrow V\}$ are Lipschitz continuous functions, and W_t is the standard Wiener process on V. $b(X_t)$ is known as the drift coefficient and $\sigma(X_t)$ is known as the diffusion coefficient. In fact, an Itô diffusion is a special type of Itô process as the drift and diffusion coefficients only rely on X_t. What's more, the probability density function $\pi(t, \cdot)$ of the Itô diffusion X_t is governed by the Fokker-Planck equation, which is a partial differential equation related to the drift field b and the diffusion field σ,

$$\frac{\partial}{\partial t}\pi(t,x) = -\sum_{i=1}^{M}\frac{\partial}{\partial x_i}\left[b_i(x)\pi(t,x)\right] + \sum_{i,j=1}^{M}\frac{\partial^2}{\partial x_i \partial x_j}\left[K_{ij}(x)\pi(t,x)\right] \tag{2.5.17}$$

where $2K(x) = \sqrt{2K(x)}\sqrt{2K(x)}^{\mathrm{T}} = \sigma(x)\sigma^T(x)$.

For a stationary diffusion process X_t, the probability density function $\pi(t, \cdot) = \pi(\cdot)$ is time-invariant. Furthermore, if the stationary process X_t has invariant diffusion coefficient $\sigma(x) = \sigma$. Then, the Fokker-Planck equation (2.5.17) becomes to a simple form,

$$0 = -\sum_{i=1}^{M}\frac{\partial}{\partial x_i}\left[b_i(x)\pi(x)\right] + \sum_{i,j=1}^{M}\frac{\partial^2}{\partial x_i \partial x_j}\left[K_{ij}\pi(x)\right] \tag{2.5.18}$$

This directly leads to the relation below,

$$b(\cdot) = K\nabla\log(\pi(\cdot)) \tag{2.5.19}$$

Substituting above formula into the Itô diffusion (2.5.16) results in the Langevin diffusion below,

$$\mathrm{d}X_t = K\nabla\log(\pi(X_t))\mathrm{d}t + \sqrt{2K}\mathrm{d}W_t \tag{2.5.20}$$

The Langevin diffusion is a stationary Itô diffusion with an invariant diffusion coefficient.

For a Langevin diffusion with Gaussian distribution $\pi \sim \mathcal{N}(m_0, \mathcal{C}_0)$, formula (2.5.20) becomes to

$$\mathrm{d}X_t = -K\mathcal{C}_0^{-1}(X_t - m_0)\mathrm{d}t + \sqrt{2K}\mathrm{d}W_t \tag{2.5.21}$$

The critical issue is how to pick the preconditioner K. For example, the simplest way is to let $K = \mathcal{I}$ be the identity operator (without preconditioning). However, the identity operator in an infinite-dimensional space is not compact, and $\mathcal{C}_0^{-1}$ and W_t are also unbounded, which causes the ill-posedness of the unconditioned setting. On the other hand, $K = \mathcal{C}_0$ results in a well-posed diffusion with finite and infinite dimensions,

$$\mathrm{d}X_t = -(X_t - m_0)\mathrm{d}t + \sqrt{2\mathcal{C}_0}\mathrm{d}W_t \tag{2.5.22}$$

As the result, the proposal $q : V \times \Sigma \to [0, 1]$ in the Metropolis-Hastings algorithm (2.5.4) (2.5.5) can be generated by discretizing the preconditioned Langevin diffusion (2.5.22). More details are discussed in the following subsection.

2.5.5 PCN-MCMC

Now, consider the Metropolis-Hastings algorithm (2.5.4) associated with the acceptance rate (2.5.5). In that formula, μ is the target measure. In Bayesian inference, it is the posterior probability measure. According to the Bayes' formula, the posterior probability measure can be written as

$$\mu(\mathrm{d}u) \propto \exp(-\Phi(u))\mu_0(\mathrm{d}u) \tag{2.5.23}$$

where Φ is the cost functional (a real-valued non-negative functional), and μ_0 is the prior measure. If the prior is a Gaussian measure $\mu_0 = \mathcal{N}(m_0, \mathcal{C}_0)$, then a proposal $q(v, \mathrm{d}w)$ of the Markov transition can be constructed by discretizing formula (2.5.22). Consider the $\theta \in [0, 1]$ scale for the discretization of (2.5.22),

$$w = v - ((1 - \theta)v + \theta w - m_0)\,\delta + \sqrt{2\delta}\zeta \tag{2.5.24}$$

where $\delta > 0$ is the step size and $\zeta \sim \mathcal{N}(0, \mathcal{C}_0)$ is a Gaussian random variable. Above formula determines the transition kernel $q(v, \mathrm{d}w)$ with respect to different values of θ.

However, as discussed in [94], only the Crank-Nicolson scale ($\theta = 1/2$) makes the Radon-Nikodym derivative in formula (2.5.5) well-defined. Therefore, with $\theta = 1/2$, formula (2.5.24) becomes to

$$w = m_0 + \sqrt{1-\beta^2}(v - m_0) + \beta\zeta \tag{2.5.25}$$

where $\beta = \frac{\sqrt{2\delta}}{1+\delta/2}$. Above formula determines a well-posed proposal $q(v, \mathrm{d}w)$, associated with the acceptance rate,

$$\alpha(v, w) = \min\{1, \exp(\Phi(v) - \Phi(w))\} \tag{2.5.26}$$

This algorithm (2.5.25) (2.5.26) is a Metropolis-adjusted Langevin algorithm conditioned to Gaussian prior using Crank-Nicolson scale. The developers [94] of this algorithm call it the **PCN**-MCMC (**p**reconditioned **C**rank-**N**icolson) method. Furthermore, it is easy to check that the proposal $q(v, \mathrm{d}w)$ generated by (2.5.25) is a prior-reversible transformation, i.e.

$$q(v, \mathrm{d}w)\mu_0(\mathrm{d}v) = q(w, \mathrm{d}v)\mu_0(\mathrm{d}w) \tag{2.5.27}$$

In comparison, the vanilla random walk MCMC method is given by,

$$w = v + \beta\zeta \tag{2.5.28}$$

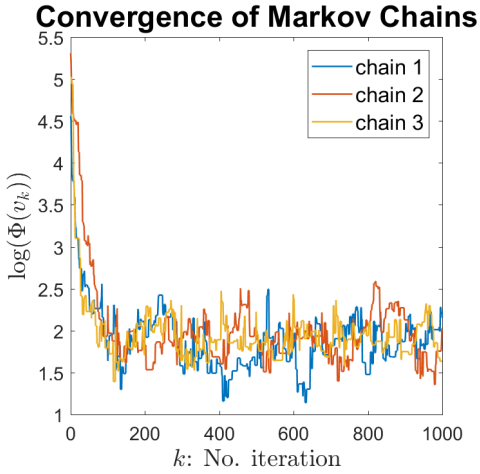
associated with the acceptance rate,

$$\alpha(v, w) = \min\left\{1, \exp\left(\Phi(v) - \Phi(w) + \frac{1}{2}\|v - m_0\|^2_{\mathcal{C}_0} - \frac{1}{2}\|w - m_0\|^2_{\mathcal{C}_0}\right)\right\} \tag{2.5.29}$$

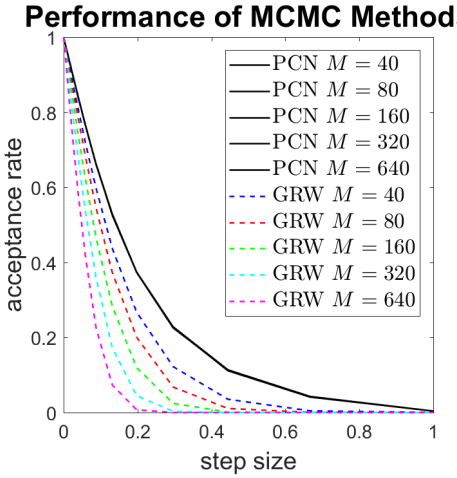
The vanilla random walk MCMC method is well-defined and workable only for finite dimensions. In function spaces, it eventually collapses in mesh refinement, because the random walk proposal leads to singularity of probability measures and the acceptance rate (2.5.29) equals to zero. In contrast, the PCN-MCMC is mesh-invariant and suitable to infinite-dimensional sampling. See figure 2.4b, where we try both Gaussian random walk MCMC (GRW-MCMC) and PCN-MCMC. It is clear that, as discrete points go to large, GRW-MCMC eventually collapses, whereas PCN-MCMC consistently works well.

2.6 A toy example for infinite-dimensional inversion

This section is to demonstrate numerical performance of optimization algorithms (for variational inversion) and sampling algorithms (for Bayesian inversion). A toy examples



(a) Convergence of Markov chains (monitoring the cost functional) from the prior distribution to the posterior distribution (using PCN-MCMC method with step size $\beta = 0.25$) within $T_0 = 1000$ iterations.



(b) The relationship between step sizes and acceptance rates of the two MCMC methods with different number M of finite elements. $T_0 = 1000$ burn-in period using PCN-MCMC with $\beta = 0.25$ has been excluded.

Figure 2.4: Compare GRW-MCMC and PCN-MCMC. The forward map is the 1D Darcy flow model (2.6.5). *The number of observations equals to $n = 10$. The noise level is $\epsilon = 0.02$. The prior is a Whittle-Matérn field on $[0, 1]$ with parameters: steady mean $\mu = -3.75$, steady standard deviation $\sigma = 0.1$, smoothness parameter $\nu = 2$ and the length-scale parameter $L = 0.15$.*

is in consideration: the 1D Darcy flow model. This example is very easy to understand, which can help readers who are not familiar with infinite-dimensional inverse problems. More complicated numerical applications can be found in chapter 5.

2.6.1 The 1D Darcy flow model

The (source-free) source-free 1D Darcy flow model for porous media comes from geophysics and fluid mechanics, which is the simplest case of the Navier-Stokes equations. It is an ODE problem formulated by

$$\left(\frac{\kappa(s)}{\nu}p'(s)\right)' = 0 \quad \forall s \in (0,1) \tag{2.6.1}$$

$$p(0) = 0 \tag{2.6.2}$$

$$\frac{\kappa(1)}{\nu}p'(1) = 1 \tag{2.6.3}$$

where $\kappa : [0,1] \to (0,+\infty)$ is the permeability of the porous media, and $\nu > 0$ is the viscosity of the fluid (e.g. $\nu = 1$ for water around 20°C), and $p : [0,1] \to \mathbb{R}$ is the fluid pressure. More clearly, the ODE problem (2.6.1)-(2.6.3) with $\nu = 1$ has the explicit solution,

$$p(s) = \int_0^s e^{-u(r)}\,\mathrm{d}r \tag{2.6.4}$$

where $u(r) = \log(\kappa(r))$ is the log permeability.

The forward problem is: given the log permeability, to calculate fluid pressures at some specified points $0 < s_1 \leq \cdots \leq s_n \leq 1$. The forward map $\mathcal{G} : L^2([0,1];\mathbb{R}) \to \mathbb{R}^n$ can be expressed as

$$\mathcal{G}(u) = [p(s_1), ..., p(s_n)]^{\mathrm{T}} = \left[\int_0^{s_1} e^{-u(r)}\,\mathrm{d}r, ..., \int_0^{s_n} e^{-u(r)}\,\mathrm{d}r\right]^{\mathrm{T}} \tag{2.6.5}$$

Conversely, the inverse problem is to recover the log permeability, given (possibly noisy) measurements of fluid pressures at the positions $0 < s_1 \leq \cdots \leq s_n \leq 1$. Both the variational approach (with Gauss-Newton algorithm) and the Bayesian approach (with PCN-MCMC method) will be applied to solve this inverse problem.

2.6.2 The 'truth' and the prior information

Let $\mathcal{N}(m_0, \mathcal{C}_0)$ be a Gaussian measure on the separable Hilbert space $L^2([0,1];\mathbb{R})$. More specifically, we consider the Whittle-Matérn class, i.e. the mean $m_0 \in L^2([0,1];\mathbb{R})$ and the covariance $\mathcal{C}_0 : L^2([0,1];\mathbb{R}) \to L^2([0,1];\mathbb{R})$ are specified: for all $s \in [0,1]$,

$$m_0(s) = \mu \tag{2.6.6}$$

and for all $h \in L^2([0,1];\mathbb{R})$, for almost every $s \in [0,1]$,

$$[\mathcal{C}_0 h](s) = \sigma \frac{2^{1-\nu}}{\Gamma(\nu)} \int_0^1 \left(\sqrt{2\nu} L^{-1}|s-t|\right)^{\nu} K_\nu \left(\sqrt{2\nu} L^{-1}|s-t|\right) h(t)\,\mathrm{d}t \tag{2.6.7}$$

where Γ is the gamma function, and K_ν is the modified Bessel function of the second kind of order ν. We fix the values of parameters: steady mean $\mu = -3.75$, steady standard deviation $\sigma = 0.5$, smoothness parameter $\nu = 2$ and the length-scale parameter $L = 0.15$.

We draw a sample from the Whittle-Matérn field, $x \sim \mathcal{N}(m_0, \mathcal{C}_0)$, and then we fix the parameter x as the true value of the log permeability. Moreover, we adopt $\mathcal{N}(m_0, \mathcal{C}_0)$ as the prior distribution for the inverse problem. This prior distribution is the natural choice as we know the truth x is 'correctly' characterized by the Gaussian measure $\mathcal{N}(m_0, \mathcal{C}_0)$, though any other prior distributions can be applied also as long as the prior distributions properly characterize the truth.

2.6.3 Experimental 'data'

Consider the forward map $\mathcal{G}$ (2.6.5). We specify the measurement positions $0 < s_1 \leq \cdots \leq s_n \leq 1$ as $s_i = \frac{i}{n}$ for all $i = 1, ..., n$, and we fix the number of measurements equal to $n = 10$. Once the truth $x \in C([0,1];\mathbb{R})$ is obtained (drawn from the Whittle-Matérn field and fixed), a clean data $y_{clean}\mathbb{R}^n$ can be simulated by formula

$$y_{clean} = \mathcal{G}(x) \tag{2.6.8}$$

Additionally, a noisy data $y \in \mathbb{R}^n$ can be simulated by formula

$$y = y_{clean} + e \tag{2.6.9}$$

where $e \in \mathbb{R}^n$ is a sample drawn from a centered non-degenerate Gaussian measure $\mathcal{N}(\mathbf{0}, \gamma)$, where the covariance γ is an $n \times n$ diagonal matrix with entries proportional to

the clean data y_{clean}, i.e.

$$\sqrt{\gamma} = \text{diag}(\epsilon \cdot \text{abs}(y_{clean})) \tag{2.6.10}$$

where ϵ will be specified as small noise level $\epsilon = 0.01$ or large noise level $\epsilon = 0.05$ in numerical experiments.

2.6.4 Variational inversion with the Gauss-Newton algorithm

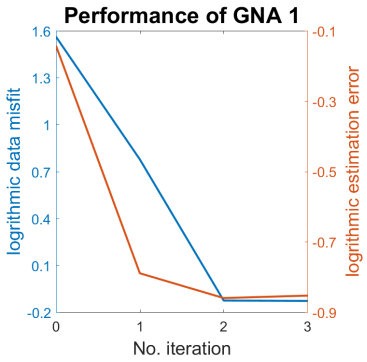
The variational inversion aims to minimize the Tikhonov regularization (2.2.4). The 1D Darcy flow model has the forward map $\mathcal{G}$ specified in formula (2.6.5), which is continuously twice differentiable with positive-definite Hessian. Since this is a convex optimization, we consider to apply Gauss-Newton algorithm (GNA). The LMA minimizing the Tikhonov regularization (2.2.4) has been shown in formula (2.4.11). The GNA is nothing more than taking $\alpha_k = 1$ in formula (2.4.11). The algorithm stops when the derivative is close to zero, as shown in formula (2.4.13), where we use $\epsilon_0 = 0.01$.

We apply the GNA on both of the two cases: one has relatively large noise level $\epsilon = 0.05$, and the other has relatively small noise level $\epsilon = 0.01$. The GNA converges within 3 or 4 iterations for large or small noise level, respectively, as shown in figure 2.5a and figure 2.5b. This is quite efficient. The final estimates (MAP) are shown in figure 2.5c, from which, we can observe that: when the noise level is high $\epsilon = 0.05$, the MAP estimate has relatively large difference from the truth; but when the noise level becomes lower $\epsilon = 0.01$, the MAP estimate is more close to the truth.

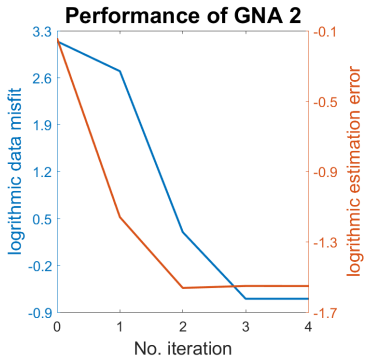
2.6.5 Bayesian inversion with the PCN-MCMC method

The Bayesian inversion aims to draw samples from the posterior probability measure determined by the Bayes' formula (2.2.5). The PCN-MCMC method (2.5.25) (2.5.26) is adopted for the infinite-dimensional sampling.

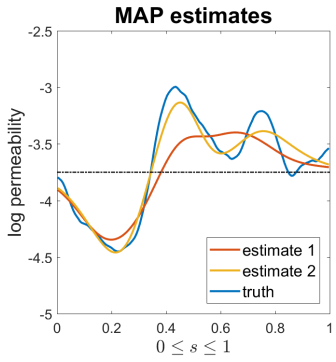
We apply the PCM-MCMC method on both of the two cases: one has relatively large noise level $\epsilon = 0.05$, and the other has relatively small noise level $\epsilon = 0.01$. The Markov chains starting from the prior distribution converge to the steady state (posterior distribution) within 10^3 or 10^4 iterations for large or small noise level, respectively, as shown in figure 2.6a and figure 2.6b. Usually, MCMC methods has low efficiency in



(a) The estimation error and data misfit in Gauss-Newton iterations (for the problem with large noise level $\epsilon = 0.05$).



(b) The estimation error and data misfit in Gauss-Newton iterations (for the problem with small noise level $\epsilon = 0.01$).

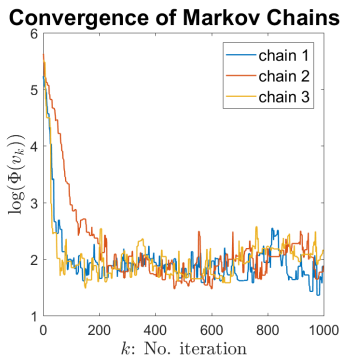


(c) The MAP estimates produced by GNA (estimate 1 has large noise; estimate 2 has small noise), compared with the truth.

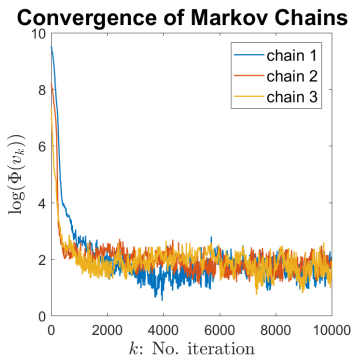
Figure 2.5: Using the GNA to solve the Tikhonov regularization of the 1D Darcy flow problem.

practice, but these kinds of sampling algorithms are accurate and universal. The posterior mean (average the final states of $J = 10^4$ independent chains) are shown in figure 2.6c, from which, we can observe that: when the noise level is high $\epsilon = 0.05$, the conditional mean has relatively large difference from the truth; but when the noise level becomes lower $\epsilon = 0.01$, the conditional mean is more close to the truth. The conditional-mean estimates shown in figure 2.6c are similar like the MAP estimates shown in figure 2.5c.

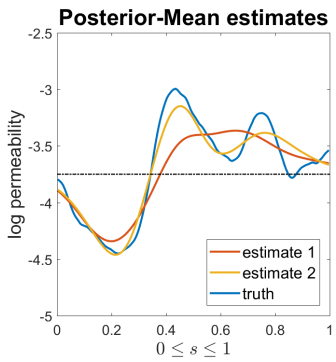
In fact, the benefit of Bayesian approach is uncertainty quantification, since the Bayesian estimation not only produces a point estimate but also implies a distribution of all possible estimates. Numerically, we can use the sample percentiles to quantify the uncertainty range of estimates. The 95% confidence interval $(2.5\% - 97.5\%)$ of the prior distribution, the posterior distribution with large noise, and the posterior distribution with small noise are shown in figure 2.7a, figure 2.7b, and figure 2.7c, respectively. The prior information has relatively large uncertainty (the confidence interval is wide in figure 2.7a). After observing data, the posterior information is updated and the uncertainty is reduced (the confidence interval in figure 2.7b is narrower than that in figure 2.7a). For lower noise level, the quality of information is further improved, and result is very close to the truth (the confidence interval in figure 2.7c is narrower than that in figure 2.7b).



(a) Convergence of several indepedent Markov chains produced by PCN-MCMC (for the problem with large noise level $\epsilon = 0.05$).

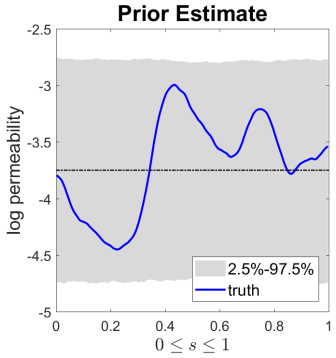


(b) Convergence of several indepedent Markov chains produced by PCN-MCMC (for the problem with small noise level $\epsilon = 0.01$).

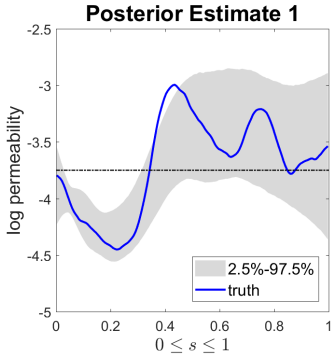


(c) The mean estimates produced by PCN-MCMC (estimate 1 has large noise; estimate 2 has small noise), compared with the truth.

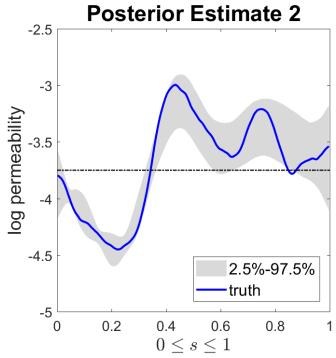
Figure 2.6: Using the PCN-MCMC method to sample from the posterior distribution of the 1D Darcy flow problem. The step size β of the PCN-MCMC method is specified as: $\beta = 0.15$ for large noise level $\epsilon = 0.05$, and $\beta = 0.03$ for small noise level $\epsilon = 0.01$; the corresponding acceptance rate is around 0.24.



(a) Prior estimate (the Whittle-Matérn field).



(b) Posterior estimate with large noise level $\epsilon_0 = 0.05$.



(c) Posterior estimate with small noise level $\epsilon_0 = 0.01$.

Figure 2.7: The Bayesian estimation with uncertainty quantification. $J = 10^4$ samples independently start from the prior distribution, and then iterate along independent Markov chains (using the PCN-MCMC method).

Let u be the posterior random estimate $u \sim \mathbb{P}(\cdot|y)$, and let x be the truth. Moreover, we consider the KL factor and the data misfit corresponding to the estimate:

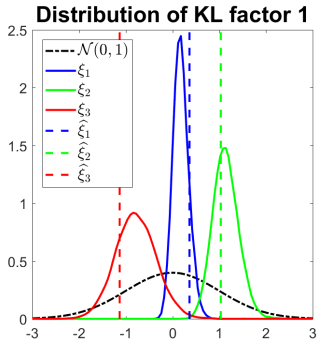
- Let ξ denote the random factor, and the let $\widehat{\xi}$ denote the true factor, where the 'factors' indicate the eigen-basis in the KL expansion of the prior field, i.e.,

$$\xi = \mathcal{C}_0^{-1/2}(u - m_0) \qquad \widehat{\xi} = \mathcal{C}_0^{-1/2}(x - m_0) \tag{2.6.11}$$

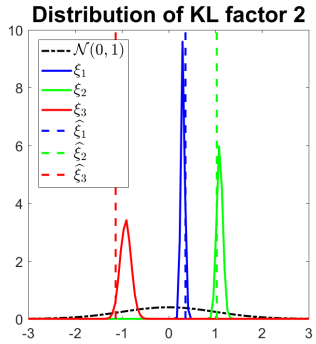
- Let X be the random misfit, and let $\widehat{X}$ be the exact amount of noise, i.e.,

$$X = \|y - \mathcal{G}(u)\|_\gamma \qquad \widehat{X} = \|y - \mathcal{G}(x)\|_\gamma \tag{2.6.12}$$

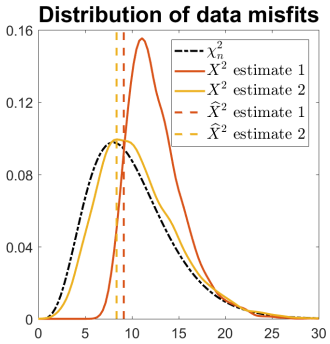
Then, we plot the (empirical) probability density functions of the KL factor ξ and the data misfit X. The KL factors with large and small noise levels are plotted in figures 2.8a and 2.8b, respectively. We can observe that, if the noise level is lower, then the distribution of the KL factor is more concentrated around the 'truth'. Moreover, the data misfits with large and small noise levels are plotted in figure 2.8c. We can observe that, if the noise level is lower, then the distribution of the squared misfit is more close to the chi-square distribution with n degrees of freedom, where $n = 10$ is the number of observations.



(a) The PDFs of KL components (for the problem with large noise level $\epsilon_0 = 0.05$). ξ_1, ξ_2 and ξ_3 are the first three components of the posterior factor ξ, and $\widehat{\xi}_1$, $\widehat{\xi}_2$ and $\widehat{\xi}_3$ are the first three components of the true factor $\widehat{\xi}$. The prior distribution of all the KL components is identity to the standard normal distribution $\mathcal{N}(0,1)$.



(b) The PDFs of KL components (for the problem with small noise level $\epsilon_0 = 0.01$). ξ_1, ξ_2 and ξ_3 are the first three components of the posterior factor ξ, and $\widehat{\xi}_1$, $\widehat{\xi}_2$ and $\widehat{\xi}_3$ are the first three components of the true factor $\widehat{\xi}$. The prior distribution of all the KL components is identity to the standard normal distribution $\mathcal{N}(0,1)$.



(c) The PDFs of squared misfits. X^2 estimate 1 has large noise level $\epsilon = 0.05$; X^2 estimate 2 has small noise level $\epsilon = 0.01$; $\widehat{X}^2$ estimate 1 is the true value of the large noise; $\widehat{X}$ estimate 2 is the true value of the small noise. It is expected that the squared misfits of the X^2 estimates should close to the χ^2_n distribution.

Figure 2.8: Empirical probability density functions (PDFs) of KL factor and data misfit.

Chapter 3

Tempering Setting and Adaptive Methods for Inverse Problems

Last chapter has introduced well-posed inverse problems via variational approach and Bayesian approach. This chapter conducts further investigation, that rewrites the original variational/Bayesian setting as a tempering setting. Inverse problems with the tempering setting can be equivalently regarded as filtering problems. Thus, approximate filtering algorithms e.g. extended Kalman filter, ensemble Kalman filter can be applied to solve inverse problems.

In this chapter, we firstly introduce the tempering setting in section 3.1, and then in section 3.2, we propose an adaptive scheme discretizing the tempering setting. Section 3.3 discusses the Kalman-like methods as approximate methods solving inverse problems with the tempering setting, and section 3.4 adopts the adaptive scheme to select discrete steps when the Kalman-like methods are numerically implemented. The last section is short summary which presents what have been done in this chapter.

3.1 Tempering setting for inverse problems

Recall the additive noise model (2.2.1). The inverse problem is to infer the unknown parameter $x \in \mathcal{H}$, given the forward map $\mathcal{G} : \mathcal{H} \to \mathbb{R}^n$ and the observation $y \in \mathbb{R}^n$. As discussed in the last chapter, there are two standard approaches setting a well-posed inverse problem via variational approach and Bayesian approach. This section discusses

a variant of the original setting. This variant is called the tempering setting.

The tempering setting is inspired from some numerical algorithms like simulated annealing and annealed importance sampling. Also, the tempering setting is closely related to Bayesian filtering. In fact, a tempered inverse problem can be regarded as a special case of filtering problems. The benefit of the tempering setting is that, it builds a continuous path from prior to posterior, such that, data misfits are reduced along the path, and complicated inverse problems can be gradually solved.

This section is divided by several subsections. Bayesian filtering, as the background, is introduced in subsection 3.1.1. The trivial motivation of the tempering setting is discussed in subsection 3.1.2. More formal definition of the tempering setting is presented in subsection 3.1.3. Some essential properties of the tempering setting are discussed in subsection 3.1.4.

3.1.1 Hidden Markov chains and Bayesian filtering

First of all, we introduce Bayesian filtering, which uses Bayesian method to extract information from a hidden Markov chain by observing sequence of data. This kind of problems commonly appear in data assimilation, signal processing, and machine learning. Bayesian filtering is closely related to the tempering setting for inverse problems. To well understand filtering problems can help readers understand the tempering setting more clearly and deeply.

We consider a model with evolution of states and observation of data. The evolution-observation model can be represented by a hidden Markov chain. The hidden states are estimated by a stochastic process with uncertainty, and the observed data are characterized by another stochastic process. We assume that the hidden states are in a separable Hilbert space $\mathcal{H}$, and that the observations are in the Euclidean space $\mathbb{R}^n$, where n is the number of dimensions of observation vector. Then, two discrete stochastic processes $\{X_i\}_{i=0}^K$ and $\{Y_i\}_{i=1}^K$ are used to represent the hidden states and observations respectively, where $\{X_i\}_{i=0}^K$ represent the hidden states (for each $i = 0, 1, ..., K$, X_i is an $\mathcal{H}$-valued random variable) and $\{Y_i\}_{i=1}^K$ represent the observations (for each $i = 1, ..., K$, Y_i is an $\mathbb{R}^n$-valued random variable). More precisely, the evolution-observation model has

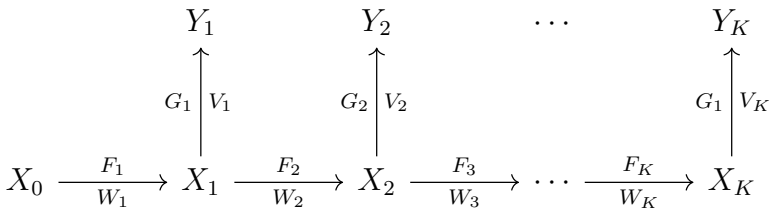


Figure 3.1: Diagram of hidden Markov chain.

the following three components (for clearance, the hidden Markov chain is also presented as a diagram in figure 3.1):

1. The probability measure of the initial state X_0.

2. The evolution equation, for any $i = 1, ..., K$,

$$X_i = F_i(X_{i-1}) + W_i \tag{3.1.1}$$

   where $F_i : \mathcal{H} \to \mathcal{H}$ is the ith evolution model, and W_i is an independent $\mathcal{H}$-valued random variable representing the ith state noise.

3. The observation equation, for any $i = 1, ..., K$,

$$Y_i = G_i(X_i) + V_i \tag{3.1.2}$$

   where $G_i : \mathcal{H} \to \mathbb{R}^n$ is the ith observation model, and V_i is an independent $\mathbb{R}^n$-valued random variable representing the ith observation noise.

After real observations $y_1, y_2, ..., y_K \in \mathbb{R}^n$ are obtained, Bayesian filtering aims to interpret conditional probabilities of hidden states given the observations. For convenience, we use the short notation $\mathcal{S}_i$ $(i = 0, 1, ..., K)$ to denote the sequence of observations, i.e.,

$$\mathcal{S}_i := \begin{cases} \emptyset & \text{if } i = 0 \\ \{y_1, ..., y_i\} & \text{if } 0 < i \leq K \end{cases} \tag{3.1.3}$$

Furthermore, we use the short notation $\mathbb{P}_{X_j}(\cdot|\mathcal{S}_i)$ $(i, j = 0, 1, ..., K)$ to denote the conditional probability measure of state X_j given the sequence of observations $\mathcal{S}_i$, i.e.,

$$\mathbb{P}_{X_j}(\cdot|\mathcal{S}_i) := \begin{cases} \mathbb{P}_{X_j}(\cdot) & \text{if } i = 0 \\ \mathbb{P}_{X_j}(\cdot|Y_1 = y_1, ..., Y_i = y_i) & \text{if } 0 < i \leq K \end{cases} \tag{3.1.4}$$

The Bayesian filtering problem means to update the conditional probability measures one by one,

$$\mathbb{P}_{X_0}(\cdot|\mathcal{S}_0) \to \mathbb{P}_{X_1}(\cdot|\mathcal{S}_1) \to \cdots \to \mathbb{P}_{X_K}(\cdot|\mathcal{S}_K) \tag{3.1.5}$$

For any iteration $i = 1, ..., K$ from $\mathbb{P}_{X_{i-1}}(\cdot|\mathcal{S}_{i-1})$ to $\mathbb{P}_{X_i}(\cdot|\mathcal{S}_i)$, the update can be split by two steps [49],

1. Evolution update: given $\mathbb{P}_{X_{i-1}}(\cdot|\mathcal{S}_{i-1})$, find $\mathbb{P}_{X_i}(\cdot|\mathcal{S}_{i-1})$ based on the Markov transition kernel $K_i(w, \cdot) = \mathbb{P}_{X_i}(\cdot|X_{i-1} = w)$, where for any $w \in \mathcal{H}$, $\mathbb{P}_{X_i}(\cdot|X_{i-1} = w)$ is the conditional probability measure of X_i given $X_{i-1} = w$.

2. Observation update: given $\mathbb{P}_{X_i}(\cdot|\mathcal{S}_{i-1})$, find $\mathbb{P}_{X_i}(\cdot|\mathcal{S}_i)$ based on the likelihood function $\mathcal{L}_i(u|y_i) = \pi(Y_i = y_i|X_i = u)$, where for any $u \in \mathcal{H}$, $\pi(Y_i = y_i|X_i = u)$ is the conditional probability density of Y_i at y_i given $X_i = u$.

The following theorem is used to determine the update equations.

Theorem 3.1.1. *(Theorem 4.2 in [49])*

1. *For evolution update, we have the Markov transition,*

$$\mathbb{P}_{X_i}(\mathrm{d}u|\mathcal{S}_{i-1}) = \int_{\mathcal{H}} \mathbb{P}_{X_i}(\mathrm{d}u|X_{i-1} = w)\mathbb{P}_{X_{i-1}}(\mathrm{d}w|\mathcal{S}_{i-1}) \tag{3.1.6}$$

2. *For observation update, we have the Bayes' formula,*

$$\mathbb{P}_{X_i}(\mathrm{d}u|\mathcal{S}_i) \propto \pi(Y_i = y_i|X_i = u)\mathbb{P}_{X_i}(\mathrm{d}u|S_{i-1}) \tag{3.1.7}$$

3.1.2 Motivation of the tempering setting

Our purpose is to infer the unknown parameter $x \in \mathcal{H}$ in the additive noise model (2.2.1), given the forward map $\mathcal{G} : \mathcal{H} \to \mathbb{R}^n$ and the observation $y \in \mathbb{R}^n$. Sometimes, directly solving this problem may be difficult due to highly nonlinear properties of the forward map. Instead, a more feasible approach is to gradually solve a sequence of simpler subproblems that finally recovers the original problem. More specifically, we can construct a sequence of sub-problems, and each sub-problem has the same mathematical structure as the original problem. Each sub-problem has lower fidelity than the original problem, but

sequentially solving all the sub-optimums recovers the high fidelity problem (the original problem).

For simplicity, we make the Gaussian assumptions. Assume that the error $e = y - \mathcal{G}(x)$ between the observation y and the prediction $\mathcal{G}(x)$ can be characterized by a centered non-degenerate Gaussian distribution $\mathcal{N}(\mathbf{0}, \gamma)$, where $\gamma : \mathbb{R}^n \to \mathbb{R}^n$ is a symmetric positive-definite bounded matrix. Moreover, assume that the prior estimation of the unknown parameter x is also given as a Gaussian distribution $\mathcal{N}(m_0, \mathcal{C}_0)$, where $m_0 \in \mathcal{H}$ is a bounded element, and $\mathcal{C}_0 : \mathcal{H} \to \mathcal{H}$ is a self-adjoint positive-semi-definite trace-class operator.

The original Bayesian method

The original Bayesian method estimating the unknown parameter x considers the one-step transition from the prior to the posterior,

$$Y = \mathcal{G}(X) + V \tag{3.1.8}$$

where $X \sim \mathcal{N}(m_0, \mathcal{C}_0)$ and $V \sim \mathcal{N}(\mathbf{0}, \gamma)$ are two independent Gaussian random variables. The independent random variable X represents prior estimates of the unknown parameter x with possibilities, and another independent random variable V represents all possible errors between real observation and mathematical prediction. Thus, the dependent random variable Y represents all possible observations under the given prior assumption and error assumption. After the observation y is obtained, the Bayesian inverse problem is to interpret the conditional probability of X given the observation $Y = y$ from the Bayes' rule,

$$\mathbb{P}(\mathrm{d}u|y) \propto \exp\left(-\frac{1}{2}\left\|\gamma^{-1/2}(y - \mathcal{G}(u))\right\|_{\mathbb{R}^n}^2\right)\mathbb{P}(\mathrm{d}u) \tag{3.1.9}$$

where $\mathbb{P} = \mathcal{N}(m_0, \mathcal{C}_0)$ is the prior probability measure, and $\mathbb{P}(\cdot|y)$ is the posterior probability measure.

The tempered Bayesian method

Sometimes, directly interpreting the conditional probability determined in formula (3.1.9) is numerically inefficient, because the forward model $\mathcal{G}$ can be very complicated which

forms a extremely sharp distribution of the weights $w = \exp\left(-\frac{1}{2}\left\|\gamma^{-1/2}(y - \mathcal{G}(X))\right\|_{\mathbb{R}^n}^2\right)$ with $X \sim \mathcal{N}(m_0, \mathcal{C}_0)$. In order to solve this issue, we can gradually solving the original problem by constructing a sequence of sub-problems.

The tempering setting is a sequence of sub-problems applying tempering parameters $0 = t_0 < t_1 < \cdots < t_K = 1$. For convenience, let $h_i = t_i - t_{i-1}$ for all $i = 1, ..., K$ be the step sizes. The tempering setting can be constructed as a Bayesian filtering problem by specifying the evolution-observation model as follows.

1. The probability measure of the initial state X_0 is the Gaussian measure $\mathcal{N}(m_0, \mathcal{C}_0)$.

2. The evolution equation is identity, for any $i = 1, ..., K$,

$$X_i = X_{i-1} \tag{3.1.10}$$

3. The observation equation is determined by the forward map $\mathcal{G}$, for any $i = 1, ..., K$,

$$Y_i = \mathcal{G}(X_i) + V_i \tag{3.1.11}$$

   where the observation noise $V_i \sim \mathcal{N}(\mathbf{0}, h_i^{-1}\gamma)$ is an independent Gaussian random variable.

Moreover, all the observations $y_1, ..., y_K$ are specified as the invariant quantity y,

$$y_1 = \cdots = y_K = y \tag{3.1.12}$$

For this special case, the update equation is very simple, derived as follows. Since evolution of the states remains identity (3.1.10), we straightforwardly have the evolution update

$$\mathbb{P}_{X_i}(\mathrm{d}u|Y_1 = y, ..., Y_{i-1} = y) = \mathbb{P}_{X_{i-1}}(\mathrm{d}u|Y_1 = y, ..., Y_{i-1} = y) \tag{3.1.13}$$

Furthermore, by using the Bayes' formula for the observation model (3.1.11), we have the observation update

$$\mathbb{P}_{X_i}(\mathrm{d}u|Y_1 = y, ..., Y_i = y) \propto \pi(Y_i = y|X_i = u)\mathbb{P}_{X_i}(\mathrm{d}u|Y_1 = y, ..., Y_{i-1} = y) \tag{3.1.14}$$

where, since the observation noise $V_i \sim \mathcal{N}(\mathbf{0}, \gamma)$ is Gaussian, the probability density $\pi(Y_i = y|X_i = u)$ is the Gaussian density

$$\pi(Y_i = y|X_i = u) \propto \exp\left(-\frac{h_i}{2}\left\|\gamma^{-1/2}(y - \mathcal{G}(u))\right\|_{\mathbb{R}^n}^2\right) \tag{3.1.15}$$

As the result, combining formulas (3.1.13) and (3.1.14) leads to, for any $i = 1, ..., K$,

$$\nu_i(\mathrm{d}u) \propto \exp\left(-\frac{h_i}{2}\left\|\gamma^{-1/2}(y - \mathcal{G}(u))\right\|_{\mathbb{R}^n}^2\right)\nu_{i-1}(\mathrm{d}u) \tag{3.1.16}$$

where the short notation $\{\nu_i\}_{i=0}^K$ denote the probability measures,

$$\nu_i(\cdot) \equiv \begin{cases} \mathbb{P}_{X_i}(\cdot) & \text{if } i = 0 \\ \mathbb{P}_{X_i}(\cdot|Y_1 = y, ..., Y_i = y) & \text{if } 0 < i \leq K \end{cases} \tag{3.1.17}$$

In brief, solving the ith sub-problem is equivalent to solving the original problem with γ replaced by $h_i^{-1}\gamma$.

It is clear that, the sequence of probability measures $\{\nu_i\}_{i=0}^K$ start from the prior probability $\mathbb{P}$, and, since $\sum_{i=1}^K h_i = 1$, end in the posterior probability $\mathbb{P}(\cdot|y)$, i.e.

$$\mathbb{P}(\cdot) = \nu_0(\cdot) \to \nu_1(\cdot) \to \cdots \to \nu_K(\cdot) = \mathbb{P}(\cdot|y) \tag{3.1.18}$$

where $\mathbb{P}(\cdot)$ and $\mathbb{P}(\cdot|y)$ are the same as those in formula (3.1.9). In many applications, the forward map $\mathcal{G}$ is nonlinear. Directly solving the original problem with one-step transition from the prior $\mathbb{P}$ to the posterior $\mathbb{P}(\cdot|y)$ may be numerically inefficient. However, sequentially solving the sub-problems in the tempering setting is more robust.

3.1.3 Formulation of the tempering setting

The tempering setting is inspired from simulated annealing [104], annealed importance sampling [88], and sequential Monte Carlo methods [3, 1]. The formal definition of the tempering setting, discussed in this subsection, applies an auxiliary parameter $t \in [0, 1]$ to rewrite the original formula (2.2.4) or (2.2.5), such that, the rewritten version indicates a class of problems with respect to the tempering parameter t. Particularly, $t = 0$ indicates the prior, and $t = 1$ indicates the posterior.

More precisely, we use mathematical formulas to define the tempering setting as follows. For convenience, we first of all define the data-misfit function $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ and the cost functional $\Phi : \mathcal{H} \to [0, +\infty)$,

$$\mathcal{Z}(\cdot) = \gamma^{-1/2}(y - \mathcal{G}(\cdot)) \qquad \Phi(\cdot) = \frac{1}{2}\|\mathcal{Z}(\cdot)\|_{\mathbb{R}^n}^2 \tag{3.1.19}$$

where $y \in \mathbb{R}^n$ is the observation, $\gamma : \mathbb{R}^n \to \mathbb{R}^n$ is the covariance matrix of observation noise, and $\mathcal{G} : \mathcal{H} \to \mathbb{R}^n$ is the forward map. Then, we add the tempering parameter $t \in [0, 1]$ to the original formulas (2.2.4) and (2.2.5). As the result, the original variational setting (2.2.4) can be rewritten as the tempered variational setting,

$$\widehat{x}_t = \arg\min_{u \in \mathcal{H}} \{t\Phi(u) + R(u)\} \quad \text{with} \quad R(\cdot) = \frac{1}{2} \|(\cdot) - m_0\|_{\mathcal{C}_0}^2 \tag{3.1.20}$$

and the original Bayesian setting (2.2.5) can be rewritten as the tempered Bayesian setting,

$$\mu_t(\mathrm{d}u) \propto \exp\left(-t\Phi(u)\right) \mathbb{P}(\mathrm{d}u) \quad \text{with} \quad \mathbb{P} = \mathcal{N}(m_0, \mathcal{C}_0) \tag{3.1.21}$$

where Φ is the cost functional defined in formula (3.1.19), $m_0 \in \mathcal{H}$ is the prior mean and $\mathcal{C}_0 : \mathcal{H} \to \mathcal{H}$ is the prior covariance operator. The objective of the tempered method is to search for the minimum point $\widehat{x}_t$ or to interpret the probability measure μ_t for $t \in (0, 1]$.

Remark 3.1.2. *The following facts should be noticed:*

1. *For any $t \in [0, 1]$, the optimum $\widehat{x}_t$ determined in formula* (3.1.20) *is the MAP point of the probability measure μ_t determined in formula* (3.1.21).

2. *For any $0 = t_0 < t_1 < \cdots < t_K = 1$, the probability measure μ_{t_i} determined in formula* (3.1.21) *exactly equals to the probability measure ν_i determined in formula* (3.1.16).

3. *The tempering setting defined in* (3.1.20) *or* (3.1.21) *has continuous parameter $t \in [0, 1]$. The benefit of the continuous setting is that, it is the theoretical limit of any discrete specifications, so that, we can keep the continuous formulation for mathematical analysis and leave the discretization at the last moment.*

The well-posedness of formula (3.1.20) and formula (3.1.21) regardless of t has been discussed in the last chapter (see theorem 2.2.4 and theorem 2.2.6 respectively). Moreover, in theorem 3.1.3 and theorem 3.1.5, we will additionally show the continuity and monotonicity with respect to the tempering parameter $t \in [0, 1]$. These properties are essential, since the continuity implies that the tempering setting is a stable process in $t \in [0, 1]$, and the monotonicity ensures that the tempering setting gradually reduces the cost functional in $t \in [0, 1]$. These two properties reveal why and how the tempering setting works well.

3.1.4 Continuity and monotonicity of the tempering setting

The theorems in this subsection reveal the essential properties of the tempering setting. We emphasizes that these properties are proved for nonlinear operators and infinite-dimensional parameters. Both the variational approach and the Bayesian approach have the similar properties.

The following theorem shows that: the tempered Bayesian setting (3.1.21) determines a trajectory along $t \in [0,1]$, such that, the average cost functional is always decreasing along this trajectory.

Theorem 3.1.3. *Let Φ, $\mathbb{P}$, μ_t denote the same symbols in formula* (3.1.21)*. For any $t \in [0,1]$, let I_t and $\langle \Phi \rangle_t$ denote*

$$I_t \equiv -\log\left(\int_{\mathcal{H}} \exp(-t\Phi(u))\,\mathbb{P}(\mathrm{d}u)\right) \qquad \langle \Phi \rangle_t \equiv \int_{\mathcal{H}} \Phi(u)\,\mu_t(\mathrm{d}u) \tag{3.1.22}$$

Assume that the forward map $\mathcal{G}$ satisfies condition 1 in assumption 2.2.3. Then I_t and $\langle \Phi \rangle_t$ are well-defined. Moreover, I_t is increasing in $t \in [0,1]$, and $\langle \Phi \rangle_t$ is decreasing in $t \in [0,1]$. Furthermore, I_t regarded as a function of t is an analytic function in $t \in [0,1]$, whose first order derivative is represented by

$$I_t' = \langle \Phi \rangle_t \tag{3.1.23}$$

Proof. Just a special case of theorem 4.2.9 with Φ specified as $\Phi(\cdot) = \frac{1}{2}\left\|\gamma^{-1/2}(y - \mathcal{G}(\cdot))\right\|_{\mathbb{R}^n}^2$. Since $\gamma^{-1/2}\mathcal{G}(\cdot)$ is assumed to have an exponential tail, then $\Phi(\cdot)$ also have an exponential tail. Consequently, we can directly use the result in theorem 4.2.9. □

Corollary 3.1.4. *For any $t \in [0,1]$, let $P_t \equiv I_t + (1-t)\langle \Phi \rangle_t$. Then P_t is decreasing in $t \in [0,1]$.*

Proof. Apply chain rule,

$$P_t' = I_t' - \langle \Phi \rangle_t + (1-t)\langle \Phi \rangle_t'$$

where $I_t' = \langle \Phi \rangle_t$ shown in theorem 3.1.3. Thus, we have

$$P_t' = (1-t)\langle \Phi \rangle_t'$$

where $(1-t) \geq 0$ and $\langle \Phi \rangle_t' \leq 0$. Thus, we have

$$P_t' \leq 0$$

Therefore, P_t is decreasing in $t \in [0,1]$. □

The following theorem shows that: the tempered variational setting (3.1.20) determines a trajectory along $t \in [0,1]$, such that, the cost functional at the minimum point is always decreasing along this trajectory.

Theorem 3.1.5. *Let Φ, R, $\widehat{x}_t$ denote the same symbols in formula* (3.1.20)*. For any $t \in [0,1]$, let J_t and ϕ_t denote*

$$J_t \equiv t\Phi(\widehat{x}_t) + R(\widehat{x}_t) \qquad \phi_t \equiv \Phi(\widehat{x}_t) \tag{3.1.24}$$

Assume that the forward map $\mathcal{G}$ satisfies condition 2 in assumption 2.2.3. Then J_t and ϕ_t are well-defined. Moreover, J_t is increasing in $t \in [0,1]$, and ϕ_t is decreasing in $t \in [0,1]$. Furthermore, J_t regarded as a function of t is Lipschitz continuous in $t \in [0,1]$ with derivative almost everywhere, represented by

$$J_t = \int_0^t \phi_s \, \mathrm{d}s \tag{3.1.25}$$

Proof. Just a special case of theorem 4.1.5 with Φ specified as $\Phi(\cdot) = \frac{1}{2}\left\|\gamma^{-1/2}(y - \mathcal{G}(\cdot))\right\|_{\mathbb{R}^n}^2$. Since $\gamma^{-1/2}\mathcal{G}(\cdot)$ is assumed to be Lipschitz continuous on any bounded subsets, then $\Phi(\cdot)$ is also Lipschitz. Consequently, we can directly use the result in theorem 4.1.5. □

Corollary 3.1.6. *For any $t \in [0,1]$, let $Q_t \equiv J_t + (1-t)\phi_t$. Then Q_t is decreasing in $t \in [0,1]$.*

Proof. According to theorem 3.1.5, it is clear that, Q_t can be represented by

$$Q_t = \int_0^t \phi_s \, \mathrm{d}s + (1-t)\phi_t$$

Since ϕ_t is decreasing, then for any $0 \le t_1 \le t_2 \le 1$, we have

$$\begin{aligned} Q_{t_2} - Q_{t_1} &= \int_{t_1}^{t_2} \phi_s \, \mathrm{d}s + (1-t_2)\phi_{t_2} - (1-t_1)\phi_{t_1} \\ &\le (t_2 - t_1)\phi_{t_1} + (1-t_2)\phi_{t_2} - (1-t_1)\phi_{t_1} \\ &= (1-t_2)(\phi_{t_2} - \phi_{t_1}) \le 0 \end{aligned}$$

Therefore, Q_t is decreasing in $t \in [0,1]$. □

The above theorems 3.1.3, 3.1.5 and corollaries 3.1.4, 3.1.6 play their roles in two aspects:

1. An inverse problem with the tempering setting can be gradually solved starting from the prior and ending in the posterior. Namely, given a sequence $0 = t_0 < t_1 < \cdots < t_K = 1$ in $[0, 1]$, conduct the iterative optimization,

$$m_0 = \widehat{x}_0 \to \widehat{x}_{t_1} \to \cdots \to \widehat{x}_{t_K} = \widehat{x}(y) \tag{3.1.26}$$

and/or conduct the sequential inference,

$$\mathcal{N}(m_0, \mathcal{C}_0) = \mu_0 \to \mu_{t_1} \to \cdots \to \mu_{t_K} = \mathbb{P}(\cdot|y) \tag{3.1.27}$$

Then, according to theorem 3.1.5, the cost functional at the minimum point must decrease,

$$\Phi(\widehat{x}_{t_0}) \geq \Phi(\widehat{x}_{t_1}) \geq \cdots \geq \Phi(\widehat{x}_{t_K}) \tag{3.1.28}$$

and according to theorem 3.1.3, the average cost functional must decrease also,

$$\int_{\mathcal{H}} \Phi(u)\, \mu_{t_0}(\mathrm{d}u) \geq \int_{\mathcal{H}} \Phi(u)\, \mu_{t_1}(\mathrm{d}u) \geq \cdots \geq \int_{\mathcal{H}} \Phi(u)\, \mu_{t_K}(\mathrm{d}u) \tag{3.1.29}$$

That means, in the iterations, the next estimate is always better than the last estimate, quantified by data misfit.

2. Sometimes in practice, it is not feasible to apply accurate numerical methods (e.g. MCMC) because accurate methods are computationally expensive. If some approximate methods (e.g. EKF and EnKF) are applied, we have to check whether the approximation is good enough or not. A possible way is to check the monotone properties in theorems 3.1.3, 3.1.5 and corollaries 3.1.4, 3.1.6, i.e. the quantities $\langle \Phi \rangle_t$, P_t, ϕ_t, Q_t should be decreasing. If the approximate method preserves the monotone properties, then we keep iterations in the algorithm, otherwise we stop the iterations in force because the approximation method cannot provide better estimates anymore.

3.2 Adaptive tempering setting

Last section has introduced the tempering setting for inverse problems. The tempering setting can be applied for both variational inversion and Bayesian inversion, shown in

formulas (3.1.20) and (3.1.21) respectively. For numerical implementation, the tempering setting should be discretization. The vital question is how to select the discrete step size reasonably and effectively. In this section, we propose an adaptive scheme, called the data-misfit controller, which monitors the data misfit in each iteration and predicts the next step size based on the data misfit.

3.2.1 Introduction of the main idea

We aim to propose an adaptive scheme discretizing the tempering setting on $t \in [0, 1]$. Informally, we use the word 'adaptive' to indicate that, discrete steps on $[0, 1]$ are selected such that, the amounts of information learned from each steps are (approximately) same. More precisely, we consider to monitor the information gain in the stepwise learning. The essential question is: what is the 'amount of information' and 'information gain'?

The amount of information is defined as information entropy by Shannon [17]. The information gain from new observations is defined as the relative entropy [93]. It is 'relative', because it accounts for the 'gain' of information from one state (random variable) to another state (random variable). In mathematical perspective, the synonym of information gain (relative entropy) is the Kullback-Leibler divergence D_{KL}, which quantifies the difference between two probability measures. The Kullback-Leibler divergence is defined as follows.

Definition 3.2.1 (Kullback-Leibler divergence)**.** *Let ν and μ be two probability measures on a measurable space (X, Σ). If ν is absolutely continuous with respect to μ, then the Kullback-Leibler divergence of ν with respect to μ is defined as,*

$$\mathrm{D}_{KL}(\nu||\mu) := \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\nu \tag{3.2.1}$$

The Kullback-Leibler divergence is asymmetric. Nevertheless, if the two probability measures ν and μ are equivalent, then a symmetric quantity can be defined as,

$$\mathrm{D}_{KL,2}(\mu, \nu) := \mathrm{D}_{KL}(\nu||\mu) + \mathrm{D}_{KL}(\mu||\nu) \tag{3.2.2}$$

Let $0 = t_0 < t_1 < \cdots < t_K = 1$ be any $K + 1$ points in interval $[0, 1]$. For any $i = 1, ..., K$, the forward information gain from $\mu_{t_{i-1}}$ to μ_{t_i} is $\mathrm{D}_{KL}(\mu_{t_i}||\mu_{t_{i-1}})$ and the

backward information gain from μ_{t_i} to $\mu_{t_{i-1}}$ is $\mathrm{D}_{KL}(\mu_{t_{i-1}}||\mu_{t_i})$,

$$\mathrm{D}_{KL}(\mu_{t_i}||\mu_{t_{i-1}}) = \int_{\mathcal{H}} \log\left(\frac{\mathrm{d}\mu_{t_i}}{\mathrm{d}\mu_{t_{i-1}}}\right) \mathrm{d}\mu_{t_i} \tag{3.2.3}$$

$$\mathrm{D}_{KL}(\mu_{t_{i-1}}||\mu_{t_i}) = \int_{\mathcal{H}} \log\left(\frac{\mathrm{d}\mu_{t_{i-1}}}{\mathrm{d}\mu_{t_i}}\right) \mathrm{d}\mu_{t_{i-1}} \tag{3.2.4}$$

where μ_t for any $t \in [0,1]$ is the probability measure defined in the tempering setting (3.1.21). Moreover, the sum of forward and backward information gain is the $\mathrm{D}_{KL,2}$ quantity,

$$\mathrm{D}_{KL,2}(\mu_{t_{i-1}}, \mu_{t_i}) = \mathrm{D}_{KL}(\mu_{t_i}||\mu_{t_{i-1}}) + \mathrm{D}_{KL}(\mu_{t_{i-1}}||\mu_{t_i}) \tag{3.2.5}$$

The symmetricity of the $\mathrm{D}_{KL,2}$ quantity indicates that, the information can be learned forwardly at the same time can be also recovered backwardly. Essentially speaking, this forward-backward property indicates the equivalence of probability measures $\mu_{t_{i-1}}$ and μ_{t_i}, that means, there is no loss of information. In fact, it will be proved later in this thesis that the probability measures determined in the tempering setting with different parameters $t \in [0,1]$ are equivalent. Thus, it is feasible to apply the symmetric quantity $\mathrm{D}_{KL,2}$ to measure the difference of the probability measures.

Moreover, controlling the $\mathrm{D}_{KL,2}$ quantity is closely related to controlling other statistics/quantities like mean, variance, and finite difference error. We will analyze the mean-variance pair, the symmetric quantity $\mathrm{D}_{KL,2}$, and the finite difference error of thermodynamic integration determined by the tempering setting. All of the analysis will finally lead to the same result: the data-misfit controller (DMisC). Thus, DMisC is adopted as the adaptive scheme for stepwise learning in the tempering setting.

3.2.2 The mean-variance pair

This subsection introduces the simplest approach to deriving the data-misfit controller. This approach assumes Gaussian error and adopts the chi-square distribution to quantify the noise level. As the result, the mean and variance of the data misfit are monitored and compared with the noise level. This approach is inspired from the Morozov's discrepancy principle [98, 80, 96] that determines the regularizing parameter in Tikohonov regularization. Similarly, the mean-variance approach can be regarded as the statistical discrepancy principle in the Bayesian framework.

Morozov's discrepancy principle

We begin from the Morozov's discrepancy principle, which has been mentioned in the last chapter and shown in formula (2.4.15). Now, we specify $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ in formula (2.4.15) as the data-misfit function (3.1.19), i.e., $\mathcal{Z}(\cdot) = \gamma^{-1/2}(y - \mathcal{G}(\cdot))$, where $y \in \mathbb{R}^n$ is the observation and $\mathcal{G} : \mathcal{H} \to \mathbb{R}^n$ is the forward model. We aim to estimate the unknown parameter whose true value is $x \in \mathcal{H}$. The Morozov's discrepancy principle suggests finding a point estimate $\widehat{x} \in \mathcal{H}$ satisfying

$$\left\|\gamma^{-1/2}\left(y - \mathcal{G}(\widehat{x})\right)\right\|_{\mathbb{R}^n} \leq \tau\delta \tag{3.2.6}$$

where $\tau > 1$ is an accuracy control parameter, and $\delta > 0$ is a given noise level satisfying

$$\delta \geq \left\|\gamma^{-1/2}\left(y - \mathcal{G}(x)\right)\right\|_{\mathbb{R}^n} \tag{3.2.7}$$

In order words, once we obtain an estimate $\widehat{x}$ satisfying the discrepancy principle (3.2.6), we do not have to conduct any further investigation.

The Morozov's discrepancy principle is suitable for point estimation via deterministic approaches. However, there are two facts that should be noticed: 1) the noise level δ sometimes is unknown in practice; 2) Morozov's discrepancy principle cannot quantify random variable estimates. Thus, some more practical principles are needed for Bayesian method.

Statistical discrepancy principle

For Gaussian noise, the cost functional has a quadratic form as the square norm of the data-misfit function, shown in formula (3.1.19). A benefit of considering Gaussian error is that, the noise level can be characterized with a chi-square distribution.

Given a fixed observation $y \in \mathbb{R}^n$, the observation-prediction error is $e = y - \mathcal{G}(x)$, where $x \in \mathcal{H}$ is the true value of the unknown parameter. Assume that the error e can be characterized by the Gaussian distribution $\mathcal{N}(\mathbf{0}, \gamma)$, i.e., the observation y can be regarded as a draw from $\mathcal{N}(\mathcal{G}(x), \gamma)$. The deterministic noise level is defined as

$$\widehat{\eta} := \left\|\gamma^{-1/2}\left(y - \mathcal{G}(x)\right)\right\|_{\mathbb{R}^n} \tag{3.2.8}$$

In fact, $\widehat{\eta}$ is not perfectly known because x is unknown in practice. Nevertheless, if n is sufficiently large, the value of $\widehat{\eta}$ can be approximately estimated using the law of large numbers, so that,

$$\widehat{\eta} \approx \sqrt{n} \tag{3.2.9}$$

Furthermore, we can conduct repeated experiments and obtain several observations. Statistically, consider the random observations $\xi = \mathcal{G}(x) + V$, where $V \sim \mathcal{N}(\mathbf{0}, \gamma)$ is the Gaussian noise. Then the random noise level $\widetilde{\eta}$ is defined as

$$\widetilde{\eta} := \left\|\gamma^{-1/2}\left(\xi - \mathcal{G}(x)\right)\right\|_{\mathbb{R}^n} = \chi_n \tag{3.2.10}$$

where χ_n^2 is a chi-square random variable with degree of freedom n.

Similar like using the Morozov's discrepancy principle to compare a point estimate with the deterministic noise level, we can also compare a random estimate with the random noise level. Given a random estimate $\widetilde{x}$ of the unknown parameter x, the estimated data-misfit $\left\|\gamma^{-1/2}\left(y - \mathcal{G}(\widetilde{x})\right)\right\|_{\mathbb{R}^n}$ can be compared relative to the random noise level $\widetilde{\eta} = \chi_n$. We propose the following discrepancy principles that measure the quality of a random estimate $\widetilde{x}$ (only for Gaussian noise):

Principle 1 check the mean of square norm of data misfit (accuracy test)

$$\mathbb{E}\left\{\left\|\gamma^{-1/2}\left(y - \mathcal{G}(\widetilde{x})\right)\right\|_{\mathbb{R}^n}^2\right\} \leq \mathbb{E}\left\{\chi_n^2\right\} = n \tag{3.2.11}$$

Principle 2 check the variance of square norm of data misfit (uncertainty test)

$$\mathrm{Var}\left\{\left\|\gamma^{-1/2}\left(y - \mathcal{G}(\widetilde{x})\right)\right\|_{\mathbb{R}^n}^2\right\} \leq \mathrm{Var}\left\{\chi_n^2\right\} = 2n \tag{3.2.12}$$

If either of the above two conditions holds, we postulate that the current estimate $\widetilde{x}$ has relatively high accuracy or relatively low uncertainty. Once either of the discrepancy principles occurs, we do not have to conduct further investigation anymore. Even if another random estimate $\widetilde{v}$ can be obtained from further work, e.g. $\mathbb{E}\left\{\left\|\gamma^{-1/2}\left(y - \mathcal{G}(\widetilde{v})\right)\right\|_{\mathbb{R}^n}^2\right\} \leq 0.1n$ or $\mathrm{Var}\left\{\left\|\gamma^{-1/2}\left(y - \mathcal{G}(\widetilde{v})\right)\right\|_{\mathbb{R}^n}^2\right\} \leq 0.2n$, this does not mean $\widetilde{v}$ is a better estimate than $\widetilde{x}$, because the observation y itself is mixed with a measurement error, which is assumed to be a Gaussian noise from $\mathcal{N}(0, \gamma)$. Since the existence of measurement error, the highest accuracy (assuming the truth x is known) of the predicted data misfit is characterized by the random noise level $\widetilde{\eta} := \left\|\gamma^{-1/2}\left(\xi - \mathcal{G}(x)\right)\right\|_{\mathbb{R}^n} = \chi_n$.

The adaptive strategy: controlling the mean-variance pair

Let $0 = t_0 < t_1 < \cdots < t_K = 1$ be any $K+1$ points in interval $[0,1]$, and let $h_i = t_i - t_{i-1}$ be the ith step size for any $i \in \{1, ..., K\}$. The adaptive strategy monitors random estimates in iterations. For the ith iteration, we need to deal with the ith sub-problem shown in formula (3.1.11). Using the Bayes' rule, we can derive the update formula (3.1.16) of the ith sub-problem. In the update formula (3.1.16), the sub-prior measure ν_{i-1} is known as the result from the last iteration, and the sub-posterior measure ν_i is the target in the current iteration. Clearly, the sub-posterior ν_i relies on two facts. One is the sub-prior ν_{i-1}, and another is the sub-noise $V_i \sim \mathcal{N}(\mathbf{0}, h_i^{-1}\gamma)$. Our purpose is to control the difference between ν_{i-1} and ν_i. Since the amount of the sub-noise V_i depends on the step size h_i, we can manually pick a small value of h_i, so that, the amount of the sub-noise V_i is large, and that the sub-posterior ν_i only has little difference from the sub-prior ν_{i-1}.

Our main idea is to use the statistical discrepancy principles (3.2.11) and (3.2.12) to determine the step size h_i, such that, the sub-prior ν_{i-1} is sufficiently accurate or concentrated relative to the mount of the sub-noise V_i. Namely, we use $h_i^{-1}\gamma$ to replace γ in formula (3.2.11) and (3.2.12) to obtain

$$\mathbb{E}\left\{\left\|h_i^{1/2}\gamma^{-1/2}\left(y - \mathcal{G}(\widetilde{x}_{i-1})\right)\right\|_{\mathbb{R}^n}^2\right\} \leq n \tag{3.2.13}$$

$$\mathrm{Var}\left\{\left\|h_i^{1/2}\gamma^{-1/2}\left(y - \mathcal{G}(\widetilde{x}_{i-1})\right)\right\|_{\mathbb{R}^n}^2\right\} \leq 2n \tag{3.2.14}$$

where $\widetilde{x}_{i-1} \sim \nu_{i-1}$ is a random estimate obeying the sub-prior distribution ν_{i-1}. As the result, the step size h_i can be chosen as the maximum value such that either of the condition (3.2.13) or (3.2.14) holds, and also the next parameter $t_i = t_{i-1} + h_i$ should not be greater than 1. Namely, h_i is determined by

$$h_i = \min\left\{\max\left\{\eta \Big/ q_{i-1}^{(1)},\ \sqrt{\eta \Big/ q_{i-1}^{(2)}}\right\},\ 1 - t_{i-1}\right\} \tag{3.2.15}$$

where $\eta = n/2$ and

$$q_{i-1}^{(1)} = \mathbb{E}\left\{\frac{1}{2}\left\|\gamma^{-1/2}\left(y - \mathcal{G}(\widetilde{x}_{i-1})\right)\right\|_{\mathbb{R}^n}^2\right\} \tag{3.2.16}$$

$$q_{i-1}^{(2)} = \mathrm{Var}\left\{\frac{1}{2}\left\|\gamma^{-1/2}\left(y - \mathcal{G}(\widetilde{x}_{i-1})\right)\right\|_{\mathbb{R}^n}^2\right\} \tag{3.2.17}$$

We call the adaptive method (3.2.15) the *data-misfit controller* for discretization of the tempering setting. 'Data-misfit controller' means it controls the mean and variance of the data misfit. In brief words, the data-misfit controller is a **stepwise regularization** method that regularizes an ill-posed inverse problem with a sequence of sub-problems iteratively, and assesses the quality of estimates in iterations by monitoring the mean-variance pair, such that, the estimates in iterations have relatively high accuracy or relatively low uncertainty.

3.2.3 The $\mathrm{D}_{KL,2}$ quantity

This subsection addresses the main idea of **stepwise learning** from the tempering setting. Namely, the (approximately) same amount of information is learned from the last state to the next state in iterations. Readers are encouraged to read the short introduction (subsection 3.2.1), in order to know the background and motivations.

There are two terminologies in this thesis: *stepwise learning* and *stepwise regularization*. These two concepts are closely related but different. On one hand, stepwise regularization has been discussed in the last subsection, that means using the statistical discrepancy principle to regularize each of the sub-problems. On the other hand, stepwise learning has been introduced in subsection 3.2.1, and will be discussed in details in this subsection, that means the same amount of information is learned from each of the sub-problems. Nevertheless, we will show that the stepwise regularization (using discrepancy principle) is an approximation of the stepwise learning (using information gain).

The analytic formula (implicit form) of the $\mathrm{D}_{KL,2}$ quantity

Now, we derivative the analytic formula representing the $\mathrm{D}_{KL,2}$ quantity. The following proposition tells that the $\mathrm{D}_{KL,2}$ quantity in formula (3.2.5) is well-defined and can be represented by the expected values of the cost functional, i.e. the sum of forward and backward information gain equals to the multiplication of step size and decrement of average cost functional.

Proposition 3.2.2. *Let* $0 = t_0 < t_1 < \cdots < t_K = 1$ *be any* $K+1$ *points in interval* $[0,1]$*, and let* $h_i = t_i - t_{i-1}$ *be the ith step size for any* $i \in \{1, ..., K\}$*. Assume that the*

forward map $\mathcal{G}$ satisfies condition 1 in assumption 2.2.3. Then, the probability measures $\{\mu_{t_i} : i = 0, 1, ..., N\}$ determined via the tempering setting (3.1.21) *are equivalent. Moreover, for any $i = 1, ..., K$, the $\mathrm{D}_{KL,2}$ quantity in formula* (3.2.5) *is well-defined, and can be represented as follows*

$$\mathrm{D}_{KL,2}(\mu_{t_{i-1}}, \mu_{t_i}) = h_i \langle \Phi \rangle_{t_{i-1}} - h_i \langle \Phi \rangle_{t_i} \tag{3.2.18}$$

where Φ is the cost functional defined in formula (3.1.19)*, and $\langle \Phi \rangle_t$ is the expected value $\langle \Phi \rangle_t \equiv \int_{\mathcal{H}} \Phi(u)\, \mu_t(\mathrm{d}u)$ for any $t \in [0, 1]$.*

Proof. Proposition 4.2.3 in section 4.2 proves the equivalence of any two probability measures. Since the probability measures $\{\mu_{t_i} : i = 0, 1, ..., K\}$ are equivalent, the $\mathrm{D}_{KL,2}$ quantity is well-defined. Next, we show the $\mathrm{D}_{KL,2}$ quantity can be represented by equation (3.2.18). For any $i = 1, ..., K$, the two probability measures $\mu_{t_{i-1}}$ and μ_{t_i} satisfy

$$\frac{\mathrm{d}\mu_{t_i}}{\mathrm{d}\mu_{t_{i-1}}}(u) = A_i^{-1} \exp\left(-h_i \Phi(u)\right) \tag{3.2.19}$$

where A_i is the normalizing constant

$$A_i = \int_{\mathcal{H}} \exp\left(-h_i \Phi(u)\right)\, \mu_{i-1}(\mathrm{d}u) \tag{3.2.20}$$

Thus, the $\mathrm{D}_{KL,2}$ quantity can be calculated as follows,

$$\begin{aligned}
\mathrm{D}_{KL,2}(\mu_{t_{i-1}}, \mu_{t_i}) =& \mathrm{D}_{KL}(\mu_{t_{i-1}} || \mu_{t_i}) + \mathrm{D}_{KL}(\mu_{t_i} || \mu_{t_{i-1}}) \\
=& \int_{\mathcal{H}} \log\left(\frac{\mathrm{d}\mu_{t_{i-1}}}{\mathrm{d}\mu_{t_i}}\right) \mathrm{d}\mu_{t_{i-1}} + \int_{\mathcal{H}} \log\left(\frac{\mathrm{d}\mu_{t_i}}{\mathrm{d}\mu_{t_{i-1}}}\right) \mathrm{d}\mu_{t_i} \\
=& \int_{\mathcal{H}} \log\left(A_i \exp\left(h_i \Phi(u)\right)\right)\, \mu_{t_{i-1}}(\mathrm{d}u) + \int_{\mathcal{H}} \log\left(A_i^{-1} \exp\left(-h_i \Phi(u)\right)\right)\, \mu_{t_i}(\mathrm{d}u) \\
=& \log(A_i) + h_i \int_{\mathcal{H}} \Phi(u)\, \mu_{t_{i-1}}(\mathrm{d}u) - \log(A_i) - h_i \int_{\mathcal{H}} \Phi(u)\, \mu_{t_i}(\mathrm{d}u) \\
=& h_i \langle \Phi \rangle_{t_{i-1}} - h_i \langle \Phi \rangle_{t_i}
\end{aligned}$$

□

Though formula (3.2.18) shows an analytic formula of the $\mathrm{D}_{KL,2}$ quantity, it seems non-practical for numerical implementation because μ_{t_i} is an implicit quantity at time t_{i-1}. Thus, some approximate methods, which only relies on information at time t_{i-1}, should be considered to approximate the $\mathrm{D}_{KL,2}$ quantity. This approximation is discussed as follows.

The approximate formula (explicit form) of the $\mathrm{D}_{KL,2}$ quantity

In practice, need to estimate the implicit formula (3.2.18) by some explicit approximations only relying on information at time t_{i-1}. Assume that the forward map $\mathcal{G}$ satisfies condition 1 in assumption 2.2.3. We use the following two approximate approaches in different circumstances:

1. If the step size $h_i = t_i - t_{i-1}$ in formula (3.2.18) is relatively large such that $\langle\Phi\rangle_{t_i} \ll \langle\Phi\rangle_{t_{i-1}}$, then $\langle\Phi\rangle_{t_i} \geq 0$ is ignored, and $\mathrm{D}_{KL,2}(\mu_{t_{i-1}}, \mu_{t_i}) = h_i \langle\Phi\rangle_{t_{i-1}} - h_i \langle\Phi\rangle_{t_i}$ is approximated by an upper bound $h_i \langle\Phi\rangle_{t_{i-1}}$.

2. If the step size $h_i = t_i - t_{i-1}$ in formula (3.2.18) is relatively small such that $\langle\Phi\rangle_{t_i}$ can be approximated by the first order expansion of $\langle\Phi\rangle_t$ (as a function of t) around $t = t_{i-1}$,
$$\langle\Phi\rangle_{t_i} \approx \langle\Phi\rangle_{t_{i-1}} + h_i \langle\Phi\rangle'_{t_{i-1}} \tag{3.2.21}$$
where $\langle\Phi\rangle'_t$ is the derivative of $\langle\Phi\rangle_t$ (as a function of t), then $\mathrm{D}_{KL,2}(\mu_{t_{i-1}}, \mu_{t_i}) = h_i \langle\Phi\rangle_{t_{i-1}} - h_i \langle\Phi\rangle_{t_i}$ is approximated by the first order expansion $-h_i^2 \langle\Phi\rangle'_{t_{i-1}}$. Furthermore, according to corollary 4.2.6 in section 4.2, for any $t \in [0,1]$ the first order derivative equals to
$$\langle\Phi\rangle'_t = -\langle\Phi,\Phi\rangle_t \tag{3.2.22}$$
where $\langle\Phi,\Phi\rangle_t$ is the variance $\langle\Phi,\Phi\rangle_t \equiv \int_{\mathcal{H}} (\Phi(u) - \langle\Phi\rangle_t)^2 \, \mu_t(\mathrm{d}u)$ for any $t \in [0,1]$.

As the result, the approximate formula is proposed as follows,
$$h_i \langle\Phi\rangle_{t_{i-1}} - h_i \langle\Phi\rangle_{t_i} \approx \begin{cases} h_i \langle\Phi\rangle_{t_{i-1}} & \text{if } h_i \text{ is large } (h_i \geq \frac{\langle\Phi\rangle_{t_{i-1}}}{\langle\Phi,\Phi\rangle_{t_{i-1}}}) \\ h_i^2 \langle\Phi,\Phi\rangle_{t_{i-1}} & \text{if } h_i \text{ is small } (h_i \leq \frac{\langle\Phi\rangle_{t_{i-1}}}{\langle\Phi,\Phi\rangle_{t_{i-1}}}) \end{cases} \tag{3.2.23}$$
where $h_i \langle\Phi\rangle_{t_{i-1}}$ is an upper bound and $h_i^2 \langle\Phi,\Phi\rangle_{t_{i-1}}$ is the first order approximation. In conclusion, we can combine the two cases in formula (3.2.23) together,
$$\mathrm{D}_{KL,2}(\mu_{t_{i-1}}, \mu_{t_i}) \approx \min\left\{ h_i \langle\Phi\rangle_{t_{i-1}}, \, h_i^2 \langle\Phi,\Phi\rangle_{t_{i-1}} \right\} \tag{3.2.24}$$

Therefore, formula (3.2.24) is an implementable formula that approximately estimates the $\mathrm{D}_{KL,2}$ quantity. This approximation is closely related to the mean-variance pair $\langle\Phi\rangle_{t_t}$–$\langle\Phi,\Phi\rangle_{t_t}$. With this approximate formula, we can develop the adaptive strategy as follow.

The adaptive strategy: controlling the $\mathrm{D}_{KL,2}$ quantity

The adaptive strategy aims to control the amount of information in the stepwise learning, i.e. the sum of forward and backward information gain is specified as a fixed number $\eta > 0$ in each step. With mathematical formulation, the $\mathrm{D}_{KL,2}$ quantity is controlled by

$$\mathrm{D}_{KL,2}(\mu_{t_{i-1}}, \mu_{t_i}) \leq \eta \tag{3.2.25}$$

where $\eta > 0$ is a user-specified accuracy control parameter that is the required maximum amount of information gain.

However, as we discussed before, the analytic formula (3.2.18) of the $\mathrm{D}_{KL,2}$ quantity is implicit. In practice, we adopt the approximate formula (3.2.24) which has an explicit form. Namely, the amount of information is approximately rather than accurately controlled as

$$\mathrm{D}_{KL,2}(\mu_{t_{i-1}}, \mu_{t_i}) \approx \min\left\{h_i \langle\Phi\rangle_{t_{i-1}}, h_i^2 \langle\Phi,\Phi\rangle_{t_{i-1}}\right\} \leq \eta \tag{3.2.26}$$

Conversely, the step size h_i can be selected as the maximum value such that condition (3.2.26) holds, and also the next parameter $t_i = t_{i-1} + h_i$ should not be greater than 1. Namely, h_i is determined by

$$h_i = \min\left\{\max\left\{\frac{\eta}{\langle\Phi\rangle_{t_{i-1}}}, \sqrt{\frac{\eta}{\langle\Phi,\Phi\rangle_{t_{i-1}}}}\right\}, 1 - t_{i-1}\right\} \tag{3.2.27}$$

Notice that, the above formula (3.2.27) has the same form as the data-misfit controller proposed in formula (3.2.15). In formula (3.2.15), the accuracy control parameter η is specified as $\eta = n/2$ because the noise level (data misfit) is quantified by (compared with) the χ^2_n-distribution. Similarly, we also suggest this choice $\eta = n/2$ in formula (3.2.27), that means, the sum of forward and backward information gain is no more than $n/2$.

The tricky point of method (3.2.27) is using the balance between the upper bound $h_i \langle\Phi\rangle_{t_{i-1}}$ and the first order approximation $h_i^2 \langle\Phi,\Phi\rangle_{t_{i-1}}$ to determine the step size h_i. If only consider one condition of the balance, it causes some issues. If only consider the upper bound, then when h_i is small, the upper bound is a bad estimate, since $\langle\Phi\rangle_{t_i}$ is close to $\langle\Phi\rangle_{t_{i-1}}$, and thus $\langle\Phi\rangle_{t_i}$ cannot be ignored. If only consider the first order approximation, then when h_i is large, the approximation is too rough, since the first order expansion around t_{i-1} does not hold for a far point $t_i = t_{i-1} + h_i$. Nevertheless,

combining the upper bound and the first order approximation together can conquer the drawbacks of each other.

3.2.4 The thermodynamic integration

In Bayesian inference, a goal is to calculate the normalizing constant. A technique raised from the tempering setting is known as the thermodynamic integration [6], which gradually calculates the normalizing constant along path sampling. Thermodynamic integration is originally studied in statistical physics. Statisticians adopt the similar form to deal with Bayesian inference problems. For more details and the generalization to infinite-dimensional framework, please in section 4.2. This subsection aims to reveal that the $\mathrm{D}_{KL,2}$ quantity is closely related to finite difference scheme of the thermodynamic integration. That means, controlling the $\mathrm{D}_{KL,2}$ quantity is equivalent to controlling the finite difference error of the thermodynamic integration.

The normalizing constant and thermodynamic integration

Let $\mathbb{P} = \mathcal{N}(m_0, \mathcal{C}_0)$ be the Gaussian prior probability on a real-valued separable Hilbert space $\mathcal{H}$. Via the original Bayesian approach (2.2.5), the posterior probability measure is determined by

$$\mathbb{P}(\mathrm{d}u|y) = \frac{1}{Z(y)} \exp\left(-\frac{1}{2}\left\|\gamma^{-1/2}(y - \mathcal{G}(u))\right\|_{\mathbb{R}^n}^2\right) \mathbb{P}(\mathrm{d}u) \tag{3.2.28}$$

where $Z(y)$ is the normalizing constant

$$Z(y) = \int_{\mathcal{H}} \exp\left(-\frac{1}{2}\left\|\gamma^{-1/2}(y - \mathcal{G}(u))\right\|_{\mathbb{R}^n}^2\right) \mathbb{P}(\mathrm{d}u) \tag{3.2.29}$$

On the other hand, via the tempered Bayesian approach (3.1.21), the tempered distributions are determined by

$$\mu_t = \frac{1}{Z_t} \exp\left(-t\Phi(u)\right) \mathbb{P}(\mathrm{d}u) \tag{3.2.30}$$

where Φ is the cost functional determined in formula (3.1.19), and Z_t is the normalizing constant

$$Z_t = \int_{\mathcal{H}} \exp\left(-t\Phi(u)\right) \mathbb{P}(\mathrm{d}u) \tag{3.2.31}$$

It is obvious that Z_t is decreasing in $t \in [0,1]$ with $Z_0 = 1$ and $Z_1 = Z(y)$. Moreover, according to theorem 3.1.3, $Z(y)$ can be represented by the thermodynamic integration along the path from $t = 0$ to $t = 1$,

$$-\log(Z(y)) = -\log(Z_1) = I_1 = \int_0^1 \langle \Phi \rangle_t \, \mathrm{d}t \tag{3.2.32}$$

where I_t and $\langle \Phi \rangle_t$ are the same in formula (3.1.22) for any $t \in [0,1]$. Clearly, the calculation of $Z(y)$ via formula (3.2.32) is different from the calculation via formula (3.2.29). The motivation of why the integration approach (3.2.32) is more preferred than the original approach (3.2.29) is simply addressed as follows.

Directly using formula (3.2.29) means conduct Monte Carlo integration under the prior probability measure $\mathbb{P}$ with the weighting function, i.e.

$$Z(y) \approx \frac{1}{J} \sum_{j=1}^{J} w_j \quad \text{with} \quad w_j = \exp\left(-\frac{1}{2} \left\| \gamma^{-1/2} \left(y - \mathcal{G}\left(u_0^{(j)} \right) \right) \right\|_{\mathbb{R}^n}^2 \right) \tag{3.2.33}$$

where J is the sample size, $\{w_j : j = 1, ..., J\}$ is the set of weights, and $\left\{ u_0^{(j)} : j = 1, ..., J \right\}$ is the set of samples independently drawn from the prior distribution $\mathbb{P} = \mathcal{N}(m_0, \mathcal{C}_0)$. However, the weights usually form a very sharp distribution with most of the weights equaling to (nearly) zeors, so that the effective sample size J_{eff} is very small relative to J ($J_{eff} \ll J$), where the effective sample size is known as the Kish's Effective Sample Size [59] for weighted data,

$$J_{eff} = \frac{\left(\sum_{j=1}^{J} w_j \right)^2}{\sum_{j=1}^{J} w_j^2} \tag{3.2.34}$$

To obtain a required minimum effective sample size $\widehat{J}$ with $J_{eff} \geq \widehat{J}$, a extremely large number J of samples is needed. Thus, the direct approach is inefficient.

On the other hand, sequential Monte Carlo simulation along the path of thermodynamic integration is a more advanced technique that conducts path sampling. Let $0 = t_0 < t_1 < \cdots < t_K = 1$ be any $K+1$ points in interval $[0,1]$, and let $h_i = t_i - t_{i-1}$ be the ith step size for any $i \in \{1, ..., K\}$. Sequential Monte Carlo method draws a sequence of samples as

$$\left\{ u_{t_0}^{(j)} \right\}_{j=1}^{J} \to \left\{ u_{t_1}^{(j)} \right\}_{j=1}^{J} \to \cdots \to \left\{ u_{t_K}^{(j)} \right\}_{j=1}^{J} \tag{3.2.35}$$

where $\left\{u_t^{(j)} : j = 1, ..., J\right\}$ is the set of samples independently drawn from the distribution μ_t for any $t \in \{t_0, t_1, ..., t_K\}$. Then, the integral (3.2.32) is estimated by

$$S_{K,0} \geq -\log(Z(y)) \geq S_{K,1} \tag{3.2.36}$$

where $S_{K,0}$ and $S_{K,1}$ are the forward and backward finite difference schemes,

$$S_{K,0} := \sum_{i=1}^{K} h_i \left\langle \Phi \right\rangle_{t_{i-1}} \tag{3.2.37}$$

$$S_{K,1} := \sum_{i=1}^{K} h_i \left\langle \Phi \right\rangle_{t_i} \tag{3.2.38}$$

and for any $t \in \{t_0, t_1, ..., t_K\}$. The inequality (3.2.36) holds because the average cost functional $\left\langle \Phi \right\rangle_t$ as a function of $t \in [0, 1]$ is always decreasing. Moreover, the expected value $\left\langle \Phi \right\rangle_t$ in formula (3.2.36) can be approximated by the Monte Carlo method

$$\left\langle \Phi \right\rangle_t \approx \frac{1}{J} \sum_{j=1}^{J} \Phi\left(u_t^{(j)}\right) \tag{3.2.39}$$

Numerically, the sequential sampling method is more efficient than the directly sampling from the prior probability.

Finite difference error v.s. information gain

Here, we aim to show the main result that controlling the Kullback-Leibler divergence of the tempering setting is equivalent to controlling the finite difference error of the thermodynamic integration.

The local forward and backward finite difference error are defined as $\delta_{i,0}$ and $\delta_{i,1}$ respectively,

$$\delta_{i,0} := h_i \left\langle \Phi \right\rangle_{t_{i-1}} - \int_{t_{i-1}}^{t_i} \left\langle \Phi \right\rangle_t \, \mathrm{d}t \tag{3.2.40}$$

$$\delta_{i,1} := \int_{t_{i-1}}^{t_i} \left\langle \Phi \right\rangle_t \, \mathrm{d}t - h_i \left\langle \Phi \right\rangle_{t_i} \tag{3.2.41}$$

The following proposition tells that, the forward (backward) finite difference error of the thermodynamic integration equals to the backward (forward) information gain of the sequential probability measures.

Proposition 3.2.3. *Let $0 = t_0 < t_1 < \cdots < t_K = 1$ be any $K+1$ points in interval $[0,1]$, and let $h_i = t_i - t_{i-1}$ be the ith step size for any $i \in \{1, ..., K\}$. Assume that the forward map $\mathcal{G}$ satisfies condition 1 in assumption 2.2.3. Then, the probability measures $\{\mu_{t_i} : i = 0, 1, ..., N\}$ determined via the tempering setting (3.1.21) are equivalent. Moreover, for any $i = 1, ..., K$, both the forward information gain $D_{KL}(\mu_{t_i}||\mu_{t_{i-1}})$ and the backward information gain $D_{KL}(\mu_{t_{i-1}}||\mu_{t_i})$ are well-defined, and they equal to the backward finite difference error $\delta_{i,1}$ and the forward finite difference error $\delta_{i,0}$ respectively,*

$$D_{KL}(\mu_{t_i}||\mu_{t_{i-1}}) = \delta_{i,1} \tag{3.2.42}$$

$$D_{KL}(\mu_{t_{i-1}}||\mu_{t_i}) = \delta_{i,0} \tag{3.2.43}$$

where $\delta_{i,0}$ and $\delta_{i,1}$ are defined in formulas (3.2.40) and (3.2.41), respectively.

Proof. This proof a rewritten version of the proof in proposition 3.2.2. In fact, we have

$$\begin{aligned}
D_{KL}(\mu_{t_{i-1}}||\mu_{t_i}) &= \int_{\mathcal{H}} \log\left(\frac{\mathrm{d}\mu_{t_{i-1}}}{\mathrm{d}\mu_{t_i}}\right) \mathrm{d}\mu_{t_{i-1}} \\
&= \int_{\mathcal{H}} \log\left(\frac{\exp(-t_{i-1}\Phi(u))/Z_{t_{i-1}}}{\exp(-t_i\Phi(u))/Z_{t_i}}\right) \mu_{t_{i-1}}(\mathrm{d}u) \\
&= \log(Z_{t_i}) - \log(Z_{t_{i-1}}) + h_i \int_{\mathcal{H}} \Phi(u)\, \mu_{t_{i-1}}(\mathrm{d}u) \\
&= -\int_{t_{i-1}}^{t_i} \langle\Phi\rangle_t \, \mathrm{d}t + h_i \langle\Phi\rangle_{t_{i-1}} = \delta_{i,0} \\
D_{KL}(\mu_{t_i}||\mu_{t_{i-1}}) &= \int_{\mathcal{H}} \log\left(\frac{\mathrm{d}\mu_{t_i}}{\mathrm{d}\mu_{t_{i-1}}}\right) \mathrm{d}\mu_{t_i} \\
&= \int_{\mathcal{H}} \log\left(\frac{\exp(-t_i\Phi(u))/Z_{t_i}}{\exp(-t_{i-1}\Phi(u))/Z_{t_{i-1}}}\right) \mu_{t_i}(\mathrm{d}u) \\
&= \log(Z_{t_{i-1}}) - \log(Z_{t_i}) - h_i \int_{\mathcal{H}} \Phi(u)\, \mu_{t_i}(\mathrm{d}u) \\
&= \int_{t_{i-1}}^{t_i} \langle\Phi\rangle_t \, \mathrm{d}t - h_i \langle\Phi\rangle_{t_i} = \delta_{i,1}
\end{aligned}$$

□

Then, the $\mathrm{D}_{KL,2}$ quantity can be also represented as the sum of the forward and backward finite difference errors,

$$\mathrm{D}_{KL,2}(\mu_{t_{i-1}}, \mu_{t_i}) = \delta_{i,0} + \delta_{i,1} \tag{3.2.44}$$

In fact, since $\langle\Phi\rangle_t$ is a monotone function (decreasing as t goes up), the sum of forward and backward finite difference errors is the upper bound of all possible finite difference errors. Namely, for any $s_i \in [t_{i-1}, t_i]$, we have

$$h_i \langle\Phi\rangle_{s_i} \leq \delta_{i,0} + \delta_{i,1} = \mathrm{D}_{KL,2}(\mu_{t_{i-1}}, \mu_{t_i}) \tag{3.2.45}$$

Thus, if the $\mathrm{D}_{KL,2}$ quantity is controlled, then the finite difference error is also controlled. Therefore, the data-misfit controller in formula (3.2.15) or (3.2.27) can be also treated as an adaptive scheme discretizing the thermodynamic integration (3.2.32). With this adaptive scheme, the local difference error of the thermodynamic integration is controlled.

3.3 Kalman-like methods for inverse problems

The standard perspectives of inverse problems are the variational approach and the Bayesian approach. With a slight modification, this thesis introduces the tempering setting. In fact, as discussed in section 3.1, an inverse problem with the tempering setting can be regarded as a filtering problem. Thus, approximate filtering algorithms (e.g. extended Kalman filter and ensemble Kalman filter) can be applied for solving the inverse problem. This procedure finally leads to Kalman-like methods for inverse problems. The details of Kalman-like methods are discussed in this section.

3.3.1 Gaussian-linear problems and Kalman filter

This subsection considers the simplest type of filtering problems with Gaussian-linear assumptions, which have the analytic solution known as the Kalman filter. The Kalman filter used for linear filtering problems is the pillar for further study in nonlinear problems. Thus, it is important to introduce the Kalman filter clearly, and then explain how to apply the Kalman filter on inverse problems with the tempering setting.

The standard form of Kalman filter for Bayesian filtering

We first of all introduce the standard form of Kalman filter used for Bayesian filtering problems. The general structure of Bayesian filtering is introduced in subsection 3.1.1. In particular, Gaussian-linear problems restrict the conditions of filtering as follows:

Assumption 3.3.1. *(Gaussian-linear assumption)*

1. *The probability measure of the initial state X_0 is a Gaussian measure $\mathcal{N}(m_0, \mathcal{C}_0)$.*

2. *The state noise $W_i \sim \mathcal{N}(d_i, \mathcal{Q}_i)$ and observation noise $V_i \sim \mathcal{N}(b_i, \gamma_i)$ in equations* (3.1.1) (3.1.2) *are independent Gaussian noises, where $d_i \in \mathcal{H}$ is a bounded element, $\mathcal{Q}_i : \mathcal{H} \to \mathcal{H}$ is a self-adjoint positive-semi-definite trace-class operator, $b_i \in \mathbb{R}^n$ is a bounded element, and $\gamma_i : \mathbb{R}^n \to \mathbb{R}^n$ is a symmetric positive-definite bounded matrix.*

3. *The evolution model F_i in equation* (3.1.1) *and the observation model G_i in equation* (3.1.2) *are bounded linear operators.*

For Gaussian-linear filtering, there exists the analytic solution, which is determined via the Kalman filter in the following theorem.

Theorem 3.3.2. *(Kalman filter, Theorem 4.3 in [49]) If the Gaussian-linear assumptions 3.3.1 hold, then the conditional probability measure $\mathbb{P}_{X_i}(\cdot|\mathcal{S}_i)$ defined in* (3.1.4) *are Gaussian measures for all $i = 1, ..., K$, with mean $m_i \in \mathcal{H}$ and covariance $\mathcal{C}_i : \mathcal{H} \to \mathcal{H}$ determined via the following update formulas*

$$m_i = m_i^{(p)} + \mathcal{C}_i^{(p)} G_i^* (G_i \mathcal{C}_i^{(p)} G_i^* + \gamma_i)^{-1} (y_i - b_i - G_i m_i^{(p)}) \tag{3.3.1}$$

$$\mathcal{C}_i = \mathcal{C}_i^{(p)} - \mathcal{C}_i^{(p)} G_i^* (G_i \mathcal{C}_i^{(p)} G_i^* + \gamma_i)^{-1} G_i \mathcal{C}_i^{(p)} \tag{3.3.2}$$

where $m_i^{(p)} \in \mathcal{H}$ and $\mathcal{C}_i^{(p)} : \mathcal{H} \to \mathcal{H}$ are the predicted mean and covariance,

$$m_i^{(p)} = F_i m_{i-1} + d_i \tag{3.3.3}$$

$$\mathcal{C}_i^{(p)} = F_i \mathcal{C}_{i-1} F_i^* + \mathcal{Q}_i \tag{3.3.4}$$

Remark 3.3.3. *Theorem 3.3.2 is an analogue of theorem 2.2.1. Both are the analytic solutions for Gaussian-linear problems. The different is that theorem 2.2.1 discusses the one-step transition from the prior to the posterior, whereas theorem 3.3.2 discusses iterative updates with sequential observations and states.*

Variant forms of Kalman method for the tempering setting

Now, remind the tempering setting for inverse problems. We consider the linear case. Namely, assume that the forward map $\mathcal{G}$ is an affine map, and then the data-misfit function $\mathcal{Z}$ defined in formula (3.1.19) can be rewritten as

$$\mathcal{Z}(\cdot) = \mathcal{K}(\cdot) + c \tag{3.3.5}$$

where $c \in \mathbb{R}^n$ is a bounded element, and $\mathcal{K} : \mathcal{H} \to \mathbb{R}^n$ a bounded linear operator. An inverse problem with the tempering setting can be regarded as a Bayesian filtering problem whose observations and underlying states are invariant. For linear problems, there exists the closed form of solutions (Kalman filter). The following propositions list different but equivalent forms of Kalman method for solving linear inverse problems with the tempering setting. It is called the Kalman method, in order to indicate that Kalmam filter is applied on the tempering setting.

There are four different forms: discrete update of mean-covariance pairs, discrete update of random variables, continuous update of mean-covariance pairs, and continuous update of random variables. Proposition 3.3.4 straightforwardly applies theorem 3.3.2 or theorem 2.2.1. Proposition 3.3.5 rewrites proposition 3.3.4 by replacing mean-covariance pairs of Gaussian measures with Gaussian random variables. Propositions 3.3.6 and 3.3.7 revel the continuous limits of the finite difference equations in propositions 3.3.4 and 3.3.5, respectively. Conversely, the analytic solutions of the ODE and SDE in propositions 3.3.6 and 3.3.7 exactly equal to the formulas in propositions 3.3.4 and 3.3.5, respectively.

Proposition 3.3.4 (discrete Kalman method of mean-covariance update)**.** *Let* $0 = t_0 < t_1 < \cdots < t_K = 1$ *be any* $K + 1$ *points. If the operator* $\mathcal{Z}$ *is represented by* (3.3.5)*, then for any* $i = 1, ..., K$*, the conditional distribution* μ_{t_i} *determined via the tempering setting* (3.1.21) *is a Gaussian measure with mean* $m_{t_i} \in \mathcal{H}$ *and covariance operator* $\mathcal{C}_{t_i} : \mathcal{H} \to \mathcal{H}$ *determined by the ordinary difference equations below (let* $h_i = t_i - t_{i-1}$ *be the step size),*

$$m_{t_i} = m_{t_{i-1}} - h_i \mathcal{C}_{t_{i-1}} \mathcal{K}^* \left(\mathbf{I} + h_i \mathcal{K} \mathcal{C}_{t_{i-1}} \mathcal{K}^* \right)^{-1} \mathcal{Z} \left(m_{t_{i-1}} \right) \tag{3.3.6}$$

$$\mathcal{C}_{t_i} = \mathcal{C}_{t_{i-1}} - h_i \mathcal{C}_{t_{i-1}} \mathcal{K}^* \left(\mathbf{I} + h_i \mathcal{K} \mathcal{C}_{t_{i-1}} \mathcal{K}^* \right)^{-1} \mathcal{K} \mathcal{C}_{t_{i-1}} \tag{3.3.7}$$

Proposition 3.3.5 (discrete Kalman method of random variable update)**.** *Let* $0 = t_0 < t_1 < \cdots < t_K = 1$ *be any* $K+1$ *points. If the operator* $\mathcal{Z}$ *is represented by* (3.3.5)*, then for any* $i = 1, ..., K$*,* u_{t_i} *admits* $u_{t_i} \sim \mathcal{N}(m_{t_i}, \mathcal{C}_{t_i})$*, where* m_{t_i} $\mathcal{C}_{t_i}$ *are determined by formulas* (3.3.6) (3.3.7)*, and* $\{u_{t_i}\}$ *is an* $\mathcal{H}$*-valued discrete stochastic process constructed as: for* $i = 0$*,* $u_0 \sim \mathcal{N}(m_0, \mathcal{C}_0)$ *is a Gaussian random variable, and for any* $i = 1, ..., K$*,* u_{t_i} *is determined by the stochastic difference equation below (let* $h_i = t_i - t_{i-1}$ *be the step size),*

$$u_{t_i} = u_{t_{i-1}} - \mathcal{C}_{t_{i-1}}\mathcal{K}^* \left(\mathbf{I} + h_i \mathcal{K}\mathcal{C}_{t_{i-1}}\mathcal{K}^*\right)^{-1} \left(h_i \mathcal{Z}\left(u_{t_{i-1}}\right) - \sqrt{h_i}\zeta_i\right) \tag{3.3.8}$$

where $\{\zeta_i\}_{i=1}^K$ *is a set of* K *independent* n*-dimensional standard Gaussian random variables.*

Proposition 3.3.6 (continuous Kalman method of mean-covariance update)**.** *If the operator* $\mathcal{Z}$ *is represented by* (3.3.5)*, then for any* $t \in (0, 1]$*, the conditional distribution* μ_t *determined via the tempering setting* (3.1.21) *is a Gaussian measure with mean* $m_t \in \mathcal{H}$ *and covariance operator* $\mathcal{C}_t : \mathcal{H} \to \mathcal{H}$ *determined by the ordinary differential equations below,*

$$\mathrm{d}m_t = -\mathcal{C}_t \mathcal{K}^* \mathcal{Z}(m_t)\mathrm{d}t \tag{3.3.9}$$

$$\mathrm{d}\mathcal{C}_t = -\mathcal{C}_t \mathcal{K}^* \mathcal{K} \mathcal{C}_t \mathrm{d}t \tag{3.3.10}$$

Proposition 3.3.7 (continuous Kalman method of random variable update)**.** *If the operator* $\mathcal{Z}$ *is represented by* (3.3.5)*, then for any* $t \in (0, 1]$*,* u_t *admits* $u_t \sim \mathcal{N}(m_t, \mathcal{C}_t)$*, where* m_t $\mathcal{C}_t$ *are determined by formulas* (3.3.9) (3.3.10)*, and* $\{u_t\}$ *is an* $\mathcal{H}$*-valued continuous stochastic process constructed as: for* $t = 0$*,* $u_0 \sim \mathcal{N}(m_0, \mathcal{C}_0)$ *is a Gaussian random variable, and for any* $t \in (0, 1]$*,* u_t *is determined by the stochastic differential equation below,*

$$\mathrm{d}u_t = -\mathcal{C}_t \mathcal{K}^* \left(\mathcal{Z}\left(u_t\right)\mathrm{d}t - \mathrm{d}W_t\right) \tag{3.3.11}$$

where W_t *is an* n*-dimensional standard Wiener process on* $t \in (0, 1]$*.*

For non-Gaussian nonlinear problems, there is no a closed form. In practice, numerical sampling algorithms like MCMC methods can be applied. However, the accurate sampling algorithms are numerically expensive. Alternatively, other candidates are heuristic

methods like extended Kalman filter (EKF) and ensemble Kalman filter (EnKF), which are also feasible in practice with benefit in computational budget.

3.3.2 Formulation of the Kalman-like methods

Consider the tempering setting shown in formulas (3.1.20) and (3.1.21). Since the tempering setting can be equivalently regarded as filtering algorithms, extended Kalman filter (EKF) and ensemble Kalman filter (EnKF) can be applied as heuristic methods to solve the tempering setting with nonlinear forward maps. There will be six different Kalman-like methods formulated in this subsection. We classify the six methods into three groups, and in each group, we have a continuous formula and a discrete formula. The three groups are extended Kalman inversion (EKI) [91, 30], mean-field limiting ensemble Kalman inversion (MFEnKI) [54], and standard ensemble Kalman inversion (EnKI) [38, 33, 65, 20, 78, 77]. The relations of the three groups are

- EKI approximates forward model by its first order Taylor expansion, so that the derivative of forward model is required. In comparison, MFEnKI and EnKI are derivative-free methods only relying on point values of forward model.
- MFEnKI constructs update of random variables, whereas EnKI constructs update of particles. More precisely, as discussed in [54], MFEnKI is the theoretical limit of EnKI as the sample size goes to infinity, and MFEnKI is the empirical sampling of EnKI with finite sample size.

For convenience, the six methods are named as the short notation:

- EKI
  - **Co**ntinuous **E**xtended **K**alman **I**nversion (CoEKI)
  - **Di**screte **E**xtended **K**alman **I**nversion (DiEKI)
- MFEnKI
  - **Co**ntinuous **M**ean-**F**ield limiting **En**semble **K**alman **I**nversion (CoMFEnKI)
  - **Di**screte **M**ean-**F**ield limiting **En**semble **K**alman **I**nversion (DiMFEnKI)

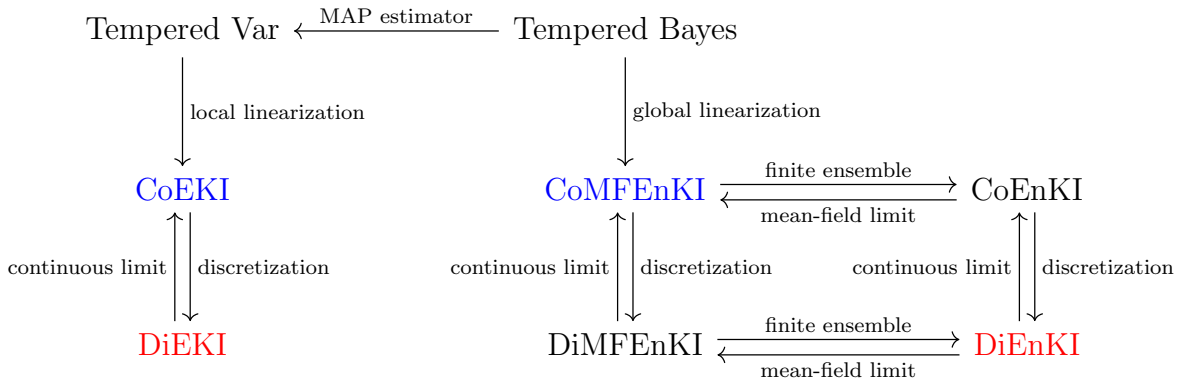


Figure 3.2: Diagram of the Kalman-like methods for inverse problems. There are totally six methods: CoEKI (definition 3.3.8), DiEKI (definition 3.3.9), CoMFEnKI (definition 3.3.10), DiMFEnKI (definition 3.3.11), CoEnKI (definition 3.3.12), DiEnKI (definition 3.3.13). The red color for the two methods DiEKI and DiEnKI implies these two are the numerically implementable methods. The blue color for the two methods CoEKI and CoMFEnKI implies these two are the theoretical limits with the higheast accuracy.

- EnKI
  - **Co**ntinuous standard **En**semble **K**alman **I**nversion (CoEnKI)
  - **Di**screte standard **En**semble **K**alman **I**nversion (DiEnKI)

The relations among these six methods can be simply viewed as a diagram in figure 3.2.

Extended Kalman inversion

EKI is a heuristic method linearizing a forward map by its first order Taylor expansion. As the result, solving the linearized problem can obtain a sub-optimum. The approximate solution is determined as follows.

Definition 3.3.8 (CoEKI). *CoEKI aims to interpret the mean-covariance pair* $\left(\widetilde{m}_t, \widetilde{\mathcal{C}}_t\right)$ *for* $t \in [0,1]$*, such that, for* $t = 0$*,* $\left(\widetilde{m}_0, \widetilde{\mathcal{C}}_0\right) = (m_0, \mathcal{C}_0)$*, and for* $t \in (0,1]$*,* $\left(\widetilde{m}_t, \widetilde{\mathcal{C}}_t\right)$ *are determined by the ordinary differential equations below,*

$$\mathrm{d}\widetilde{m}_t = -\widetilde{\mathcal{C}}_t \mathrm{D}\mathcal{Z}\left(\widetilde{m}_t\right)^* \mathcal{Z}\left(\widetilde{m}_t\right) \mathrm{d}t \tag{3.3.12}$$

$$\mathrm{d}\widetilde{\mathcal{C}}_t = -\widetilde{\mathcal{C}}_t \mathrm{D}\mathcal{Z}\left(\widetilde{m}_t\right)^* \mathrm{D}\mathcal{Z}\left(\widetilde{m}_t\right) \widetilde{\mathcal{C}}_t \mathrm{d}t \tag{3.3.13}$$

where $\mathcal{Z}:\mathcal{H}\to\mathbb{R}^n$ *is the data-misfit function defined in formula* (3.1.19)*, and* $\mathrm{D}\mathcal{Z}(u):\mathcal{H}\to\mathbb{R}^n$ *is the Fréchet derivative of* $\mathcal{Z}$ *at* $u\in\mathcal{H}$*.*

Definition 3.3.9 (DiEKI)**.** *DiEKI aims to interpret the mean-covariance pair* $\left(\widetilde{m}_t^h,\widetilde{\mathcal{C}}_t^h\right)$ *for* $t\in\{t_i:0=t_0<t_1<\cdots<t_K=1\}$*, such that, for* $i=0$*,* $\left(\widetilde{m}_0^h,\widetilde{\mathcal{C}}_0^h\right)=(m_0,\mathcal{C}_0)$*, and for* $i=1,...,K$*,* $\left(\widetilde{m}_{t_i}^h,\widetilde{\mathcal{C}}_{t_i}^h\right)$ *are determined by the ordinary difference equations below (let* $h_i=t_i-t_{i-1}$ *be the step size),*

$$\widetilde{m}_{t_i}^h=\widetilde{m}_{t_{i-1}}^h-h_i\widetilde{\mathcal{C}}_{t_{i-1}}^h\mathrm{D}\mathcal{Z}\left(\widetilde{m}_{t_{i-1}}^h\right)^*\left(\mathbf{I}+h_i\mathrm{D}\mathcal{Z}\left(\widetilde{m}_{t_{i-1}}^h\right)\widetilde{\mathcal{C}}_{t_{i-1}}^h\mathrm{D}\mathcal{Z}\left(\widetilde{m}_{t_{i-1}}^h\right)^*\right)^{-1}\mathcal{Z}\left(\widetilde{m}_{t_{i-1}}^h\right) \tag{3.3.14}$$

$$\widetilde{\mathcal{C}}_{t_i}^h=\widetilde{\mathcal{C}}_{t_{i-1}}^h-h_i\widetilde{\mathcal{C}}_{t_{i-1}}^h\mathrm{D}\mathcal{Z}\left(\widetilde{m}_{t_{i-1}}^h\right)^*\left(\mathbf{I}+h_i\mathrm{D}\mathcal{Z}\left(\widetilde{m}_{t_{i-1}}^h\right)\widetilde{\mathcal{C}}_{t_{i-1}}^h\mathrm{D}\mathcal{Z}\left(\widetilde{m}_{t_{i-1}}^h\right)^*\right)^{-1}\mathrm{D}\mathcal{Z}\left(\widetilde{m}_{t_{i-1}}^h\right)\widetilde{\mathcal{C}}_{t_{i-1}}^h \tag{3.3.15}$$

where $\mathcal{Z}:\mathcal{H}\to\mathbb{R}^n$ *is the data-misfit function defined in formula* (3.1.19)*, and* $\mathrm{D}\mathcal{Z}(u):\mathcal{H}\to\mathbb{R}^n$ *is the Fréchet derivative of* $\mathcal{Z}$ *at* $u\in\mathcal{H}$*.*

CoEKI is formulated by an ODE system, and DiEKI is a discrete scheme for CoEKI. After the discretization, DiEKI is a numerically implementable algorithm searching for a sub-optimum $\widetilde{m}_1^h$. DiEKI is very similar to the Levenberg-Marquardt algorithm (LMA), and the tempering parameter in DiEKI plays the similar role as the damping factor in LMA. Just like LMA, the behavior of DiEKI is between Gauss-Newton and gradient descent. The difference is that LMA aims to search for a local optimum that is a stationary point, whereas DiEKI aims to search for a sub-optimum that is not necessarily to be stationary. Another difference is that DiEKI has the covariance update, which does not appear in LMA.

Ensemble Kalman inversion (infinite sample size)

Ensemble Kalman filter [33, 65, 61, 20, 78, 77] is a popular method in applied mathematics. Numerically, it is a particle filter with finite sample size. Moreover, [54] also considers the mean-field limit of ensemble Kalman filter as the sample size goes to infinity. The mean-field limit thus forms a Markov process of random variables, and conversely, the ensemble Kalman method with finite sample size can be regarded as the finite ensemble of the mean-field limit. We firstly define the mean-field limiting ensemble Kalman inversion as continuous/discrete stochastic process in the follow definitions.

Definition 3.3.10 (CoMFEnKI). *CoMFEnKI aims to interpret the random variable $\widetilde{u}_t$ for $t \in [0,1]$, such that, for $t = 0$, $\widetilde{u}_0 \sim \mathcal{N}(m_0, \mathcal{C}_0)$ is an $\mathcal{H}$-valued Gaussian random variable, and for $t \in (0,1]$, $\widetilde{u}_t$ is determined by the stochastic differential equation below,*

$$\mathrm{d}\widetilde{u}_t = -\widetilde{\mathcal{C}}_{uz,t}\left(\mathcal{Z}\left(\widetilde{u}_t\right)\mathrm{d}t - \mathrm{d}W_t\right) \tag{3.3.16}$$

where $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ is the data-misfit function defined in formula (3.1.19), W_t is an n-dimensional standard Wiener process on $t \in (0,1]$, and for any $t \in [0,1]$,

$$\widetilde{\mathcal{C}}_{uz,t} \equiv \mathrm{COV}\left\{\widetilde{u}_t, \widetilde{z}_t\right\} \tag{3.3.17}$$

is the covariance of random variables, where $\widetilde{z}_t$ is the random variable

$$\widetilde{z}_t \equiv \mathcal{Z}\left(\widetilde{u}_t\right) \tag{3.3.18}$$

Definition 3.3.11 (DiMFEnKI). *DiMFEnKI aims to interpret the random variable $\widetilde{u}_t^h$ for $t \in \{t_i : 0 = t_0 < t_1 < \cdots < t_K = 1\}$, such that, for $i = 0$, $\widetilde{u}_0^h \sim \mathcal{N}(m_0, \mathcal{C}_0)$ is an $\mathcal{H}$-valued Gaussian random variable, and for $i = 1, ..., K$, $\widetilde{u}_{t_i}^h$ is determined by the stochastic difference equation below (let $h_i = t_i - t_{i-1}$ be the step size),*

$$\widetilde{u}_{t_i}^h = \widetilde{u}_{t_{i-1}}^h - \widetilde{\mathcal{C}}_{uz,t_{i-1}}^h \left(\mathbf{I} + h_i \widetilde{\mathcal{C}}_{zz,t_{i-1}}^h\right)^{-1} \left(h_i \mathcal{Z}\left(\widetilde{u}_{t_{i-1}}^h\right) - \sqrt{h_i}\zeta_i\right) \tag{3.3.19}$$

where $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ is the data-misfit function defined in formula (3.1.19), $\{\zeta_i\}_{i=1}^K$ is a set of K independent n-dimensional standard Gaussian random variables, and for any $t \in \{t_0, t_1, ..., t_K\}$,

$$\widetilde{\mathcal{C}}_{uz,t}^h \equiv \mathrm{COV}\left\{\widetilde{u}_t^h, \widetilde{z}_t^h\right\} \qquad \widetilde{\mathcal{C}}_{zz,t}^h \equiv \mathrm{COV}\left\{\widetilde{z}_t^h, \widetilde{z}_t^h\right\} \tag{3.3.20}$$

are the covariances of random variables, where $\widetilde{z}_t^h$ is the random variable

$$\widetilde{z}_t^h \equiv \mathcal{Z}\left(\widetilde{u}_t^h\right) \tag{3.3.21}$$

Both CoMFEnKI and DiMFEnKI are not numerically implementable, because the covariance $\widetilde{\mathcal{C}}_{uz,t}$ in formula (3.3.16) and covariances $\widetilde{\mathcal{C}}_{uz,t_{i-1}}^h$ $\widetilde{\mathcal{C}}_{zz,t_{i-1}}^h$ in formula (3.3.19) are not perfectly known, except for linear problems ($\mathcal{Z}$ is an affine transformation). However, CoMFEnKI and DiMFEnKI have the theoretical importance, since they are the mean-field limits of the standard ensemble Kalman inversion with finite sample size. The standard ensemble Kalman inversion are discussed as follows.

Ensemble Kalman inversion (finite sample size)

Ensemble Kalman inversion with finite sample size is the standard form that can be found in literature [33, 65, 61, 20, 78, 77]. EnKI (particle) can be regarded as the finite ensemble of MFEnKI (random variable). Thus, by replacing random variables and covariances in DiMFEnKI (3.3.19) and CoMFEnKI (3.3.16) with samples and sample covariances, we can straightforwardly make the following definitions.

Definition 3.3.12 (CoEnKI [20])**.** *CoEnKI aims to interpret the particle U_t for $t \in [0,1]$, such that, for $t = 0$, U_0 is a set containing J samples independently drawn from the Gaussian measure $\mathcal{N}(m_0, \mathcal{C}_0)$, and for $t \in (0,1]$, U_t is determined by the differential equations below, for all $j = 1, ..., J$,*

$$\mathrm{d}U_t(j) = -C_{uz,t}\left(\mathcal{Z}\left(U_t(j)\right)\mathrm{d}t - \mathrm{d}B_t(j)\right) \tag{3.3.22}$$

where $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ is the data-misfit function defined in formula (3.1.19), $B_t(j)$ for $j = 1, ..., J$ are J Brownian motion paths/trajectories on $t \in (0,1]$ independently drawn from the n-dimensional standard Wiener process, and for any $t \in [0,1]$,

$$C_{uz,t} \equiv \mathrm{cov}\left(U_t, Z_t\right) \tag{3.3.23}$$

is the sample coavriance of particles, where Z_t is the particle such that, for all $j = 1, ..., J$,

$$Z_t(j) \equiv \mathcal{Z}(U_t(j)) \tag{3.3.24}$$

Definition 3.3.13 (DiEnKI [61])**.** *DiEnKI aims to interpret the particle U_t^h for $t \in \{t_i : 0 = t_0 < t_1 < \cdots < t_K = 1\}$, such that, for $i = 0$, U_0^h is a set containing J samples independently drawn from the Gaussian measure $\mathcal{N}(m_0, \mathcal{C}_0)$, and for $i = 1, ..., K$, $U_{t_i}^h$ is determined by the difference equations below (let $h_i = t_i - t_{i-1}$ be the step size), for all $j = 1, ..., J$,*

$$U_{t_i}^h(j) = U_{t_{i-1}}^h(j) - C_{uz,t_{i-1}}^h\left(\mathbf{I} + h_i C_{zz,t_{i-1}}^h\right)^{-1}\left(h_i \mathcal{Z}\left(U_{t_{i-1}}^h(j)\right) - \sqrt{h_i}V_{ij}\right) \tag{3.3.25}$$

where $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ is the data-misfit function defined in formula (3.1.19), V_{ij} for $i = 1, ..., K$ and $j = 1, ..., J$ are $K \times J$ samples independently drawn from the n-dimensional standard normal distribution, and for any $t \in \{t_0, t_1, ..., t_K\}$,

$$C_{uz,t}^h \equiv \mathrm{cov}\left(U_t^h, Z_t^h\right) \qquad C_{zz,t}^h \equiv \mathrm{cov}\left(Z_t^h, Z_t^h\right) \tag{3.3.26}$$

are the sample covariances of particles, where Z_t^h *is the particle such that, for all* $j = 1, ..., J$,

$$Z_t^h(j) \equiv \mathcal{Z}\left(U_t^h(j)\right) \tag{3.3.27}$$

As the result, DiEnKI is the numerically implementable algorithm with discrete step size and finite sample size. Higher accuracy can be obtained by using smaller step size and larger sample size. Theoretically, the highest accuracy can be obtained by using infinitesimal step size and finite sample size. This theoretical limit is CoMFEnKI. DiEnKI produces finite samples forming an empirical distribution as the approximation of the exact distribution determined by CoMFEnKI, and the exact distribution determined by CoMFEnKI is further regarded as a heuristic solution of the original Bayesian inverse problem.

3.3.3 Intuitive derivation of the Kalman-like methods

For linear problems, the Kalman filter is the analytic solution. For nonlinear problems, propositions 3.3.4-3.3.7 do not hold. Nevertheless, in practice, EKF and EnKF are heuristic algorithms using linearization of nonlinear forward maps. To explain how to conduct the linearization, we first of all notice the following facts: if the data-misfit function $\mathcal{Z}$ is an affine transformation (3.3.5), then we have

- the Fréchet derivative $\mathrm{D}\mathcal{Z}(u) : \mathcal{H} \to \mathbb{R}^n$ of $\mathcal{Z}$ at any $u \in \mathcal{H}$ equals to $\mathcal{K}$,

$$\mathrm{D}\mathcal{Z}(u) = \mathcal{K} \tag{3.3.28}$$

- the covarinace $\mathcal{C}_{uz,t} : \mathbb{R}^n \to \mathcal{H}$ of u_t and $\mathcal{Z}(u_t)$ equals to $\mathcal{C}_t\mathcal{K}^*$ for any $t \in [0, 1]$,

$$\mathcal{C}_{uz,t} \equiv \mathrm{COV}\left\{u_t, \mathcal{Z}(u_t)\right\} = \mathrm{COV}\left\{u_t, \mathcal{K}u_t + c\right\} = \mathrm{COV}\left\{u_t, u_t\right\}\mathcal{K}^* = \mathcal{C}_t\mathcal{K}^* \tag{3.3.29}$$

- the covarinace $\mathcal{C}_{uz,t} : \mathbb{R}^n \to \mathbb{R}^n$ of $\mathcal{Z}(u_t)$ and $\mathcal{Z}(u_t)$ equals to $\mathcal{K}\mathcal{C}_t\mathcal{K}^*$ for any $t \in [0, 1]$,

$$\mathcal{C}_{zz,t} \equiv \mathrm{COV}\left\{\mathcal{Z}(u_t), \mathcal{Z}(u_t)\right\} = \mathrm{COV}\left\{\mathcal{K}u_t + c, \mathcal{K}u_t + c\right\} = \mathcal{K}\,\mathrm{COV}\left\{u_t, u_t\right\}\mathcal{K}^* = \mathcal{K}\mathcal{C}_t\mathcal{K}^* \tag{3.3.30}$$

Now, the Kalman-like methods (DiEKI, CoEKI, DiMFEnKI, CoMFEnKI, DiEnKI, and CoEnKI) can be derived as follows:

1. EKI. Substituting formula (3.3.28) into formulas (3.3.6) (3.3.7) in proposition 3.3.4 leads to the DiEKI (definition 3.3.9). Substituting formula (3.3.28) into formulas (3.3.9) (3.3.10) in proposition 3.3.6 leads to the CoEKI (definition 3.3.8).

2. MFEnKI. Substituting formulas (3.3.29) (3.3.30) into formula (3.3.8) in proposition 3.3.5 leads to the DiMFEnKI (definition 3.3.11). Substituting formula (3.3.29) into formula (3.3.11) in proposition 3.3.7 leads to the CoMFEnKI (definition 3.3.10).

3. EnKI. DiEnKI (definition 3.3.13) and CoEnKI (definition 3.3.12) are the finite ensemble of DiMFEnKI and CoMFEnKI, respectively. Namely, the random variables and covariances in DiMFEnKI and CoMFEnKI are replaced by samples and sample covariances.

These variants are treated as heuristic methods for nonlinear problems. Thus, we use the tilde notation like $\widetilde{m}$, $\widetilde{\mathcal{C}}$ and $\widetilde{u}$ in definitions 3.3.8-3.3.13, in order to indicate that these quantities are only approximations.

3.3.4 Deeper understanding of the Kalman-like methods

The last subsection shows an intuitive way to deriving the Kalman-like methods by substitution. It would be better if the ODE system (CoEKI) and the SDE system (CoMFEnKI) can be directly derived from the tempering setting of inverse problems. This subsection reveals that, CoEKI and CoMFEnKI use different linearization methods to simplify the variational inversion and the Bayesian inversion, respectively.

Linearization of the forward map

The Kalman-like methods are approximations of the original variational/Bayesian methods. In fact, these approximation methods can be derived by linearization of the forward map. In order to derive the extended Kalman inversion (EKI) and the ensemble Kalman inversion (EnKI), two different linearization methods are needed. To avoid confusion, we call the two linearization methods as the local linearization (LL) and the global linearization (GL), respectively. They are called 'local' and 'global', because LL is the linearization around a local point using Taylor expansion and GL is the linearization using average

values over the entire space under a probability measure. On one hand, LL is used to derive EKI as an approximation of the variational approach. On the other hand, GL is used to derive EnKI as an approximation of the Bayesian approach.

Definition 3.3.14 (local linearization (LL)). *Let $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ be a function. The LL of $\mathcal{Z}$ on a point $u \in \mathcal{H}$ is a map $\widetilde{\mathcal{Z}}_l(\cdot; u) : U \to \mathbb{R}^n$, such that for all u' in a neighborhood $U \subset \mathcal{H}$ around u, we have*

$$\widetilde{\mathcal{Z}}_l(u'; u) := [\mathrm{D}\mathcal{Z}(u)]\,(u' - u) + \mathcal{Z}(u) \tag{3.3.31}$$

where $\mathrm{D}\mathcal{Z}(x) : \mathcal{H} \to \mathbb{R}^n$ is the Fréchet derivative of $\mathcal{Z}$ at $x \in \mathcal{H}$.

Definition 3.3.15 (global linearization (GL)). *Let $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ be a function. The GL of $\mathcal{Z}$ with respect to a probability measure $\mu : \mathcal{B}(\mathcal{H}) \to [0, 1]$ is a map $\widetilde{\mathcal{Z}}_g(\cdot; \mu) : \mathcal{H} \to \mathbb{R}^n$, such that for almost every $u' \in \mathcal{H}$, we have*

$$\widetilde{\mathcal{Z}}_g(u'; \mu) := [\mathbb{E}\{\mathrm{D}\mathcal{Z}(X)\}]\,(u' - \mathbb{E}\{X\}) + \mathbb{E}\{\mathcal{Z}(X)\} \ \text{ with } X \sim \mu \tag{3.3.32}$$

where $\mathrm{D}\mathcal{Z}(x) : \mathcal{H} \to \mathbb{R}^n$ is the Fréchet derivative of $\mathcal{Z}$ at $x \in \mathcal{H}$.

The motivations of LL and GL are discussed as follows.

- Firstly, LL plays its role in mathematical optimization. The exact Newton's method adopts a quadratic form to approximate the objective functional locally. The quadratic form is composed with the gradient vector (first-order derivative) and the Hessian matrix (second-order derivative). For least-squares problems, the Gauss-Newton algorithm makes a modification that the Hessian matrix is approximately estimated by the Jacobian matrix (first-order derivative). Computing the Jacobian is much more efficient than computing the Hessian. In the Gauss-Newton algorithm, LL directly implies the core idea that the Hessian is only approximated by the Jacobian.

- Secondly, GL plays its role in Bayesian inference. In fact, GL is summerized from this thesis in order to derive EnKI. We do not find much literature about the GL. Nevertheless, we emphasize two points: 1) GL is used to simplify integrals when the Bayes' formula is used; 2) GL in Bayesian inference can be regarded as

the analogue of LL in mathematical optimization, i.e. GL implies the core idea that the expected value of Hessian is only approximated by the expected value of Jacobian. The similarity of GL and LL can be clearly viewed in their definitions (3.3.31) and (3.3.32).

The ODE method (CoEKI) and the SDE method (CoMFEnKI)

The ODE method (CoEKI) is a heuristic approach simplifying the variational tempering setting (3.1.20) by using the local linearization (LL), and CoMFEnKI is a heuristic approach simplifying the Bayesian tempering setting (3.1.21) by using the global linearization (GL). In fact, the simplification from the tempering settings to the Kalman-like methods help us realize the following facts:

1. The Kalman-like methods solving inverse problems are approximate methods using linearization of forward maps, so that, the original inverse problem with a nonlinear forward map can be simplified.

2. Different linearization techniques lead to different methods, i.e. EKI can be derived from the variational method by using LL of the forward map, and EnKI can be derived from the Bayesian method by using GL of the forward map.

3. EKI is a point estimation method, which is more like a mathematical optimization algorithm via variational approach. EnKI is a particle filtering method, which is more like a statistical sampling algorithm via Bayesian approach.

After we realize the essential roles of LL and GL played in the Kalman-like methods, we can predetermine applicable conditions to use the Kalman-like methods. Some practical principles are proposed:

EKI CoEKI in definition 3.3.8 produces the point estimate $\widetilde{m}_t$ as an approximation of the optimum $\widehat{x}_t$ determined in formula (3.1.20). This approximation holds on the condition that the data-misfit function $\mathcal{Z}$ can approximated by the LL of $\mathcal{Z}$. In practice, this usually requires continuous differentiability of $\mathcal{Z}$. That means, EKI works for inverse problems whose forward maps are continuously differentiable.

In this case, EKI is similar like Levenberg-Marquardt algorithm solving inverse problems with variational setting.

EnKI CoMFEnKI in definition 3.3.10 produces the random estimate $\widetilde{u}_t$ that approximately obeys the probability measure μ_t determined in formula (3.1.21). This approximation holds on the condition that the data-misfit function $\mathcal{Z}$ can approximated by the GL of $\mathcal{Z}$. In practice, this usually requires strong linear dependence between $\widetilde{u}_t$ and $\mathcal{Z}(\widetilde{u}_t)$. That means, EnKI works for inverse problems with strong correlations between parameters and observations. In this case, EnKI is like a linear regression algorithm.

3.4 Adaptive Kalman-like methods

The last section discusses Kalman-like methods solving inverse problems with the tempering setting. Two types of Kalman-like methods has been introduced. The first one is the extended Kalman inversion (EKI), which is expressed as an ODE system (CoEKI in definition 3.3.8), and the ODE system can be numerically solved by discretization (DiEKI in definition 3.3.9). The second one is the ensemble Kalman inversion (EnKI), which is expressed as a SDE system (CoMFEnKI in definition 3.3.10), and the SDE system can be numerically solved by discretization associated with empirical sampling (DiEnKI in definition 3.3.13). The question is: how to select the step size in DiEKI and DiEnKI? We aim to apply the data-misfit controller (3.2.15) or (3.2.27) to determine the step size adaptively.

However, the data-misfit controller is designed for the tempering setting. The Kalman-like methods (both EKI and EnKI) are only the approximations of the tempering setting. For this reason, the data-misfit controller cannot be directly applied on the Kalman-like methods. Nevertheless, after some modification, a modified data-misfit controller can be easily obtained that is suitable for the Kalman-like methods. The main idea of the modification is to use Gaussian measures to approximate the exact probability measures determined via the tempering setting. Moreover, since Kalman-like methods are approximate algorithms, so these methods may not preserve the monotone properties of the tempering setting (the monotone properties are discussed in subsection 3.1.4). Thus, the

monotone properties should be monitored when Kalman-like methods are implemented. As the result, the early stop criterion for the Kalman-like methods can be established by checking the monotone properties.

3.4.1 Kalman-like methods with Gaussian approximation

Numerically, using sampling algorithms (like sequential Monte Carlo) to solve inverse problems with the tempering setting may cost too much computational recourse. Alternatively, Kalman-like methods, as heuristic algorithms, only requires much fewer computational budget. EKI updates mean-covariance pairs, and EnKI updates random variables (or particles). If EKI is implemented, we just take the mean and the covariance. If EnKI is implemented, we proceed as in random variable (or particle), but after the update we retain only the mean and the covariance of the random variable (or particle). Thus, for either of the two Kalman-like methods, we can always obtain the mean and covariance. Then, we use the mean and covariance to construct a Gaussian measure as an approximation. This subsection explains how the accurate probability measures determined via the tempering setting are approximated by Gaussian measures.

Consider the conditional probability μ_t determined via the tempering setting (3.1.21). This probability μ_t is 'accurate' in the sense that it rigorously obeys the Bayes' rule. Let u_t denote the (random) estimate under the conditional probability measure, i.e. for any $t \in [0, 1]$,

$$u_t \sim \mu_t \tag{3.4.1}$$

In addition, let $z_t \equiv \mathcal{Z}(u_t)$ be the predicted (random) data misfit, where $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ is the data-misfit function determined in formula (3.1.19). Then, let $\mathbb{Q}_t$ denote the joint probability of the random pair (u_t, z_t), i.e. for any $t \in [0, 1]$,

$$\begin{bmatrix} u_t \\ z_t \end{bmatrix} \sim \mathbb{Q}_t \tag{3.4.2}$$

The joint probability $\mathbb{Q}_t$ gathers all information at time $t \in [0, 1]$ including the estimate and the predicted data-misfit.

With Gaussian-linear approximation, the original joint probability (3.4.2) is approximated by a Gaussian distribution (the approximate equal becomes to the exact equal for

linear problems), for any $t \in [0,1]$,

$$\mathbb{Q}_t \approx \widetilde{\mathbb{Q}}_t^G := \mathcal{N}\left(\begin{bmatrix}\widetilde{m}_{u,t}\\ \widetilde{m}_{z,t}\end{bmatrix}, \begin{bmatrix}\widetilde{\mathcal{C}}_{uu,t} & \widetilde{\mathcal{C}}_{uz,t}\\ \widetilde{\mathcal{C}}_{zu,t} & \widetilde{\mathcal{C}}_{zz,t}\end{bmatrix}\right) \tag{3.4.3}$$

where $\mathbb{Q}_t$ is the original joint probability at t, $\widetilde{\mathbb{Q}}_t^G$ is called the Gaussian-regularized (GR) joint probability at t, and the mean and covariance of the GR joint distribution can be specified by the Kalman-like methods. For example,

CoEKI The mean and covariance of the GR joint distribution are specified by, for any $t \in [0,1]$,

$$\begin{bmatrix}\widetilde{m}_{u,t}\\ \widetilde{m}_{z,t}\end{bmatrix} = \begin{bmatrix}\widetilde{m}_t\\ \mathcal{Z}\left(\widetilde{m}_t\right)\end{bmatrix} \qquad \begin{bmatrix}\widetilde{\mathcal{C}}_{uu,t} & \widetilde{\mathcal{C}}_{uz,t}\\ \widetilde{\mathcal{C}}_{zu,t} & \widetilde{\mathcal{C}}_{zz,t}\end{bmatrix} = \begin{bmatrix}\widetilde{\mathcal{C}}_t & \widetilde{\mathcal{C}}_t \mathrm{D}\mathcal{Z}\left(\widetilde{m}_t\right)^*\\ \mathrm{D}\mathcal{Z}\left(\widetilde{m}_t\right)\widetilde{\mathcal{C}}_t & \mathrm{D}\mathcal{Z}\left(\widetilde{m}_t\right)\widetilde{\mathcal{C}}_t \mathrm{D}\mathcal{Z}\left(\widetilde{m}_t\right)^*\end{bmatrix}$$

where $\left(\widetilde{m}_t, \widetilde{\mathcal{C}}_t\right)$ is the pair determined via the ODEs (3.3.12) and (3.3.13).

DiEKI The mean and covariance of the GR joint distribution are specified by, for any $t \in \{t_i : i = 0, 1, ..., K\}$,

$$\begin{bmatrix}\widetilde{m}_{u,t}\\ \widetilde{m}_{z,t}\end{bmatrix} = \begin{bmatrix}\widetilde{m}_t^h\\ \mathcal{Z}\left(\widetilde{m}_t^h\right)\end{bmatrix} \qquad \begin{bmatrix}\widetilde{\mathcal{C}}_{uu,t} & \widetilde{\mathcal{C}}_{uz,t}\\ \widetilde{\mathcal{C}}_{zu,t} & \widetilde{\mathcal{C}}_{zz,t}\end{bmatrix} = \begin{bmatrix}\widetilde{\mathcal{C}}_t^h & \widetilde{\mathcal{C}}_t^h \mathrm{D}\mathcal{Z}\left(\widetilde{m}_t^h\right)^*\\ \mathrm{D}\mathcal{Z}\left(\widetilde{m}_t^h\right)\widetilde{\mathcal{C}}_t^h & \mathrm{D}\mathcal{Z}\left(\widetilde{m}_t^h\right)\widetilde{\mathcal{C}}_t^h \mathrm{D}\mathcal{Z}\left(\widetilde{m}_t^h\right)^*\end{bmatrix}$$

where $\left(\widetilde{m}_t^h, \widetilde{\mathcal{C}}_t^h\right)$ is the pair determined via the difference equations (3.3.14) and (3.3.15).

CoMFEnKI The mean and covariance of the GR joint distribution are specified by, for any $t \in [0,1]$,

$$\begin{bmatrix}\widetilde{m}_{u,t}\\ \widetilde{m}_{z,t}\end{bmatrix} = \begin{bmatrix}\mathbb{E}\left\{\widetilde{u}_t\right\}\\ \mathbb{E}\left\{\widetilde{z}_t\right\}\end{bmatrix} \qquad \begin{bmatrix}\widetilde{\mathcal{C}}_{uu,t} & \widetilde{\mathcal{C}}_{uz,t}\\ \widetilde{\mathcal{C}}_{zu,t} & \widetilde{\mathcal{C}}_{zz,t}\end{bmatrix} = \begin{bmatrix}\mathrm{COV}\left\{\widetilde{u}_t, \widetilde{u}_t\right\} & \mathrm{COV}\left\{\widetilde{u}_t, \widetilde{z}_t\right\}\\ \mathrm{COV}\left\{\widetilde{z}_t, \widetilde{u}_t\right\} & \mathrm{COV}\left\{\widetilde{z}_t, \widetilde{z}_t\right\}\end{bmatrix}$$

where $\widetilde{u}_t$ is the (random) estimate determined via the SDE (3.3.16), and $\widetilde{z}_t$ is the predicted (random) data misfit,

$$\widetilde{z}_t \equiv \mathcal{Z}\left(\widetilde{u}_t\right)$$

DiEnKI The mean and covariance of the GR joint distribution are specified by, for any $t \in \{t_i : i = 0, 1, ..., K\}$,

$$\begin{bmatrix} \widetilde{m}_{u,t} \\ \widetilde{m}_{z,t} \end{bmatrix} = \begin{bmatrix} \text{mean}\left(U_t^h\right) \\ \text{mean}\left(Z_t^h\right) \end{bmatrix} \qquad \begin{bmatrix} \widetilde{\mathcal{C}}_{uu,t} & \widetilde{\mathcal{C}}_{uz,t} \\ \widetilde{\mathcal{C}}_{zu,t} & \widetilde{\mathcal{C}}_{zz,t} \end{bmatrix} = \begin{bmatrix} \text{cov}\left(U_t^h, U_t^h\right) & \text{cov}\left(U_t^h, Z_t^h\right) \\ \text{cov}\left(Z_t^h, U_t^h\right) & \text{cov}\left(Z_t^h, Z_t^h\right) \end{bmatrix}$$

where U_t^h is the particle of estimates determined via the difference equation (3.3.25), and Z_t^h is the particle of predicted data misfits,

$$Z_t^h(j) \equiv \mathcal{Z}(U_t^h(j)) \qquad \forall j = 1, ..., J$$

3.4.2 Adaptive scheme of updating the GR joint distributions

In this subsection, we consider how to select the step size for updating the GR joint distributions. Let $0 = t_0 < t_1 < \cdots < t_K = 1$ be any $K + 1$ points in interval $[0, 1]$, and let $h_i = t_i - t_{i-1}$ be the ith step size for any $i \in \{1, ..., K\}$.

Firstly, we consider the ith update of the original joint distribution $\mathbb{Q}_t$ from $t = t_{i-1}$ to $t = t_i$:

$$\left(\begin{bmatrix} u_{t_{i-1}} \\ z_{t_{i-1}} \end{bmatrix} \sim \mathbb{Q}_{t_{i-1}} \right) \xrightarrow{h_i} \left(\begin{bmatrix} u_{t_i} \\ z_{t_i} \end{bmatrix} \sim \mathbb{Q}_{t_i} \right) \tag{3.4.4}$$

The original data-misfit controller has been proposed in formula (3.2.15) or (3.2.27). It is as an adaptive scheme for the stepwise regularization/learning, such that, the step size h_i in formula (3.4.4) can be determined by

$$h_i = \min \left\{ \max \left\{ \eta \Big/ q_{i-1}^{(1)} \, , \, \sqrt{\eta \Big/ q_{i-1}^{(2)}} \right\} , \, 1 - t_{i-1} \right\} \tag{3.4.5}$$

where $\eta = n/2$ and

$$q_{i-1}^{(1)} = \mathbb{E}\left\{ \frac{1}{2} \left\| z_{t_{i-1}} \right\|_{\mathbb{R}^n}^2 \right\} \tag{3.4.6}$$

$$q_{i-1}^{(2)} = \text{Var}\left\{ \frac{1}{2} \left\| z_{t_{i-1}} \right\|_{\mathbb{R}^n}^2 \right\} \tag{3.4.7}$$

Similarly, we consider the ith update of the GR joint distribution $\widetilde{\mathbb{Q}}_t^G$ from $t = t_{i-1}$ to $t = t_i$:

$$\left(\begin{bmatrix} \widetilde{u}_{t_{i-1}}^G \\ \widetilde{z}_{t_{i-1}}^G \end{bmatrix} \sim \widetilde{\mathbb{Q}}_{t_{i-1}}^G \right) \xrightarrow{h_i} \left(\begin{bmatrix} \widetilde{u}_{t_i}^G \\ \widetilde{z}_{t_i}^G \end{bmatrix} \sim \widetilde{\mathbb{Q}}_{t_i}^G \right) \tag{3.4.8}$$

If the GR joint distribution $\widetilde{\mathbb{Q}}_t^G$ is sufficiently close to the original joint distribution $\mathbb{Q}_t$. Then, the data-misfit controller designed for updating $\mathbb{Q}_t$ can be also applied for updating $\widetilde{\mathbb{Q}}_t^G$. Namely, we just replace the accurate joint distribution $\mathbb{Q}_{t_{i-1}}$ by the GR joint distribution $\widetilde{\mathbb{Q}}_{t_{i-1}}^G$. This replacement holds as long as $\widetilde{\mathbb{Q}}_{t_{i-1}}^G$ is sufficiently close to $\mathbb{Q}_{t_{i-1}}$. Then, the Gaussian-regularized data-misfit controller determines the step size h_i in formula (3.4.8) as follows,

$$h_i = \min\left\{\max\left\{\eta \Big/ q_{i-1}^{(1)}\,,\ \sqrt{\eta \Big/ q_{i-1}^{(2)}}\right\},\ 1 - t_{i-1}\right\} \tag{3.4.9}$$

where $\eta = n/2$ and

$$q_{i-1}^{(1)} = \mathbb{E}\left\{\frac{1}{2}\left\|\widetilde{z}_{t_{i-1}}^G\right\|_{\mathbb{R}^n}^2\right\} \tag{3.4.10}$$

$$q_{i-1}^{(2)} = \mathrm{Var}\left\{\frac{1}{2}\left\|\widetilde{z}_{t_{i-1}}^G\right\|_{\mathbb{R}^n}^2\right\} \tag{3.4.11}$$

Furthermore, since $\widetilde{z}_{t_{i-1}}^G$ in formula (3.4.10) and formula (3.4.11) is a Gaussian random variable, these formulas (3.4.10) and (3.4.11) can be explicitly expressed as

$$q_{i-1}^{(1)} = \frac{1}{2}\,\mathrm{Tr}\left(\widetilde{\mathcal{C}}_{zz,t_{i-1}}\right) + \frac{1}{2}\widetilde{m}_{z,t_{i-1}}^*\widetilde{m}_{z,t_{i-1}} \tag{3.4.12}$$

$$q_{i-1}^{(2)} = \frac{1}{2}\,\mathrm{Tr}\left(\widetilde{\mathcal{C}}_{zz,t_{i-1}}^2\right) + \widetilde{m}_{z,t_{i-1}}^*\widetilde{\mathcal{C}}_{zz,t_{i-1}}\widetilde{m}_{z,t_{i-1}} \tag{3.4.13}$$

where $\widetilde{m}_{z,t_{i-1}}$ and $\widetilde{\mathcal{C}}_{zz,t_{i-1}}$ are the mean and covariance of $\widetilde{z}_{t_{i-1}}^G$, respectively. The derivation of formulas (3.4.12) and (3.4.13) from formulas (3.4.10) and (3.4.11) is the direct result in the following proposition.

Proposition 3.4.1 (generalized chi-square distribution)**.** *Let $z \sim \mathcal{N}(m_z, \mathcal{C}_{zz})$ be an n-dimensional Gaussian random variable, where $m_z \in \mathbb{R}^n$ and $\mathcal{C}_{zz} : \mathbb{R}^n \to \mathbb{R}^n$ are the given mean and covariance. Then the generalized chi-square random variable $X_n^2 := \|z\|_{\mathbb{R}^n}^2$ has mean and variance expressed by,*

$$\mathbb{E}\left\{\frac{1}{2}X_n^2\right\} = \frac{1}{2}\,\mathrm{Tr}(\mathcal{C}_{zz}) + \frac{1}{2}m_z^*m_z \tag{3.4.14}$$

$$\mathrm{Var}\left\{\frac{1}{2}X_n^2\right\} = \frac{1}{2}\,\mathrm{Tr}(\mathcal{C}_{zz}^2) + m_z^*\mathcal{C}_{zz}m_z \tag{3.4.15}$$

Proof. Transform $m_z \in \mathbb{R}^n$ and $\mathcal{C}_{zz} : \mathbb{R}^n \to \mathbb{R}^n$ to the eigensystem $k_z \in \mathbb{R}^n$ and $\Lambda_{zz} : \mathbb{R}^n \to \mathbb{R}^n$,

$$\mathcal{C}_{zz} \equiv U_z\Lambda_{zz}U_z^{\mathrm{T}}, \qquad m_z \equiv U_z k_z \tag{3.4.16}$$

where U_z is a unitary matrix and Λ_{zz} is diagonal matrix. Thus, we have

$$\begin{aligned}\frac{1}{2}\|z\|^2_{\mathbb{R}^n} &= \frac{1}{2}\left\|\mathcal{C}_{zz}^{1/2}\zeta + m_z\right\|^2_{\mathbb{R}^n} \\ &= \frac{1}{2}\left\|U_z\Lambda_{zz,t}^{1/2}U_z^{\mathrm{T}}\zeta + U_z k_z\right\|^2_{\mathbb{R}^n} \\ &= \frac{1}{2}\left\|\Lambda_{zz}^{1/2}\xi + k_z\right\|^2_{\mathbb{R}^n} \\ &= \frac{1}{2}\sum_{i=1}^{n}\left(\lambda_i^{1/2}\xi_i + \kappa_i\right)^2\end{aligned}$$

where ζ is an n-dimensional standard Gaussian random variable, $\xi := U_z^{\mathrm{T}}\zeta$ is also an n-dimensional standard Gaussian random variable, λ_i is the ith diagonal element of Λ_{zz}, ξ_i is the ith component of ξ, and κ_i is the ith component of k_z. Therefore, the mean and variance of $\frac{1}{2}\|z_t\|^2_{\mathbb{R}^n}$ can be expressed by

$$\begin{aligned}\mathbb{E}\left\{\frac{1}{2}\|z\|^2_{\mathbb{R}^n}\right\} &= \mathbb{E}\left\{\frac{1}{2}\sum_{i=1}^{n}\left(\lambda_i^{1/2}\xi_i + \kappa_i\right)^2\right\} \\ &= \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left\{\left(\lambda_i^{1/2}\xi_i + \kappa_i\right)^2\right\} \\ &= \frac{1}{2}\sum_{i=1}^{n}\left(\lambda_i + \kappa_i^2\right) \\ &= \frac{1}{2}\operatorname{Tr}(\mathcal{C}_{zz}) + \frac{1}{2}m_z^* m_z\end{aligned}$$

$$\begin{aligned}\operatorname{Var}\left\{\frac{1}{2}\|z\|^2_{\mathbb{R}^n}\right\} &= \operatorname{Var}\left\{\frac{1}{2}\sum_{i=1}^{n}\left(\lambda_i^{1/2}\xi_i + \kappa_i\right)^2\right\} \\ &= \frac{1}{4}\sum_{i=1}^{n}\operatorname{Var}\left\{\left(\lambda_i^{1/2}\xi_i + \kappa_i\right)^2\right\} \\ &= \frac{1}{4}\sum_{i=1}^{n}\left(2\lambda_i^2 + 4\lambda_i\kappa_i^2\right) \\ &= \frac{1}{2}\operatorname{Tr}(\mathcal{C}_{zz}^2) + m_z^*\mathcal{C}_{zz}m_z\end{aligned}$$

(Note: in above calculation, $\{\xi_i : i = 1, ..., n\}$ are i.i.d. from $\mathcal{N}(0, 1)$.) □

3.4.3 Stop criteria of updating the GR joint distributions

When we use sampling algorithms (like sequential Monte Carlo) to update the original joint distributions $\mathbb{Q}_t$, the natural stop criterion is $t = 1$, i.e., the algorithms should stop once the tempering parameter $t \in [0, 1]$ reaches the end $t = 1$. However, when we apply approximate algorithms (like DiEKI and DiEnKI) to update the GR joint distributions $\widetilde{\mathbb{Q}}_t^G$, the stop criteria could be more complicated. This is because the GR joint distributions are only approximations of the original joint distributions. Theoretically, it seems too difficult to establish error bounds and/or propose some preconditions that can ensure a well behavior of the Kalman-like methods. Since then, this thesis considers more practical approaches to checking some properties of the Kalman-like inversion, in

order to determine weather continue or stop the iterations. As we discussed in subsection 3.1.4, there exists monotone properties under probability $\mathbb{Q}_t$. Thus, it is hoped (but not guaranteed) that $\widetilde{\mathbb{Q}}_t^G$ as an approximation of $\mathbb{Q}_t$ should preserve the monotone properties also. If the monotone properties does not hold under probability $\widetilde{\mathbb{Q}}_s^G$ at some $s \in [0,1]$, then we should stop the algorithm in force, because $\widetilde{\mathbb{Q}}_s^G$ is not a good approximation of $\mathbb{Q}_s$ anymore.

Consider the update of the GR joint distributions,

$$\widetilde{\mathbb{Q}}_{t_0}^G \to \widetilde{\mathbb{Q}}_{t_1}^G \to \cdots \to \widetilde{\mathbb{Q}}_{t_K}^G \tag{3.4.17}$$

The stop criteria are proposed as follows:

1. The *natural stop rule*: for some positive integer K, if the tempering parameter reaches

   $$t_K = 1 \tag{3.4.18}$$

   then, we stop the algorithm, and produce $\widetilde{\mathbb{Q}}_1^G$ as the final result.

2. The *enforced stop rule* that validates the monotonicity of forward update: after the ith ($i = 1, ..., K$) iteration from $\widetilde{\mathbb{Q}}_{t_{i-1}}^G$ to $\widetilde{\mathbb{Q}}_{t_i}^G$, the cost functional and the objective functional should be decreased,

   $$\left\|\widetilde{m}_{z,t_{i-1}}\right\|_{\mathbb{R}^n}^2 \geq \|\widetilde{m}_{z,t_i}\|_{\mathbb{R}^n}^2 \tag{3.4.19}$$

   $$\left\|\widetilde{m}_{z,t_{i-1}}\right\|_{\mathbb{R}^n}^2 + \left\|\widetilde{m}_{u,t_{i-1}} - m_0\right\|_{\mathcal{C}_0}^2 \geq \|\widetilde{m}_{z,t_i}\|_{\mathbb{R}^n}^2 + \|\widetilde{m}_{u,t_i} - m_0\|_{\mathcal{C}_0}^2 \tag{3.4.20}$$

   If both of the above two conditions hold, then we continue the algorithm; otherwise we stop the algorithm, and produce $\widetilde{\mathbb{Q}}_{t_{i-1}}^G$ as the result.

The tricky point here is the enforced stop rule, see formula (3.4.19) and formula (3.4.20). Remind the fact that the tempering setting must satisfy the monotone properties, see theorems 3.1.5 3.1.3 and corollaries 3.1.6 3.1.4. Thus, it is hoped, but not guaranteed, that the Kalman-like methods should also preserve the monotone properties; otherwise, we think the Kalman-like approach is not a good approximation. In practice, the enforced stop rule improves the robustness of the Kalman-like algorithms (avoid divergence). The enforced stop rule leads to early stop of the Kalman-like algorithms. It is called 'early stop', because the resulting estimate neither touches the stopping time $t = 1$ nor closes to a stationary point.

3.4.4 List of the DMisC-Kalman-like algorithms

Let a real-valued separable Hilbert space $\mathcal{H}$ be the space of parameters. Let an Euclidean space $\mathbb{R}^n$ be the space of observations. Let $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ be the data-misfit function. Let a Gaussian measure $\mathcal{N}(m_0, \mathcal{C}_0)$ on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ be the prior probability of the unknown parameter.

Provided with the data-misfit function $\mathcal{Z}$ and the prior probability measure $\mathcal{N}(m_0, \mathcal{C}_0)$, we consider to use extended Kalman filter (EKF) and ensemble Kalman filter (EnKF) to solve the inverse problem. The Kalman-like filters solving inverse problems are explained in section 3.3. As the results, there are two kinds of implementable algorithms: DiEKI in definition 3.3.9 and DiEnKI in definition 3.3.13.

Furthermore, we can apply the data-misfit controller to select the step size, which performs well in practice. The extended Kalman inversion (EKI) and ensemble Kalman inversion (EnKI) associated with the data-misfit controller (DMisC) are called DMisC-EKI and DMisC-EnKI. DMisC-EKI is computationally efficient, but it requires the first order derivative of the data-misfit function; DMisC-EnKI is derivative-free, though it requires more computational costs for simulation of particles. DMisC-EKI and DMisC-EnKI are listed using pseudocode, see algorithm 1 and algorithm 2 in the end of this chapter. In addition, the early stop criterion can be applied to improve robustness of the Kalman-like methods. DMisC-EKI and DMisC-EnKI associated with the early stop criterion are listed in algorithm 3 and algorithm 4 in the end of this chapter.

3.5 Brief notes and summary

In this chapter, we introduce the tempering setting and adaptive methods. In brief, the contents in this chapter is about: 1) how to transform the standard setting of inverse problems to the tempering setting, 2) how the tempering setting can be simplified (using linearization) into the Kalman-like methods, and 3) how to design adaptive strategy to discretize the tempering setting and the Kalman-like methods. More details are listed as follows.

1. Mathematically, inverse problems are set via two standard approaches, i.e., the

variational setting (2.2.4) and the Bayesian setting (2.2.5). In section 3.1, we introduce a variant setting which adds a tempering parameter $t \in [0,1]$ to rewrite the variational setting as the variational tempering setting (3.1.20), and rewrite the Bayesian setting as the Bayesian tempering setting (3.1.21). The difference between the standard setting and the tempering setting is that: the standard setting is the one-step transition from prior to posterior, whereas the tempering setting gradually evolves from prior to posterior along a continuous trajectory indexed by the tempering parameter $t \in [0,1]$. The tempering setting has good properties like continuity and monotonecity, such that, the cost functional along the trajectory is continuously decreasing.

2. Numerically, the tempering setting on $t \in [0,1]$ needs to be discretized. In section 3.2, we propose an adaptive strategy selecting discrete tempering parameter $0 = t_0 < t_1 < \cdots < t_K = 1$. This adaptive method is called the data-misfit controller, see formula (3.2.15) or (3.2.27), which (precisely) controls the mean and variance of data misfits and also (approximately) controls the information gain in each step.

3. Inverse problems with the tempering setting can be equivalently regarded as filtering problems with invariant underlying states and observations. Then, we can apply approximate filtering algorithms like extended Kalman filter and ensemble Kalman filter to solve the tempered inverse problems. The Kalman-like methods for tempered inverse problems are introduced in section 3.3. Different filters can be defined depending on facts that they are continuous or discrete, derivative-required or derivative-free, mean-field limit or finite ensemble. As the results, there are six different filters in definitions 3.3.8-3.3.13, from which, CoEKI and CoMFEnKI are the theoretical limits, and DiEKI and DiEnKI are the numerically implementable versions. The relations among these methods is clearly presented as a diagram in figure 3.2.

4. For nonlinear inverse problems, if parameters and observations still have strong linear correlations, then the Kalman-like methods can be applied as approximations in practice. If so, the data-misfit controller designed for the tempering setting can be also applied as the adaptive strategy selecting discrete steps for the Kalman-

like methods. Under the Gaussian-linear approximation, section 3.4 propose the data-misfit controller for the Kalman-like methods in formula (3.4.5). Furthermore, an early stop criterion is also proposed in order to check if the Kalman-like methods are good approximations or not. If the the Kalman-like methods are bad approximations that means they cannot preserve the monotone properties of the tempering setting, then the Kalman-like filtering should be early stopped because this approximate filtering is too rough which cannot improves estimates any more.

Algorithm 1 DMisC-EKI (without early stop) for data-misfit function $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ with Gaussian prior $\mathcal{N}(m_0, \mathcal{C}_0)$

Let $\eta \equiv n/2$.

Assign the initial state $t \leftarrow 0$, $m \leftarrow m_0$, $\mathcal{C} \leftarrow \mathcal{C}_0$, $z \leftarrow \mathcal{Z}(m)$, $D \leftarrow \mathrm{D}\mathcal{Z}(m)$, $Q \leftarrow \mathcal{C}D^*$, $C \leftarrow DQ$.

while $t < 1$ **do**

Predict the step size h with the data-misfit controller,

$$h \leftarrow \min\left\{\max\left\{\eta\,/q^{(1)}\,,\,\sqrt{\eta\,/q^{(2)}}\right\},\,1-t\right\} \tag{3.5.1}$$

where

$$q^{(1)} \equiv \frac{1}{2}\operatorname{Tr}(C) + \frac{1}{2}z^*z \qquad q^{(2)} \equiv \frac{1}{2}\|C\|_F^2 + z^*Cz \tag{3.5.2}$$

where $\operatorname{Tr}(\cdot)$ is the trace of a matrix, and $\|\cdot\|_F$ is the Frobenius norm of a matrix.

Predict the mean-covariance pair $(m_p, \mathcal{C}_p)$ with the extended Kalman filter,

$$m_p \leftarrow m - hQ\left(\mathbf{I} + hC\right)^{-1} z \qquad \mathcal{C}_p \leftarrow \mathcal{C} - hQ\left(\mathbf{I} + hC\right)^{-1} Q^* \tag{3.5.3}$$

Calculate

$$z_p \leftarrow \mathcal{Z}(m_p) \qquad D_p \leftarrow \mathrm{D}\mathcal{Z}(m_p) \qquad Q_p \leftarrow \mathcal{C}_pD_p^* \qquad C_p \leftarrow D_pQ_p \tag{3.5.4}$$

Renew the state $t \leftarrow t + h$, $m \leftarrow m_p$, $\mathcal{C} \leftarrow \mathcal{C}_p$, $z \leftarrow z_p$, $D \leftarrow D_p$, $Q \leftarrow Q_p$, $C \leftarrow C_p$.

end while

return $(m, \mathcal{C})$ as the mean-covariance pair of the posterior distribution (approximately).

Algorithm 2 DMisC-EnKI (without early stop) for data-misfit function $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ with Gaussian prior $\mathcal{N}(m_0, \mathcal{C}_0)$

Provide the sample size J. Let $\eta \equiv n/2$.

Assign $t \leftarrow 0$. Draw J samples independently from the prior distribution $\mathcal{N}(m_0, \mathcal{C}_0)$, and let these K samples into a particle U. Calculate the particle Z such that, $Z(j) \leftarrow \mathcal{Z}(U(j))$ for all $j = 1, ..., J$. Calculate the sample means and sample covariances, $\overline{u} \leftarrow \text{mean}(U)$, $\overline{z} \leftarrow \text{mean}(Z)$, $C_{uz} \leftarrow \text{cov}(U, Z)$, $C_{zz} \leftarrow \text{cov}(Z, Z)$.

while $t < 1$ **do**

Predict the step size h with the data-misfit controller,

$$h \leftarrow \min\left\{\max\left\{\eta / q^{(1)}\ ,\ \sqrt{\eta / q^{(2)}}\right\},\ 1 - t\right\} \tag{3.5.5}$$

where

$$q^{(1)} \equiv \frac{1}{2}\,\text{Tr}\,(C_{zz}) + \overline{z}^* \overline{z} \qquad q^{(2)} \equiv \frac{1}{2}\,\|C_{zz}\|_F^2 + \overline{z}^* C_{zz} \overline{z} \tag{3.5.6}$$

where $\text{Tr}(\cdot)$ is the trace of a matrix, and $\|\cdot\|_F$ is the Frobenius norm of a matrix.

Draw J samples independently from the n-dimensional standard normal distribution, and let these J samples into a particle V. Predict the particle U_p with the ensemble Kalman filter, such that, for all $j = 1, ..., J$,

$$U_p(j) \leftarrow U(j) - C_{uz}\,(\mathbf{I} + hC_{zz})^{-1}\left(Z(j)h - V(j)\sqrt{h}\right) \tag{3.5.7}$$

Calculate the particle Z_p such that, $Z_p(j) \leftarrow \mathcal{Z}(U_p(j))$ for all $j = 1, ..., J$. Calculate the sample means and sample covariances, $\overline{u}_p \leftarrow \text{mean}(U_p)$, $\overline{z}_p \leftarrow \text{mean}(Z_p)$, $C_{uz,p} \leftarrow \text{cov}(U_p, Z_p)$, $C_{zz,p} \leftarrow \text{cov}(Z_p, Z_p)$.

Renew the state $t \leftarrow t + h$, $U \leftarrow U_p$, $Z \leftarrow Z_p$, $\overline{u} \leftarrow \overline{u}_p$, $\overline{z} \leftarrow \overline{z}_p$, $C_{uz} \leftarrow C_{uz,p}$, $C_{zz} \leftarrow C_{zz,p}$.

end while

return U as the particle under the posterior distribution (approximately).

Algorithm 3 DMisC-EKI (with early stop) for data-misfit function $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ with Gaussian prior $\mathcal{N}(m_0, \mathcal{C}_0)$

Let $\eta \equiv n/2$.

Assign the initial state $t \leftarrow 0$, $m \leftarrow m_0$, $\mathcal{C} \leftarrow \mathcal{C}_0$, $z \leftarrow \mathcal{Z}(m)$, $D \leftarrow \mathrm{D}\mathcal{Z}(m)$, $Q \leftarrow \mathcal{C}D^*$, $C \leftarrow DQ$.

while $t < 1$ **do**

Predict the step size h with the data-misfit controller,

$$h \leftarrow \min\left\{\max\left\{\eta\,/q^{(1)}\,,\,\sqrt{\eta\,/q^{(2)}}\right\},\,1-t\right\} \tag{3.5.8}$$

where

$$q^{(1)} \equiv \frac{1}{2}\operatorname{Tr}(C) + \frac{1}{2}z^*z \qquad q^{(2)} \equiv \frac{1}{2}\|C\|_F^2 + z^*Cz \tag{3.5.9}$$

where $\operatorname{Tr}(\cdot)$ is the trace of a matrix, and $\|\cdot\|_F$ is the Frobenius norm of a matrix.

Predict the mean-covariance pair $(m_p, \mathcal{C}_p)$ with the extended Kalman filter,

$$m_p \leftarrow m - hQ\left(\mathbf{I} + hC\right)^{-1} z \qquad \mathcal{C}_p \leftarrow \mathcal{C} - hQ\left(\mathbf{I} + hC\right)^{-1} Q^* \tag{3.5.10}$$

Calculate

$$z_p \leftarrow \mathcal{Z}(m_p) \qquad D_p \leftarrow \mathrm{D}\mathcal{Z}(m_p) \qquad Q_p \leftarrow \mathcal{C}_pD_p^* \qquad C_p \leftarrow D_pQ_p \tag{3.5.11}$$

if $\|z_p\|_{\mathbb{R}^n}^2 > \|z\|_{\mathbb{R}^n}^2$ **or** $\|z_p\|_{\mathbb{R}^n}^2 + \|m_p - m_0\|_{\mathcal{C}_0}^2 > \|z\|_{\mathbb{R}^n}^2 + \|m - m_0\|_{\mathcal{C}_0}^2$ **then**

break (early stop).

end if

Renew the state $t \leftarrow t + h$, $m \leftarrow m_p$, $\mathcal{C} \leftarrow \mathcal{C}_p$, $z \leftarrow z_p$, $D \leftarrow D_p$, $Q \leftarrow Q_p$, $C \leftarrow C_p$.

end while

return $(m, \mathcal{C})$ as the mean-covariance pair of the posterior distribution (approximately).

Algorithm 4 DMisC-EnKI (with early stop) for data-misfit function $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ with Gaussian prior $\mathcal{N}(m_0, \mathcal{C}_0)$

Provide the sample size J. Let $\eta \equiv n/2$.

Assign $t \leftarrow 0$. Draw J samples independently from the prior distribution $\mathcal{N}(m_0, \mathcal{C}_0)$, and let these K samples into a particle U. Calculate the particle Z such that, $Z(j) \leftarrow \mathcal{Z}(U(j))$ for all $j = 1, ..., J$. Calculate the sample means and sample covariances, $\overline{u} \leftarrow \text{mean}(U)$, $\overline{z} \leftarrow \text{mean}(Z)$, $C_{uz} \leftarrow \text{cov}(U, Z)$, $C_{zz} \leftarrow \text{cov}(Z, Z)$.

while $t < 1$ **do**

Predict the step size h with the data-misfit controller,

$$h \leftarrow \min\left\{\max\left\{\eta / q^{(1)}\ ,\ \sqrt{\eta / q^{(2)}}\right\},\ 1 - t\right\} \tag{3.5.12}$$

where

$$q^{(1)} \equiv \frac{1}{2}\,\text{Tr}\,(C_{zz}) + \overline{z}^* \overline{z} \qquad q^{(2)} \equiv \frac{1}{2}\,\|C_{zz}\|_F^2 + \overline{z}^* C_{zz} \overline{z} \tag{3.5.13}$$

where $\text{Tr}(\cdot)$ is the trace of a matrix, and $\|\cdot\|_F$ is the Frobenius norm of a matrix.

Draw J samples independently from the n-dimensional standard normal distribution, and let these J samples into a particle V. Predict the particle U_p with the ensemble Kalman filter, such that, for all $j = 1, ..., J$,

$$U_p(j) \leftarrow U(j) - C_{uz}\,(\mathbf{I} + hC_{zz})^{-1}\left(Z(j)h - V(j)\sqrt{h}\right) \tag{3.5.14}$$

Calculate the particle Z_p such that, $Z_p(j) \leftarrow \mathcal{Z}(U_p(j))$ for all $j = 1, ..., J$. Calculate the sample means and sample covariances, $\overline{u}_p \leftarrow \text{mean}(U_p)$, $\overline{z}_p \leftarrow \text{mean}(Z_p)$, $C_{uz,p} \leftarrow \text{cov}(U_p, Z_p)$, $C_{zz,p} \leftarrow \text{cov}(Z_p, Z_p)$.

if $\|\overline{z}_p\|_{\mathbb{R}^n}^2 > \|\overline{z}\|_{\mathbb{R}^n}^2$ **or** $\|\overline{z}_p\|_{\mathbb{R}^n}^2 + \|\overline{u}_p - m_0\|_{\mathcal{C}_0}^2 > \|\overline{z}\|_{\mathbb{R}^n}^2 + \|\overline{u} - m_0\|_{\mathcal{C}_0}^2$ **then**

break (early stop).

end if

Renew the state $t \leftarrow t + h$, $U \leftarrow U_p$, $Z \leftarrow Z_p$, $\overline{u} \leftarrow \overline{u}_p$, $\overline{z} \leftarrow \overline{z}_p$, $C_{uz} \leftarrow C_{uz,p}$, $C_{zz} \leftarrow C_{zz,p}$.

end while

return U as the particle under the posterior distribution (approximately).

Chapter 4

Theoretical Analysis and Proofs

This chapter gathers theoretical analysis, which is used to prove theorems and propositions in the previous two chapters. This chapter are divided into several topics:

1. Section 4.1 analyzes solutions of Tikhonov regularization depending on the regularizing parameters. Results in this section are used for the tempered variational inversion (3.1.20).

2. Section 4.2 generalizes statistical thermodynamics from finite dimensions (sample distributions) to infinite dimensions (probability measures). Results in this section are used for the tempered Bayesian inversion (3.1.21).

3. Section 4.3 discusses the technique of integration by parts with respect to Gaussian probability measures on separable Hilbert spaces. This is a useful tool for Bayesian inference with Gaussian priors.

4.1 Tikhonov regularization on RKHS

Tikhonov regularization is originally named after the Russian mathematician Andrey Nikolayevich Tikhonov. Now, it is regarded as the fundamental method in inverse problems. Tikhonov's collection book was published in 1977 [10]. The convergence rate of Tikhonov methods for ill-posed nonlinear problems in general Hilbert or Banach spaces was densely studies in 1989 [97, 42]. Further investigation about the regularizing parameter determined via the Morozov's discrepancy principle was published in [80, 96].

Tikhonov regularization is closely related to the tempered variational inversion (3.1.20). In fact, the tempered variational inversion can be regarded as Tikhonov regularization with continuous regularizing parameter. The reciprocal of the regularizing parameter in the Tikhonov regularization is exactly the tempering parameter in the tempered variational inversion. This section aims to analyze some properties of the tempered variational inversion.

4.1.1 Formulation of the problem

We start from the standard form of Tikhonov regularization. Consider the following optimization problem in a separable Hilbert space $\mathcal{H}$,

$$\widetilde{x}_\alpha = \arg\min_{x\in\mathcal{H}} \left\{ \Phi\left(u(x)\right) + \frac{\alpha}{2}\|x\|_{\mathcal{H}}^2 \right\} \tag{4.1.1}$$

where $u : \mathcal{H} \to \mathcal{H}$ is a coordinate transformation, $\Phi : \mathcal{H} \to [0, +\infty)$ is a cost functional (typically, Φ can be specified as a quadratic form $\Phi(\cdot) = \frac{1}{2}\|\mathcal{Z}(\cdot)\|_{\mathbb{R}^n}^2$ associated with a function $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$), and $\alpha > 0$ is the regularizing parameter of Tikhonov regularization.

We analyze a special but typical setting of the Tikhonov regularization that constrains parameters from the separable Hilbert space $\mathcal{H}$ to a compact subspace. Namely, consider the affine transformation,

$$u = u(x) = \mathcal{C}_0^{1/2} x + m_0 \tag{4.1.2}$$

where $\mathcal{C}_0 : \mathcal{H} \to \mathcal{H}$ is a self-adjoint positive-semi-definite trace-class operator, and $m_0 \in \mathcal{H}$ is a bounded element for translation. $\mathcal{C}_0$ is trace class, so $\mathcal{C}_0^{1/2}$ is Hilbert-Schmidt, which implies that $\mathcal{C}_0^{1/2}$ is a compact operator. Therefore, u is in a compact set for any bounded $x \in \mathcal{H}$.

Now, we can derive the tempered variational inversion (3.1.20) by applying the following variable substitution into formula (4.1.1).

$$x = \mathcal{C}_0^{-1/2}(u - m_0) \qquad \alpha = 1/t \tag{4.1.3}$$

As the result, for any $\alpha > 0$, $\widetilde{x}_\alpha$ in formula (4.1.1) can be represented by

$$\widetilde{x}_\alpha = \mathcal{C}_0^{-1/2}(\widehat{x}_{1/\alpha} - m_0) \tag{4.1.4}$$

where $\widehat{x}_t$ for any bounded $t \geq 0$ is determined by

$$\widehat{x}_t = \arg\min_{u \in \mathcal{H}} \left\{ t\Phi(u) + \frac{1}{2}\|u - m_0\|_E^2 \right\} \tag{4.1.5}$$

where $E = \mathrm{Ran}\left(\mathcal{C}_0^{1/2}\right) \subset \mathcal{H}$ is the reproducing kernel Hilbert space (RKHS) equipped with inner product, for all $v, w \in E$,

$$\langle v, w \rangle_E = \left\langle \mathcal{C}_0^{-1/2} v, \mathcal{C}_0^{-1/2} w \right\rangle_{\mathcal{H}} \tag{4.1.6}$$

The tempered variational inversion (3.1.20) is exactly formula (4.1.5) with the cost functional Φ specified as $\Phi(\cdot) = \frac{1}{2}\|\mathcal{Z}(\cdot)\|_{\mathbb{R}^n}^2$, where $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ is the data-misfit function.

For convenience, let $O_t : \mathcal{H} \to [0, +\infty)$ be the objective functional in the optimization (4.1.5),

$$O_t(\cdot) = t\Phi(\cdot) + R(\cdot) \quad \text{with} \quad R(\cdot) = \frac{1}{2}\|(\cdot) - m_0\|_E^2 \tag{4.1.7}$$

4.1.2 Existence and stability of solution

The existence of solution of Tikhonov regularization is well discussed in [10, 97, 42, 80, 96]. We consider a special case of Tikhonov regularization with parameters in a compact set and observations in an Euclidean space. Thus, we just conduct trivial derivation to show the existence of the solution in our case. Our purpose is to prove theorem 4.1.5 more naturally, which is one of the main theorems proposed in this thesis. Theorem 4.1.5 shows the monotone property of cost functional with respect to the regularizing parameter.

The following two definitions describe 'balls' in space $\mathcal{H}$ and space $m_0 + E$.

Definition 4.1.1. *For any $r > 0$, let $B_0(r)$ denote the open ball in $\mathcal{H}$ centered at 0, i.e.*

$$B_0(r) := \{u \in \mathcal{H} : \ \|u\|_{\mathcal{H}} < r\} \tag{4.1.8}$$

Moreover, let $\overline{B_0(r)}$ denote the closure of $B_0(r)$, i.e.

$$\overline{B_0(r)} := \{u \in \mathcal{H} : \ \|u\|_{\mathcal{H}} \leq r\} \tag{4.1.9}$$

Definition 4.1.2. *For any $M > 0$, let $E_0(M)$ denote the open ball in $m_0 + E$ centered at m_0, i.e.*

$$E_0(r) := \{u \in \mathcal{H} : \ \|u - m_0\|_E < M\} \tag{4.1.10}$$

Moreover, let $\overline{E_0(M)}$ *denote the closure of* $E_0(M)$*, i.e.*

$$\overline{E_0(M)} := \{u \in \mathcal{H} : \|u - m_0\|_E \leq M\} \tag{4.1.11}$$

The following lemma shows that any bounded ball in m_0+E is contained in a bounded ball in $\mathcal{H}$, i.e. the RKHS E is continuously embedded in the Hilbert space $\mathcal{H}$.

Lemma 4.1.3. *For any bounded* $M > 0$*, there exists an* $r = r(M) > 0$ *such that,*

$$\overline{E_0(M)} \subseteq \overline{B_0(r(M))} \tag{4.1.12}$$

where

$$r(M) := \|m_0\|_{\mathcal{H}} + \|\mathcal{C}_0\|_{op}^{1/2} M \tag{4.1.13}$$

Equivalently, E *is continuously embedded in* $\mathcal{H}$*, i.e. for any* $u \in E$*,*

$$\|u\|_E \leq \|\mathcal{C}_0\|_{op}^{1/2} \|u\|_{\mathcal{H}} \tag{4.1.14}$$

Proof. For any $u \in m_0 + E$, we have

$$\begin{aligned}
\|u\|_{\mathcal{H}} &= \|u - m_0 + m_0\|_{\mathcal{H}} && (4.1.15)\\
&\leq \|u - m_0\|_{\mathcal{H}} + \|m_0\|_{\mathcal{H}} && (4.1.16)\\
&= \left\|\mathcal{C}_0^{1/2}\mathcal{C}_0^{-1/2}(u - m_0)\right\|_{\mathcal{H}} + \|m_0\|_{\mathcal{H}} && (4.1.17)\\
&\leq \left\|\mathcal{C}_0^{1/2}\right\|_{op} \left\|\mathcal{C}_0^{-1/2}(u - m_0)\right\|_{\mathcal{H}} + \|m_0\|_{\mathcal{H}} && (4.1.18)\\
&= \|\mathcal{C}_0\|_{op}^{1/2} \|u - m_0\|_E + \|m_0\|_{\mathcal{H}} && (4.1.19)
\end{aligned}$$

Thus,

$$\|u - m_0\|_E \leq M \implies \|u\|_{\mathcal{H}} \leq \|\mathcal{C}_0\|_{op}^{1/2} M + \|m_0\|_{\mathcal{H}} \tag{4.1.20}$$

The equivalent statement is nothing more than using $m_0 = 0$. □

The existence of the minimum point $\widehat{x}_t$ in formula (4.1.5) is ensured by the following theorem.

Theorem 4.1.4. *Assume that the cost functional* $\Phi : \mathcal{H} \to [0, +\infty)$ *in formula* (4.1.7) *is Lipschitz continuous on any bounded and closed subsets, i.e. for every* $r > 0$ *there is a* $K = K(r) > 0$ *such that, for all* $u_1, u_2 \in \mathcal{H}$ *with* $\max\{\|u_1\|_{\mathcal{H}}, \|u_2\|_{\mathcal{H}}\} \leq r$*,*

$$|\Phi(u_1) - \Phi(u_2)| \leq K\|u_1 - u_2\|_{\mathcal{H}} \tag{4.1.21}$$

Then for any bounded $t \geq 0$, the minimum $\widehat{x}_t$ determined in formula (4.1.5) *exists and satisfies*

$$R(\widehat{x}_t) \leq t\Phi(m_0) \tag{4.1.22}$$

where R and Φ are the same as those in formula (4.1.7).

Proof. It is clear that the objective functional $O_t(u) = t\Phi(u) + R(u)$ in formula (4.1.7) is bounded below $O_t(u) \geq 0$ for any $t \geq 0$ and $u \in \mathcal{H}$, so there exists the unique infimum,

$$\inf \{O_t(u) : u \in \mathcal{H}\} \tag{4.1.23}$$

For any bounded $t \geq 0$ and any bounded $M > 0$, define a positive real number $C_t(M) \geq M$,

$$C_t(M) := \sqrt{M^2 + 2tK(r(M))\|\mathcal{C}_0\|_{op}^{1/2} M + 2t\Phi(m_0)} \tag{4.1.24}$$

This number $C_t(M)$ will play the essential role in the following proof.

On one hand, for all $u \in \overline{E_0(M)} \subseteq \overline{B_0(r(M))}$, where $r(M)$ is given in formula (4.1.13), we have

$$\begin{aligned}
O_t(u) &= R(u) + t\Phi(u) && (4.1.25)\\
&\leq R(u) + t|\Phi(u) - \Phi(m_0)| + t\Phi(m_0) && (4.1.26)\\
&\leq R(u) + tK(r(M))\|u - m_0\|_{\mathcal{H}} + t\Phi(m_0) && (4.1.27)\\
&\leq R(u) + tK(r(M))\|\mathcal{C}_0\|_{op}^{1/2}\|u - m_0\|_E + t\Phi(m_0) && (4.1.28)\\
&\leq \frac{1}{2}M^2 + tK(r(M))\|\mathcal{C}_0\|_{op}^{1/2} M + t\Phi(m_0) = \frac{1}{2}C_t(M)^2 && (4.1.29)
\end{aligned}$$

On the other hand, for for all $u \in \mathcal{H} \setminus \overline{E_0(C_t(M))}$, we have

$$\begin{aligned}
O_t(u) &= R(u) + t\Phi(u) && (4.1.30)\\
&\geq R(u) && (4.1.31)\\
&> \frac{1}{2}C_t(M)^2 && (4.1.32)
\end{aligned}$$

In conclusion, $\forall u_1 \in \overline{E_0(M)} \subseteq \overline{E_0(C_t(M))}$ and $\forall u_2 \in \mathcal{H} \setminus \overline{E_0(C_t(M))}$,

$$O_t(u_1) < O_t(u_2) \tag{4.1.33}$$

Thus, we have

$$\inf\left\{O_t(u) : u \in \overline{E_0(C_t(M))}\right\} \tag{4.1.34}$$

$$\leq \inf\left\{O_t(u) : u \in \overline{E_0(M)} \subseteq \overline{E_0(C_t(M))}\right\} \tag{4.1.35}$$

$$\leq \inf\left\{O_t(u) : u \in \mathcal{H} \setminus \overline{E_0(C_t(M))}\right\} \tag{4.1.36}$$

Namely,

$$\inf\left\{O_t(u) : u \in \mathcal{H}\right\} \tag{4.1.37}$$

$$= \min\left\{\inf\left\{O_t(u) : u \in \overline{E_0(C_t(M))}\right\}, \inf\left\{O_t(u) : u \in \mathcal{H} \setminus \overline{E_0(C_t(M))}\right\}\right\} \tag{4.1.38}$$

$$= \inf\left\{O_t(u) : u \in \overline{E_0(C_t(M))}\right\} \tag{4.1.39}$$

Moreover, since $O_t : \mathcal{H} \to [0, +\infty)$ is a continuous function on the compact subset $\overline{E_0(C_t(M))} \subset \mathcal{H}$, the minimum exists (according to the extreme value theorem), i.e.

$$\inf\left\{O_t(u) : u \in \overline{E_0(C_t(M))}\right\} = \min\left\{O_t(u) : u \in \overline{E_0(C_t(M))}\right\} \tag{4.1.40}$$

In addition, the minimum $\widehat{x}_t$ must be in the compact sets $\overline{E_0(C_t(M))}$ for all $M > 0$, that means

$$\|\widehat{x}_t - m_0\|_E \leq \inf_{M>0}\{C_t(M)\} = \sqrt{2t\Phi(m_0)} \tag{4.1.41}$$

or expressed by

$$R(\widehat{x}_t) = \frac{1}{2}\|\widehat{x}_t - m_0\|_E^2 \leq t\Phi(m_0) \tag{4.1.42}$$

□

Moreover, the following theorem reveals the stability of minimum point $\widehat{x}_t$ relying on parameter t, such that, the minimum value $O_t(\widehat{x}_t)$ regarded as a function of t is Lipschitz continuous on any closed interval $t \in [0, c]$ $(c > 0)$, and the cost functional at the minimum point $\Phi(\widehat{x}_t)$ is decreasing as a function of t.

Theorem 4.1.5. *For any $t \geq 0$, define the following quantity*

$$J_t := O_t(\widehat{x}_t) \qquad \phi_t := \Phi(\widehat{x}_t) \tag{4.1.43}$$

where $O_t : \mathcal{H} \to [0, +\infty)$ and $\Phi : \mathcal{H} \to [0, +\infty)$ are the same as those in formula (4.1.7)*, and $\widehat{x}_t$ is the minimum point of O_t determined in formula* (4.1.5)*. If Φ is Lipschitz*

continuous on any bounded and closed subsets, then for any bounded $c > 0$, J_t is increasing in $t \in [0, c]$, and ϕ_t is decreasing in $t \in [0, c]$. Furthermore, J_t is Lipschitz continuous in $t \in [0, c]$ with derivative almost everywhere, represented by

$$J_t = \int_0^t \phi_s \, \mathrm{d}s \tag{4.1.44}$$

Proof. First of all, since Φ is Lipschitz continuous on any bounded and closed subsets, the minimum point $\widehat{x}_t$ exists according to theorem 4.1.4. Thus, J_t and ϕ_t are well-defined.

Furthermore, consider that, given any $0 \leq t < t + \Delta \leq 1$, the relations below hold,

$$J_t = \min_{u \in \mathcal{H}} \{O_t(u)\} \leq O_t\left(\widehat{x}_{t+\Delta}\right) \tag{4.1.45}$$

$$J_{t+\Delta} = \min_{u \in \mathcal{H}} \{O_{t+\Delta}(u)\} \leq O_{t+\Delta}\left(\widehat{x}_t\right) \tag{4.1.46}$$

The right hand sides of formulas (4.1.45) and (4.1.46) can be represented by

$$O_t\left(\widehat{x}_{t+\Delta}\right) = O_{t+\Delta}\left(\widehat{x}_{t+\Delta}\right) - \Delta\Phi\left(\widehat{x}_{t+\Delta}\right) = J_{t+\Delta} - \Delta\phi_{t+\Delta} \tag{4.1.47}$$

$$O_{t+\Delta}\left(\widehat{x}_t\right) = O_t\left(\widehat{x}_t\right) + \Delta\Phi\left(\widehat{x}_t\right) = J_t + \Delta\phi_t \tag{4.1.48}$$

Substitute equations (4.1.47) and (4.1.48) into formulas (4.1.45) and (4.1.46), respectively, to obtain

$$J_{t+\Delta} - J_t \geq \Delta\phi_{t+\Delta} \tag{4.1.49}$$

$$J_{t+\Delta} - J_t \leq \Delta\phi_t \tag{4.1.50}$$

As the results, we have the following statements:

1. J_t is increasing, because formula (4.1.49) determines that for all $\Delta > 0$, we have $J_{t+\Delta} - J_t \geq \Delta\phi_{t+\Delta} \geq 0$.

2. ϕ_t is decreasing, because formulas (4.1.49) and (4.1.50) determine that for all $\Delta > 0$, we have $\Delta\phi_{t+\Delta} \leq J_{t+\Delta} - J_t \leq \Delta\phi_t$.

3. ϕ_t is decreasing, so $\phi_t \leq \phi_0$ is bounded for all $t \in [0, c]$. As the result, the Riemann integral on interval $[0, c]$ is bounded,

$$\int_0^c \phi_t \, \mathrm{d}t \leq c\phi_0 < \infty \tag{4.1.51}$$

Since then, formulas (4.1.49) and (4.1.50) are the backward and forward finite difference schemes, and the infinite Riemann sum on interval $[0, c]$ is bounded. Thus, J_t is absolutely continuous with derivative ϕ_t almost everywhere, equivalently characterized by the fundamental theorem of integral calculus,

$$J_t = J_0 + \int_0^t \phi_s \, \mathrm{d}s \tag{4.1.52}$$

where $J_0 = 0$. Furthermore, since ϕ_t is bounded for any $t \in [0, c]$, J_t is also Lipschitz continuous on $[0, c]$.

□

Remark 4.1.6. *Theorem 4.1.5 implies stronger arguments than theorem 4.1.4, though the two theorems are based on the same assumption (Lipschitz continuity). In fact, formula* (4.1.22) *in theorem 4.1.4 is implied by formula* (4.1.44) *in theorem 4.1.5, since*

$$R(\widehat{x}_t) \leq O_t(\widehat{x}_t) = \int_0^t \Phi(\widehat{x}_s) \, \mathrm{d}s \leq t\Phi(m_0) \tag{4.1.53}$$

4.1.3 Uniqueness and searching line of solution

This subsection aims to derive the ordinary differential equation (4.1.58), which can be regarded as a variant of Newton's method. The difference is that, the original Newton's method is an iterative approach, but the ordinary differential equation (4.1.58) is continuous. Similar like Newton's method that produces a stationary point in optimization, equation (4.1.58) provides a trajectory of stationary points of the objective functional O_t along $t \in [0, c]$ for a bounded $c > 0$.

Twice differentialbility of the objective functional is usually required in mathematical optimization using Newton's methods. The first derivative is the gradient vector, and the second derivative is the Hessian matrix. In infinite-dimensional spaces, we define the gradient and Hessian by using the first and second order Fréchet derivatives as follows.

Definition 4.1.7 (gradient and Hessian on Hilbert spaces)**.** *Let $f : \mathcal{X} \to \mathbb{R}$ be a twice differentiable functional on a Hilbert space $\mathcal{X}$. Let $\mathrm{D}f(x) \in \mathcal{X}^*$ and $\mathrm{D}^2 f(x) : \mathcal{X} \to \mathcal{X}^*$ denote the first and second order Fréchet derivatives of f at $x \in \mathcal{X}$. More conveniently, equivalent notation of the Fréchet derivatives can be defined as* ***gradient*** *and* ***Hessian****.*

Let $\nabla f(x) \in \mathcal{X}$ and $\mathrm{H}f(x) : \mathcal{X} \to \mathcal{X}$ denote the gradient and Hessian of f at $x \in \mathcal{X}$, such that, for all bounded $v, w \in \mathcal{X}$, the following equations hold,

$$\langle \nabla f(x), v \rangle_{\mathcal{X}} = [\mathrm{D}f(x)](v) \tag{4.1.54}$$

$$\langle [\mathrm{H}f(x)](w), v \rangle_{\mathcal{X}} = [\mathrm{D}^2 f(x)](v)(w) \tag{4.1.55}$$

where $\nabla f(x)$ and $\mathrm{H}f(x)$ are uniquely determined via the Riesz representation theorem.

Moreover, we generalize the concept of derivative from Euclidean spaces to Hilbert spaces. The following generalization is similar as the generalization of gradient and Hessian in definition 4.1.7.

Definition 4.1.8 (differential on Hilbert spaces)**.** *Let $\{x_t \in \mathcal{X} : t \in \mathbb{R}\}$ be a sequence on a Hilbert space $\mathcal{X}$. x_t is called **differentiable** at $t \in \mathbb{R}$, if there exists an element $x_t' \in \mathcal{H}$ such that, for all bounded $h \in \mathcal{H}$, the following equation holds,*

$$\langle x_t', h \rangle_{\mathcal{X}} = \lim_{\Delta \to 0} \frac{\langle x_{t+\Delta} - x_t, h \rangle_{\mathcal{X}}}{\Delta} \tag{4.1.56}$$

If the limit in the right hand side exists for all h, then x_t' is uniquely determined via the Riesz representation theorem. Then, x_t' is the derivative of x_t at t, denoted by the differential equation below,

$$\mathrm{d}x_t = x_t' \mathrm{d}t \tag{4.1.57}$$

*If x_t is differentiable for all $t \in \Omega$, where Ω is an interval in $\mathbb{R}$, then we say x_t has a **differentiable path** in Ω.*

The following theorem shows that there exists a differentiable path of stationary points of the objective functional O_t. Furthermore, for convex optimization, the stationary points are also the global minimums. The following method can be regarded as a variant of the Newton's method with continuous trajectory (the original Newton's method has discrete iterations).

Theorem 4.1.9. *Let $O_t : \mathcal{H} \to [0, +\infty)$ and $\Phi : \mathcal{H} \to [0, +\infty)$ be the objective functional and the cost functional in formula (4.1.7), respectively. Assume Φ is twice differentiable. Consider the following initial value problem,*

$$\widehat{m}_0 = m_0, \qquad \mathrm{d}\widehat{m}_t = -\left(t\mathrm{H}\Phi(\widehat{m}_t) + \mathcal{C}_0^{-1}\right)^{-1} \nabla\Phi(\widehat{m}_t)\mathrm{d}t \quad t > 0 \tag{4.1.58}$$

where $\nabla\Phi(x) \in \mathcal{H}$ *and* $\mathrm{H}\Phi(x) : \mathcal{H} \to \mathcal{H}$ *are the gradient and Hessian of* Φ *at* $x \in \mathcal{H}$*, respectively. Then* $\widehat{m}_t$ *is a stationary point of the objective functional* O_t *for any bounded* $t \geq 0$*. Moreover, if the Hessian* $\mathrm{H}\Phi(x)$ *is a self-adjoint positive-semi-definite operator for all* $x \in \mathcal{H}$*, then* $\widehat{m}_t$ *is the unique global minimum of the objective functional* O_t *for any bounded* $t \geq 0$*.*

Proof. Φ is twice differentiale, which implies Lipschitz continuity on any bounded and closed subsets. Thus, theorem 4.1.4 ensures the existence of minimums. Moreover, in this proof, we aim to show that $\widehat{m}_t$ determined by the ODE (4.1.58) is indeed the unique minimum of the objective functional O_t, on the condition that the Hessian of cost functional is positive-semi-definite.

Firstly, we address a necessary condition. Since the objective functional O_t is twice differentiable, any minimum points must be stationary points, i.e. the derivative of O_t at $\widehat{m}_t$ must be zero. Therefore, $\widehat{m}_t$ should necessarily satisfy, for all bounded $h \in \mathcal{H}$,

$$\left[\mathrm{D}O_t\left(\widehat{m}_t\right)\right](h) = \left[\mathrm{D}_u\left(t\Phi(u) + \frac{1}{2}\left\|u - m_0\right\|_E^2\right)\Big|_{u=\widehat{m}_t}\right](h) \tag{4.1.59}$$

$$= t\left[\mathrm{D}\Phi(\widehat{m}_t)\right](h) + \left\langle \widehat{m}_t - m_0, h\right\rangle_E \tag{4.1.60}$$

$$= \left\langle t\nabla\Phi(\widehat{m}_t), h\right\rangle_{\mathcal{H}} + \left\langle \mathcal{C}_0^{-1}(\widehat{m}_t - m_0), h\right\rangle_{\mathcal{H}} = 0 \tag{4.1.61}$$

For convenience, let $r_t : \mathcal{H} \to \mathcal{H}$ denote

$$r_t(\cdot) \equiv t\nabla\Phi(\cdot) + \mathcal{C}_0^{-1}((\cdot) - m_0) \tag{4.1.62}$$

Then equation (4.1.61) becomes to

$$\left\langle r_t(\widehat{m}_t), h\right\rangle_{\mathcal{H}} = 0 \tag{4.1.63}$$

Above formula holds for all bounded $h \in \mathcal{H}$, which implies

$$r_t(\widehat{m}_t) = 0 \tag{4.1.64}$$

Now, we check that $\widehat{m}_t$ determined by the ODE (4.1.58) satisfies $r_t(\widehat{m}_t) = 0$, so $\widehat{m}_t$ is a stationary point. In fact, the ODE implies even stronger arguments such that, $r_0(\widehat{m}_0) = 0$ and $\mathrm{d}r_t(\widehat{m}_t) = 0$ for any $t > 0$, which make $r_t(\widehat{m}_t) = 0$ consistently hold for all $t \geq 0$. In

the following, we check the two conditions: 1) $r_0(\widehat{m}_0) = 0$, and 2) $\mathrm{d}r_t(\widehat{m}_t) = 0$ for any $t > 0$. On one hand, let $t = 0$, $r_0(\widehat{m}_t)$ equals

$$r_0(\widehat{m}_0) = \mathcal{C}_0^{-1}(\widehat{m}_0 - m_0) \tag{4.1.65}$$

However, the initial condition in the ODE (4.1.58) is $\widehat{m}_0 = m_0$, so we have

$$r_0(\widehat{m}_0) = 0 \tag{4.1.66}$$

On the other hand, for any $t > 0$, $\mathrm{d}r_t(\widehat{m}_t)$ can be explicitly calculated by chain rule,

$$\mathrm{d}r_t(\widehat{m}_t) = \nabla\Phi(\widehat{m}_t)\mathrm{d}t + t\mathrm{H}\Phi(\widehat{m}_t)\mathrm{d}\widehat{m}_t + \mathcal{C}_0^{-1}\mathrm{d}\widehat{m}_t \tag{4.1.67}$$

However, substitute the ODE (4.1.58) into above formula leads to

$$\mathrm{d}r_t(\widehat{m}_t) = 0 \tag{4.1.68}$$

In conclusion, the two conditions $r_0(\widehat{m}_0) = 0$ and $\mathrm{d}r_t(\widehat{m}_t) = 0$ hold, so $r_t(\widehat{m}_t) = 0$ consistently hold for $t \geq 0$. Thus, $\widehat{m}_t$ is a stationary point of O_t for any bounded $t \geq 0$.

Moreover, we show that, for any bounded $t \geq 0$, if the Hessian of cost functional is positive-semi-definite, then the Hessian of objective functional is positive-definite. (The Hessian of objective functional is positive-definite rather than positive-semi-definite, because there exists the penalty functional.) To show the positive-definiteness of the Hessian of the objective functional O_t at $x \in \mathcal{H}$, we have to show, for all bounded $h \in \mathcal{H} \wedge h \neq 0$, the following inequality holds,

$$\langle \mathrm{H}O_t(x)h, h\rangle_{\mathcal{H}} = [\mathrm{D}^2 O_t(x)](h)(h) > 0 \tag{4.1.69}$$

In fact, we have

$$\left[\mathrm{D}^2 O_t\left(x\right)\right](h)(h) = \left[\mathrm{D}_u^2\left(t\Phi(u) + \frac{1}{2}\left\|u - m_0\right\|_E^2\right)\Big|_{u=x}\right](h)(h) \tag{4.1.70}$$

$$= t\left[\mathrm{D}^2\Phi(x)\right](h)(h) + \langle h, h\rangle_E \tag{4.1.71}$$

$$= t\langle \mathrm{H}\Phi(x)h, h\rangle_{\mathcal{H}} + \langle h, h\rangle_E > 0 \tag{4.1.72}$$

Above formula is greater than 0 since $\mathrm{H}\Phi(x)$ is self-adjoint positive-semi-definite. Since the Hessian of objective functional is positive-definite at $\widehat{m}_t$, the stationary point $\widehat{m}_t$ must be a minimal extremum. Furthermore, since the Hessian of objective functional is positive-definite at all $x \in \mathcal{H}$, that means, the objective functional is strictly convex. Thus, the local minimum $\widehat{m}_t$ is the unique global optimum. □

Sometimes, the convex assumption in theorem 4.1.9 is too strong, so we additionally propose the following theorem with weaker statement. This statement is not for numerical implementation, because the required condition in the statement is non-practical. However, theoretically, the following statement implies another explanation of the minimum points: if there exists a differentiable path of minimums, then the differentiable path is unique and determined by the ODE (4.1.58).

Theorem 4.1.10. *Let $O_t : \mathcal{H} \to [0, +\infty)$ and $\Phi : \mathcal{H} \to [0, +\infty)$ be the objective functional and the cost functional in formula* (4.1.7)*, respectively. For any $t \in [0, c]$ ($c > 0$), let X_t denote the set of all minimums of O_t. Assume that the following conditions hold:*

1. *Φ is twice differentiable.*
2. *The initial value problem* (4.1.58) *is well-defined on a closed interval $t \in [0, c]$, such that, there exists the unique solution $\widehat{m}_t \in \mathcal{H}$ for any $t \in [0, c]$.*
3. *There exists a sequence a minimums $\{\widehat{x}_t \in X_t : t \in [0, c]\}$ indexed by t, such that, $\widehat{x}_t$ forms a differentiable path in $[0, c]$.*

Then, the differentiable path of minimums is uniquely determined by the ODE (4.1.58)*, i.e. $\widehat{x}_t = \widehat{m}_t$ for any $t \in [0, c]$.*

Proof. Φ is twice differentiale, which implies Lipschitz continuity on any bounded and closed subsets. Thus, theorem 4.1.4 ensures the existence of minimums.

By assumption, there exists a differentiable path of minimums $\{x_t : t \in [0, 1]\}$, so we consider the minimums along this differentiable path. Since these minimums must be stationary points, i.e. the following equation must hold for all bounded $t \in [0, c]$,

$$r_t(\widehat{x}_t) = 0 \tag{4.1.73}$$

where $r_t : \mathcal{H} \to \mathcal{H}$ is determined in formula (4.1.62). However, $\widehat{x}_t$ is differentiable, thus we have

$$r_0(\widehat{x}_0) = 0 \qquad \mathrm{d}r_t(\widehat{x}_t) = 0, \quad t \in (0, c] \tag{4.1.74}$$

On one hand, $r_0(\widehat{x}_0) = 0$ implies

$$0 = r_0(\widehat{x}_0) = \mathcal{C}_0^{-1}(\widehat{x}_0 - m_0) \tag{4.1.75}$$

Therefore, we have

$$\widehat{x}_0 = m_0 \tag{4.1.76}$$

On the other hand, $\mathrm{d}r_t(\widehat{x}_t) = 0$ implies

$$0 = \mathrm{d}r_t(\widehat{x}_t) = \nabla\Phi(\widehat{x}_t)\mathrm{d}t + t\mathrm{H}\Phi(\widehat{x}_t)\mathrm{d}\widehat{x}_t + \mathcal{C}_0^{-1}\mathrm{d}\widehat{x}_t \tag{4.1.77}$$

Therefore, we have,

$$\mathrm{d}\widehat{x}_t = -\left(t\mathrm{H}\Phi(\widehat{x}_t) + \mathcal{C}_0^{-1}\right)^{-1}\nabla\Phi(\widehat{x}_t)\mathrm{d}t \tag{4.1.78}$$

Thus, $\widehat{x}_t$ is identical to $\widehat{m}_t$ in formula (4.1.58) for any $t \in [0, c]$, as long as the ODE is well-defined and has a unique solution for any $t \in [0, c]$. □

In brief, theorem 4.1.10 proves that: on the condition that the ODE (4.1.58) is well-defined, then the following two statements are equivalent:

1. The stationary points determined via the ODE (4.1.58) are global minimums.

2. There exists a differentiable paths consisting of global minimums.

A more tricky question is how to ensure the ODE (4.1.58) only produces global minimums, i.e. how to constructively build the differentiable path of global minimums. The simplest example has been shown in theorem 4.1.9, i.e. if the Hessian of cost functional is positive-semi-definite on the entire space, then the differentiable path of minimums is uniquely determined via the ODE (4.1.58). The construction for non-convex problems is more challenging, which is beyond this PhD thesis, but it is interesting and will be investigated in future work.

Though the ODE (4.1.58) indeed provides a path of stationary points, there are still two difficulties: theoretically, it is not easy to prove the positive-definiteness of operator $t\mathrm{H}\Phi(\widehat{m}_t)+\mathcal{C}_0^{-1}$, where $t\mathrm{H}\Phi(\widehat{m}_t)+\mathcal{C}_0^{-1}$ is the Hessian of the objective functional O_t; practically, computing the Hessian (second order derivative) requires too much computational cost. This why we prefer to apply the first order approximation. After linearization, the ODE (4.1.58) becomes to the continuous extended Kalman inversion (EKI) in definition 3.3.8. This linearization technique is just like the simplification of Newton's method to Gauss-Newton method. We state that, the ODE (4.1.58) is a continuous variant of Newton's method, and that, the EKI in definition 3.3.8 is a continuous variant of Gauss-Newton method.

4.2 Tempering setting on Hilbert spaces

Let $\mathcal{H}$ be a separable Hilbert space. For any $t \in (0,1]$, let $\mu_t : \mathcal{B}(\mathcal{H}) \to (0,1]$ denote the probability measure determined via

$$\mu_t(\mathrm{d}u) \propto \exp(-t\Phi(u))\mu_0(\mathrm{d}u) \tag{4.2.1}$$

where $\Phi : \mathcal{H} \to [0,+\infty)$ is the cost functional, and $\mu_0 : \mathcal{B}(\mathcal{H}) \to [0,1]$ is the prior probability measure.

Formula (4.2.1) is just like the canonical ensemble in statistical mechanics. Statistisians develope algorithms with similar mathematical structure like the canonical ensemble. These algorithms can be found in simulated annealing [104], annealed importance sampling [88], sequential Monte Carlo method [3, 1], etc. The words like 'annealing'/'annealed' indicate that these mathematical algorithms are related to statistical thermodynamics.

Mathematically, formula (4.2.1) is designed for iterative optimization and/or sequential sampling. The benefit is that, the algorithm starts with initial guesses in a wide range, and then produces more and more accurate estimates as the 'temperature' goes down (the 'temperature' here is $T = 1/t$). In other words, $t = 0$ indicates the prior distribution, and $t = 1$ indicates the posterior distribution.

This section aims to generalize the mathematical structure from finite dimensions (ensemble probability or empirical probability) to infinite dimensions (probability measures on Hilbert spaces), and analyze its properties in Hilbert spaces. This generalization, of course, requires careful descriptions as well as proofs, that are presented in this section.

4.2.1 Notation

This subsection defines notation that will be used later.

Definition 4.2.1 (tempering setting)**.** *The mathematical structure of the tempering setting are listed below:*

1. *Let $\mathcal{H}$ be a real-valued separable Hilbert space.*

2. *Let $\Phi : \mathcal{H} \to [0,+\infty)$ be a cost functional.*

3. *Let $t \in [0, c]$ be the tempering parameter, where $c > 0$ is a bounded real number.*

4. *Let $\{\mu_t : t \in [0, c]\}$ be a sequence of probability measures such that, for $t = 0$, μ_0 is an arbitrary probability measure on the measurable space $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$, and for any $t \in (0, c]$, μ_t is the probability measure absolutely continuous with respect to μ_0 and the Radon-Nikodym derivative satisfies, for almost every $u \in \mathcal{H}$,*

$$\frac{\mathrm{d}\mu_t}{\mathrm{d}\mu_0}(u) = \frac{1}{Z_t} \exp(-t\Phi(u)) \tag{4.2.2}$$

5. *Let $\{Z_t : t \in [0, c]\}$ be a sequence of normalizing constants such that, for any $t \in [0, c]$,*

$$Z_t = \int_{\mathcal{H}} \exp(-t\Phi(u))\, \mu_0(\mathrm{d}u) \tag{4.2.3}$$

6. *Let $\{(\langle\Phi\rangle_t, \langle\Phi, \Phi\rangle_t) : t \in [0, c]\}$ be a sequence of pairs of characterizing quantities such that, for any $t \in [0, c]$,*

$$\langle\Phi\rangle_t := \int_{\mathcal{H}} \Phi(u)\, \mu_t(\mathrm{d}u) \qquad \langle\Phi, \Phi\rangle_t := \int_{\mathcal{H}} (\Phi(u) - \langle\Phi\rangle_t)^2\, \mu_t(\mathrm{d}u) \tag{4.2.4}$$

where $\langle\Phi\rangle_t$ is the average cost functional, and $\langle\Phi, \Phi\rangle_t$ is the variance of cost functional.

4.2.2 Under the $L^1(\mathcal{H}, \mu_0)$ condition

First of all, the tempering setting is analyzed with the L^1 condition, i.e. it is assumed that the cost functional $\Phi : \mathcal{H} \to [0, +\infty)$ is absolutely integrable,

$$\int_{\mathcal{H}} |\Phi(u)|\, \mu_0(\mathrm{d}u) \leq M_0 < \infty \tag{4.2.5}$$

Proposition 4.2.2. *Consider the same notation in definition 4.2.1. If the cost functional Φ is absolutely integrable, then for any bounded $t \geq 0$, the normalizing constant Z_t is strictly greater than 0.*

Proof.

$$\begin{aligned} Z_t &= \int_{\mathcal{H}} \exp(-t\Phi(u))\, \mu_0(\mathrm{d}u) && (4.2.6)\\ &\geq \exp\left(-t \int_{\mathcal{H}} \Phi(u)\, \mu_0(\mathrm{d}u)\right) \qquad \textit{use Jensen's inequality} && (4.2.7)\\ &\geq \exp\left(-tM_0\right) > 0 \qquad \textit{use formula } (4.2.5) && (4.2.8) \end{aligned}$$

□

Proposition 4.2.3. *Consider the same notation in definition 4.2.1. If the cost functional* Φ *is absolutely integrable, then for any bounded* $t \geq 0$*, the probability measure* μ_t *is equivalent to the prior probability measure* μ_0*.*

Proof. By definition, μ_t is absolutely continuous to μ_0. Thus we only have to prove the converse: $\forall \mathcal{X} \in \mathcal{B}(\mathcal{H})$, $u_t(\mathcal{X}) = 0 \implies u_0(\mathcal{X}) = 0$. We prove this statement by contradiction. Assume that there exists $\mathcal{X} \in \mathcal{B}(\mathcal{H})$ such that $u_t(\mathcal{X}) = 0$ and $u_0(\mathcal{X}) > 0$. Since $u_0(\mathcal{X}) > 0$, we can define a another probability measure $\nu_{\mathcal{X}}$ on the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that for all $\mathcal{S} \in \mathcal{B}(\mathcal{X})$,

$$\nu_{\mathcal{X}}(\mathcal{S}) := \frac{u_0(\mathcal{S})}{u_0(\mathcal{X})} \tag{4.2.9}$$

Thus,

$$\begin{aligned}
u_t(\mathcal{X}) :=& \int_{\mathcal{X}} \mu_t(du) && (4.2.10)\\
=& \int_{\mathcal{X}} \frac{\mathrm{d}\mu_t}{\mathrm{d}\mu_0}(u)\, \mu_0(du) && (4.2.11)\\
=& \frac{1}{Z_t} \int_{\mathcal{X}} \exp(-t\Phi(u))\, \mu_0(du) && (4.2.12)\\
=& \frac{\mu_0(\mathcal{X})}{Z_t} \left(\int_{\mathcal{X}} \exp(-t\Phi(u))\, \nu_{\mathcal{X}}(\mathrm{d}u) \right) \qquad \textit{use formula } (4.2.9) && (4.2.13)\\
\geq& \frac{\mu_0(\mathcal{X})}{Z_t} \exp\left(-t \int_{\mathcal{X}} \Phi(u)\, \nu_{\mathcal{X}}(\mathrm{d}u) \right) \qquad \textit{use Jensen's inequality} && (4.2.14)\\
=& \frac{\mu_0(\mathcal{X})}{Z_t} \exp\left(-\frac{t}{\mu_0(\mathcal{X})} \int_{\mathcal{X}} \Phi(u)\, \mu_0(\mathrm{d}u) \right) \qquad \textit{use formula } (4.2.9) && (4.2.15)\\
\geq& \frac{\mu_0(\mathcal{X})}{Z_t} \exp\left(-\frac{t}{\mu_0(\mathcal{X})} \int_{\mathcal{H}} \Phi(u)\, \mu_0(\mathrm{d}u) \right) && (4.2.16)\\
\geq& \frac{\mu_0(\mathcal{X})}{Z_t} \exp\left(-\frac{tN_0}{\mu_0(\mathcal{X})} \right) > 0 \qquad \textit{use formula } (4.2.5) && (4.2.17)
\end{aligned}$$

However, the assumption tells that $u_t(\mathcal{X}) = 0$. Contradiction occurs. □

Theorem 4.2.4 (thermodynamic integration 1)**.** *Consider the same notation in definition 4.2.1. If the cost functional* Φ *is absolutely integrable, then the normalizing constant* Z_t *as a function of* $t \geq 0$ *is differentiable with derivative represented by*

$$\frac{Z_t'}{Z_t} = -\left\langle \Phi \right\rangle_t \tag{4.2.18}$$

Proof. For convenience, let

$$f_u(t) = \exp(-t\Phi(u)) \tag{4.2.19}$$

By definition, Z_t is represented by

$$Z_t = \int_{\mathcal{H}} f_u(t)\, \mu_0(\mathrm{d}u) \tag{4.2.20}$$

Notice that the derivative of f_u is bounded by $\Phi(u)$ for any $t \geq 0$, since

$$|f_u'(t)| = |-\Phi(u)\exp(-t\Phi(u))| \leq \Phi(u) \tag{4.2.21}$$

Remind that Φ is absolutely integrable, so according to the dominated convergence theorem, the derivative of Z can be calculated by

$$Z_t' = \int_{\mathcal{H}} f_u'(t)\, \mu_0(\mathrm{d}u) = -\int_{\mathcal{H}} \Phi(u)\exp(-t\Phi(u))\, \mu_0(\mathrm{d}u) \tag{4.2.22}$$

According to proposition 4.2.2, Z_t is strictly greater than 0 for any bounded $t \geq 0$, so Z_t' can be divided by Z_t,

$$\frac{Z_t'}{Z_t} = -\frac{1}{Z_t}\int_{\mathcal{H}} \Phi(u)\exp(-t\Phi(u))\, \mu_0(\mathrm{d}u) \tag{4.2.23}$$

According to proposition 4.2.3, μ_t and μ_0 are equivalent for any bounded $t \geq 0$, so the measure for integral can be changed from μ_0 to μ_t,

$$\frac{Z_t'}{Z_t} = -\int_{\mathcal{H}} \Phi(u)\mu_t(\mathrm{d}u) \tag{4.2.24}$$

□

4.2.3 Under the $L^2(\mathcal{H}, \mu_0)$ condition

Moreover, the tempering setting is analyzed with the L^2 condition, i.e. it is assumed that the cost functional $\Phi : \mathcal{H} \to [0, +\infty)$ is square integrable,

$$\int_{\mathcal{H}} |\Phi(u)|^2\, \mu_0(\mathrm{d}u) < \infty \tag{4.2.25}$$

Notice that the L^2 condition (4.2.25) is stronger than the L^1 condition (4.2.5), so all the previous results still hold.

Theorem 4.2.5 (thermodynamic integration 2)**.** *Consider the same notation in definition 4.2.1. Let $u_t \sim \mu_t$ be a random variable for any bounded $t \geq 0$. If the cost functional Φ is square integrable, then for any square integrable functional $g : \mathcal{H} \to \mathbb{R}$, the expected value $\mathbb{E}\{g(u_t)\}$ as a function of $t \geq 0$ is differentiable and its derivative equals to*

$$\mathbb{E}\{g(u_t)\}' = -\operatorname{COV}\{g(u_t), \Phi(u_t)\} \tag{4.2.26}$$

Proof. Since μ_t is absolutely continuous to μ_0, $\mathbb{E}\{g(u_t)\}$ can be rewritten by changing the measure from μ_t to μ_0,

$$\mathbb{E}\{g(u_t)\} = \int_{\mathcal{H}} g(u)\,\mu_t(\mathrm{d}u) = \int_{\mathcal{H}} g(u)\frac{\mathrm{d}\mu_t}{\mathrm{d}\mu_0}(u)\,\mu_0(\mathrm{d}u) = \frac{1}{Z_t}\int_{\mathcal{H}} g(u)\exp(-t\Phi(u))\,\mu_0(\mathrm{d}u) \tag{4.2.27}$$

For convenience, let

$$f_u(t) = g(u)\exp(-t\Phi(u)) \tag{4.2.28}$$

Thus, the expected value can be expressed by

$$\mathbb{E}\{g(u_t)\} = \frac{1}{Z_t}\int_{\mathcal{H}} f_u(t)\,\mu_0(\mathrm{d}u) \tag{4.2.29}$$

By applying the chain rule, the derivative is given by

$$\mathbb{E}\{g(u_t)\}' = \left(\frac{1}{Z_t}\int_{\mathcal{H}} f_u(t)\,\mu_0(\mathrm{d}u)\right)' = -\frac{Z_t'\int_{\mathcal{H}} f_u(t)\,\mu_0(\mathrm{d}u)}{Z_t^2} + \frac{\left(\int_{\mathcal{H}} f_u(t)\,\mu_0(\mathrm{d}u)\right)'}{Z_t} \tag{4.2.30}$$

Notice that the derivative of f_u is bounded by $|g(u)\Phi(u)|$ for any $t \geq 0$, since

$$|f_u'(t)| = |-g(u)\Phi(u)\exp(-t\Phi(u))| \leq |g(u)\Phi(u)| \tag{4.2.31}$$

Remind that $g \cdot \Phi$ is absolutely integrable as both of g and Φ are square integrable (Cauchy-Schwarz inequality), so according to the dominated convergence theorem, we have

$$\left(\int_{\mathcal{H}} f_u(t)\,\mu_0(\mathrm{d}u)\right)' = \int_{\mathcal{H}} f_u'(t)\,\mu_0(\mathrm{d}u) = -\int_{\mathcal{H}} \Phi(u) f_u(t)\,\mu_0(\mathrm{d}u) \tag{4.2.32}$$

On the other hand, theorem 4.2.4 tells that

$$\frac{Z_t'}{Z_t} = -\int_{\mathcal{H}} \Phi(u)\,\mu_t(\mathrm{d}u) \tag{4.2.33}$$

Thus, substitute formulas (4.2.32) and (4.2.33) into formula (4.2.30), we have

$$\mathbb{E}\{g(u_t)\}' = \frac{\int_{\mathcal{H}} \Phi(u)\mu_t(\mathrm{d}u)\int_{\mathcal{H}} f_u(t)\,\mu_0(\mathrm{d}u)}{Z_t} - \frac{\int_{\mathcal{H}} \Phi(u) f_u(t)\,\mu_0(\mathrm{d}u)}{Z_t} \tag{4.2.34}$$

According to proposition 4.2.3, μ_t and μ_0 are equivalent for any bounded $t \geq 0$, so the measure for integral can be changed from μ_0 to μ_t,

$$\mathbb{E}\left\{g(u_t)\right\}' = \int_{\mathcal{H}} \Phi(u)\mu_t(\mathrm{d}u) \int_{\mathcal{H}} g(u)\,\mu_t(\mathrm{d}u) - \int_{\mathcal{H}} \Phi(u)g(u)\,\mu_t(\mathrm{d}u) \tag{4.2.35}$$

$$= -\operatorname{COV}\left\{g(u_t), \Phi(u_t)\right\} \tag{4.2.36}$$

□

Corollary 4.2.6 (energy fluctuations)**.** *Consider the same notation in definition 4.2.1. If the cost functional Φ is square integrable, then the expected value $\langle\Phi\rangle_t$ as a function of $t \geq 0$ is differentiable and its derivative equals to the negative of variance $\langle\Phi,\Phi\rangle_t$,*

$$\langle\Phi\rangle_t' = -\langle\Phi,\Phi\rangle_t \tag{4.2.37}$$

Proof. Directly apply theorem 4.2.5, in which let $g = \Phi$. □

Corollary 4.2.7 (dynamic of mean)**.** *Consider the same notation in definition 4.2.1. Particularly, assume that the probability measure μ_0 has the second moment,*

$$\int_{\mathcal{H}} \|u\|_{\mathcal{H}}^2\,\mu_0(\mathrm{d}u) < \infty \tag{4.2.38}$$

Let $u_t \sim \mu_t$ be a random variable for any bounded $t \geq 0$. In this case, if the cost functional Φ is square integrable, then the equation below holds for any bounded $t \geq 0$ and for any bounded $h \in \mathcal{H}$,

$$\mathbb{E}\left\{\langle u_t, h\rangle_{\mathcal{H}}\right\}' = -\operatorname{COV}\left\{\langle u_t, h\rangle_{\mathcal{H}}, \Phi(u_t)\right\} \tag{4.2.39}$$

Proof. Directly apply theorem 4.2.5, in which let $g(\cdot) = \langle\cdot, h\rangle_{\mathcal{H}}$ for any bounded $h \in \mathcal{H}$. □

Corollary 4.2.8 (dynamic of covariance)**.** *Consider the same notation in definition 4.2.1. Particularly, assume that the probability measure μ_0 has the fourth moment,*

$$\int_{\mathcal{H}} \|u\|_{\mathcal{H}}^4\,\mu_0(\mathrm{d}u) < \infty \tag{4.2.40}$$

Let $u_t \sim \mu_t$ be a random variable for any bounded $t \geq 0$. In this case, if the cost functional Φ is square integrable, then the equation below holds for any bounded $t \geq 0$ and for any bounded $v, w \in \mathcal{H}$,

$$\mathbb{E}\left\{\langle v, u_t - m_t\rangle_{\mathcal{H}} \langle u_t - m_t, w\rangle_{\mathcal{H}}\right\}' = -\operatorname{COV}\left\{\langle v, u_t - m_t\rangle_{\mathcal{H}} \langle u_t - m_t, w\rangle_{\mathcal{H}}, \Phi(u_t)\right\} \tag{4.2.41}$$

where $m_t \in \mathcal{H}$ is the element such that for all bounded $h \in \mathcal{H}$,

$$\langle m_t, h\rangle_{\mathcal{H}} = \mathbb{E}\left\{\langle u_t, h\rangle_{\mathcal{H}}\right\} \tag{4.2.42}$$

Proof. First of all, apply theorem 4.2.5, in which let $g(\cdot) = \langle v, \cdot\rangle_{\mathcal{H}} \langle \cdot, w\rangle_{\mathcal{H}}$ for any bounded $v, w \in \mathcal{H}$. Thus, we have

$$\mathbb{E}\left\{\langle v, u_t\rangle_{\mathcal{H}} \langle u_t, w\rangle_{\mathcal{H}}\right\}' = -\operatorname{COV}\left\{\langle v, u_t\rangle_{\mathcal{H}} \langle u_t, w\rangle_{\mathcal{H}}, \Phi(u)\right\} \tag{4.2.43}$$

Then, conduct the calculation below,

$$\begin{aligned}
&\mathbb{E}\left\{\langle v, u_t - m_t\rangle_{\mathcal{H}} \langle u_t - m_t, w\rangle_{\mathcal{H}}\right\}' && (4.2.44)\\
=&\left(\mathbb{E}\left\{\langle v, u_t\rangle_{\mathcal{H}} \langle u_t, w\rangle_{\mathcal{H}}\right\} - \langle v, m_t\rangle_{\mathcal{H}} \langle m_t, w\rangle_{\mathcal{H}}\right)' && (4.2.45)\\
=&\mathbb{E}\left\{\langle v, u_t\rangle_{\mathcal{H}} \langle u_t, w\rangle_{\mathcal{H}}\right\}' - \langle v, m_t'\rangle_{\mathcal{H}} \langle m_t, w\rangle_{\mathcal{H}} - \langle v, m_t\rangle_{\mathcal{H}} \langle m_t', w\rangle_{\mathcal{H}} && (4.2.46)\\
=&-\operatorname{COV}\left\{\langle v, u_t\rangle_{\mathcal{H}} \langle u_t, w\rangle_{\mathcal{H}}, \Phi(u)\right\} \quad \textit{use formula } (4.2.43) && (4.2.47)\\
&+\operatorname{COV}\left\{\langle v, u_t\rangle_{\mathcal{H}}, \Phi(u)\right\} \langle m_t, w\rangle_{\mathcal{H}} \quad \textit{use formula } (4.2.39) && (4.2.48)\\
&+\langle v, m_t\rangle_{\mathcal{H}} \operatorname{COV}\left\{\langle u_t, w\rangle_{\mathcal{H}}, \Phi(u)\right\} \quad \textit{use formula } (4.2.39) && (4.2.49)\\
=&-\operatorname{COV}\left\{\langle v, u_t\rangle_{\mathcal{H}} \langle u_t, w\rangle_{\mathcal{H}} - \langle v, u_t\rangle_{\mathcal{H}} \langle m_t, w\rangle_{\mathcal{H}} - \langle v, m_t\rangle_{\mathcal{H}} \langle u_t, w\rangle_{\mathcal{H}}, \Phi(u)\right\} && (4.2.50)\\
=&-\operatorname{COV}\left\{\langle v, u_t - m_t\rangle_{\mathcal{H}} \langle u_t - m_t, w\rangle_{\mathcal{H}}, \Phi(u)\right\} && (4.2.51)
\end{aligned}$$

□

4.2.4 Under the Gaussian condition

Finally, the tempering setting is analyzed with the Gaussian condition, i.e. it is assumed that the prior probability measure μ_0 is Gaussian, and the cost functional $\Phi : \mathcal{H} \to [0, +\infty)$ has exponential tails, namely for every $\epsilon > 0$ there is an $M = M(\epsilon) \in \mathbb{R}$, such that, for almost every $u \in \mathcal{H}$,

$$|\Phi(u)| \le \exp(\epsilon \|u\|_{\mathcal{H}}^2 + M) \tag{4.2.52}$$

According to the Fernique's theorem, Φ is integrable with any orders of moments under Gaussian measures, so all the previous results still hold.

Since μ_0 is assumed to be a Gaussian measure, and Φ is Gaussian integrable with any orders of moments, we can iteratively apply theorem 4.2.5 with $g = \Phi$. As the result, we have the following theorem.

Theorem 4.2.9. *Consider the same notation in definition 4.2.1. Let I_t denote $I_t \equiv -\log(Z_t)$ for any bounded $t \geq 0$. If the cost functional Φ has an exponential tail and the prior measure μ_0 is a Gaussian measure, then I_t as a function of $t \geq 0$ is an analytic function, whose first and second order derivatives are represented by*

$$I_t' = \langle \Phi \rangle_t \qquad \langle \Phi \rangle_t' = -\langle \Phi, \Phi \rangle_t \tag{4.2.53}$$

Proof. By assumption, Φ has an exponential tail and μ_0 is Gaussian, so Φ is integrable with any orders of moments under the Gaussian measure μ_0, ensured by the Fernique's theorem.

Then, according to theorem 4.2.4, we have

$$\frac{Z_t'}{Z_t} = -\langle \Phi \rangle_t \tag{4.2.54}$$

Equivalently,

$$I_t' = -\log(Z_t)' = -\frac{Z_t'}{Z_t} = \langle \Phi \rangle_t \tag{4.2.55}$$

On the other hand, according to corollary 4.2.6, we have

$$\langle \Phi \rangle_t' = -\langle \Phi, \Phi \rangle_t \tag{4.2.56}$$

Moreover, theorem 4.2.5 can be iteratively applied, so that, the ith order derivative of I_t exists as long as the ith order moment of Φ exists. Thus, I_t has any order of derivatives, as long as Φ is integrable with any order of moments under measure μ_t. Since for any bounded $t > 0$, proposition 4.2.3 ensures that μ_t is equivalent to the prior probability measure μ_0, then Φ is integrable with any order of moments under μ_t also. Thus, I_t is analytic. □

4.3 Gaussian integration by parts on Hilbert spaces

Integration by parts is simple in Euclidean spaces as there there exist the Lebesgue measures. Under the Lebesgue measures, we can analyze the Remain integrals straightforwardly. However, integration by parts is not trivial in infinite-dimensional spaces. This section aims to propose the integration by parts with respect to Gaussian measures on

separable Hilbert spaces. This technique has significance for infinite-dimensional Bayesian inference with Gaussian priors.

The main theorems in this section will be proved in this logic: firstly, we prove integration by parts with the standard normal distribution over the real line; secondly, we generalize the results from single variable to countable variables; thirdly, we transform countable variables into separable Hilbert spaces using the Karhunen-Loéve theorem.

4.3.1 On the real line

In this subsection, we consider function $f : \mathbb{R} \to \mathbb{R}$ defined on the real line.

Definition 4.3.1. *$f : \mathbb{R} \to \mathbb{R}$ is called absolutely continuous on $\mathbb{R}$, if and only if f is differentiable almost everywhere, and for any $-\infty < a < b < +\infty$,*

$$f(b) - f(a) = \int_a^b f'(x)\,\mathrm{d}x \tag{4.3.1}$$

Lemma 4.3.2. *Let $f : \mathbb{R} \to \mathbb{R}$ be an absolutely continuous function satisfying*

$$\int_{\mathbb{R}} f(x)^2\,\mu(\mathrm{d}x) < \infty \tag{4.3.2}$$

where $\mu = \mathcal{N}(0,1)$ is the standard Gaussian measure on $\mathbb{R}$. Then the following equation holds

$$\int_{\mathbb{R}} f'(x)\,\mu(\mathrm{d}x) = \int_{\mathbb{R}} xf(x)\,\mu(\mathrm{d}x) \tag{4.3.3}$$

Proof. Since f is absolutely continuous, f' exits almost everywhere, and apply integration by parts on any interval $[a, b] \subset \mathbb{R}$,

$$\int_a^b f'(x)\,\mu(\mathrm{d}x) = \frac{1}{\sqrt{2\pi}} f(b)\exp(-b^2/2) - \frac{1}{\sqrt{2\pi}} f(a)\exp(-a^2/2) + \int_a^b xf(x)\,\mu(\mathrm{d}x) \tag{4.3.4}$$

Let $a \to -\infty$ and $b \to +\infty$, the limits in the right hand side of the above equation exist. On one hand,

$$\int_{\mathbb{R}} |xf(x)|\,\mu(\mathrm{d}x) \le \sqrt{\int_{\mathbb{R}} x^2\,\mu(\mathrm{d}x) \int_{\mathbb{R}} f(x)^2\,\mu(\mathrm{d}x)} < \infty \tag{4.3.5}$$

On the other hand,

$$\int_{\mathbb{R}} |f(x)|\,\mu(\mathrm{d}x) \le \sqrt{\int_{\mathbb{R}} f(x)^2\,\mu(\mathrm{d}x)} < \infty \tag{4.3.6}$$

and $\int_{\mathbb{R}} |f(x)|\,\mu(\mathrm{d}x) < \infty$ implies

$$\lim_{x\to\pm\infty}\left(\frac{1}{\sqrt{2\pi}}f(x)\exp(-x^2/2)\right) = 0 \tag{4.3.7}$$

Thus,

$$\int_{\mathbb{R}} f'(x)\,\mu(\mathrm{d}x) = \int_{\mathbb{R}} xf(x)\,\mu(\mathrm{d}x) \tag{4.3.8}$$

□

Lemma 4.3.3. *Let $f : \mathbb{R} \to \mathbb{R}$ be an absolutely continuous function satisfying*

$$\int_{\mathbb{R}} f(x)^2\,\mu(\mathrm{d}x) < \infty \tag{4.3.9}$$

where $\mu = \mathcal{N}(0,1)$ is the standard Gaussian measure on $\mathbb{R}$. Then the following equation holds

$$\int_{\mathbb{R}} xf'(x)\,\mu(\mathrm{d}x) = \int_{\mathbb{R}} (x^2-1)f(x)\,\mu(\mathrm{d}x) \tag{4.3.10}$$

Proof. Since f is absolutely continuous, f' exits almost everywhere, and apply integration by parts on any interval $[a,b] \subset \mathbb{R}$,

$$\int_a^b xf'(x)\,\mu(\mathrm{d}x) = \frac{1}{\sqrt{2\pi}}bf(b)\exp(-b^2/2) - \frac{1}{\sqrt{2\pi}}af(a)\exp(-a^2/2) + \int_a^b (x^2-1)f(x)\,\mu(\mathrm{d}x) \tag{4.3.11}$$

Let $a \to -\infty$ and $b \to +\infty$, the limits in the right hand side of the above equation exist. On one hand,

$$\int_{\mathbb{R}} |(x^2-1)f(x)|\,\mu(\mathrm{d}x) \le \sqrt{\int_{\mathbb{R}} (x^2-1)^2\,\mu(\mathrm{d}x)\int_{\mathbb{R}} f(x)^2\,\mu(\mathrm{d}x)} < \infty \tag{4.3.12}$$

On the other hand,

$$\int_{\mathbb{R}} |xf(x)|\,\mu(\mathrm{d}x) \le \sqrt{\int_{\mathbb{R}} x^2\,\mu(\mathrm{d}x)\int_{\mathbb{R}} f(x)^2\,\mu(\mathrm{d}x)} < \infty \tag{4.3.13}$$

and $\int_{\mathbb{R}} |xf(x)|\,\mu(\mathrm{d}x) < \infty$ implies

$$\lim_{x\to\pm\infty}\left(\frac{1}{\sqrt{2\pi}}xf(x)\exp(-x^2/2)\right) = 0 \tag{4.3.14}$$

Thus,

$$\int_{\mathbb{R}} xf'(x)\,\mu(\mathrm{d}x) = \int_{\mathbb{R}} (x^2-1)f(x)\,\mu(\mathrm{d}x) \tag{4.3.15}$$

□

4.3.2 On countable real numbers

In this subsection, we consider function $r : \mathbb{R}^\infty \to \mathbb{R}$ with countable inputs.

Definition 4.3.4. *Let $x \in \mathbb{R}^\infty$, where x can be represented by $x = \{x_i \in \mathbb{R} : i \in \mathbb{N}_1\}$. For any $i \in \mathbb{N}_1$, let $\widehat{x}_i = x \setminus \{x_i\}$. $f : \mathbb{R}^\infty \to \mathbb{R}$ is called absolutely continuous on $\mathbb{R}^\infty$, if and only f is differentiable almost everywhere in $\mathbb{R}^\infty$, and for any $i \in \mathbb{N}_1$, for any fixed $\widehat{x}_i$, for any $-\infty < a < b < +\infty$,*

$$f(b; \widehat{x}_i) - f(a; \widehat{x}_i) = \int_a^b \frac{\partial f(x_i; \widehat{x}_i)}{\partial x_i} \, \mathrm{d}x_i \tag{4.3.16}$$

Lemma 4.3.5. *Let $r : \mathbb{R}^\infty \to \mathbb{R}$ be an absolutely continuous function satisfying*

$$\mathbb{E}\left\{r(\xi)^2\right\} < \infty \tag{4.3.17}$$

and for any $i \in \mathbb{N}_1$, the partial derivative is absolutely integrable,

$$\mathbb{E}\left\{\left|\frac{\partial r(\xi)}{\partial x_i}\right|\right\} < \infty \tag{4.3.18}$$

where $\xi = \{\xi_i : i \in \mathbb{N}_1\}$ is a set of countable i.i.d. standard Gaussian random variables on $\mathbb{R}$. Then, for any $i \in \mathbb{N}_1$,

$$\mathbb{E}\left\{\frac{\partial r(\xi)}{\partial x_i}\right\} = \mathbb{E}\left\{\xi_i r(\xi)\right\} \tag{4.3.19}$$

Proof. For any $i \in \mathbb{N}_1$, $r(\xi)$ can be represented by $r(\xi) = r(\xi_i; \widehat{\xi}_i)$, where $\widehat{\xi}_i = \xi \setminus \{\xi_i\}$. In fact, ξ_i is independent on $\widehat{\xi}_i$, so the conditional expectations below can be written as

$$\mathbb{E}\left\{\frac{\partial r(\xi)}{\partial x_i}\middle|\widehat{\xi}_i\right\} = \mathbb{E}\left\{\frac{\partial r(\xi_i; \widehat{\xi}_i)}{\partial x_i}\middle|\widehat{\xi}_i\right\} = \int_{\mathbb{R}} \frac{\partial r(x_i; \widehat{\xi}_i)}{\partial x_i} \, \mu(\mathrm{d}x_i) \tag{4.3.20}$$

$$\mathbb{E}\left\{\xi_i r(\xi)\middle|\widehat{\xi}_i\right\} = \mathbb{E}\left\{\xi_i r(\xi_i; \widehat{\xi}_i)\middle|\widehat{\xi}_i\right\} = \int_{\mathbb{R}} x_i r(x_i; \widehat{\xi}_i) \, \mu\left(\mathrm{d}x_i\right) \tag{4.3.21}$$

where $\mu = \mathcal{N}(0, 1)$ is the standard Gaussian measure on $\mathbb{R}$. Moreover, the function r is absolutely continuous and square-integrable, so lemma 4.3.2 tells that

$$\int_{\mathbb{R}} \frac{\partial r(x_i; \widehat{\xi}_i)}{\partial x_i} \, \mu\left(\mathrm{d}x_i\right) = \int_{\mathbb{R}} x_i r(x_i; \widehat{\xi}_i) \, \mu\left(\mathrm{d}x_i\right) \tag{4.3.22}$$

Therefore, we have

$$\mathbb{E}\left\{\frac{\partial r(\xi)}{\partial x_i}\middle|\widehat{\xi}_i\right\} = \mathbb{E}\left\{\xi_i r(\xi)\middle|\widehat{\xi}_i\right\} \tag{4.3.23}$$

Furthermore, $\frac{\partial r(\xi)}{\partial x_i}$ and $\xi_i r(\xi)$ are absolutely integrable,

$$\mathbb{E}\left\{\left|\frac{\partial r(\xi)}{\partial x_i}\right|\right\} < \infty \tag{4.3.24}$$

$$\mathbb{E}\left\{|\xi_i r(\xi)|\right\} \leq \sqrt{\mathbb{E}\left\{\xi_i^2\right\}\mathbb{E}\left\{r(\xi)^2\right\}} = \sqrt{\mathbb{E}\left\{r(\xi)^2\right\}} < \infty \tag{4.3.25}$$

so the law of total expectation tells

$$\mathbb{E}\left\{\frac{\partial r(\xi)}{\partial x_i}\right\} = \mathbb{E}\left\{\mathbb{E}\left\{\frac{\partial r(\xi)}{\partial x_i}\middle|\widehat{\xi}_i\right\}\right\} = \mathbb{E}\left\{\mathbb{E}\left\{\xi_i r(\xi)\middle|\widehat{\xi}_i\right\}\right\} = \mathbb{E}\left\{\xi_i r(\xi)\right\} \tag{4.3.26}$$

□

Lemma 4.3.6. *Let $r : \mathbb{R}^\infty \to \mathbb{R}$ be an absolutely continuous function satisfying*

$$\mathbb{E}\left\{r(\xi)^2\right\} < \infty \tag{4.3.27}$$

and for any $i \in \mathbb{N}_1$, the partial derivative is square-integrable,

$$\mathbb{E}\left\{\left(\frac{\partial r(\xi)}{\partial x_i}\right)^2\right\} < \infty \tag{4.3.28}$$

where $\xi = \{\xi_i : i \in \mathbb{N}_1\}$ is a set of countable i.i.d. standard Gaussian random variables on $\mathbb{R}$. Then, for any $i, j \in \mathbb{N}_1$,

$$\mathbb{E}\left\{\xi_i \frac{\partial r(\xi)}{\partial x_j}\right\} = \mathrm{COV}\left\{\xi_i \xi_j, r(\xi)\right\} \tag{4.3.29}$$

Proof. For any $j \in \mathbb{N}_1$, $r(\xi)$ can be represented by $r(\xi) = r(\xi_j; \widehat{\xi}_j)$, where $\widehat{\xi}_j = \xi \setminus \{\xi_j\}$. In fact, ξ_j is independent on $\widehat{\xi}_j$, so the conditional expectation below can be written as

$$\mathbb{E}\left\{\xi_i \frac{\partial r(\xi)}{\partial x_j}\middle|\widehat{\xi}_j\right\} = \mathbb{E}\left\{\xi_i \frac{\partial r(\xi_j;\widehat{\xi}_j)}{\partial x_j}\middle|\widehat{\xi}_j\right\} = \begin{cases} \int_{\mathbb{R}} x_j \frac{\partial r(x_j;\widehat{\xi}_j)}{\partial x_j}\,\mu(\mathrm{d}x_j) & \text{if } i = j \\ \int_{\mathbb{R}} \xi_i \frac{\partial r(x_j;\widehat{\xi}_j)}{\partial x_j}\,\mu(\mathrm{d}x_j) & \text{if } i \neq j \end{cases} \tag{4.3.30}$$

where $\mu = \mathcal{N}(0,1)$ is the standard Gaussian measure on $\mathbb{R}$. Furthermore, r is absolutely continuous and square-integrable, so lemma 4.3.3 ($i = j$) and lemma 4.3.2 ($i \neq j$) tell that

$$\int_{\mathbb{R}} x_j \frac{\partial r(x_j;\widehat{\xi}_j)}{\partial x_j}\,\mu(\mathrm{d}x_j) = \int_{\mathbb{R}} \left(x_j^2 - 1\right) r(x_j;\widehat{\xi}_j)\,\mu(\mathrm{d}x_j), \quad i = j \tag{4.3.31}$$

$$\int_{\mathbb{R}} \xi_i \frac{\partial r(x_j;\widehat{\xi}_j)}{\partial x_j}\,\mu(\mathrm{d}x_j) = \int_{\mathbb{R}} \xi_i x_j r(x_j;\widehat{\xi}_j)\,\mu(\mathrm{d}x_j), \quad i \neq j \tag{4.3.32}$$

Again, apply the independence of ξ_j and $\widehat{\xi}_j$,

$$\int_{\mathbb{R}} \left(x_j^2 - 1\right) r(x_j; \widehat{\xi}_j)\, \mu\left(\mathrm{d}x_j\right) = \mathbb{E}\left\{ \left(\xi_j^2 - 1\right) r(\xi_j; \widehat{\xi}_j) \,\middle|\, \widehat{\xi}_j \right\} = \mathbb{E}\left\{ \left(\xi_j^2 - 1\right) r(\xi) \,\middle|\, \widehat{\xi}_j \right\}, \quad i = j \tag{4.3.33}$$

$$\int_{\mathbb{R}} \xi_i x_j r(x_j; \widehat{\xi}_j)\, \mu\left(\mathrm{d}x_j\right) = \mathbb{E}\left\{ \xi_i \xi_j r(\xi_j; \widehat{\xi}_j) \,\middle|\, \widehat{\xi}_j \right\} = \mathbb{E}\left\{ \xi_i \xi_j r(\xi) \,\middle|\, \widehat{\xi}_j \right\}, \quad i \neq j \tag{4.3.34}$$

Therefore, we have

$$\mathbb{E}\left\{ \xi_i \frac{\partial r(\xi)}{\partial x_j} \,\middle|\, \widehat{\xi}_j \right\} = \begin{cases} \mathbb{E}\left\{ \left(\xi_j^2 - 1\right) r(\xi) \,\middle|\, \widehat{\xi}_j \right\} & \text{if } i = j \\ \mathbb{E}\left\{ \xi_i \xi_j r(\xi) \,\middle|\, \widehat{\xi}_j \right\} & \text{if } i \neq j \end{cases} \tag{4.3.35}$$

Furthermore, $\xi_i \frac{\partial r(\xi)}{\partial x_j}$, $\left(\xi_j^2 - 1\right) r(\xi)$ and $\xi_i \xi_j r(\xi)$ are absolutely integrable,

$$\mathbb{E}\left\{ \left| \xi_i \frac{\partial r(\xi)}{\partial x_j} \right| \right\} \leq \sqrt{\mathbb{E}\left\{\xi_i^2\right\} \mathbb{E}\left\{ \left(\frac{\partial r(\xi)}{\partial x_j} \right)^2 \right\}} = \sqrt{\mathbb{E}\left\{ \left(\frac{\partial r(\xi)}{\partial x_j} \right)^2 \right\}} < \infty \tag{4.3.36}$$

$$\mathbb{E}\left\{ \left| \left(\xi_j^2 - 1\right) r(\xi) \right| \right\} \leq \sqrt{\mathbb{E}\left\{ \left(\xi_j^2 - 1\right)^2 \right\} \mathbb{E}\left\{ r(\xi)^2 \right\}} = \sqrt{2\mathbb{E}\left\{ r(\xi)^2 \right\}} < \infty, \quad i = j \tag{4.3.37}$$

$$\mathbb{E}\left\{ \left| \xi_i \xi_j r(\xi) \right| \right\} \leq \sqrt{\mathbb{E}\left\{ (\xi_i \xi_j)^2 \right\} \mathbb{E}\left\{ r(\xi)^2 \right\}} = \sqrt{\mathbb{E}\left\{ r(\xi)^2 \right\}} < \infty, \quad i \neq j \tag{4.3.38}$$

so the law of total expectation tells

$$\mathbb{E}\left\{ \xi_i \frac{\partial r(\xi)}{\partial x_j} \right\} = \mathbb{E}\left\{ \mathbb{E}\left\{ \xi_i \frac{\partial r(\xi)}{\partial x_j} \,\middle|\, \widehat{\xi}_j \right\} \right\} \tag{4.3.39}$$

$$= \begin{cases} \mathbb{E}\left\{ \mathbb{E}\left\{ \left(\xi_j^2 - 1\right) r(\xi) \,\middle|\, \widehat{\xi}_j \right\} \right\} = \mathbb{E}\left\{ \left(\xi_j^2 - 1\right) r(\xi) \right\} & \text{if } i = j \\ \mathbb{E}\left\{ \mathbb{E}\left\{ \xi_i \xi_j r(\xi) \,\middle|\, \widehat{\xi}_j \right\} \right\} = \mathbb{E}\left\{ \xi_i \xi_j r(\xi) \right\} & \text{if } i \neq j \end{cases} \tag{4.3.40}$$

In addition, the right hand side of above formula are two cases, which can be combined in the form of

$$\mathrm{COV}\left\{ \xi_i \xi_j, r(\xi) \right\} = \mathbb{E}\left\{ \xi_i \xi_j r(\xi) \right\} - \mathbb{E}\left\{ \xi_i \xi_j \right\} \mathbb{E}\left\{ r(\xi) \right\} \tag{4.3.41}$$

$$= \begin{cases} \mathbb{E}\left\{ \xi_j^2 r(\xi) \right\} - \mathbb{E}\left\{ \xi_j^2 \right\} \mathbb{E}\left\{ r(\xi) \right\} = \mathbb{E}\left\{ (\xi_j^2 - 1) r(\xi) \right\} & \text{if } i = j \\ \mathbb{E}\left\{ \xi_i \xi_j r(\xi) \right\} - \mathbb{E}\left\{ \xi_i \right\} \mathbb{E}\left\{ \xi_j \right\} \mathbb{E}\left\{ r(\xi) \right\} = \mathbb{E}\left\{ \xi_i \xi_j r(\xi) \right\} & \text{if } i \neq j \end{cases} \tag{4.3.42}$$

□

4.3.3 On real-valued separable Hilbert spaces

In this subsection, we consider function $f : \mathcal{H} \to \mathbb{R}$ defined on a real-valued separable Hilbert space $\mathcal{H}$.

Definition 4.3.7. *f is called absolutely continuous on $\mathcal{H}$, if and only if f is differentiable almost everywhere in $\mathcal{H}$, and for any normalized vector $\varphi \in \mathcal{H}$ with $\|\varphi\|_{\mathcal{H}} = 1$, for any fixed $c \in \mathcal{H}$, for any $-\infty < a < b < +\infty$,*

$$f(b\varphi + c) - f(a\varphi + c) = \int_a^b \frac{\partial f(s\varphi + c)}{\partial s} \, \mathrm{d}s \tag{4.3.43}$$

Theorem 4.3.8. *Let $f : \mathcal{H} \to \mathbb{R}$ be an absolutely continuous function satisfying*

$$\int_{\mathcal{H}} \left(f(x)^2 + \|\mathrm{D}f(x)\|_{op} \right) \mu(\mathrm{d}x) < \infty \tag{4.3.44}$$

where $\mathrm{D}f(x) : \mathcal{H} \to \mathbb{R}$ is the Fréchet derivative of f at $x \in \mathcal{H}$, $\mu = \mathcal{N}(m, \mathcal{C})$ is the Gaussian measure on $\mathcal{H}$, $m \in \mathcal{H}$ is the mean and $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ is the (trace-class semi-positive-definite self-adjoint) covariance operator. Then the equation below holds for all bounded $h \in \mathcal{H}$,

$$\mathbb{E}\left\{ [\mathrm{D}f(u)] (\mathcal{C}h) \right\} = \mathrm{COV}\left\{ \langle u, h \rangle_{\mathcal{H}}, f(u) \right\} \tag{4.3.45}$$

where $u \sim \mu = \mathcal{N}(m, \mathcal{C})$ is a Gaussian random variable.

Proof. This theorem will be proved many some steps.

1. $\mathcal{C}$ is a self-adjoint compact operator on $\mathcal{H}$, so let $\{\lambda_i : i = 1, 2, ...\}$ and $\{\phi_i : i = 1, 2, ...\}$ be the countable eigenvalues and (orthonormal) eigenfunctions of operator $\mathcal{C}$. Then $\mathcal{C}h$ equals to

$$\mathcal{C}h = \sum_{i=1}^{\infty} \lambda_i \langle \phi_i, h \rangle_{\mathcal{H}} \phi_i \tag{4.3.46}$$

   Thus we have 1):

$$\mathbb{E}\left\{ [\mathrm{D}f(u)] (\mathcal{C}h) \right\} = \mathbb{E}\left\{ [\mathrm{D}f(u)] \left(\sum_{i=1}^{\infty} \lambda_i \langle \phi_i, h \rangle_{\mathcal{H}} \phi_i \right) \right\} \tag{4.3.47}$$

2. The expectation and infinite sum in formula (4.3.47) is Fubini, since

$$\sum_{i=1}^{\infty} \mathbb{E}\left\{\left|\lambda_i \left\langle \phi_i, h\right\rangle_{\mathcal{H}} \left[\mathrm{D}f(u)\right](\phi_i)\right|\right\} \tag{4.3.48}$$

$$\leq \sum_{i=1}^{\infty} \lambda_i \|h\|_{\mathcal{H}} \|\phi_i\|_{\mathcal{H}}^2 \mathbb{E}\left\{\|\mathrm{D}f(u)\|_{op}\right\} \tag{4.3.49}$$

$$= \left(\sum_{i=1}^{\infty} \lambda_i\right) \|h\|_{\mathcal{H}} \mathbb{E}\left\{\|\mathrm{D}f(u)\|_{op}\right\} < \infty \tag{4.3.50}$$

Thus we have 2):

$$\mathbb{E}\left\{\left[\mathrm{D}f(u)\right](\mathcal{C}h)\right\} = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \left\langle \phi_i, h\right\rangle_{\mathcal{H}} \mathbb{E}\left\{\left[\mathrm{D}f(u)\right]\left(\sqrt{\lambda_i}\phi_i\right)\right\} \tag{4.3.51}$$

3. However, the random variable u in formula (4.3.51) can be represented by the KL expansion,

$$u = u(\xi) := m + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\xi_i\phi_i \tag{4.3.52}$$

Then the chain rule provides that for any $i \in \mathbb{N}_1$,

$$\frac{\partial f(u(\xi))}{\partial x_i} = \left[\mathrm{D}f(u(\xi))\right]\left(\frac{\partial u(\xi)}{\partial x_i}\right) = \left[\mathrm{D}f(u(\xi))\right]\left(\sqrt{\lambda_i}\phi_i\right) \tag{4.3.53}$$

Thus we have 3):

$$\mathbb{E}\left\{\left[\mathrm{D}f(u)\right](\mathcal{C}h)\right\} = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \left\langle \phi_i, h\right\rangle_{\mathcal{H}} \mathbb{E}\left\{\frac{\partial f(u(\xi))}{\partial x_i}\right\} \tag{4.3.54}$$

4. Furthermore, lemma 4.3.5 can be applied on formula (4.3.54) for $r(\xi) = f(u(\xi))$, because f is absolutely continuous and square-integrable, and for any $i \in \mathbb{N}_1$,

$$\mathbb{E}\left\{\left|\frac{\partial f(u(\xi))}{\partial x_i}\right|\right\} = \mathbb{E}\left\{\left|\left[\mathrm{D}f(u(\xi))\right]\left(\sqrt{\lambda_i}\phi_i\right)\right|\right\} \tag{4.3.55}$$

$$\leq \mathbb{E}\left\{\|\mathrm{D}f(u(\xi))\|_{op}\sqrt{\lambda_i}\|\phi_i\|_{\mathcal{H}}\right\} \tag{4.3.56}$$

$$= \sqrt{\lambda_i}\mathbb{E}\left\{\|\mathrm{D}f(u)\|_{op}\right\} < \infty \tag{4.3.57}$$

Thus we have 4):

$$\mathbb{E}\left\{\left[\mathrm{D}f(u)\right](\mathcal{C}h)\right\} = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \left\langle \phi_i, h\right\rangle_{\mathcal{H}} \mathbb{E}\left\{\xi_i f(u(\xi))\right\} \tag{4.3.58}$$

5. In order to deal with formula (4.3.58), define a sum

$$S_N = \sum_{i=1}^{N} \sqrt{\lambda_i}\xi_i \left\langle \phi_i, h \right\rangle_{\mathcal{H}} \tag{4.3.59}$$

According to the KL theorem, as $N \to \infty$, the sum S_N is mean-square convergent to $\langle u - m, h \rangle_{\mathcal{H}}$. Therefore, $S_N f(u)$ is absolutely integrable

$$\mathbb{E}\left\{|S_N f(u)|\right\} \leq \sqrt{\mathbb{E}\left\{S_N^2\right\}\mathbb{E}\left\{f(u)^2\right\}} < \infty \tag{4.3.60}$$

and it converges to $\langle u - m, h \rangle_{\mathcal{H}} f(u)$ in mean as $N \to \infty$,

$$\mathbb{E}\left\{\left|\langle u - m, h \rangle_{\mathcal{H}} f(u) - S_N f(u)\right|\right\} \leq \sqrt{\mathbb{E}\left\{\left(\langle u - m, h \rangle_{\mathcal{H}} - S_N\right)^2\right\}\mathbb{E}\left\{f(u)^2\right\}} \to 0 \tag{4.3.61}$$

The dominated convergence theorem determines

$$\begin{aligned}
\mathbb{E}\left\{\langle u - m, h \rangle_{\mathcal{H}} f(u)\right\} &= \lim_{N\to\infty} \mathbb{E}\left\{S_N f(u)\right\} && (4.3.62)\\
&= \lim_{N\to\infty} \mathbb{E}\left\{\sum_{i=1}^{N} \sqrt{\lambda_i}\xi_i \left\langle \phi_i, h \right\rangle_{\mathcal{H}} f(u)\right\} && (4.3.63)\\
&= \lim_{N\to\infty} \sum_{i=1}^{N} \sqrt{\lambda_i} \left\langle \phi_i, h \right\rangle_{\mathcal{H}} \mathbb{E}\left\{\xi_i f(u(\xi))\right\} && (4.3.64)\\
&= \sum_{i=1}^{\infty} \sqrt{\lambda_i} \left\langle \phi_i, h \right\rangle_{\mathcal{H}} \mathbb{E}\left\{\xi_i f(u(\xi))\right\} && (4.3.65)
\end{aligned}$$

Thus we have 5):

$$\begin{aligned}
\mathbb{E}\left\{\left[\mathrm{D}f(u)\right](\mathcal{C}h)\right\} &= \mathbb{E}\left\{\langle u - m, h \rangle_{\mathcal{H}} f(u)\right\} && (4.3.66)\\
&= \mathrm{COV}\left\{\langle u, h \rangle_{\mathcal{H}}, f(u)\right\} && (4.3.67)
\end{aligned}$$

□

Corollary 4.3.9. *Let $F : \mathcal{H} \to \mathbb{R}^n$ be an absolutely continuous function satisfying*

$$\int_{\mathcal{H}} \left(\|F(x)\|_{\mathbb{R}^n}^2 + \|\mathrm{D}F(x)\|_{op}\right) \mu(\mathrm{d}x) < \infty \tag{4.3.68}$$

where $\mathrm{D}F(x) : \mathcal{H} \to \mathbb{R}$ is the Fréchet derivative of F at $x \in \mathcal{H}$, $\mu = \mathcal{N}(m, \mathcal{C})$ is the Gaussian measure on $\mathcal{H}$, $m \in \mathcal{H}$ is the mean and $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ is the (trace-class semi-positive-definite self-adjoint) covariance operator. Then the equation below holds for all bounded $h \in \mathcal{H}$,

$$\mathbb{E}\left\{\left[\mathrm{D}F(u)\right](\mathcal{C}h)\right\} = \mathcal{C}_{zu}h \tag{4.3.69}$$

where $u \sim \mu = \mathcal{N}(m, \mathcal{C})$ *is a Gaussian random variable, and* $\mathcal{C}_{zu} : \mathcal{H} \to \mathbb{R}^n$ *is the covariance between* $z := F(u)$ *and* u *such that,* $\mathcal{C}_{zu} := \mathrm{COV}\{z, u\}$.

Proof. Apply theorem (4.3.8) for each element of $F_i(u)$, $i = 1, ..., n$,

$$\mathbb{E}\left\{\left[\mathrm{D}F_i(u)\right](\mathcal{C}h)\right\} = \mathrm{COV}\left\{\langle u, h\rangle_{\mathcal{H}}, F_i(u)\right\} = \mathrm{COV}\left\{F_i(u), u\right\} h \tag{4.3.70}$$

Therefore

$$\mathbb{E}\left\{\left[\mathrm{D}F(u)\right](\mathcal{C}h)\right\} = \mathrm{COV}\left\{\langle u, h\rangle_{\mathcal{H}}, F(u)\right\} = \mathrm{COV}\left\{F(u), u\right\} h \tag{4.3.71}$$

□

Theorem 4.3.10. *Let* $f : \mathcal{H} \to \mathbb{R}$ *be an absolutely continuous function satisfying*

$$\int_{\mathcal{H}} \left(f(x)^2 + \|\mathrm{D}f(x)\|_{op}^2\right) \mu(\mathrm{d}x) < \infty \tag{4.3.72}$$

where $\mathrm{D}f(x) : \mathcal{H} \to \mathbb{R}$ *is the Fréchet derivative of* f *at* $x \in \mathcal{H}$, $\mu = \mathcal{N}(m, \mathcal{C})$ *is the Gaussian measure on* $\mathcal{H}$, $m \in \mathcal{H}$ *is the mean and* $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ *is the (trace-class semi-positive-definite self-adjoint) covariance operator. Then the equation below holds for all bounded* $v, w \in \mathcal{H}$,

$$\mathbb{E}\left\{\langle v, u - m\rangle_{\mathcal{H}} \left[\mathrm{D}f(u)\right](\mathcal{C}w)\right\} = \mathrm{COV}\left\{\langle v, u - m\rangle_{\mathcal{H}} \langle u - m, w\rangle_{\mathcal{H}}, f(u)\right\} \tag{4.3.73}$$

where $u \sim \mu = \mathcal{N}(m, \mathcal{C})$ *is a Gaussian random variable.*

Proof. This theorem will be proved by some steps.

1. $\mathcal{C}$ is a self-adjoint compact operator on $\mathcal{H}$, so let $\{\lambda_i : i = 1, 2, ...\}$ and $\{\phi_i : i = 1, 2, ...\}$ be the countable eigenvalues and (orthonormal) eigenfunctions of operator $\mathcal{C}$. Then $\mathcal{C}w$ equals to

$$\mathcal{C}w = \sum_{j=1}^{\infty} \lambda_j \langle \phi_j, w\rangle_{\mathcal{H}} \phi_j \tag{4.3.74}$$

Thus we have 1):

$$\mathbb{E}\left\{\langle v, u - m\rangle_{\mathcal{H}} \left[\mathrm{D}f(u)\right](\mathcal{C}w)\right\} = \mathbb{E}\left\{\langle v, u - m\rangle_{\mathcal{H}} \left[\mathrm{D}f(u)\right]\left(\sum_{j=1}^{\infty} \lambda_j \langle \phi_j, w\rangle_{\mathcal{H}} \phi_j\right)\right\} \tag{4.3.75}$$

2. The expectation and infinite sum in formula (4.3.75) is Fubini, since

$$\sum_{j=1}^{\infty} \mathbb{E}\left\{\left|\lambda_j \left\langle v, u-m\right\rangle_{\mathcal{H}} \left\langle \phi_j, w\right\rangle_{\mathcal{H}} \left[\mathrm{D}f(u)\right](\phi_j)\right|\right\} \tag{4.3.76}$$

$$\leq \sum_{j=1}^{\infty} \lambda_j \|v\|_{\mathcal{H}} \|w\|_{\mathcal{H}} \|\phi_j\|_{\mathcal{H}}^2 \sqrt{\mathbb{E}\left\{\|u-m\|_{\mathcal{H}}^2\right\} \mathbb{E}\left\{\|\mathrm{D}f(u)\|_{op}^2\right\}} \tag{4.3.77}$$

$$= \left(\sum_{j=1}^{\infty} \lambda_j\right) \|v\|_{\mathcal{H}} \|w\|_{\mathcal{H}} \sqrt{\mathbb{E}\left\{\|u-m\|_{\mathcal{H}}^2\right\} \mathbb{E}\left\{\|\mathrm{D}f(u)\|_{op}^2\right\}} < \infty \tag{4.3.78}$$

Thus we have 2):

$$\mathbb{E}\left\{\left\langle v, u-m\right\rangle_{\mathcal{H}} \left[\mathrm{D}f(u)\right](\mathcal{C}w)\right\} = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \left\langle \phi_j, w\right\rangle_{\mathcal{H}} \mathbb{E}\left\{\left\langle v, u-m\right\rangle_{\mathcal{H}} \left[\mathrm{D}f(u)\right]\left(\sqrt{\lambda_j}\phi_j\right)\right\} \tag{4.3.79}$$

3. However, the random variable u in formula (4.3.79) can be represented by the KL expansion,

$$u = u(\xi) := m + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\xi_i\phi_i \tag{4.3.80}$$

Then the chain rule provides that for any $j \in \mathbb{N}_1$,

$$\frac{\partial f(u(\xi))}{\partial x_j} = \left[\mathrm{D}f(u(\xi))\right]\left(\frac{\partial u(\xi)}{\partial x_j}\right) = \left[\mathrm{D}f(u(\xi))\right]\left(\sqrt{\lambda_j}\phi_j\right) \tag{4.3.81}$$

Thus we have 3):

$$\mathbb{E}\left\{\left\langle v, u-m\right\rangle_{\mathcal{H}} \left[\mathrm{D}f(u)\right](\mathcal{C}w)\right\} = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \left\langle \phi_j, w\right\rangle_{\mathcal{H}} \mathbb{E}\left\{\sum_{i=1}^{\infty} \sqrt{\lambda_i}\xi_i \left\langle v, \phi_i\right\rangle_{\mathcal{H}} \frac{\partial f(u(\xi))}{\partial x_j}\right\} \tag{4.3.82}$$

4. In order to deal with formula (4.3.82), define a sum

$$S_N = \sum_{i=1}^{N} \sqrt{\lambda_i}\xi_i \left\langle v, \phi_i\right\rangle_{\mathcal{H}} \tag{4.3.83}$$

According to the KL theorem, as $N \to \infty$, the sum S_N is mean-square convergent to $\left\langle v, u-m\right\rangle_{\mathcal{H}}$. Therefore, $S_N \frac{\partial f(u(\xi))}{\partial x_j}$ is absolutely integrable

$$\mathbb{E}\left\{\left|S_N \frac{\partial f(u(\xi))}{\partial x_j}\right|\right\} \leq \sqrt{\mathbb{E}\left\{S_N^2\right\} \mathbb{E}\left\{\left(\frac{\partial f(u(\xi))}{\partial x_j}\right)^2\right\}} < \infty \tag{4.3.84}$$

and it converges to $\langle v, u-m\rangle_{\mathcal{H}} \frac{\partial f(u(\xi))}{\partial x_j}$ in mean as $N \to \infty$,

$$\mathbb{E}\left\{\left|\langle v, u-m\rangle_{\mathcal{H}} \frac{\partial f(u(\xi))}{\partial x_j} - S_N \frac{\partial f(u(\xi))}{\partial x_j}\right|\right\} \tag{4.3.85}$$

$$\leq \sqrt{\mathbb{E}\left\{\left(\langle v, u-m\rangle_{\mathcal{H}} - S_N\right)^2\right\} \mathbb{E}\left\{\left(\frac{\partial f(u(\xi))}{\partial x_j}\right)^2\right\}} \to 0 \tag{4.3.86}$$

The dominated convergence theorem determines

$$\mathbb{E}\left\{\langle v, u-m\rangle_{\mathcal{H}} \frac{\partial f(u(\xi))}{\partial x_j}\right\} = \lim_{N\to\infty} \mathbb{E}\left\{S_N \frac{\partial f(u(\xi))}{\partial x_j}\right\} \tag{4.3.87}$$

$$= \lim_{N\to\infty} \mathbb{E}\left\{\sum_{i=1}^{N} \sqrt{\lambda_i}\xi_i \langle v, \phi_i\rangle_{\mathcal{H}} \frac{\partial f(u(\xi))}{\partial x_j}\right\} \tag{4.3.88}$$

$$= \lim_{N\to\infty} \sum_{i=1}^{N} \sqrt{\lambda_i} \langle v, \phi_i\rangle_{\mathcal{H}} \mathbb{E}\left\{\xi_i \frac{\partial f(u(\xi))}{\partial x_j}\right\} \tag{4.3.89}$$

$$= \sum_{i=1}^{\infty} \sqrt{\lambda_i} \langle v, \phi_i\rangle_{\mathcal{H}} \mathbb{E}\left\{\xi_i \frac{\partial f(u(\xi))}{\partial x_j}\right\} \tag{4.3.90}$$

Thus we have 4):

$$\mathbb{E}\left\{\langle v, u-m\rangle_{\mathcal{H}} \left[\mathrm{D}f(u)\right](\mathcal{C}w)\right\} = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \langle \phi_j, w\rangle_{\mathcal{H}} \sum_{i=1}^{\infty} \sqrt{\lambda_i} \langle v, \phi_i\rangle_{\mathcal{H}} \mathbb{E}\left\{\xi_i \frac{\partial f(u(\xi))}{\partial x_j}\right\} \tag{4.3.91}$$

5. Furthermore, lemma 4.3.6 can be applied on formula (4.3.91) for $r(\xi) = f(u(\xi))$, because f is absolutely continuous and square-integrable, and for any $i \in \mathbb{N}_1$,

$$\mathbb{E}\left\{\left(\frac{\partial f(u(\xi))}{\partial x_i}\right)^2\right\} = \mathbb{E}\left\{\left(\left[\mathrm{D}f(u(\xi))\right]\left(\sqrt{\lambda_i}\phi_i\right)\right)^2\right\} \tag{4.3.92}$$

$$\leq \mathbb{E}\left\{\|\mathrm{D}f(u(\xi))\|_{op}^2 \lambda_i \|\phi_i\|_{\mathcal{H}}^2\right\} \tag{4.3.93}$$

$$= \mathbb{E}\left\{\lambda_i \|\mathrm{D}f(u)\|_{op}^2\right\} < \infty \tag{4.3.94}$$

Thus we have 5):

$$\mathbb{E}\left\{\left[\mathrm{D}f(u)\right](\mathcal{C}h)\right\} = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \langle \phi_j, w\rangle_{\mathcal{H}} \sum_{i=1}^{\infty} \sqrt{\lambda_i} \langle v, \phi_i\rangle_{\mathcal{H}} \mathrm{COV}\left\{\xi_i\xi_j, f(u(\xi))\right\} \tag{4.3.95}$$

$$= \sum_{j=1}^{\infty} \sqrt{\lambda_j} \langle \phi_j, w\rangle_{\mathcal{H}} \sum_{i=1}^{\infty} \sqrt{\lambda_i} \langle v, \phi_i\rangle_{\mathcal{H}} \mathbb{E}\left\{\xi_i\xi_j \left(f(u) - \mathbb{E}(f(u))\right)\right\} \tag{4.3.96}$$

6. In order to deal with formula (4.3.96), define two sums

$$S_{N_1}^{(1)} = \sum_{i=1}^{N_1} \sqrt{\lambda_i}\xi_i \left\langle v, \phi_i \right\rangle_{\mathcal{H}} \qquad S_{N_2}^{(2)} = \sum_{j=1}^{N_2} \sqrt{\lambda_j}\xi_j \left\langle \phi_j, w \right\rangle_{\mathcal{H}} \tag{4.3.97}$$

According to the KL theorem, as $N_1, N_2 \to \infty$, the sums $S_{N_1}^{(1)}$ and $S_{N_2}^{(2)}$ are mean-square convergent to $\left\langle v, u-m \right\rangle_{\mathcal{H}}$ and $\left\langle u-m, w \right\rangle_{\mathcal{H}}$. Therefore, $S_{N_1}^{(1)} S_{N_2}^{(2)} \left(f(u) - \mathbb{E}\left\{f(u)\right\}\right)$ is absolutely integrable

$$\mathbb{E}\left\{\left| S_{N_1}^{(1)} S_{N_2}^{(2)} \left(f(u) - \mathbb{E}\left\{f(u)\right\}\right)\right|\right\} \tag{4.3.98}$$

$$\leq \sqrt{\mathbb{E}\left\{\left(S_{N_1}^{(1)} S_{N_2}^{(2)}\right)^2\right\} \mathbb{E}\left\{\left(f(u) - \mathbb{E}\left\{f(u)\right\}\right)^2\right\}} < \infty \tag{4.3.99}$$

and it converges to $\left\langle v, u-m \right\rangle_{\mathcal{H}} \left\langle u-m, w \right\rangle_{\mathcal{H}} \left(f(u) - \mathbb{E}\left\{f(u)\right\}\right)$ in mean as $N_1, N_2 \to \infty$,

$$\mathbb{E}\left\{\left| \left\langle v, u-m \right\rangle_{\mathcal{H}} \left\langle u-m, w \right\rangle_{\mathcal{H}} \left(f(u) - \mathbb{E}\left\{f(u)\right\}\right) - S_{N_1}^{(1)} S_{N_2}^{(2)} \left(f(u) - \mathbb{E}\left\{f(u)\right\}\right)\right|\right\} \tag{4.3.100}$$

$$\leq \sqrt{\mathbb{E}\left\{\left(\left\langle v, u-m \right\rangle_{\mathcal{H}} \left\langle u-m, w \right\rangle_{\mathcal{H}} - S_{N_1}^{(1)} S_{N_2}^{(2)}\right)^2\right\} \mathbb{E}\left\{\left(f(u) - \mathbb{E}\left\{f(u)\right\}\right)^2\right\}} \to 0 \tag{4.3.101}$$

The dominated convergence theorem determines

$$\mathbb{E}\left\{\left\langle v, u-m \right\rangle_{\mathcal{H}} \left\langle u-m, w \right\rangle_{\mathcal{H}} \left(f(u) - \mathbb{E}\left\{f(u)\right\}\right)\right\} \tag{4.3.102}$$

$$= \lim_{N_1, N_2 \to \infty} \mathbb{E}\left\{S_{N_1}^{(1)} S_{N_2}^{(2)} \left(f(u) - \mathbb{E}\left\{f(u)\right\}\right)\right\} \tag{4.3.103}$$

$$= \lim_{N_1, N_2 \to \infty} \mathbb{E}\left\{\sum_{i=1}^{N_1} \sqrt{\lambda_j}\xi_i \left\langle \phi_i, v \right\rangle_{\mathcal{H}} \sum_{j=1}^{N_2} \sqrt{\lambda_j}\xi_j \left\langle \phi_j, w \right\rangle_{\mathcal{H}} \left(f(u) - \mathbb{E}\left\{f(u)\right\}\right)\right\} \tag{4.3.104}$$

$$= \lim_{N_1, N_2 \to \infty} \sum_{i=1}^{N_1} \sqrt{\lambda_i} \left\langle v, \phi_i \right\rangle_{\mathcal{H}} \sum_{j=1}^{N_2} \sqrt{\lambda_j} \left\langle \phi_j, w \right\rangle_{\mathcal{H}} \mathbb{E}\left\{\xi_i \xi_j \left(f(u) - \mathbb{E}\left\{f(u)\right\}\right)\right\} \tag{4.3.105}$$

$$= \sum_{j=1}^{\infty} \sqrt{\lambda_j} \left\langle \phi_j, w \right\rangle_{\mathcal{H}} \sum_{i=1}^{\infty} \sqrt{\lambda_i} \left\langle v, \phi_i \right\rangle_{\mathcal{H}} \mathbb{E}\left\{\xi_i \xi_j \left(f(u) - \mathbb{E}(f(u))\right)\right\} \tag{4.3.106}$$

Thus, we have 6):

$$\mathbb{E}\left\{\left[\mathrm{D}f(u)\right](\mathcal{C}h)\right\} = \mathbb{E}\left\{\left\langle u-m, w \right\rangle_{\mathcal{H}} \left\langle v, u-m \right\rangle_{\mathcal{H}} \left(f(u) - \mathbb{E}\left\{f(u)\right\}\right)\right\} \tag{4.3.107}$$

$$= \mathrm{COV}\left\{\left\langle u-m, w \right\rangle_{\mathcal{H}} \left\langle v, u-m \right\rangle_{\mathcal{H}}, f(u)\right\} \tag{4.3.108}$$

□

Corollary 4.3.11. *Let $f : \mathcal{H} \to \mathbb{R}$ be a differentiable function whose derivative is absolutely continuous, and f satisfies*

$$\int_{\mathcal{H}} \left(f(x)^2 + \|\mathrm{D}f(x)\|_{op}^2 + \|\mathrm{D}^2 f(x)\|_{op} \right) \mu(\mathrm{d}x) < \infty \tag{4.3.109}$$

where $\mathrm{D}f(x) : \mathcal{H} \to \mathbb{R}$ is the Fréchet derivative of f at $x \in \mathcal{H}$, $\mathrm{D}^2 f(x) : \mathcal{H} \to \mathcal{H}^*$ is the second order Fréchet derivative of f at $x \in \mathcal{H}$, $\mu = \mathcal{N}(m, \mathcal{C})$ is the Gaussian measure on $\mathcal{H}$, $m \in \mathcal{H}$ is the mean and $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ is the (trace-class semi-positive-definite self-adjoint) covariance operator. Then the equations below holds for all bounded $v, w \in \mathcal{H}$,

$$\mathbb{E}\left\{ \left[\mathrm{D}^2 f(u)\right] (\mathcal{C}v)(\mathcal{C}w) \right\} = \mathrm{COV}\left\{ \langle v, u - m \rangle_{\mathcal{H}} \langle u - m, w \rangle_{\mathcal{H}}, f(u) \right\} \tag{4.3.110}$$

where $u \sim \mathcal{N}(m, \mathcal{C})$ is a Gaussian random variable.

Proof. Firstly apply theorem 4.3.10, and then apply theorem 4.3.8,

$$\begin{aligned}
&\mathrm{COV}\left\{ \langle v, u - m \rangle_{\mathcal{H}} \langle u - m, w \rangle_{\mathcal{H}}, f(u) \right\} && (4.3.111)\\
=&\mathbb{E}\left\{ \langle v, u - m \rangle_{\mathcal{H}} \left[\mathrm{D}f(u)\right] (\mathcal{C}w) \right\} && (4.3.112)\\
=&\mathrm{COV}\left\{ \langle v, u \rangle_{\mathcal{H}}, \left[\mathrm{D}f(u)\right] (\mathcal{C}w) \right\} && (4.3.113)\\
=&\mathbb{E}\left\{ \left[\mathrm{D}^2 f(u)\right] (\mathcal{C}v)(\mathcal{C}w) \right\} && (4.3.114)
\end{aligned}$$

□

4.4 Brief notes and summary

In this chapter, we conduct theoretical analysis. Our main outcomes are listed as follows.

1. Traditionally, the regularizing parameter of Tikhonov regularization is fixed for a inverse problem, and the regularized inverse problem is solved by optimization methods, e.g. Newton's method or Gauss-Newton method. In this thesis, we consider a continuous regularizing parameter. The reciprocal of the regularizing parameter is the tempering parameter. We prove that the cost functional Φ at the minimum point is always decreasing with respect to the continuous tempering parameter, see

theorem 4.1.5. Furthermore, we shows that, the ODE (4.1.58) provides stationary points indexed by t of the objective functional O_t for $t \in [0, c]$ $(c > 0)$. The ODE (4.1.58) is regarded as a continuous variant of Newton's method. If the cost functional Φ in (4.1.58) has a quadratic form $\Phi(\cdot) = \frac{1}{2}\|\mathcal{Z}(\cdot)\|_{\mathbb{R}^n}$ associated with a data-misfit function $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$, then the ODE (4.1.58) can be simplified into the EKI in definition 3.3.8 by applying the local linearization (3.3.31). This simplification is similar like the simplification from Newton's method to Gauss-Newton method. Thus, the EKI is regarded as a continuous variant of Gauss-Newton method.

2. The Bayesian tempering setting (3.1.21) has the same/similar mathematical structure as canonical ensemble in statistical mechanics. The difference is that, canonical ensemble has finite discrete states, whereas, the Bayesian tempering setting represents infinite-dimensional probability measures on separable Hilbert spaces. For this reason, section 4.2 generalizes the existing results in statistical mechanics from finite-dimensional probability spaces to infinite-dimensional probability spaces. The main results are in theorems 4.2.4 and 4.2.5 which validate the thermodynamic integration in infinite-dimensional space, followed by the corollary 4.2.6 which reveals the behavior of energy fluctuation. Energy fluctuation shows that the average cost functional is always decreasing, and the decreasing rate equals to the variance of the cost functional. This is a pillar to develop the data-misfit controller and the early stop criterion proposed in this thesis.

3. Section 4.3 generalizes integration by parts with respect to Gaussian measures from finite-dimensional Euclidean spaces to infinite-dimensional Hilbert spaces. This is used to deal with Gaussian priors existing in the Bayesian tempering setting (3.1.21), so that, the integration with respect to Gaussian priors can be represented via the technique of integration by parts. Gaussian integration by parts, together with the dynamic mean (4.2.39), the dynamic covariance (4.2.41), and the global linearization (3.3.32), can simplify the Bayesian tempering setting (3.1.21) into the continuous mean-field limit ensemble Kalman filter in definition 3.3.10.

Chapter 5

Numerical Strategies and Applications

In previous chapters, we have discussed inverse problems theoretically. In this chapter, we develop integrated strategies to numerically solving inverse problems. For linear problems, there exists the closed solution (see theorem 2.2.1), so we do not discuss it anymore. For nonlinear problems, if the parameters and observations still have significant linear correlations, then it is feasible to apply Kalman-like methods to solve the problems in practice. There are two different algorithms: extended Kalman inversion (EKI) and ensemble Kalman inversion (EnKI). Theoretically, the Kalman-like methods can be regarded as continuous filtering algorithms in $t \in [0, 1]$. Numerically, the Kalman-like methods should be discretized $0 = t_0 < t_1 < \cdots < t_K = 1$, where K is the total number of discrete steps. We develop the data-misfit controller (DMisC) as an adaptive strategy choosing the step sizes $h_i = t_i - t_{i-1}$ when the Kalman-like methods are implemented. Additionally, we propose the early stop criterion for the Kalman-like methods. Early stop means that the filtering algorithms may stop at some $t = s$ with $s < 1$ rather than $t = 1$. This early stop criterion is proposed to monitor the quality of estimates in iterations, since sometimes (if forward model is highly nonlinear) the Kalman-like methods may be bad approximations. The robustness of the Kalman-like methods should be improved.

In this chapter, we aim to test the Kalman-like methods via conducting numerical experiments. There are three levels of objectives in tests:

1. The first level is to show that the EKI and EnKI work. In this level, we adopt fixed step sizes. The values of the step sizes are user-specified and very small, such that, the discretization errors of the Kalman-like methods are very low and the discrete Kalman-like methods are sufficiently close to their continuous limits. This level aims to eliminate the effect of discretization, only to show that the Kalman-like methods are feasible for approximately solving inverse problems with the tempering setting.

2. The second level is to apply the DMisC as an adaptive strategy choosing step sizes when the Kalman-like methods are implemented. Then the results produced in this level can be compared with the benchmark produced in the first level. Then, we can observe that DMisC works quite well in both accuracy and efficiency. Thus, we believe that the DMisC picks proper step sizes for the Kalman-like methods.

3. The third level is to show that the early stop criterion improves the robustness of the Kalman-like methods for highly nonlinear inverse problems. Since this situation is far from the Gaussian-linear assumptions, the Kalman-like methods could be bad approximations. Then, we adopt the DMisC associated with the early stop criterion which cuts off many 'useless' iterations. Further iterations cannot produce better estimates for the highly nonlinear inverse problems.

To conduct numerical experiments, we adopt a classical application: electrical impedance tomography (EIT). The first and the second levels are tested with a basic model whose underling parameters are homogeneous (single phase/pattern). This model is nonlinear but has strong linear correlations between parameters and observations. The third level is tested with an advanced model whose underling parameters are heterogeneous (multiple phases/patterns). This model is highly nonlinear, as there exists the indicator functions for classifying different phases/patterns of the underling parameters.

5.1 Electrical impedance tomography

Our testing model is electrical impedance tomography (EIT). EIT is a noninvasive type of medical imaging, which applies electrical current from surface electrodes into body, and

the inside electrical conductivity/permittivity/impedance can be inferred. In applied mathematics, EIT is also used as a benchmark to test inverse algorithms.

5.1.1 Statement of the forward/inverse problem

As we have discussed in chapter 1, EIT model is a simplification of the Maxwell's equations under electrostatics, that leads to an elliptic PDE. Mathematically, the complete electrode model of EIT [31] is represented by:

- Forward problem: given σ, $\{z_m\}_{m=1}^{n_e}$ and $I = \{I_m\}_{m=1}^{n_e}$, to compute v and $V = \{V_m\}_{m=1}^{n_e}$

$$\nabla \cdot \sigma \nabla v = 0 \qquad \text{in } D \tag{5.1.1}$$

$$v + z_m \sigma \nabla v \cdot \nu = V_m \qquad \text{on } e_m,\ m = 1, ..., n_e \tag{5.1.2}$$

$$\sigma \nabla v = 0 \qquad \text{on } \partial D \setminus \cup_{m=1}^{n_e} e_m \tag{5.1.3}$$

$$\int_{e_m} \sigma \nabla v \cdot \nu \, \mathrm{d}s = I_m \qquad m = 1, ..., n_e \tag{5.1.4}$$

- Inverse problem: given $I^{(1)}, \ldots, I^{(n)}$ and the observations of voltages $V^{(1)}, \ldots, V^{(n)}$, to find σ (and possibly z_m).

5.1.2 EIDORS codes for the forward simulation

EIDORS (Electrical Impedance Tomography and Diffuse Optical Tomography Reconstruction Software, see website [108]) is a online MATLAB package for EIT problems. We adopt this package to conduct the forward simulation in our numerical tests.

We use the notation $G(\cdot) : L^\infty(\overline{D}; (0, +\infty)) \to \mathbb{R}^n$ to denote the forward simulation: given a conductivity $\sigma \in L^\infty(\overline{D}; (0, +\infty))$, the surface voltages $V \in \mathbb{R}^n$ at electrodes are simulated,

$$V = G(\sigma) \tag{5.1.5}$$

Furthermore, we use the notation $\mathrm{D}G(\sigma) : L^\infty(\overline{D}; (0, +\infty)) \to \mathbb{R}^n$ to denote the Fréchet derivative of G at any $\sigma \in L^\infty(\overline{D}; (0, +\infty))$. In order to ensure the positiveness of conductivity, it is more convenient to apply the log conductivity such that, for all $s \in \overline{D}$,

$$\sigma(s) = \exp(u(s)) \tag{5.1.6}$$

where $u \in L^\infty(\overline{D};\mathbb{R})$ is the log conductivity. Thus, the forward map $\mathcal{G} : L^\infty(\overline{D};\mathbb{R}) \to \mathbb{R}^n$ is composed by

$$V = G(\sigma) = G(\exp(u)) := \mathcal{G}(u) \tag{5.1.7}$$

Numerically, the forward simulation uses finite element method solving the PDE system. EIDORS provides the finite element solver of the forward model (function 'fwd_solve' in EIDORS) as well as its numerical Jacobian (function 'calc_jacobian' in EIDORS). After discretization, the conductivity $\sigma : \overline{D} \to (0, +\infty)$ is specified as a piece-wise constant function with invariant values in each of the finite elements, i.e., for all $x \in \overline{D}$,

$$\sigma(x) = \sum_{i=1}^{M} 1_{D_i}(x)\sigma_i \tag{5.1.8}$$

where M is the total number of finite elements, $\sigma_i \in \mathbb{R}$ is the value of conductivity in the ith finite element, $D_i \subset \overline{D}$ is the domain of the ith finite element, and 1_A is the indicator function of any subset $A \subset \overline{D}$. Thus, we have

$$G(\sigma) = G\left(\sum_{i=1}^{M} 1_{D_i}(x)\sigma_i\right) = F([\sigma_i]) \tag{5.1.9}$$

where $F : \mathbb{R}^M \to \mathbb{R}^n$ is the finite element solver (method fwd_solve in EIDORS), and $[\sigma_i]$ is an M-dimensional column vector with entries σ_i for $i = 1, ..., M$. Furthermore, for any two piece-wise constant conductivitys $\sigma^{(1)}$ and $\sigma^{(2)}$, we have

$$\left[\mathrm{D}G\left(\sigma^{(1)}\right)\right]\left(\sigma^{(2)}\right) = \left[\mathrm{D}G\left(\sum_{i=1}^{M} 1_{D_i}(x)\sigma_i^{(1)}\right)\right]\left(\sum_{i=1}^{M} 1_{D_i}(x)\sigma_i^{(2)}\right) \tag{5.1.10}$$

$$= \left[\mathcal{J}\left(\left[\sigma_i^{(1)}\right]\right)\right] \cdot \left[\sigma_i^{(2)}\right] \tag{5.1.11}$$

where $\mathcal{J}$ is the Jacabian calculator (method calc_jacobian in EIDORS), whose input $\left[\sigma_i^{(1)}\right]$ is an M-dimensional column vector and the output $\left[\mathcal{J}\left(\left[\sigma_i^{(1)}\right]\right)\right]$ is an $n \times M$ matrix which is the numerical Jacobian of G at $\sigma^{(1)}$.

5.1.3 Simulation experiment

For the EIT problem, we arrange 16 electrodes around a 2D circle (the boundary of domain $D = \{x \in \mathbb{R}^2 : \|x\| < 1\}$), apply adjacent current drive into the domain, and measure voltages at each pair of adjacent electrodes (except the current carrying electrodes) on the 2D circle. Therefore, there are totally $16 \times 13 = 208$ measured voltages

in one experiment. Once we obtain the measured voltages, we can implement inverse algorithms to estimate the inside conductivity.

In this thesis, we dot not use real data from clinical medicine, but we conduct simulation experiment. Firstly, we assign values of the inside conductivity, then simulate boundary voltages by using the forward model, and then add noises into the simulated voltages to obtain noisy data. Inversely, we pretend to know nothing about the inside conductivity, and hope to estimate the inside conductivity from the simulated noisy data.

The noisy data $y \in \mathbb{R}^n$ is simulated as follow,

$$y = \widehat{y} + e \qquad \widehat{y} = \mathcal{G}(x) \tag{5.1.12}$$

where $\widehat{y} \in \mathbb{R}^n$ is the clean data which is a column vector containing simulated voltages, $\mathcal{G} : L^\infty(\overline{D}; \mathbb{R}) \to \mathbb{R}^n$ is the forward model for simulation, $x \in L^\infty(D; \mathbb{R})$ is the inside log conductivity whose values are user-specified, and $e \in \mathbb{R}^n$ is the measurement noise which is drawn as a sample from a user-specified probability distribution. We adopt an n-dimensional noise distribution with independent components whose expected values are zeros and standard deviations are proportional to the clean data. Namely, if let $\pi : \mathbb{R}^n \to [0, +\infty)$ denote the probability density function of the noise and let $\pi_i : \mathbb{R} \to [0, +\infty)$ denote the probability density function of the ith component of the noise, then, for any vector $e \in \mathbb{R}^n$ with components e_i $(i = 1, ..., n)$, we have

$$\pi(e) = \prod_{i=1}^{n} \pi_i(e_i) \tag{5.1.13}$$

with

$$\int_{\mathbb{R}} s\pi_i(s)\, \mathrm{d}s = 0 \qquad \sqrt{\int_{\mathbb{R}} s^2 \pi_i(s)\, \mathrm{d}s} = \epsilon \left|\widehat{y}_i\right| \tag{5.1.14}$$

where $\epsilon > 0$ is a user-specified proportional rate, that is the coefficient of variation of the noisy data.

5.2 Basic tests with single-phase conductivity

This subsection considers single-phase conductivity of the EIT problem. Namely, the inside conductivity on the whole domain can be characterized with a single phase/pattern. For example, consider a conductivity as a sample drawn from a stationary random field.

Then, the conductivity on the whole domain is characterized as the single stationary random field.

In this section, we aim to test the methods: DMisC-EKI (without early stop) and DMisC-EnKI (without early stop), which are listed in algorithm 1 and algorithm 2 in the end of chapter 3. In this section, we do not apply the early stop criterion, because the EIT model with single-phase parameters characterized by a stationary random filed is differentiable and has strong linear correlations between the parameters and observations, so there is no much difference from adopting early stop. The early stop criterion will be applied in the next section with multi-phases conductivity.

5.2.1 Experimental configurations

In this subsection, we explain how to assign experimental configurations of our simulation experiments. There are two components that need to be specified: one is the prior and another is the noise. More details are listed as follows.

- The inside log conductivity is assigned as a sample drawn from a Whittle-Matérn field on D, and this Whittle-Matérn field is also adopted as the prior field. Remind that (please see subsection 2.3.3), a Whittle-Matérn field is a stationary Gaussian field with four parameters: $\mu \in \mathbb{R}$ which is the mean of the steady state, $\sigma > 0$ which is the standard deviation of the steady state, $\nu > 0$ which is the smoothness parameter of the autocorrelation function, and $L : \mathbb{R}^d \to \mathrm{R}^d$ which is the length-scale parameter of the autocorrelation function. We will fix the steady mean as

$$\mu = -7.5 \tag{5.2.1}$$

  and fix the smoothness parameter as

$$\nu = 2; \tag{5.2.2}$$

  but we will test different uncertainty levels

$$\sigma \in \{0.2, 0.4, 0.6, 0.8, 1\} \tag{5.2.3}$$

  and we will test different length scales of the autocorrelation

$$L = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}, \qquad L = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.4 \end{bmatrix}, \qquad L = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.4 \end{bmatrix} \tag{5.2.4}$$

- We have assumed that the addictive noise $e \in \mathbb{R}^n$ is a sample drawn from an n-dimensional distribution π with independent components $\pi(e) = \prod_{i=1}^{n} \pi_i(e_i)$. We will test different noise types: Gaussian noise, uniformly distributed noise, and exponentially distributed noise. The first is the normal distribution, i.e. the probability density function π_i of the ith component of the noise is given by

$$\pi_i(s) = \frac{1}{\sqrt{2\pi}\epsilon\,|\widehat{y}_i|} \exp\left(-\frac{s^2}{2\epsilon^2\widehat{y}_i^2}\right) \tag{5.2.5}$$

the second is the uniform distribution, i.e. the probability density function π_i of the ith component of the noise is given by

$$\pi_i(s) = \begin{cases} \frac{1}{2\sqrt{3}\epsilon|\widehat{y}_i|} & if\ s \in \left[-\sqrt{3}\epsilon\,|\widehat{y}_i|\,, \sqrt{3}\epsilon\,|\widehat{y}_i|\right] \\ 0 & if\ s \in \mathbb{R} \setminus \left[-\sqrt{3}\epsilon\,|\widehat{y}_i|\,, \sqrt{3}\epsilon\,|\widehat{y}_i|\right] \end{cases} \tag{5.2.6}$$

and the third is the exponential distribution, i.e. the probability density function π_i of the ith component of the noise is given by

$$\pi_i(s) = \begin{cases} \frac{1}{\epsilon|\widehat{y}_i|} \exp\left(-\frac{s}{\epsilon|\widehat{y}_i|} - 1\right) & if\ s \geq -\epsilon\,|\widehat{y}_i| \\ 0 & if\ s < -\epsilon\,|\widehat{y}_i| \end{cases} \tag{5.2.7}$$

Furthermore, we will test different noise levels: the first group of noise levels are the small amounts of noise

$$\epsilon \in \{0.4\%, 0.8\%, 1.2\%, 1.6\%, 2\%\} \tag{5.2.8}$$

the second group of noise levels are the moderate amounts of noise

$$\epsilon \in \{2\%, 4\%, 6\%, 8\%, 10\%\} \tag{5.2.9}$$

and the third group of noise levels are the large amounts of noise

$$\epsilon \in \{10\%, 20\%, 30\%, 40\%, 50\%\} \tag{5.2.10}$$

where ϵ is the coefficient of variation of the noisy data.

5.2.2 DMisC compared with other methods

In this subsection, we fix the prior length scale $l_x = 0.2$, $l_y = 0.2$, fix the prior uncertainty level $\sigma = 1$, fix the noise type as Gaussian noise, and fix the noise level $\epsilon = 0.02$. With this fixed experimental configuration, we aim to test the DMisC-EKI (without early stop) and DMisC-EnKI (without early stop), compared with other methods.

DMisC vs fixed step sizes

In this thesis, we write the Kalman-like methods as continuous forms (PDE/SDE), and propose the data-misfit controller (DMisC) as an adaptive strategy selecting step sizes when the Kalman-like methods are numerically implemented to solve inverse problems. We aim to show the good performance of this adaptive strategy. This can be tested by using fixed step sizes as comparison. The fixed step sizes are specified as very small values such that the discrete error is sufficiently low. Then, the discretized Kalman-like methods with the very small step sizes can be referred as the benchmark, since they are sufficiently close to the continuous limits. After that, we adopt the DMisC to discretize the Kalman-like methods and to observe the difference from the benchmark.

From our practical experience, the step sizes usually increase and nearly form a geometric sequence. Therefore, we choose the fixed step sizes as, for $i = 1, ..., K$,

$$h_i = \frac{q^i - q^{i-1}}{q^K - 1} \tag{5.2.11}$$

where $q > 1$ is the increasing rate, and K is the total number of steps. q and K are user-specified and fixed. In comparison, DMisC is an adaptive method that automatically chooses step sizes relying on the current underlying states in the iterative algorithm. More clearly, the DMisC choosing step sizes for the Kalman-like methods has been proposed in formula (3.4.9).

Extended Kalman inversion (EKI) and ensemble Kalman inversion (EnKI) associated with data-misfit controller (DMisC) and geometric step sizes (GSS) are noted by DMisC-EKI, DMisC-EnKI, GSS-EKI, and GSS-EnKI, respectively. DMisC-EKI (without early stop) and DMisC-EnKI (without early stop) have been listed with pseudocode in algorithms 1 and 2 in the end of chapter 3. GSS-EKI and GSS-EnKI are listed in algorithms 5 and 6 with pseudocode in the end of this chapter.

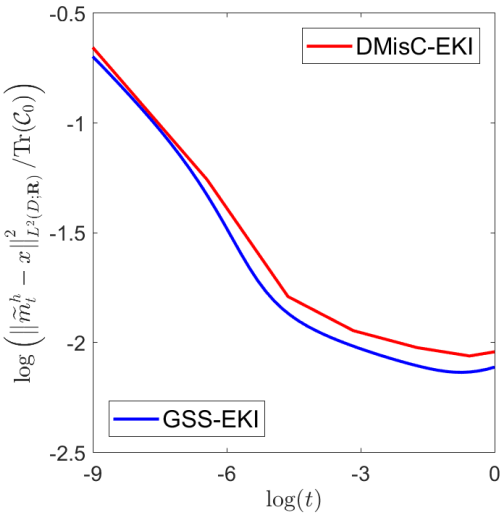
Now, we numerically test these inverse algorithms. Firstly, we implement GSS-EKI and GSS-EnKI as the benchmarks. We adopt increasing rate $q = 1.2$ and total number of iterations $K = 50$ for GSS-EKI and GSS-EnKI to make sufficiently small step sizes, such that, GSS-EKI and GSS-EnKI are very close to the continuous limits CoEKI and CoEnKI. Therefore, the results produced by GSS-EKI and GSS-EnKI can be referred as the benchmarks. After that, we implement DMisC-EKI and DMisC-EnKI, which

are compared with GSS-EKI and GSS-EnKI, respectively. The numerical results are presented in figure 5.1. We can observe that, the estimation errors and data misfits determined via DMisC-EKI/EnKI are close to the results determined via GSS-EKI/EnKI. This means that DMisC-EKI and DMisC-EnKI are sufficiently accurate. At the same time, DMisC-EKI and DMisC-EnKI are very efficient only requiring 7 and 9 iterations, respectively. Therefore, we believe that the data-misfit controller determines proper step sizes for EKI and EnKI.

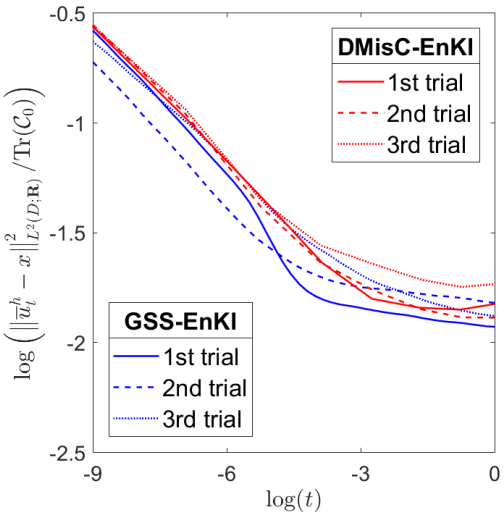
One more fact should be noticed. Numerical implementation of EnKI requires ensemble of particles. If the particle size is too small, there could be large statistical errors. We test different sample sizes $J \in \{64, 144, 256, 400\}$, and for each sample size, we conduct several trials of EnKI to check the amount of statistical errors. Theses tests are conducted by applying GSS-EnKI with very small step size, so the effect of step size is minor in these tests. The results of these tests are shown in figure 5.2, where we plot the logarithm of data misfits in the left yaxis and the logarithm of estimation errors in the right yaxis. For smaller sample size like $J = 64$ or $J = 144$, EnKI only produce unstable or inaccurate results, as the estimation error can even be increasing in the filtering. However, this issue does not happen for larger sample size like $J = 256$ or $J = 400$. When $J = 400$, the statistical errors are relatively small and the estimates are relatively accurate. Therefore, we adopt sample size $J = 400$ for our numerical implementation of EnKI.

DMisC vs RLMS

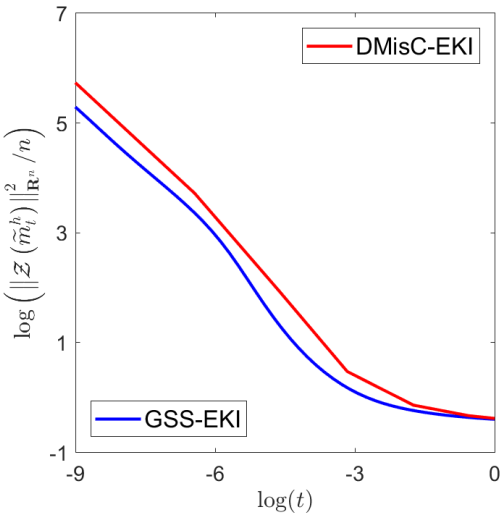
The two methods DMisC-EKI and DMisC-EnKI proposed in this thesis are similar like another group of existing methods: RLMS-LMA and RLMS-EnKI. In fact, DMisC-EKI can be regarded as a variant of RLMS-LMA with additional covariance update. RLMS-LMA keeps the prior covariance operator unchanged in iterations since it is an optimization algorithm, while DMisC-EKI updates the covariance operator step by step since it is a filtering algorithm. Numerically, RLMS-LMA and DMisC-EKI have different adaptive strategies and stop criteria. On the other hand, both DMisC-EnKI and RLMS-EnKI apply the ensemble Kalman filter, but the difference is that: DMisC-EnKI applies the data-misfit controller developed from the Bayesian filtering framework, whereas RLMS-EnKI borrows the adaptive strategy from RLMS-LMA.



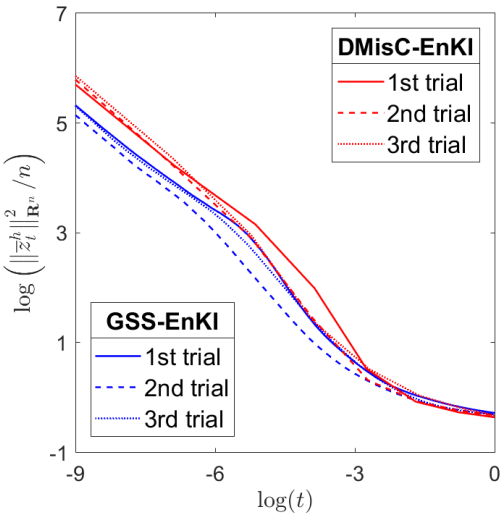
(a) The estimation error in filtering, where x is the true value of parameter, and $\widetilde{m}_t^h$ is the estimate at $t \in (0,1]$ produced by EKI.



(b) The estimation error in filtering, where $\overline{u}_t^h$ is the average estimate at $t \in (0,1]$ produced by EnKI with sample size $J = 400$.

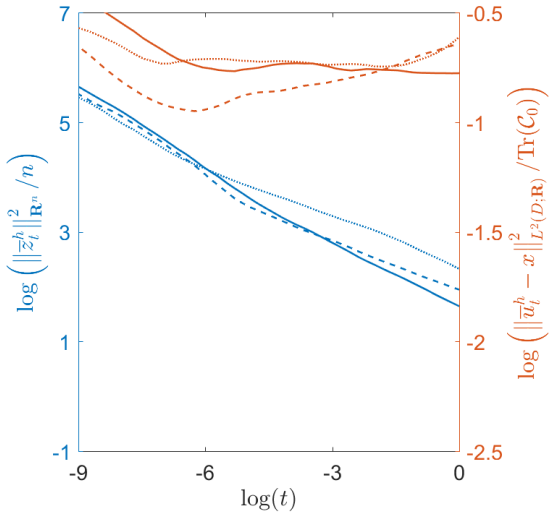


(c) The data misfit in filtering, where $\mathcal{Z}$ is the data-mifit function, and $\widetilde{m}_t^h$ (produced by EKI) is the estimate of x at $t \in (0,1]$.

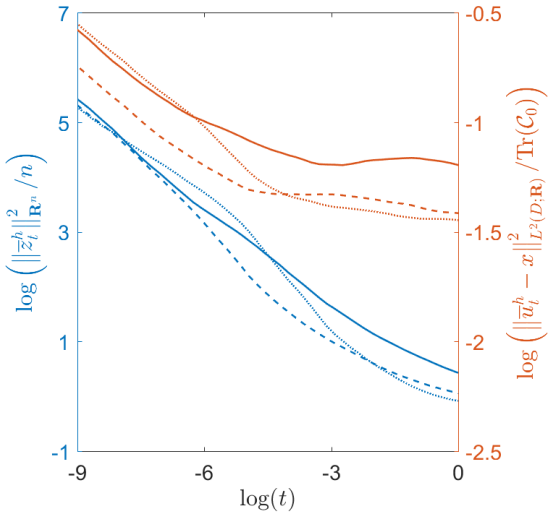


(d) The data misfit in filtering, where $\overline{z}_t^h$ (produced by EnKI with sample size $J = 400$) is the average data misift at $t \in (0,1]$.

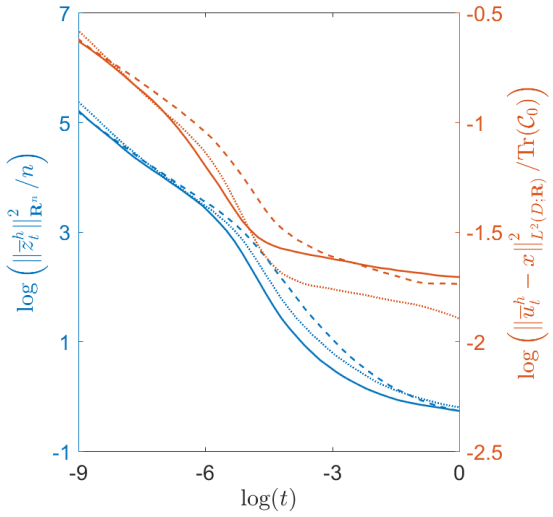
Figure 5.1: The preformaces of EKI/EnKI solving the EIT problem. GSS-EKI/EnKI adopts very small step size, sufficiently close to the continuous limit, so GSS-EKI/EnKI is reffered as the benchmark. DMisC-EKI/EnKI uses adaptive step size, much larger than the fixed step size, but DMisC-EKI/EnKI still leads to results quite close to those produced by GSS-EKI/EnKI. Therefore, DMisC-EKI/EnKI keeps both accuracy and efficiency in this test.



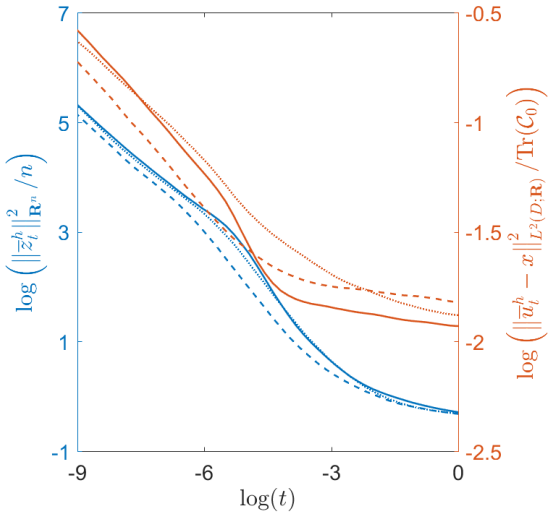
(a) The estimation errors (red lines) and data misfits (blue lines) of three trials of EnKI with particle size $J = 64$.



(b) The estimation errors (red lines) and data misfits (blue lines) of three trials of EnKI with particle size $J = 144$.



(c) The estimation errors (red lines) and data misfits (blue lines) of three trials of EnKI with particle size $J = 256$.



(d) The estimation errors (red lines) and data misfits (blue lines) of three trials of EnKI with particle size $J = 400$.

Figure 5.2: The sample size effect in implementation of EnKI. The estimation errors (red lines) and data misfits (blue lines) in filtering are plotted with respect to the different sample sizes $J = 64$, $J = 144$, $J = 256$, and $J = 400$. As a particle filter, numerically implementation of EnKI with finite sample size causes statistical errors. If sample size is too small such as $J = 64$, the statistical error can be very large, and EnKI only produces unstable and inaccurate results. If sample size is larger such as $J = 400$, then EnKI performs well.

The Levenberg-Marquardt algorithm (LMA) and the regularizing Levenberg-Marquardt scheme (RLMS) has been introduced in section 2.4. Moreover, Iglesias [61] borrowed the algorithmic idea similar to RLMS-LMA to implement EnKI, i.e., the step size $h_i \in (0, +\infty)$ of EnKI in formula (3.3.25) is determined by solving equation

$$\left\| \left(\mathbf{I} + h_i C^h_{zz,t_{i-1}} \right)^{-1} \overline{z}^h_{t_{i-1}} \right\|_{\mathbb{R}^n} = \rho \left\| \overline{z}^h_{t_{i-1}} \right\|_{\mathbb{R}^n} \tag{5.2.12}$$

where $\rho \in (0, 1)$ is a control parameter, and $\overline{z}^h_{t_{i-1}}$ and $C^h_{zz,t_{i-1}}$ are the sample mean and sample covariance of $\left\{ Z^h_t(j) \equiv \mathcal{Z}\left(U^h_t(j)\right) \right\}_{j=1}^J$. The above equation (5.2.12) is the analogue of the RLMS (2.4.24). In addition, the RLMS-EnKI stops for some $t \in [0, +\infty)$ once

$$\|\overline{z}_t\|_{\mathbb{R}^n} \leq \tau \|\mathcal{Z}(x)\|_{\mathbb{R}^n} \tag{5.2.13}$$

where $\tau > \rho^{-1}$ is an accuracy control parameter and $x \in \mathcal{H}$ is the true value of the unknown parameter. The above criterion (5.2.13) of RLMS-EnKI is the analogue of the stop rule (2.4.15) of RLMS-LMA. If $\|\mathcal{Z}(x)\|_{\mathbb{R}^n}$ is unknown, Iglesias suggests $\sqrt{n}$ as an approximation of $\|\mathcal{Z}(x)\|_{\mathbb{R}^n}$. This approximation comes from the law of large numbers, which has been discussed in formulas (3.2.8) and (3.2.9). The RLMS-LMA and RLMS-EnKI are listed with pseudocode in algorithm 7 and algorithm 8 in the end of this chapter.

Now, we test DMisC-EKI (without early stop) and DMisC-EnKI (without early stop), which are compared with RLMS-LMA and RLMS-EnKI, respectively. The sample size of EnKI are chosen as $J = 400$. In practice, DMisC-EKI and DMisC-EnKI have no tuning parameters, so users can directly apply these algorithms. In comparison, RLMS-LMA and RLMS-EnKI has a tuning parameter $\rho \in (0, 1)$, which should be adjusted by trials. We try different values $\rho \in \{0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ to catch the overall performance of RLMS-LMA and RLMS-EnKI. Usually, for $\rho < 0.3$ is inaccurate, and for $\rho > 0.8$ is inefficient, so values out of the interval $[0.3, 0.8]$ rarely appear in practice.

We implement all the inverse algorithms and present the results in figure 5.3. Firstly, DMisC-EKI/DMisC-EnKI has lower estimation errors and lower data misfits than RLMS-LMA/RLMS-EnKI for all the tuning parameters $0.3 \leq \rho \leq 0.8$. This is because the stop criteria of DMisC-EKI/DMisC-EnKI and RLMS-LMA/RLMS-EnKI are different. RLMS-LMA/RLMS-EnKI stops when the Morozov's discrepancy principle occurs. Different from that, DMisC-EKI/DMisC-EnKI accounts for deeper filtering until the tem-

pering parameter goes to $t = 1$, which leads to more accurate results. Secondly, when ρ is around 6.0, RLMS-LMA/RLMS-EnKI has the similar computational cost as DMisC-EKI/DMisC-EnKI, but RLMS-LMA/RLMS-EnKI has much higher estimation error and data misfit. It is clearly that DMisC-EKI/DMisC-EnKI is more efficient. Thirdly, as a statistical approach, DMisC-EnKI than RLMS-EnKI has much less standard deviation of the data misfits. In fact, DMisC-EnKI consistently stops at $t = 1$, which leads to the same sate (numerically with computational errors). However, RLMS-EnKI stops at the first time when the Morozov's discrepancy principle occurs. This stop rule leads to inconsistent states relying on iterations in the algorithm. Thus, the variance caused by RLMS-EnKI has two components, the difference of states and the statistical error. In comparison, the variance caused by DMC-EnKI only has the statistical error without the difference of states.

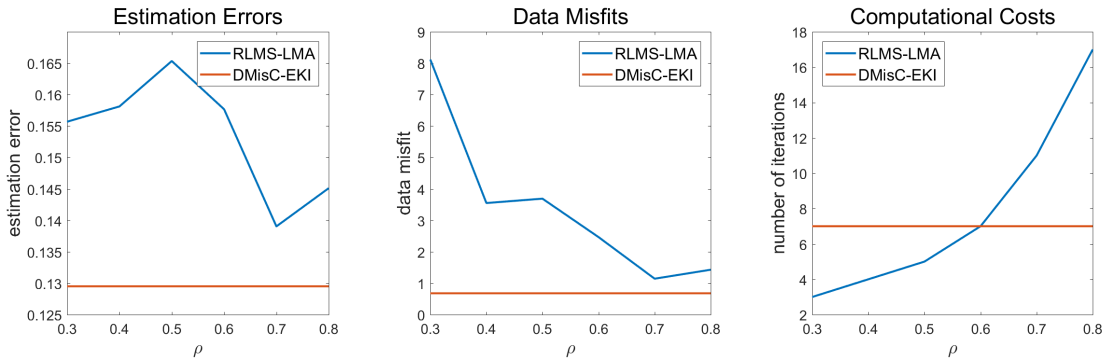
5.2.3 DMisC used in several circumstances

In this subsection, we conduct robustness testing of DMisC-EKI (without early stop) and DMisC-EnKI (without early stop) by using different experimental configurations. We will apply different length scales of the prior field, different distribution types of the noise, different uncertainty levels of the prior field, and different amounts of the noise.

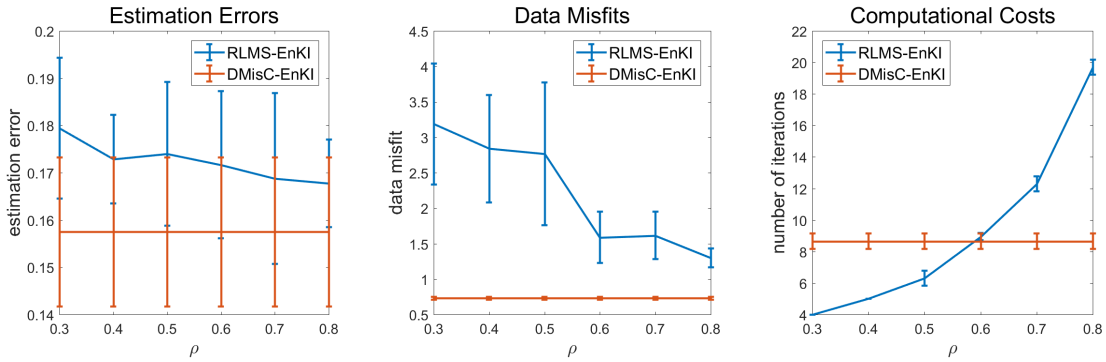
Effect of prior length scales

Now, we fix the prior uncertainty level $\sigma = 1$, fix the noise type as Gaussian noise, and fix the noise level $\epsilon = 0.02$. However, we apply different length scale parameters: case1 $l_x = 0.2$, $l_y = 0.2$, case2 $l_x = 0.2$, $l_y = 0.4$, case3 $l_x = 0.4$, $l_y = 0.4$. As the length-scale increases, the autocorrelation of the prior filed is also increases. Consequently, within fixed accuracy level, less eigenvalues are required to characterize the random fields. In other words, higher autocorrelation implies less modes of uncertainty. Thus, we expect that, more accurate estimates can be obtained with larger length scales. In the next, we will check this hypothesis.

We use DMisC-EKI and DMisC-EnKI (sample size $J = 400$) to solve the EIT problems with the different length scales. The estimates produced by the two methods compared



(a) Comparison of DMisC-EKI with RLMS-LMA. RLMS-LMA has a tuning parameter ρ. We try $\rho = 0.3$, $\rho = 0.4$, $\rho = 0.5$, $\rho = 0.6$, $\rho = 0.7$, and $\rho = 0.8$. For all the values of the tuning parameter, the leading results are compared with that produced by DMisC-EnKI. We present the estimation errors, data misfits, and computational costs of the two kinds of algorithms.



(b) Comparison of DMisC-EnKI with RLMS-EnKI (sample size $J = 400$). RLMS-EnKI has a tuning parameter ρ. We try $\rho = 0.3$, $\rho = 0.4$, $\rho = 0.5$, $\rho = 0.6$, $\rho = 0.7$, and $\rho = 0.8$. For all the values of the tuning parameter, the leading results are compared with that produced by DMisC-EnKI. We present the estimation errors, data misfits, and computational costs of the two kinds of algorithms. In above figures, the solid line and the error bar represent the mean and standard devation, respectively, calcualted from independent 20 implemenations.

Figure 5.3: Comparison of data-misfit controller and regularizing Levenberg-Marquardt scheme.

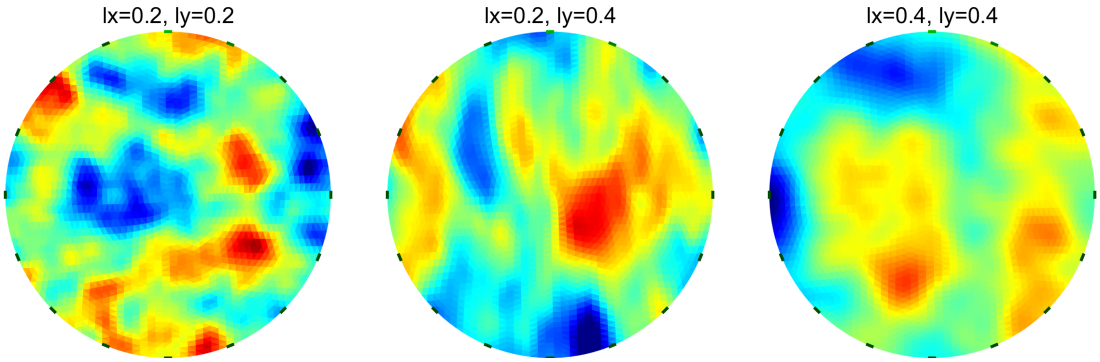
with the true values of the log conductivity are shown in figure 5.4, from which, we can observe that both DMisC-EKI and DMisC-EnKI works well for different length scales. Regardless the length scale is large or small, isotropic or anisotropic, the estimates produced by DMisC-EKI and DMisC-EnKI capture the main features of the truths. More clearly, we present the cross-sectional data of the estimates in figure 5.5, which shows that the estimates is more close to the truth when the length scale becomes larger. This fact validates our hypothesis. Finally, we present the estimation errors corresponding to different length scales in figure 5.6, which clearly shows that larger length scale implies lower uncertainty and higher accuracy.

Effect of noise types

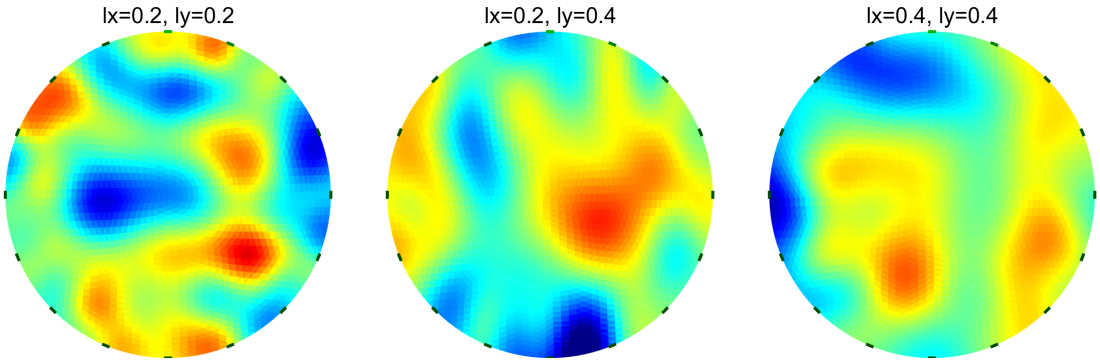
In theory, the Kalman-like methods and the data-misfit controller assume Gaussian noises. However, in the real world, noisy data is not always Gaussian. Practically, we hope to know whether the Kalman-like methods and the data-misfit controller works for non-Gaussian data or not. Then, in this test, we fix the prior uncertainty level $\sigma = 1$, fix the length-scale parameter $l_x = 0.2$, $l_y = 0.2$, and fix the noise level $\epsilon = 0.02$, but we apply different types of noise distributions: normal distribution, uniform distribution, and exponential distribution.

We apply both DMisC-EKI and DMisC-EnKI (sample size $J = 400$) to solve the EIT problems with the different noise types. The results are shown in figure 5.7. We can observe that, there is only little difference in estimation errors with respect to different types of noise distributions. Nevertheless, from figure 5.7a, we can still found the fact that, when we apply DMisC-EKI, the normal distribution leads to lower estimation errors than the exponential distribution, and lower than the uniform distribution. A possible reason is that, the normal distribution is endless on the real line, the exponential distribution has an end in one side, and the uniform distribution has two ends, so the exponential distribution than the uniform distribution is more close to the normal distribution. However, from figure 5.7b, this fact does not hold when DMisC-EnKI is applied. This is because EnKI is a statistical approach, the effect of statistical errors has taken over the effect of noise types. Therefore, the result is not distinguish when DMisC-EnKI is applied.

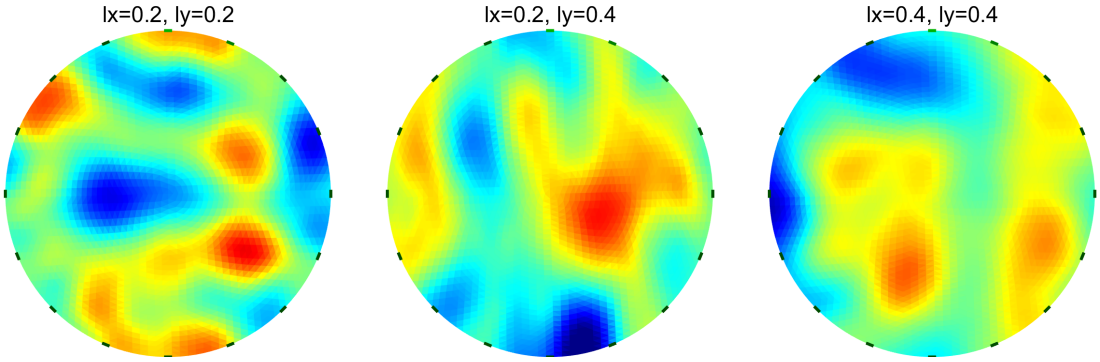
In summary, the effect of noise types is limited, that means the Kalman-like methods



(a) The true values of the log conductivity for the different length scales.

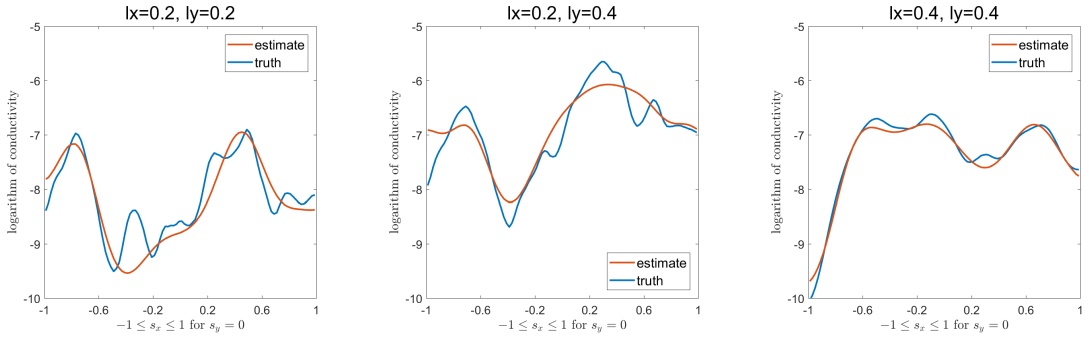


(b) The estimates produced by DMisC-EKI for the different length scales.

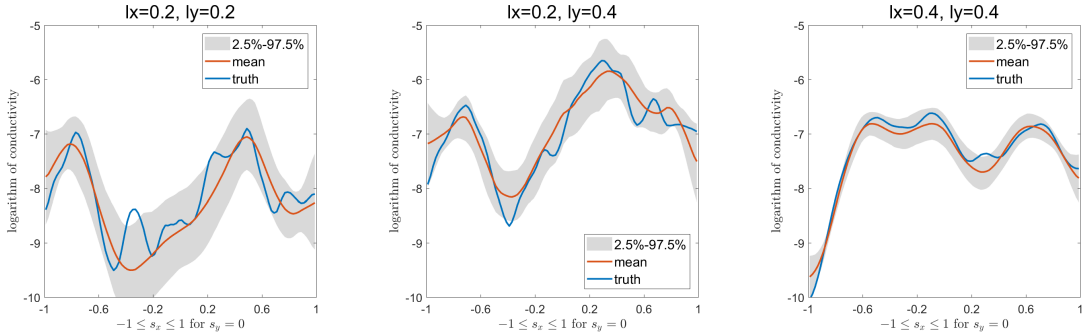


(c) The estimates produced by DMisC-EnKI (sample size $J = 400$) for the different length scales. EnKI is a statistical approach, so we repeat each of the estimation for 3 times, and each estimate is the mean of the 3 particle means.

Figure 5.4: The estimates produced by the DMisC-Kalman-like methods solving the EIT problem.

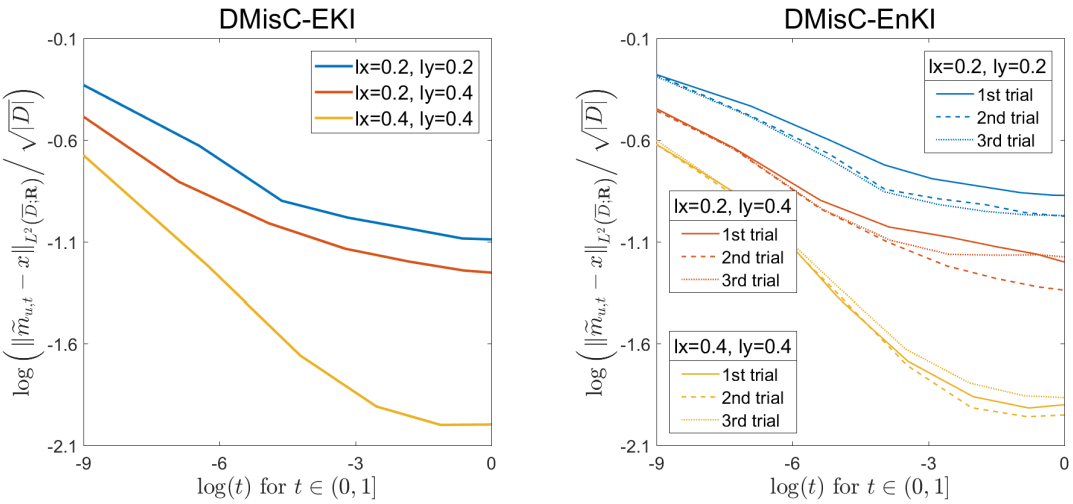


(a) The cross-sectional data at $s_y = 0$ of the estimates produced by DMisC-EKI for the different length scales.



(b) The cross-sectional data at $s_y = 0$ of the estimates produced by DMisC-EnKI (sample size $J = 400$) for the different length scales. EnKI is a statistical approach, so we repeat each of the estimation for 3 times, and each estimate is the mean of the 3 particle means.

Figure 5.5: The cross-sectional data at $s_y = 0$ of the estimates produced by the DMisC-Kalman-like methods solving the EIT problem.



(a) The estimation errors in the filtering of DMisC-EKI for the different length scales.

(b) The estimation errors in the filtering of DMisC-EnKI for the different length scales.

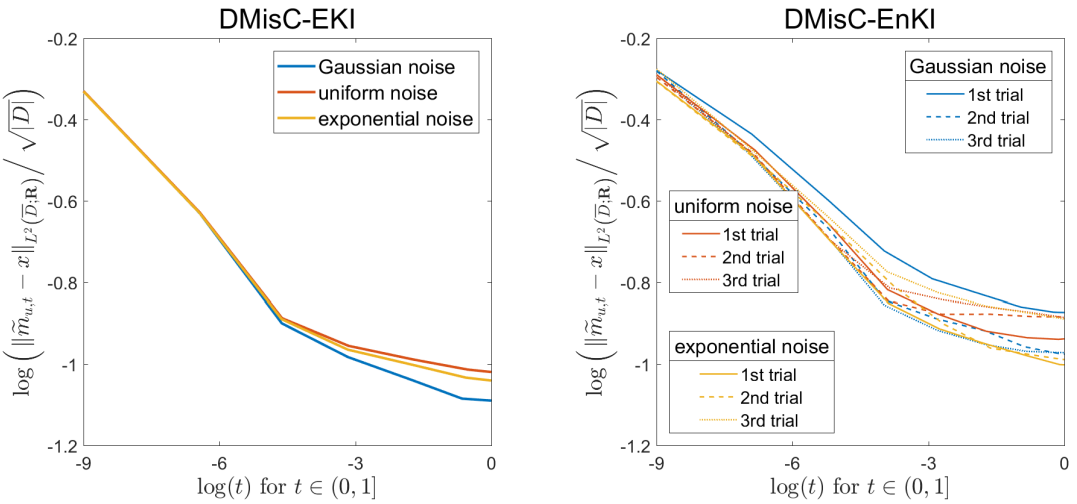
Figure 5.6: The estimation errors agsinst the prior length scales.

and the data-misfit controller are still feasible for non-Gaussian noisy data in practice.

Effect of prior uncertainty

In this test, we fix the prior length-scale parameter $l_x = 0.2$, $l_y = 0.2$, fix the noise level $\epsilon = 0.02$, and fix the type of noise distribution as Gaussian noise. Then, we try different prior uncertainty levels $\sigma = 0.2, 0.4, 0.6, 0.8, 1$. It is expected that, the uncertainty of the posterior estimate should increase as the prior uncertainty rises up.

Then, we apply DMisC-EKI and DMisC-EnKI (sample size $J = 400$) to solve the EIT problems with the different prior uncertainty levels. The results are shown in figure. For DMisC-EnKI, since it is a statistical approach, we conduct independent and repeated tests for 20 times. For each time, we obtain a posterior particle mean, and this mean is treated as an estimate of the truth. Then, we use the L_2 norm to calculate the estimation error of the particle mean from the truth. Since EnKI is statistical method, the particle means and the estimation errors for the independent and repeated 20 tests are different. Then, we calculate the mean and standard deviation of 20 estimation errors. We show the results of posterior estimation errors against different values of σ in figure 5.8. We can clearly observe that the posterior estimation error is nearly proportional to the prior



(a) The estimation errors in the filtering of DMisC-EKI for the different length scales.

(b) The estimation errors in the filtering of DMisC-EnKI for the different length scales.

Figure 5.7: The estimation errors against the types of noise ditribution.

uncertainty level. Also, the required number of iterations increases in the similar rate as the estimation error have.

Effect of noise level

Finally, we test the noise level ϵ. We fix the prior length-scale parameter $l_x = 0.2$, $l_y = 0.2$, fix the prior uncertainty level $\sigma = 1$, and fix the type of noise distribution as Gaussian noise, but we try different values of the noise level ϵ. It is expected that, the estimation error increases as the noise level goes up. However, the behavior of noise level is more complicated. We classify the noise levels into three groups: the small amounts of noise

$$\epsilon \in \{0.4\%, 0.8\%, 1.2\%, 1.6\%, 2\%\} \tag{5.2.14}$$

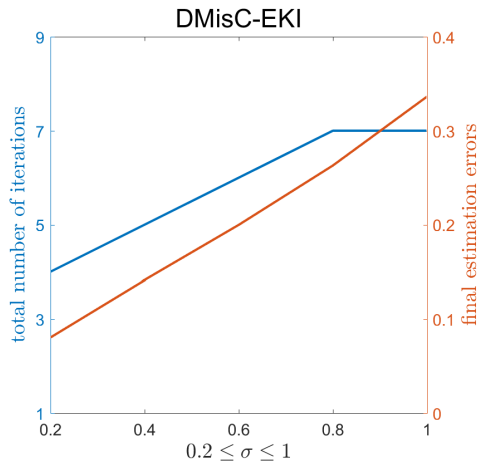
the moderate amounts of noise

$$\epsilon \in \{2\%, 4\%, 6\%, 8\%, 10\%\} \tag{5.2.15}$$

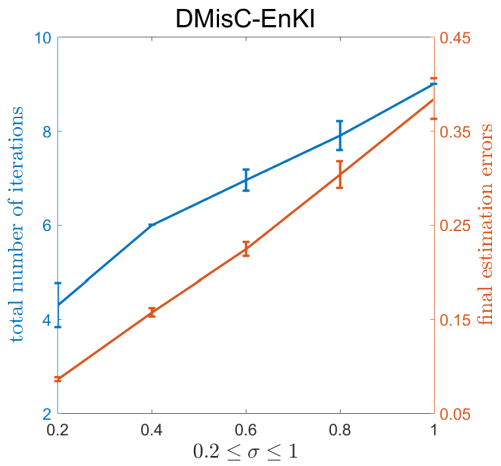
and the large amounts of noise

$$\epsilon \in \{10\%, 20\%, 30\%, 40\%, 50\%\} \tag{5.2.16}$$

We conduct DMisC-EKI and DMisC-EnKI (sample size $J = 400$) to solve the EIT problems with the different noise levels. The results are shown in figure 5.9, figure 5.10,



(a) The estimation errors in the filtering of DMisC-EKI for the different prior uncertainty levels.

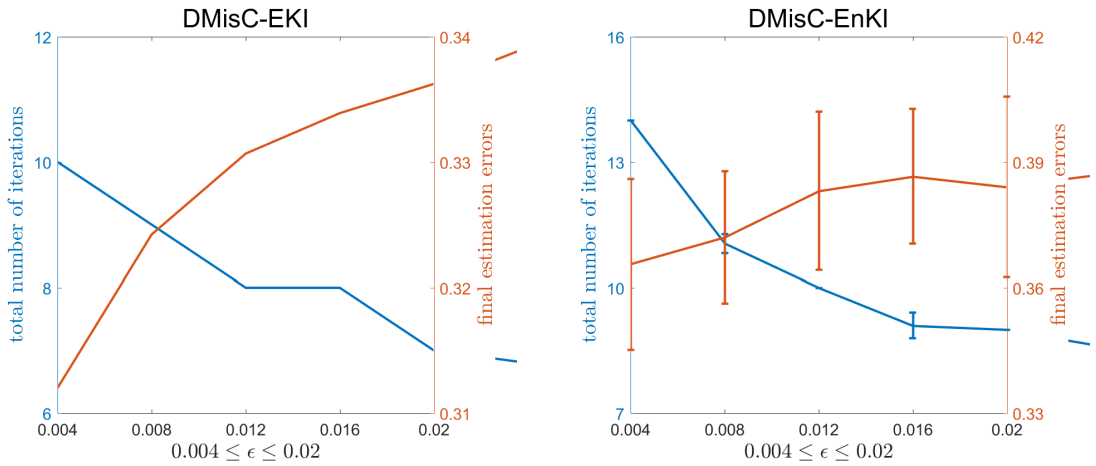


(b) The estimation errors in the filtering of DMisC-EnKI for the different prior uncertainty levels.

Figure 5.8: The estimation errors against the prior uncertainty levels.

and figure 5.11. Form figures 5.10 and 5.11, we can observe that, for moderate and large amounts of noise, for both DMisC-EKI and DMisC-EnKI, the estimation error increases (almost in a linear rate) as the noise level goes up. This conforms our expectation. However, the behavior becomes different for small noise levels. In the range of small noise levels, the estimation error of DMisC-EKI shows a pattern of accelerated reduction as the noise level tends to zero, but the estimation error of DMisC-EnKI is not stable in this range. We repeatedly conduct 20 times of DMisC-EKI, and plot the mean and standard deviation of the 20 estimation errors in figure 5.9b. We can clearly observe that the statistical error of DMisC-EnKI rather than the noise level play the dominate role. This comes from two facts: 1) the standard deviation of the estimation errors is larger than the change of the estimation errors with different noise levels; 2) the mean of estimation errors at $\epsilon = 0.02$ even smaller than the mean of estimation errors at $\epsilon = 0.16$. This violates our expectation.

An interesting fact is that: the critical value seems to be $\epsilon = 0.02$. For any $\epsilon > 0.02$, for both DMisC-EKI and DMisC-EnKI, there is a stably increasing pattern (nearly in linear rate) of the estimation errors as ϵ rises up. For any $\epsilon < 0.2$, when DMisC-EKI is applied, the estimation errors are acceleratedly reduced as ϵ tends to zero, and when



(a) The estimation errors in the filtering of DMisC-EKI for the different noise levels.

(b) The estimation errors in the filtering of DMisC-EnKI for the different noise levels.

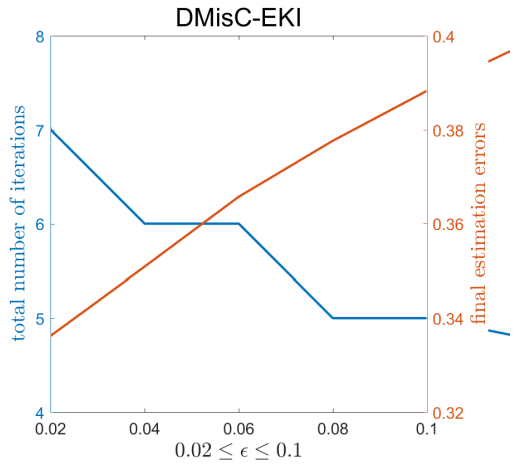
Figure 5.9: The estimation errors against the small noise levels.

DMisC-EnKI is applied, there is no stable tendency. This is why $\epsilon = 0.02$ is fixed as the critical value in our numerical tests. For implementation of EnKI, we do not suggest using too small noise level since there exists the statistical error of EnKI. This statistical error should be accounted as a part of the noise source. Namely, the value of ϵ should contain the effect of the statistical error caused by the inverse algorithm.

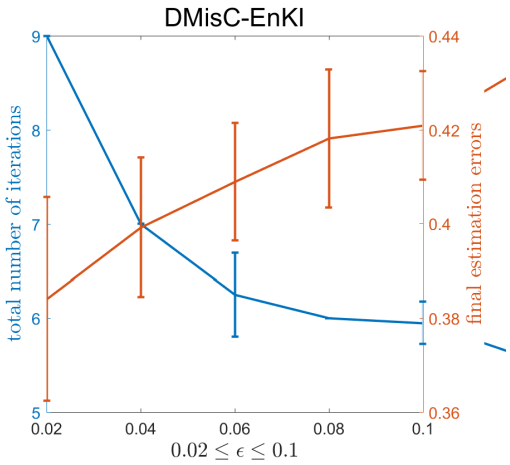
5.3 Advanced tests with multi-phases conductivity

In previous experiments, we have applied single-phase priors. However, in the real world, unknown parameters could be more likely to have multi-phases. For example, consider EIT in medical imaging. There are contrasts between lung/brain and tissue fluid, causing inhomogeneity of the underlying parameter. Thus, it would be more realistic to classify the inhomogeneous media into two patterns: one is the target (lung/brain), and another is the background (tissue fluid). Applying level-set priors can improve resolution of imaging, as level-set priors use indicator functions for classification that provides more identifiable features of underlying parameters.

In this section, our main purpose is to show that, EnKI also works for level-set priors which are non-differentiable and highly nonlinear (because of the existing of indicator

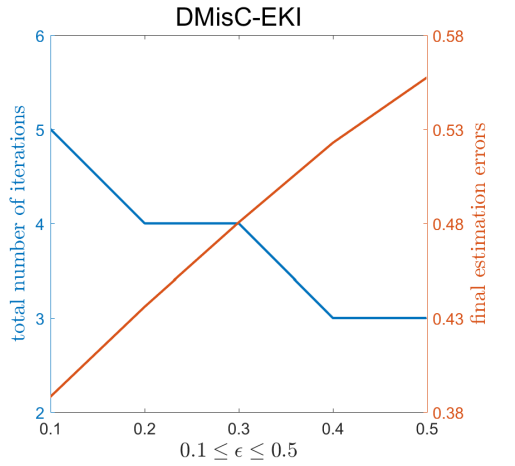


(a) The estimation errors in the filtering of DMisC-EKI for the different noise levels.

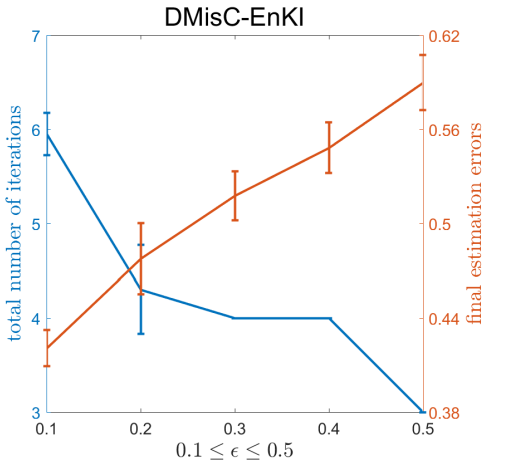


(b) The estimation errors in the filtering of DMisC-EnKI for the different noise levels.

Figure 5.10: The estimation errors against the moderate noise levels.



(a) The estimation errors in the filtering of DMisC-EKI for the different noise levels.



(b) The estimation errors in the filtering of DMisC-EnKI for the different noise levels.

Figure 5.11: The estimation errors against the large noise levels.

functions). Furthermore, the early stop criterion significantly improves the robustness of the Kalman-like methods when they are applied for highly nonlinear problems. DMisC-EKI (with early stop) and DMisC-EnKI (with early stop) are listed in algorithm 3 and algorithm 4 in the end of chapter 3.

5.3.1 Introduction of level-set priors

In practice, forward models are not always differentiable, or the derivatives are sometimes intractable, e.g. a black box model. Even though it is feasible to calculate the derivatives of forward models, prior models can be non-differentiable. Consider the chain of models involving hyper parameters,

$$y = \mathcal{G}(u), \quad u = g_1(x_1), \quad x_1 = g_2(x_2), \quad x_2 = g_3(x_3), \quad \cdots \tag{5.3.1}$$

where $\mathcal{G}$ is the forward model, and $g_1 \circ g_2 \circ g_3 \circ \cdots$ is the prior model. The hierarchical structure of the prior model allows any sophisticated construction, which provides sufficient parameters fitting observations and making predictoins. [75] investigates how deep is deep enough.

A simple example of non-differentiablity of prior model is the level set [74] of random fields which involves $p > 1$ components, i.e. the prior random field u on D is represented by, for all $s \in D$,

$$u(s) = \sum_{i=1}^{p} w_i(s) 1_{A_i}(v(s)) \tag{5.3.2}$$

where w_1,..., w_p, v are real-valued random fields on D, and 1_{A_1},..., 1_{A_p} are the indicator functions with indicator sets A_1,..., A_p, where for any $i = 1, ..., p$, $A_i = (a_{i-1}, a_i)$ with $-\infty = a_0 < a_1 < \cdots < a_p = +\infty$. Clearly, $u = u(w, v)$ is non-differentiable with respect to v.

With non-differentiable forward models or hierarchical prior models, EKI is not feasible anymore, since EKI requires the first order derivative. However, EnKI also works in practice, as EnKI treats the models as black boxes and updates parameters with a simple algorithm. Compared with other derivative-free methods, e.g. MCMC methods, the EnKI avoids many simulations of long Markov chains. EnKI is very popular in inverse problems, because of the derivative-free property.

5.3.2 The simulated truth and data

For our numerical experiments of the EIT problem. We draw a truth from a random field with two phases, so that, each phase represents either the lung/brain or the tissue fluid. The true value of log conductivity x is drawn as a sample from a random field $\widehat{u}$ on $D = \{s \in \mathbb{R}^2 : \|s\| \leq 1\}$, such that, for all $s \in D$,

$$\widehat{u}(s) = \begin{cases} w_1(s) & \text{if } f(s) \leq 1 \\ w_2(s) & \text{if } f(s) > 1 \end{cases} \tag{5.3.3}$$

where w_1 and w_2 are two Whittle-Matérn fields on D, and $f(s)$ is the elliptic curve, let $s = (x, y)$,

$$f(s) = f(x, y) = \frac{(|x| - 0.4)^2}{0.3^2} + \frac{y^2}{0.7^2} \tag{5.3.4}$$

Parameters of the Whittle-Matérn fields w_1 and w_2 are specified below,

$$\mu_{w_1} = -8.5; \qquad \sigma_{w_1} = 0.5; \qquad \nu_{w_1} = 2; \qquad L_{w_1} = \begin{bmatrix} 0.15 & 0 \\ 0 & 0.3 \end{bmatrix} \tag{5.3.5}$$

$$\mu_{w_2} = -6.5; \qquad \sigma_{w_2} = 0.5; \qquad \nu_{w_2} = 2; \qquad L_{w_2} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \tag{5.3.6}$$

where μ is the steady mean, σ is the steady standard deviation, ν is the smoothness parameter, and L is the length-scale parameter of Whittle-Matérn fields.

We draw a sample $\omega \in \Omega$ (Ω is the event set) from the random field $\widehat{u}$, and then assign $x(s) = \widehat{u}(s, \omega)$ for all $s \in D$. Then we let x be the true value of the log conductivity on D. The truth is shown in the top row in figure 5.12. Clearly, there are two phases, inside and outside of the elliptic curves. Then, we simulate a noisy data $y = \mathcal{G}(x) + 0.02 \cdot \text{abs}(\mathcal{G}(x)) \odot \zeta$, where $\mathcal{G}$ is the forward model, ζ is a sample drawn from the n-dimensional standard normal distribution, $\text{abs}(\cdot)$ takes the entry-wise absolute values, and $\odot$ is the entry-wise multiplication. Then, we pretend to know nothing about the truth x, and use the noisy observation y to estimate the truth x as the inverse problem.

5.3.3 Objectives of numerical experiments

Our purpose is to show the following facts from numerical experiments.

1. For multi-phases underlying parameters, level-set prior rather than single-phase prior can improve the imaging resolution.

2. DMisC-EnKI, as a derivative-free method with adaptive step sizes, works well when level-set prior is applied.

3. The early stop criterion further improves robustness of DMisC-EnKI. As the result, DMisC-EnKI associated with the early stop criterion is quite efficient.

We will compare a single-phase prior field and a two-phases prior field for the EIT problem. For the single-phase prior, we will apply both the DMisC-EKI and the DMisC-EnKI; for the level-set prior, only DMisC-EnKI is applicable.

On one hand, the single-phase prior is chosen as a Whittle-Matérn field with steady mean $\mu = -7.5$, steady standard deviation $\sigma = 1$, smoothness parameter $\nu = 2$, and length-scale parameter $L = \text{diag}([0.3, 0.3])$. Whittle-Matérn field is discussed in subsection 2.3.3.

On the other hand, the two-phases prior is specified as the level-set field with two components,

$$u(s) = \begin{cases} w_1(s) & \text{if } v(s) < 0 \\ w_2(s) & \text{if } v(s) \geq 0 \end{cases} \tag{5.3.7}$$

where w_1 and w_2 are the two Whittle-Matérn fields same as the specifications in (5.3.5) and (5.3.6), and v is the third Whittle-Matérn field which has steady mean $\mu = 0$, steady standard deviation $\sigma = 1$, smoothness parameter $\nu = 2$, and length-scale parameter $L = \text{diag}([0.3, 0.3])$. We adopt u as the level-set prior field.

EIT usually has lower spatial resolution than computed tomography scan and magnetic resonance imaging. However, its resolution can be improved by using 32 instead of 16 electrodes. We will also compare the effect of number of electrodes.

In summery, we will conduct 6 experiments:

1. 16 electrodes, single-phase stationary prior, DMisC-EKI;

2. 16 electrodes, single-phase stationary prior, DMisC-EnKI;

3. 16 electrodes, two-phases level-set prior, DMisC-EnKI;

4. 32 electrodes, single-phase stationary prior, DMisC-EKI;

5. 32 electrodes, single-phase stationary prior, DMisC-EnKI;

6. 32 electrodes, two-phases level-set prior, DMisC-EnKI.

5.3.4 Results of the numerical tests

We conduct all the six numerical experiments. When DMisC-EnKI is applied, since it is a statistical approach, we independently repeat implementations of DMisC-EnKI for 3 times to check if the results are close to each other or not. When level-set prior is applied, since it is a highly nonlinear model involving the indicator function, we adopt the early stop criterion monitoring the performance of DMisC-EnKI in the filtering. It is not necessary to apply early stop for the stationary prior, because in this case the early stop criterion makes no difference or only little difference.

The final estimates obtained in the 6 numerical tests are shown in figure 5.12, where the truth is also presented as the benchmark. When the single-phase prior is applied, all the images are blurred. However, using the level-set prior significantly improves the imaging resolution. Furthermore, when the level-set prior is applied, using 32 electrodes can be even better than using 16 electrodes, as the underlying parameter is more identifiable given more observations. In contrast, if only increase the electrodes from 16 to 32, without applying level-set prior, then there is no much improvement, because the stationary prior does not provide proper characterization of the underlying parameter.

From the cross-sectional data in figure 5.13, it is notable that the estimates based on the single-phase prior do not capture the pattern of the two-phases truth. When the level-set prior is applied, the estimates are much better as they cover the main part of the truth. Especially, when the level-set prior is associated with 32 electrodes, most area of the truth is contained in the 95% confidence interval.

In the next, we plot the data misfits and estimation errors against the tempering parameter $t \in (0, 1]$ in figure 5.14. This figure shows the adequate performance of the data-misfit controller associated with the early stop criterion. Here is about the key discussion points. We expand in the following two paragraphs.

Firstly, consider the left column and the middle column in figure 5.14. In these columns, the single-phase stationary prior field is applied. Under this prior, DMisC-EKI and DMisC-EnKI are efficient, only requiring 7–9 iterations in the filtering from $t = 0$ to $t = 1$, but the resulting estimates have low fidelity. The low resolution of imaging is not the issue of the algorithm, but the issue of the prior information, since the stationary prior does not characterize the truth properly.

Secondly, consider the right column in figure 5.14. In this column, the level-set prior is applied. Since level-set model is highly nonlinear, we need to consider two possibilities: adopting the early stop criterion or not. Even if we refuse the early stop criterion and just conduct the filtering until the tempering parameter touches the final point $t = 1$, there is no much difference, as shown in the right column in figure 5.14, the estimation errors only fluctuates but not decreases from the early stop point to the final point $t = 1$. The early stop criterion remains the similar accuracy of estimates, but has much more efficiency. More clearly, if the early stop criterion is not used, then DMisC-EnKI requires average 40.67 iterations for 16 electrodes, and average 22.33 iterations for 32 electrodes. However, once the early stop criterion is adopted, DMisC-EnKI only needs average 12.67 and 10 iterations for 16 and 32 electrodes, respectively.

The early stop criterion significantly improves the robustness of the Kalman-like methods. The reason is that, the level-set prior is highly nonlinear with an indicator function, so in this case, the accuracy of the Kalman-like approximation may be insufficient. It is under consideration to validate whether the Kalman-like approximation is good or not. The early stop criterion thus monitors the quality of the estimates by checking the monotone properties. If the cost functional or the objective functional does not decrease anymore, we should stop the algorithm, since the Kalman-like approximation has been far from the accurate tempering setting which is provided with the monotone properties.

5.4 Brief notes and summary

From the numerical tests in this chapter, we conclude the following notations.

1. The Kalman-like methods are feasible for nonlinear inverse problems, whose forward maps are continuously differntiable (choose EKI), or whose parameters and

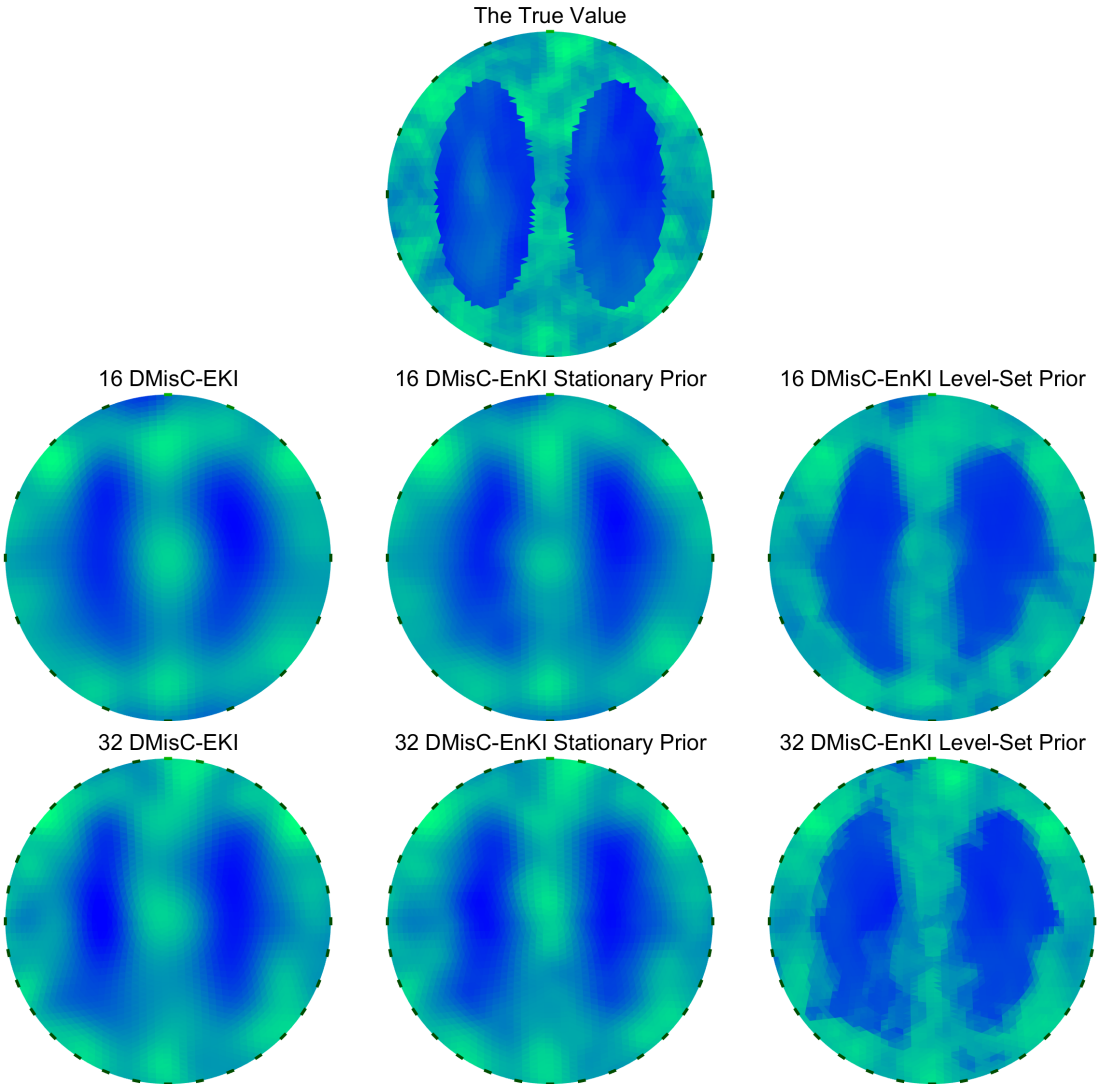


Figure 5.12: Comparison of the truth and the estimates produced by different algorithms (DMisC-EKI or DMisC-EnKI) with different priors (stationary field or level-set field) and different numbers of electrodes (16 or 32).

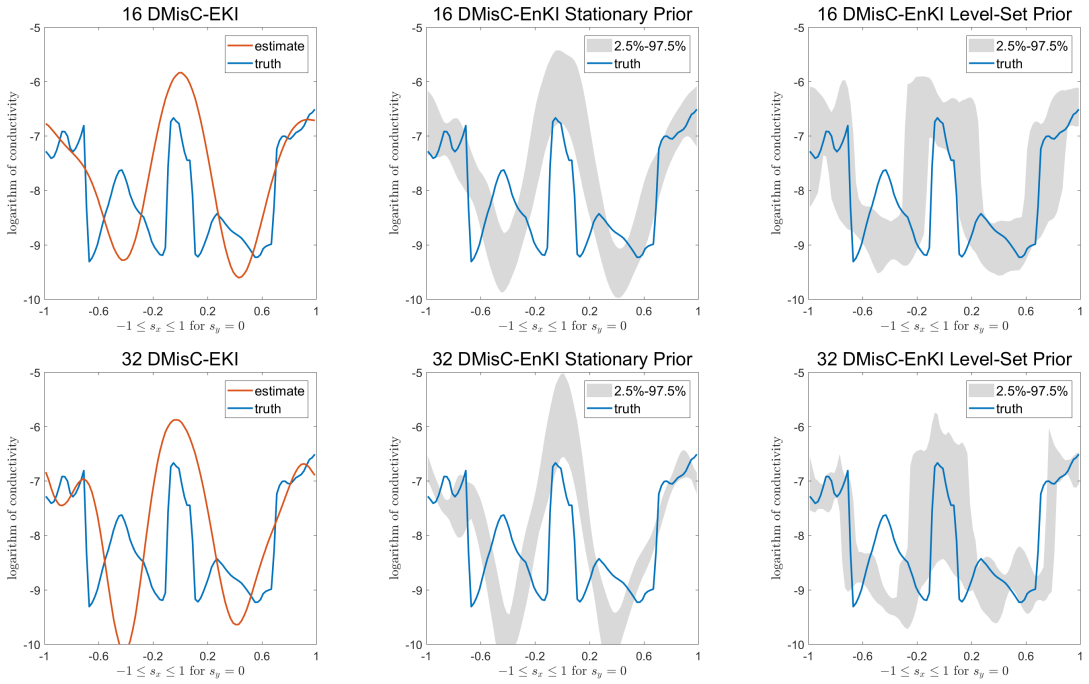


Figure 5.13: The cross-sectional data of the estimates at $s_y = 0$ produced by different algorithms (DMisC-EKI or DMisC-EnKI) with different priors (stationary field or level-set field) and different numbers of electrodes (16 or 32).

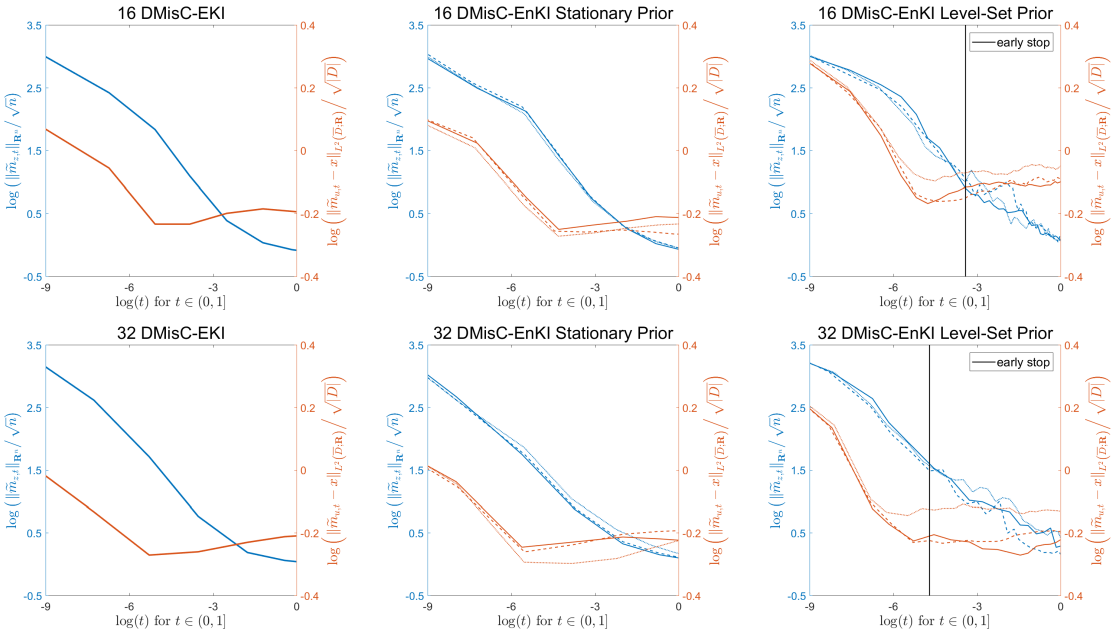


Figure 5.14: The performace of different algorithms (DMisC-EKI or DMisC-EnKI) in filtering under different priors (stationary field or level-set field) and different numbers of electrodes (16 or 32).

observations still have strong linear correlations (choose EnKI).

2. The data-misfit controller proposed in this thesis, as an adaptive strategy, works well for the Kalman-like methods, and keeps the balance between accuracy and efficiency.

3. Practically, the Kalman-like methods and the data-misfit controller are also feasible for non-Gaussian data, even though they use the Gaussian assumption in theory.

4. For highly nonlinear problems, the early stop criterion proposed in this thesis is recommended to adopt, which can improve the robustness of approximate filtering methods like the Kalman-like methods.

5. In conclusion, we recommend the following numerical strategy for nonlinear inverse problems: the Kalman-like methods + the data-misfit controller + the early stop criterion. If this strategy fails, then universal methods like MCMC methods have to be in consideration, although they may cost much more computational recourse.

Algorithm 5 GSS-EKI for data-misfit function $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ with Gaussian prior $\mathcal{N}(m_0, \mathcal{C}_0)$

Provide the increasing rate $q > 1$ and the total number of discrete steps K.

Assign the initial state $i \leftarrow 0$, $m \leftarrow m_0$, $\mathcal{C} \leftarrow \mathcal{C}_0$, $z \leftarrow \mathcal{Z}(m)$, $D \leftarrow \mathrm{D}\mathcal{Z}(m)$, $Q \leftarrow \mathcal{C}D^*$, $C \leftarrow DQ$.

while $i < K$ **do**

Predict the step size h as the ith item in a geometric sequence,

$$h_i = \frac{q^i - q^{i-1}}{q^K - 1} \tag{5.4.1}$$

Predict the mean-covariance pair $(m_p, \mathcal{C}_p)$ with the extended Kalman filter,

$$m_p \leftarrow m - hQ\left(\mathbf{I} + hC\right)^{-1} z \qquad \mathcal{C}_p \leftarrow \mathcal{C} - hQ\left(\mathbf{I} + hC\right)^{-1} Q^* \tag{5.4.2}$$

Calculate

$$z_p \leftarrow \mathcal{Z}(m_p) \qquad D_p \leftarrow \mathrm{D}\mathcal{Z}(m_p) \qquad Q_p \leftarrow \mathcal{C}_p D_p^* \qquad C_p \leftarrow D_p Q_p \tag{5.4.3}$$

Renew the state $i \leftarrow i + 1$, $m \leftarrow m_p$, $\mathcal{C} \leftarrow \mathcal{C}_p$, $z \leftarrow z_p$, $D \leftarrow D_p$, $Q \leftarrow Q_p$, $C \leftarrow C_p$.

end while

return $(m, \mathcal{C})$ as the mean-covariance pair of the posterior distribution (approximately).

Algorithm 6 GSS-EnKI for data-misfit function $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ with Gaussian prior $\mathcal{N}(m_0, \mathcal{C}_0)$

Provide the sample size J, the increasing rate $q > 1$, and the total number of discrete steps K.

Assign $i \leftarrow 0$. Draw J samples independently from the prior distribution $\mathcal{N}(m_0, \mathcal{C}_0)$, and let these K samples into a particle U. Calculate the particle Z such that, $Z(j) \leftarrow \mathcal{Z}(U(j))$ for all $j = 1, ..., J$. Calculate the sample means and sample covariances, $\overline{u} \leftarrow \text{mean}(U)$, $\overline{z} \leftarrow \text{mean}(Z)$, $C_{uz} \leftarrow \text{cov}(U, Z)$, $C_{zz} \leftarrow \text{cov}(Z, Z)$.

while $i < K$ **do**

Predict the step size h as the ith item in a geometric sequence,

$$h_i = \frac{q^i - q^{i-1}}{q^K - 1} \tag{5.4.4}$$

Draw J samples independently from the n-dimensional standard normal distribution, and let these J samples into a particle V. Predict the particle U_p with the ensemble Kalman filter, such that, for all $j = 1, ..., J$,

$$U_p(j) \leftarrow U(j) - C_{uz}\left(\mathbf{I} + hC_{zz}\right)^{-1}\left(Z(j)h - V(j)\sqrt{h}\right) \tag{5.4.5}$$

Calculate the particle Z_p such that, $Z_p(j) \leftarrow \mathcal{Z}(U_p(j))$ for all $j = 1, ..., J$. Calculate the sample means and sample covariances, $\overline{u}_p \leftarrow \text{mean}(U_p)$, $\overline{z}_p \leftarrow \text{mean}(Z_p)$, $C_{uz,p} \leftarrow \text{cov}(U_p, Z_p)$, $C_{zz,p} \leftarrow \text{cov}(Z_p, Z_p)$.

Renew the state $i \leftarrow i + 1$, $U \leftarrow U_p$, $Z \leftarrow Z_p$, $\overline{u} \leftarrow \overline{u}_p$, $\overline{z} \leftarrow \overline{z}_p$, $C_{uz} \leftarrow C_{uz,p}$, $C_{zz} \leftarrow C_{zz,p}$.

end while

return U as the particle under the posterior distribution (approximately).

Algorithm 7 RLMS-LMA for data-misfit function $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ with Gaussian prior $\mathcal{N}(m_0, \mathcal{C}_0)$

Provide two accuracy control parameters $\rho \in (0,1)$ and $\epsilon > \rho^{-1}\|\mathcal{Z}(x)\|$, where $x \in \mathcal{H}$ is the true value of the unknown parameter (practically, use $\epsilon = \rho^{-1}\sqrt{n}$ if x is not given). Assign the initial state $m \leftarrow m_0$, $z \leftarrow \mathcal{Z}(m)$, $D \leftarrow \mathrm{D}\mathcal{Z}(m)$, $C \leftarrow D\mathcal{C}_0 D^*$.

while $\|z\|_{\mathbb{R}^n} > \epsilon$ **do**

Predict the step size h with the RLMS, i.e, h is determined by solving the following equation,

$$\left\|(\mathbf{I} + hC)^{-1} z\right\|_{\mathbb{R}^n} = \rho \|z\|_{\mathbb{R}^n} \tag{5.4.6}$$

Predict the estimate m_p with the LMA,

$$m_p \leftarrow m - hQ\,(\mathbf{I} + hC)^{-1} z \tag{5.4.7}$$

Calculate

$$z_p \leftarrow \mathcal{Z}(m_p) \qquad D_p \leftarrow \mathrm{D}\mathcal{Z}(m_p) \qquad C_p \leftarrow D_p \mathcal{C}_0 D_p^* \tag{5.4.8}$$

Renew the state $m \leftarrow m_p$, $z \leftarrow z_p$, $D \leftarrow D_p$, $C \leftarrow C_p$.

end while

return m as the estimate of the unknown parameter.

Algorithm 8 RLMS-EnKI for data-misfit function $\mathcal{Z} : \mathcal{H} \to \mathbb{R}^n$ with Gaussian prior $\mathcal{N}(m_0, \mathcal{C}_0)$

Provide sample size J, and two accuracy control parameters $\rho \in (0,1)$ and $\epsilon > \rho^{-1}\|\mathcal{Z}(x)\|$, where $x \in \mathcal{H}$ is the true value of the unknown parameter (practically, use $\epsilon = \rho^{-1}\sqrt{n}$ if x is not given).

Draw J samples independently from the prior distribution $\mathcal{N}(m_0, \mathcal{C}_0)$, and let these K samples into a particle U. Calculate the particle Z such that, $Z(j) \leftarrow \mathcal{Z}(U(j))$ for all $j = 1, ..., J$. Calculate the sample means and sample covariances, $\overline{u} \leftarrow \text{mean}(U)$, $\overline{z} \leftarrow \text{mean}(Z)$, $C_{uz} \leftarrow \text{cov}(U, Z)$, $C_{zz} \leftarrow \text{cov}(Z, Z)$.

while $\|\overline{z}\|_{\mathbb{R}^n} > \epsilon$ **do**

Predict the step size h with the RLMS, i.e, h is determined by solving the following equation,

$$\left\|\left(\mathbf{I} + hC_{zz}\right)^{-1}\overline{z}\right\|_{\mathbb{R}^n} = \rho\,\|\overline{z}\|_{\mathbb{R}^n} \tag{5.4.9}$$

Draw J samples independently from the n-dimensional standard normal distribution, and let these J samples into a particle V. Predict the particle U_p with the ensemble Kalman filter, such that, for all $j = 1, ..., J$,

$$U_p(j) \leftarrow U(j) - C_{uz}\left(\mathbf{I} + hC_{zz}\right)^{-1}\left(Z(j)h - V(j)\sqrt{h}\right) \tag{5.4.10}$$

Calculate the particle Z_p such that, $Z_p(j) \leftarrow \mathcal{Z}(U_p(j))$ for all $j = 1, ..., J$. Calculate the sample means and sample covariances, $\overline{u}_p \leftarrow \text{mean}(U_p)$, $\overline{z}_p \leftarrow \text{mean}(Z_p)$, $C_{uz,p} \leftarrow \text{cov}(U_p, Z_p)$, $C_{zz,p} \leftarrow \text{cov}(Z_p, Z_p)$.

Renew the state $U \leftarrow U_p$, $Z \leftarrow Z_p$, $\overline{u} \leftarrow \overline{u}_p$, $\overline{z} \leftarrow \overline{z}_p$, $C_{uz} \leftarrow C_{uz,p}$, $C_{zz} \leftarrow C_{zz,p}$.

end while

return U as the particle under the posterior distribution (approximately).

Chapter 6

Summary

We have conducted study on numerical algorithms solving inverse problems. Our main idea is to apply the tempering setting, which rewrites the one-step transition from prior to posterior as the continuous transition, such that, inverse problems can be equivalently regarded as Bayesian filtering algorithm as the tempering parameter $t \in [0, 1]$ goes up. $t = 0$ indicates the prior and $t = 1$ indicates the posterior. Our main contribution is to develop the adaptive strategy (called data-misfit controller in this thesis) used to discretize the continuous tempering setting. This adaptive strategy has both theoretical and numerical advantages. In theory, the data-misfit controller determines step sizes for the tempering setting, from which, the accuracy (mean) and uncertainty (variance) of estimates in the discrete filtering are controlled, and at the same time, the sum of forward and backward information gain of any two successive probability measures is also controlled. In practice, this method has no tuning parameters, so users can directly apply it. This is much more convenient than other existing algorithms. In many complicated applications, they suffer from that they need to tune algorithmic parameters. Then data-misfit controller thus conquers this issue. Moreover, data-misfit controller keeps the balance of accuracy and efficiency.

In practice, it is usually numerically too expensive to solve Bayesian inverse problems with an accurate method such as MCMC. Then in realistic applications, we resort to apply some approximate methods, such as Kalman-like methods, if the approximate methods can provide satisfactory results. When we implement Kalman-like methods, we can still apply the data-misfit controller to select step sizes. However, Kalman-like methods are

heuristic algorithms, that means, sometimes they work, sometimes not. We need to validate if the approximation is acceptable. For this purpose, we additionally propose the early stop criterion for the Kalman-like methods. That means the filtering stops at $t = s$ for some $s < 1$ before it touches $t = 1$. The early stop criterion proposed in this thesis is based on the monotone properties of the tempering setting. The suggested criterion is simple to implement: since the tempering setting theoretically has monotone properties, once any approximation of the tempering setting violates the monotone properties, it cannot be a good approximation and then we should refuse it. Applying the early stop criterion can cut off many unnecessary or inaccurate iterations, and improve robustness of Kalman-like methods for highly nonlinear problems.

There exists the natural connection between Tikhonov regularization and Bayesian inversion under Gaussian measures: the former is the MAP estimator of the latter. For both the variational inversion and the Bayesian inversion, we prove that the tempering setting determines a trajectory, such that, the cost functional at the MAP point and the average cost functional are always decreasing along this trajectory from the prior to the posterior. The proofs are done for nonlinear infinite-dimensional problems. Furthermore, under the local linearization, the tempered Tikhonov regularization can be simplified as the extended Kalman filter (EKF); and under the global linearization, the tempered Bayesian inversion with Gaussian prior and Gaussian noise can be simplified as the ensemble Kalman filter (EnKF). Although these simplifications are heuristic approach, they lead to practical benefits: solving the EKF and EnKF is just solving the ODE and SDE with finite difference schemes, where the step size can be automatically selected by the data-misfit controller.

We have used the EIT model to test the proposed algorithms. Data-misfit controller performs well in several circumstances: for different autocorrelation functions, different prior uncertainty levels, different noise types, and different noise levels. It picks step sizes effectively and produces results accurately, which is better than some other existing algorithms (RLMS-LMA, RLMS-EnKI). Furthermore, we additionally apply the level-set prior, which involves indicator functions for classification of multiple patterns of underlying parameters. The level-set prior is non-differentiable, with which, we can show the good properties of EnKI as a derivative-free method. Also, we emphasize that, for this

highly nonlinear (indicator functions) problem, the early stop criterion saves more than half computational costs. Therefore, we recommend to apply the data-misfit controller associated with early stop criterion, especially for highly nonlinear problems. For nearly linear problems, there is no much difference if the early stop criterion is adopted or not.

In the future, the interesting research focus will lie on the combination of Kalman approach and MCMC approach simultaneously solving the tempering setting. The main purpose will be that, since Kalman-like methods are efficient and MCMC methods are accurate, it is desired to develop integrated numerical strategies with benefits from the both sides, and then have robust properties for more general and complicated applications. Our logic is to firstly apply the switch update of Kalman approach and MCMC approach to approximately search the posterior mean or MAP estimator, and secondly use the early stop criterion for Kalman-like methods to approximately find the posterior covariance operator. The combination of Kalman-like methods and MCMC methods is possible and meaningful.

Bibliography

[1] A. Beskos, A. Jasra, E. A. Muzaffer and A. M. Stuart, *Sequential Monte Carlo Methods for Bayesian Elliptic Inverse Problems*, 2015, Stat Comput.

[2] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM (1996), ISBN 978-0-89871-360-2.

[3] A. Doucet, S. Godsill and C. Andrieu, *On Sequential Monte Carlo Sampling Methods for Bayesian Filtering*, Statistics and Computing (2000) 10, 197208.

[4] A. Doicu, T. Trautmann and F. Schreier, *Numerical Regularization for Atmospheric Inverse Problems*, Springer-Verlag, Berlin Heidelberg, 2010.

[5] A. E. Gelfand, and A. F. M. Smith, *Sampling-based Approaches to Calculating Marginal Densities*, 1990, J. Amer. Statist. Assoc. 85 398409.

[6] A. Gelman and X. Meng, *Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path Sampling*, Statistical Science, Vol. 13, No. 2. (May, 1998), pp. 163-185.

[7] A. Kemna, J. Vanderborght, B. Kulessa and H. Vereecken, *Imaging and Characterisation of Subsurface Solute Transport Using Electrical Resistivity Tomography (ERT) and Equivalent Transport Models*, Journal of Hydrology 267 (2002) 125146.

[8] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numerica (2010), pp. 451559.

[9] A. M. Stuart, P. Wiberg and J. Voss, *Conditional Path Sampling of SDEs and The Langevin MCMC Method*, Communications in Mathematical Sciences 2(4) (Dec 2004) 685-697.

[10] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-posed Problems*, 1977, Winston and Sons, Washington DC.

[11] A. Sommerfeld, *Partial Differential Equation in Physics*, Lectures on Theoretical Physics - Pure and Applied Mathematics, New York: Academic Press, 1949.

[12] B. Borden, *Mathematical Problems in Radar Inverse Scattering*, Inverse Problems 18 (2002) R1R28.

[13] B. C. Barish, *LIGO and the Detection of Gravitational Waves*, Physics Today 52, 10, 44 (1999).

[14] B. L. Ellerbroek and C. R. Vogel, *Inverse Problems in Astronomical Adaptive Optics*, Inverse Problems 25 (2009) 063001 (37pp).

[15] B. THIDÉ, *Electromagnetic Potentials* In *Electromagnetic Field Theory*, Upsilon Books, Communa AB, Uppsala, Sweden, 2017.

[16] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, 2006, the MIT Press.

[17] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, 1949, Univ of Illinois Press.

[18] C. L. Epstein, *The Mathematics of Medical Imaging*, 2001, ISBN: 978-0-898716-42-9.

[19] C. Robert and G. Casella, *A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data*, Statistical Science 2011, Vol. 26, No. 1, 102115.

[20] C. Schillings and A.M. Stuart, *Analysis of The Ensemble Kalman Filter For Inverse Problems*, SIAM J Numerical Analysis 55(3) (2017), 1264-1290.

[21] C. Steger, M. Ulrich and C. Wiedemann, *Machine Vision Algorithms and Applications*, John Wiley & Sons, 2nd Edition, 2017.

[22] C. T. Kelley, *Iterative Methods for Optimization*, SIAM Frontiers in Applied Mathematics, no 18, 1999, ISBN 0-89871-433-8.

[23] D. Colton, *Inverse Acoustic and Electromagnetic Scattering Theory*, Inverse Problems, MSRI Publications, Volume 47, 2003.

[24] D. Gamerman and H.F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition*, 2006, Chapman and Hall/CRC. ISBN 9781584885870

[25] D. Jurafsky (Stanford University) and J. H. Martin (University of Colorado at Boulder), *Speech and Language Processing*, 3rd Edition draft, 2017.

[26] D. Kraaijpoel, *Seismic Ray Fields and Ray Field Maps: Theory and Algorithms*, Doctoral thesis, University Utrecht, 2003.

[27] D. Nicholson, *Inverse Problems in Gravitational Wave Astronomy*, Inverse Problems 11 (1995) 677-686.

[28] D. W. Marquardt, *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*, 1963, Journal of the Society for Industrial and Applied Mathematics, 11(2), 431441.

[29] F. Riesz, and B. Szőkefalvi-Nagy, *Functional Analysis (Dover ed)*, New York: Dover Publications (1990, first published in 1955).

[30] G. A. Einicke, *Smoothing, Filtering and Prediction: Estimating the Past, Present and Future (2nd ed.)*, Amazon Prime Publishing. ISBN 978-0-6485115-0-2.

[31] G. Boverman, B. S. Kim, D. Isaacson and J. C. Newell, *The Complete Electrode Model For Imaging and Electrode Contact Compensation in Electrical Impedance Tomography*, 2007, 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, 2007, pp. 3462-3465.

[32] G. D. Prato, and J. Zabczyk, *Stochastic Equations in Infinite Dimensions (Encyclopedia of Mathematics and its Applications)*, Cambridge: Cambridge University Press (2014, first published in 1992).

[33] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, 2009, Springer.

[34] G. Fourestey and M. Moubachir, *Solving Inverse Problems Involving the Navier-Stokes Equations Discretized by a LagrangeGalerkin Method*, Comput. Methods Appl. Mech. Engrg. 194 (2005) 877906.

[35] G. N. Milstein and M. V. Tretyakov, *The Simplest Random Walks for The Dirichlet Problem*, Theory of Probability & Its Applications, 2003, Vol. 47, No. 1 : pp. 53-68.

[36] G. N. Milstein and M. V. Tretyakov, *Practical Variance Reduction via Regression for Simulating Diffusions*, SIAM J. Numer. Anal., 47(2), 887910. (24 pages), 2009.

[37] G. N. Milstein and M. V. Tretyakov, *Solving the Dirichlet problem for NavierStokes equations by probabilistic approach*, BIT Numer Math (2012) 52: 141153.

[38] G. Welch and G. Bishop, *An Introduction to the Kalman Filter*, 1995, Technical Report.

[39] G. Potvin, *General Rytov Approximation*, Journal of the Optical Society of America A, Vol. 32, No. 10, 2015.

[40] H. C. Torrey, *Bloch Equations with Diffusion Terms*, Physical Review Volume 104, Number 3 November 1, 1956.

[41] H. W Engl, C. Flamm, P. Kügler, J. Lu, S. Müller and P. Schuster, *Inverse Problems in Systems Biology*, Inverse Problems 25 (2009) 123014 (51pp).

[42] H. W. Engl, K. Kunisch and A. Neubauer, *Convergence Rates for Tikhonov Regularisation of Non-linear Ill-posed Problems*, 1989 Inverse Problems 5 523.

[43] I. G. Currie, *Fundamental Mechanics of Fluids*, Marcel Dekker, Inc., New York, 3rd edition, 2003.

[44] J. Bell, *Trace Class Operators and Hilbert-Schmidt Operators*, 2016, *http://individual.utoronto.ca/jordanbell/notes/traceclass.pdf*

[45] J. Deteix, A. Jendoubi and D. Yakoubi, *A Coupled Prediction Scheme for Solving the Navier–Stokes and Convection-Diffusion Equations*, SIAM J. Numer. Anal., 52(5), (2014) 24152439.

[46] J. G. Hernández, *Representing Functional Data in Reproducing Kernel Hilbert Spaces with Applications to Clustering, Classication and Time Series Problems*, Universidad Carlos III de Madrid, Department of Statistics, (2010), *https://core.ac.uk/download/pdf/30042929.pdf*

[47] J. Hadamard, *Sur les Problèmes Aux Dérivées Partielles Et Leur Signification Physique*, pp. 4952, Princeton University Bulletin.

[48] J. Karwowski, *Inverse Problems in Quantum Chemistry*, International Journal of Quantum Chemistry, Vol 109, 24562463 (2009).

[49] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, 2005, Vol. 160 of Applied Mathematical Sciences, Springer.

[50] J. Mercer, *Functions of Positive and Negative type, and Their Connection The Theory of Integral Equations*, The Royal Society (1909). *https://doi.org/10.1098/rsta.1909.0016*

[51] J. Trampert, *Global Seismic Tomography: the Inverse Problem and Beyond*, Inverse Problems 14 (1998) 371385.

[52] K. G. van der Zee, E. H. van Brummelen, I. Akkerman and R. de Borst, *Goal-oriented Error Estimation and Adaptivity for Fluidstructure Interaction Using Exact Linearized Adjoints*, Comput. Methods Appl. Mech. Engrg. 200 (2011) 27382757.

[53] K. H. Matlack, J.-Y. Kim, L. J. Jacobs and J. Qu, *Review of Second Harmonic Generation Measurement Techniques for Material State Determination in Metals*, J Nondestruct Eval (2015).

[54] K. J. H. Law, H. Tembine, and R. Tempone, *Deterministic Mean-Field Ensemble Kalman Filtering*, SIAM J. Sci. Comput., 38(3), A1251A1279. (29 pages), 2016.

[55] K. K. Kwong, J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, B. P. Poncelet, D. N. Kennedy, B. E. Hoppel, M. S. Cohen, R. Turner, H.-M. Cheng, T. J. Brady and B. R. Rosen, *Dynamic Magnetic Resonance Imaging of Human Brain*

Activity during Primary Sensory Stimulation, Proc. Nadl. Acad. Sci. USA, Vol. 89, pp. 5675-5679, June 1992, Neurobiology.

[56] K. Shakenov, *Solution of Parametric Inverse Problem of Atmospheric Optics by Monte Carlo Methods*, TWMS Jour. Pure Appl. Math. V.3, N.2, 2012, pp.220-230.

[57] L. Borcea, *Electrical Impedance Tomography*, Inverse Problems 18 (2002) R99R136.

[58] L. Gilles, C. Vogel and J. Bardsley, *Computational Methods for a Large-scale Inverse Problem Arising in Atmospheric Optics*, Inverse Problems 18 (2002) 237252.

[59] L. Kish, *Survey Sampling*, 1965, New York: Wiley.

[60] L. Landweber, *An Iteration Formula for Fredholm Integral Equations of the First Kind*, Vol. 73, No. 3 (Jul., 1951), pp. 615-624 (10 pages), American Journal of Mathematics.

[61] M. A. Iglesias, *Iterative Regularization for Ensemble Data Assimilation in Reservoir Models*, Comput Geosci (2015) 19:177212.

[62] M. A. Iglesias and C. Dawson, *An Iterative Representer-based Scheme for Data Inversion in Reservoir Modeling*, Inverse Problems 25 (2009) 035006 (34pp).

[63] M. A. Iglesias and C. Dawson, *The Regularizing Levenberg-Marquardt Scheme for History Matching of Petroleum Reservoirs*, Comput Geosci (2013) 17:10331053.

[64] M. A. Iglesias and D. McLaughlin, *Level-set Techniques for Facies Identification in Reservoir Modeling*, Inverse Problems 27 (2011) 035008 (36pp).

[65] M. A. Iglesias, K. J. H. Law and A. M. Stuart, *Ensemble Kalman Methods for Inverse Problems*, Inverse Problems, 29(2013) 045001.

[66] M. A. Iglesias, K. J. H. Law and A. M. Stuart, *Evaluation of Gaussian Approximations for Data Assimilation in Reservoir Models*, Comput Geosci (2013) 17:851885.

[67] M. Benning and M. Burger, *Modern Regularization Methods for Inverse Problems*, 2018, to appear in Acta Numerica.

[68] M. Bizzarri, *Systems Biology*, Springer Science+Business Media LLC, 2018.

[69] M. Cheney, D. Isaacson and J. C. Newell, *Electrical Impedance Tomography*, SIAM REVIEW (1999) Vol. 41, No. 1, pp. 85101.

[70] M. Dashti and A. M Stuart, *The Bayesian Approach to Inverse Problems*, Notes, 2015.

[71] M. Hanke, *A Regularizing Levenberg–Marquardt Scheme, with Applications to Inverse Groundwater Filtration Problems*, 1997, Inverse Problems 13 79.

[72] M. J. Brammer, E. T. Bullmore, A. Simmons, S. C. R. Williams, P. M. Grasby, R. J. Howard, P. W. R. Woodruff and S. Rabe-Hesketh, *Generic Brain Activation Mapping in Functional Magnetic Resonance Imaging: a Nonparametric Approach*, Magnetic Resonance Imaging, Vol. 15, No.7, pp. 763-770, 1997.

[73] M. K. Transtrum and J. P. Sethn, *Improvements to The Levenberg–Marquardt Algorithm for Nonlinear Least-Squares Minimization*, 2012, *https://arxiv.org/abs/1201.5885*

[74] M. M. Dunlop, M. A. Iglesias and A. M. Stuart, *Hierarchical Bayesian Level Set Inversion*, Statistics and Computing (2016).

[75] M. M. Dunlop, M. A. Girolami, A. M. Stuart, and A. L. Teckentrup, *How Deep Are Deep Gaussian Processes?*, Journal of Machine Learning Research 19 (2018) 1-4.

[76] M. van Gerven and S. Bohte, *Artificial Neural Networks as Models of Neural Information Processing*, Frontiers in Computational Neuroscience, 2018.

[77] N. B. Kovachki and A. M. Stuart, *Ensemble Kalman Inversion: A Derivative-Free Technique For Machine Learning Tasks*, Inverse Problems, Volume 35, Number 9, 2019.

[78] N. K. Chada, M. A. Iglesias, L. Roininen and A. M. Stuart, *Parameterizations for Ensemble Kalman Inversion*, Inverse Problems, 34 (2018).

[79] N. S. Pillai, A. M. Stuart and A. H. Thiery, *Optimal Scaling and Diffusion Limits for The Langevin Algorithm in High Dimensions*, Annals of Applied Probability 22 (2012) 2320-2356.

[80] O. Scherzer, H. W. Engl, and K. Kunisch, *Optimal A Posteriori Parameter Choice for Tikhonov Regularization for Solving Nonlinear Ill-posed problems*, 1993, SIAM J. NUMER. ANAL. Vol. 30, No. 6, pp. 1796-1838.

[81] P. C. Hansen, *The Truncated SVD as A Method for Regularization*, December 1987, Volume 27, Issue 4, pp 534553

[82] P. M. Shearer, *The Seismic Wave Equation* In *Introduction to Seismology* (pp. 39-64), Cambridge University Press, Cambridge, 2nd edition, 2009.

[83] P. Walters, *An Introduction to Ergodic Theory*, 1982, Springer.

[84] R. E. Hansen, *Introduction to Sonar*, Course materiel to INF-GEO4310, University of Oslo, (Dated: October 7, 2009).

[85] R. Ghanem, D. Higdon, and H. Owhadi, *Handbook of Uncertainty Quantification*, 2017, Springer.

[86] R. Hanke, T. Fuchs and N. Uhlmann, *X-ray Based Methods for Non-destructive Testing and Material Characterization*, Nuclear Instruments and Methods in Physics Research A 591 (2008) 1418.

[87] R. Kibble, *Introduction to Natural Language Processing*, Goldsmiths, University of London, 2013.

[88] R. M. Neal, *Annealed Importance Sampling*, 1998, Technical Report No. 9805, Department of Statistics, University of Toronto.

[89] R. Nakamura, I. Nishimura and T. Watanabe, *Determination of Hypocenter Using Seismic Intensity Distributions and 3-D Attenuation Structure*, Elsevier Science Ltd., Paper No. 737, 1996.

[90] R. S. C. Cobbold, *Foundations of Biomedical Ultrasound*, Oxford University Press, 2006.

[91] S. Haykin, *Kalman Filtering and Neural Networks*, 2001, John Wiley & Sons, Inc. ISBN 9780471369981

[92] S. H. Schot, *Eighty Years of Sommerfelds Radiation Condition*, Historia Mathematica 19 (1992) 385-401.

[93] S. Kullback and R. A. Leibler, *On Information and Sufficiency*, 1951, Annals of Mathematical Statistics. 22 (1): 7986.

[94] S. L. Cotter, G. O. Roberts, A. M. Stuart and D. White, *MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster*, Statistical Science 2013, Vol. 28, No. 3, 424446.

[95] S. Ogawa, T. M. Lee, A. R. Kay and D. W. Tank, *Brain Magnetic Resonance Imaging with Contrast Dependent on Blood Oxygenation*, Proc. NatI. Acad. Sci. USA, Vol. 87, pp. 9868-9872, December 1990, Biophysics.

[96] T. Bonesky, *Morozovs Discrepancy Principle and Tikhonov-type Functionals*, Inverse Problems 25 (2009).

[97] T. I. Seidman and C. R. Vogel, *Well Posedness and Convergence of Some Regularisation Methods for Non-linear Ill Posed Problems*, 1989, Inverse Problems 5 227.

[98] V. A. Morozov, *Methods for Solving Incorrectly Posed Problems*, Spring (1984).

[99] V. Mlynárik, *Introduction to Nuclear Magnetic Resonance*, Analytical Biochemistry 529 (2017) 4-9.

[100] W. B. Beydoun and A. Tarantola, *First Born and Rytov Approximations: Modeling and Inversion Conditions in a Canonical Example*, The Journal of the Acoustical Society of America 83, 1045 (1988).

[101] W. B. Muniz, Haroldo F. de Campos Velho and F. M. Ramos, *A Comparison of Some Inverse Methods for Estimating the Initial Condition of the Heat Equation*, Journal of Computational and Applied Mathematics 103 (1999) 145-163.

[102] W. Daily, A. Ramirez, D. Labrecque and J. Nitao, *Electrical Resistivity Tomography of Vadose Water Movement*, Water Resources Research (1992) Vol. 28, NO. 5, Pages 1429-1442.

[103] W. Hastings, *Monte Carlo Sampling Methods Using Markov Chains and Their Application*, (1970), Biometrika 57 97109.

[104] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *"Section 10.12. Simulated Annealing Methods". Numerical Recipes: The Art of Scientific Computing (3rd ed.)*, 2007, New York: Cambridge University Press.

[105] W. Munk, P. W. and C. Wunsch, *Ocean Acoustic Tomography*, Cambridge University Press, Cambridge, 1995.

[106] W. Rudin, *Functional analysis (2nd ed)*, McGraw-Hill, New York (1991).

[107] W. W. Symes, *The Seismic Reflection Inverse Problem*, Inverse Problems 25 (2009) 123008 (39pp).

[108] website of Electrical Impedance Tomography and Diffuse Optical Tomography Reconstruction Software (EIDORS), *http://eidors3d.sourceforge.net*

[109] website of European Centre for Medium-Range Weather Forecasts (ECMWF), *https://www.ecmwf.int/*