CHAPTER 1

Basics of seismic inversion

Seismic inversion attempts to extract spatially variable physical parameters from measured seismic data. These physical parameters may be representative of the Earth's subsurface media, and have physical and geological meanings, and thus seismic inversion is a quantitative interpretation of seismic measurement. The inversion procedure is generally nonlinear, as the entire inversion engine to solve the inverse problem, at least partly, depends upon the solution. In practice, the inverse problem is often linearised, and the final nonlinear solution is obtained through the iterative application of linearised solvers. Therefore, this book will focus on the linear inverse problem.

1.1 The linear inverse problem

Linear seismic inversion may include at least three basic steps such as the following:

1) Setting up an objective function, which describes how well a model estimate represents the seismic observation and meets our human expectation;
2) Optimising the objective function based on a minimal variation principle, which leads to a linear system of equations, if the objective function is defined as a quadratic function;
3) Solving this linear system, to obtain a quantitative solution.

Data fitting is a principal part of the objective function. Seismic inversion

Seismic Inversion: Theory and Applications, First Edition. Yanghua Wang.

uses forward modelling to generate synthetic seismic data that will match the observed seismic data. Forward modelling can be presented in a linear form:

$$\mathbf{Gm} = \mathbf{d}, \tag{1.1}$$

where $\mathbf{G}$ is a geophysical operator (a matrix), $\mathbf{m}$ is the 'model' vector, and $\mathbf{d}$ is the 'data' vector. Both vectors $\mathbf{m}$ and $\mathbf{d}$ are defined in the Hilbert space in which the structure (the length and angle) of an inner product of vectors can be measured. Row vectors and column vectors of matrix $\mathbf{G}$ are also defined in the Hilbert space.

For example, for a two-dimensional (2-D) velocity model defined in the $x-z$ domain, we cannot straightforwardly include a fracture or a fault in the model parameterisation. Instead, we shall use an equivalent velocity model, which takes into account the effect of the fracture or the fault, so that the model vector $\mathbf{m}$ is in the Hilbert space and can be involved in any inner product in seismic inversion.

Then, we can define the data-fitting objective function as

$$\phi(\mathbf{m}) = \| \tilde{\mathbf{d}} - \mathbf{Gm} \|^2, \tag{1.2}$$

where $\tilde{\mathbf{d}}$ is the observed data vector, and $\| \mathbf{r} \|^2 = \mathbf{r}^{\mathrm{T}}\mathbf{r} = (\mathbf{r}, \mathbf{r}) = \sum_i r_i^2$ is the inner product of a single vector $\mathbf{r}$. The symbol $\| \cdot \|$ represents the L_2 norm of a vector.

The optimisation working on the objective function is not necessarily a minimisation only. Depending on the set-up of the objective function, it can also be maximisation. For example, minimising the data misfit is equivalent to the maximisation of the probability. Minimal variation $\partial\phi/\partial\mathbf{m} = \mathbf{0}$, where $\mathbf{0}$ is a null vector, can find either the minimal or maximal extremer of an objective function.

In the objective function of Equation 1.2, $\| \tilde{\mathbf{d}} - \mathbf{Gm} \|^2$ is the energy of data residuals. The least-squares solution using a minimal variation principle, that is, setting $\partial\phi / \partial\mathbf{m} = \mathbf{0}$, leads to the following linear system:

$$\mathbf{G}^{\mathrm{T}}\mathbf{Gm} = \mathbf{G}^{\mathrm{T}}\tilde{\mathbf{d}}, \tag{1.3}$$

where $\mathbf{G}^{\mathrm{T}}$ is the transpose of the rectangular matrix $\mathbf{G}$, and $\mathbf{G}^{\mathrm{T}}\mathbf{G}$ is a square matrix. A simplified version of Equation 1.3 is

$$\mathbf{Gm} = \tilde{\mathbf{d}}. \tag{1.4}$$

Generally speaking, the inverse problem corresponds to calculating the inverse of the rectangular matrix $\mathbf{G}$ in Equation 1.4. However, this matrix inverse cannot be calculated directly. The problem ultimately corresponds to the least-squares solution of Equation 1.3, which leads to calculating the inverse of the square matrix $\mathbf{G}^{\mathrm{T}}\mathbf{G}$.

In practice, the matrix inverse does not always exist. It means that either the operator $\mathbf{G}$ or $\mathbf{G}^{\mathrm{T}}\mathbf{G}$ are singular. Any modification to the operator is called regularisation, which makes the inverse mapping, from the data space to the model space, happen in a stable and unique way.

For solving the linear system with a large-sized matrix, in order to avoid the direct calculation of the matrix inverse, an iterative method can be used. Each iteration can also be treated as a linear inverse problem, in which the objective function is defined by an error function, and the solution estimate is updated along the (negative) gradient direction. This is called a gradient-based method.

1.2 Data, model and mapping

Let us compare two linear systems presented in Equations 1.1 and 1.4, respectively. Equation 1.1 is a direct problem:

$$(\text{model } M) \Rightarrow [\text{direct mapping } G] \Rightarrow \text{data } D\,. \tag{1.5}$$

Given an Earth model M, defined as a set of Earth parameters, and a mapping operator G, the object is to find a set of data D containing all possible measurements in the data space.

Even for this direct problem, M and G may not be unique for a practical problem. For instance, the acoustic, the elastic or the viscoelastic wave equations all can be used for the problem of generating synthetic reflection seismograms. The selection is made based on the practical requirements and *a priori* information.

For a correctly defined physical problem, the direct problem is usually well-posed, which means this mathematical model that describes physical phenomena does have a unique solution, and the solution depends continuously on the model. The continuous dependency means that a small variation $\Delta\mathbf{m}$ in the model space M causes a small perturbation $\Delta\mathbf{d}$ in the data space D.

The inverse problem in Equation 1.4 states that, given a data set D and the mapping operator G, to find a model M:

$$\text{model } M \Leftarrow [\text{direct mapping } G]^{-1} \Leftarrow (\text{data } D)\,. \tag{1.6}$$

The inversion theory aims to guide the study of inverse problems in order to extract all the information contained in data, while controlling artefacts introduced through the inversion. There are two main kinds of study, as follows:

The first kind is the exact study with perfect data, that is, the study of the existence and uniqueness of solutions and constructing the exact inverse mapping operator. This is a beautiful exercise for classical mathematical analysis, by knowing the statement that the inverse problem is well-posed and readily solved if G is bicontinuous and bijection occurred between spaces M and D. Bicontinuous means that a continuous function also has a continuous inverse function. Bijection means that, for every $\mathbf{d}$ in D, there is exactly one $\mathbf{m}$ in M such that $G(\mathbf{m}) = \mathbf{d}$, and vice versa. However, this inverse problem is not of much interest to applied scientists and engineers.

The other kind is the study of the definitions of generalised solutions and methods for inexact and incomplete data. For the study of this kind of inverse problems, which geophysicists are interested in, the following three remarks should be noted:

1) The inverse mapping operator G^{-1} may not exist. In this case, one should check the definition of the Earth parameters, considering equivalent mapping and *a priori* information.

2) The solution of an inverse problem is often not unique. Hence, the term *solution estimate* instead of solution should be used for the generalised inversion.

3) Different from a direct problem, an inverse problem is usually ill-posed, that is, a small variation of the data leads to uncontrollable perturbation in the solution (Hadamard, 1902).

Mathematically non-continuous mapping operator G^{-1} causes the problem to be ill-posed. For the linear case presented in the previous section, $G(\mathbf{m}) = \mathbf{Gm}$, the study will be on singularity and condition number of $\mathbf{G}$. Non-singularity suggests the existence of the matrix inverse, $\mathbf{G}^{-1}$, and a low condition number indicates that $\mathbf{Gm} = \tilde{\mathbf{d}}$ is a well-posed problem.

1.3 General solutions

Let $\mathbf{d} = G(\mathbf{m})$ be a predicted data set, and $\tilde{\mathbf{d}}$ be an observed data set; a solution $\mathbf{m}$ may be estimated by minimising the distance between $\tilde{\mathbf{d}}$ and $\mathbf{d}$. This type of solution is called a quasi-solution for $\mathbf{m}$.

A quadratic distance is frequently employed in seismic inversion. The quadratic distance between two vectors is measured as

$$\text{dis}(\mathbf{r}_1, \mathbf{r}_2) = \| \mathbf{r}_1 - \mathbf{r}_2 \|, \tag{1.7}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are two vectors in the same space. The objective function of Equation 1.2 provides a quasi-solution, as it is related to errors in the observed data set, $\tilde{\mathbf{d}} - G(\mathbf{m}) \neq \mathbf{0}$.

As an observed data set is not only inexact, but incomplete as well, an approximate solution, beside the quasi-solution, is also needed. An approximate solution is obtained by minimising the combination of the data-fitting quality criterion and the model choice criterion. The objective function is defined as

$$\phi(\mathbf{m}) = \text{dis}_\text{D}(\tilde{\mathbf{d}}, G(\mathbf{m})) + \mu \text{dis}_\text{M}(\mathbf{m}, \mathbf{m}_\text{ref}), \tag{1.8}$$

where $\text{dis}_\text{D}(\tilde{\mathbf{d}}, G(\mathbf{m}))$ is the distance defined in the data space D, $\text{dis}_\text{M}(\mathbf{m}, \mathbf{m}_\text{ref})$ is the distance defined in the model space M, and μ is a trade-off parameter balancing the contribution of two criteria in the objective function.

In seismic inverse problems, $\mathbf{m}_\text{ref}$ can be an expected solution. For instance, in the linear case $G(\mathbf{m}) = \mathbf{G}\mathbf{m}$, an objective function can be defined as

$$\phi(\mathbf{m}) = \| \tilde{\mathbf{d}} - \mathbf{G}\mathbf{m} \|^2 + \mu \| \mathbf{m} - \mathbf{m}_\text{ref} \|^2. \tag{1.9}$$

The quadratic distance defined by the L_2 norm is a special case of distance measurement, although it is often used in seismic inversion. The distance can be measured in different ways, such as a weighted quadratic distance, frequently employed in seismic inversion.

The objective function in Equations 1.8 or 1.9 is a constrained inverse problem, in which $\text{dis}_\text{M}(\mathbf{m}, \mathbf{m}_\text{ref})$ is a typical model constraint. Different forms of model constraints can be used. Any constraints in the objective function are regularisation working on the geophysical mapping operator, as shown in the following section.

1.4 Regularisation

Regularisation means to suppress singularities that make the problems ill-posed and will cause difficulties in computation. The approximate solution mentioned above is just a practical way to consider the inexact and incomplete data. Regularisation considers the properties of the mapping operator G from the mathematical viewpoint: Whether the numerical instability comes from the singularity, and whether the singular operator can be modified to stabilise the computation.

The stability behaviour means that a small variation in data causes a small perturbation in the solution estimate, and thus depends on the property of the mapping operator. But how strongly is it dependent? Let $\boldsymbol{\varepsilon}$ be the vector of the data errors, and $\Delta\mathbf{m}$ the perturbation in the model solution caused by the errors. There are three types of dependences:

1) Linear: $\| \Delta\mathbf{m} \| = \alpha \| \boldsymbol{\varepsilon} \|$;

2) Power law: $\| \Delta\mathbf{m} \| \le A \| \boldsymbol{\varepsilon} \|^{\alpha}$, $0 < \alpha \le 1$ and constant A;

3) Logarithmic: $\| \Delta\mathbf{m} \| \propto \left(\ln \dfrac{1}{\| \boldsymbol{\varepsilon} \|} \right)^{-\alpha} = \left(-\ln \| \boldsymbol{\varepsilon} \| \right)^{-\alpha}$.

With a linear dependency, it is a well-posed problem. For logarithmic dependency, it is an ill-posed problem. In order to have a stable inversion, at least an operator of power-law dependency, with the exponent α less than 1, should be employed. Unfortunately, the inverse operators in geophysical problems are usually ill-posed with a logarithmic dependency, and thus need to be regularised.

Figure 1.1 displays the dependence of model perturbation $\| \Delta\mathbf{m} \|$ on the data error $\| \boldsymbol{\varepsilon} \|$, with the three relationships: linear (solid curves), $\| \Delta\mathbf{m} \| = \alpha \| \boldsymbol{\varepsilon} \|$; power law (dotted curves), $\| \Delta\mathbf{m} \| = \| \boldsymbol{\varepsilon} \|^{\alpha}$, $0 < \alpha \le 1$; and logarithmic (dashed curves), $\| \Delta\mathbf{m} \| = (-\ln \| \boldsymbol{\varepsilon} \|)^{-\alpha}$, in which $\| \boldsymbol{\varepsilon} \|$ can be treated as a pre-normalised data error (with the maximum probable error of 1). The three panels (left to right) are cases with $\alpha = 0.3,\ 0.6,\ 0.9$, respectively.

Related to this stability issue in the inverse problem, there is a property of the operator, called the condition number. It is defined by the maximum value of the ratio of the relative errors in the model solution to the relative error in the data. If the condition number is small then the error in $\mathbf{m}$ will not be much bigger than the error in $\tilde{\mathbf{d}}$. On the other hand, if the

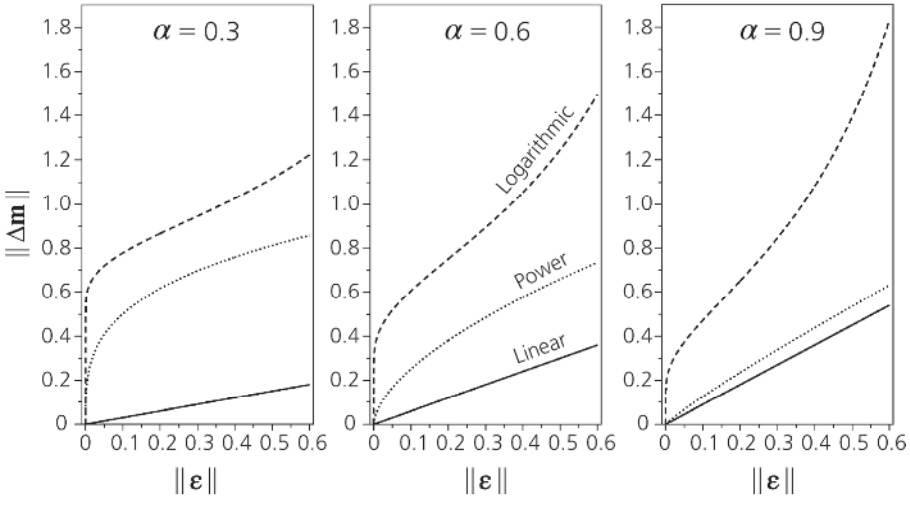

Figure 1.1 The dependence of model perturbation $\| \Delta \mathbf{m} \|$ on the data errors $\| \boldsymbol{\varepsilon} \|$. There are three types of dependence: linear (solid curves), power (dotted curves) and logarithmic (dashed curves). The three panels (left to right) are cases with $\alpha = 0.3,\ 0.6,\ 0.9,$ respectively.

condition number is large, even a small error in data may cause a large error in the model solution. A problem with a low condition number is said to be well-conditioned, whereas a problem with a high condition number is said to be ill-conditioned.

In order to stabilise the inverse problem, by reducing the condition number of the operator, regularisation can be realised as model constraints added to the objective function. It first defines a stabilising function $R(\mathbf{m})$, which satisfies $R(\mathbf{m}) \le E$ for any real number E, and then incorporates $R(\mathbf{m})$ into the objective function, as

$$\phi(\mathbf{m}) = Q(\mathbf{m}) + \mu R(\mathbf{m}), \tag{1.10}$$

where $Q(\mathbf{m})$ is the data fit quality criterion, and $R(\mathbf{m})$ stands for the model regularisation term.

To understand this stabilisation, let us see an example objective function,

$$\phi(\mathbf{m}) = \| \tilde{\mathbf{d}} - \mathbf{G}\mathbf{m} \|^2 + \mu \| \mathbf{m} \|^2 . \tag{1.11}$$

Compared to the objective function in Equation 1.9, $\mathbf{m}_{\text{ref}} = \mathbf{0}$ is set here. Minimisation by setting $\partial \phi / \partial \mathbf{m} = \mathbf{0}$ yields the following equation:

$$[\mathbf{G}^{\mathrm{T}}\mathbf{G} + \mu\mathbf{I}]\,\mathbf{m} = \mathbf{G}^{\mathrm{T}}\tilde{\mathbf{d}}\,. \tag{1.12}$$

If the matrix $\mathbf{G}^{\mathrm{T}}\mathbf{G}$ was singular, the modified operator $[\mathbf{G}^{\mathrm{T}}\mathbf{G} + \mu\mathbf{I}]$ is no longer singular, and the solution of Equation 1.12 exists. Hence, μ is also called the stabilisation factor. The solution estimate $\mathbf{m}$ is unique, as well as continuously dependent on the averaged data, $\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{d}}$. Therefore, constraining the objective function is in fact regularising the geophysical mapping operator, so as to stabilise the inverse problem.

Tikhonov regularisation (Tikhonov, 1935; Tikhonov and Arsenin, 1977; Tikhonov *et al.*, 1995) is expressed as

$$R(\mathbf{m}) = \int_{r_a}^{r_b} \left(\mu_1(r) \,\|\, \mathbf{m}(r) \,\|^2 + \mu_2(r) \left\| \frac{\partial \mathbf{m}(r)}{\partial r} \right\|^2 \right) \mathrm{d}r\,, \tag{1.13}$$

where r is the spatial position, and $\mu_1(r)$ and $\mu_2(r)$ are positive weighting functions, defined within the range $[r_a,\ r_b]$.

Moreover, regularisation can also be applied directly to the geophysical operation, for depressing any singularity. Let us see a simple example, differentiating a continuous function. Assume that $f(r)$ is a real continuous function, but its derivative might not exist. To subjugate this singularity, regularisation can be achieved by convolving $f(r)$ with a continuous and differentiable function $h(r)$,

$$\tilde{f}(r) = f(r) * h(r)\,. \tag{1.14}$$

This processed function $\tilde{f}(r)$ is differentiable without singularities.

In order to make $\tilde{f}(r)$ a good approximation for $f(r)$, the following conditions should be satisfied:

1) Finite range of h: $h(r) \equiv 0$ for r outside a small range;
2) Unimodular:

$$\int_{-\infty}^{\infty} h(r)\,\mathrm{d}r = 1\,; \tag{1.15}$$

3) Approximation:

$$\frac{1}{r_b - r_a} \int_{r_a}^{r_b} |\,\tilde{f}(r) - f(r)\,|\,\mathrm{d}r < \varepsilon\,. \tag{1.16}$$

The first condition means a localised regularisation, the second condition requires that this process does not change the power of the original function, and the last condition, of course, requires the approximation being sufficiently close to the original function. These three conditions are the basic requirement of a regularisation, if the regularisation is directly applied to the geophysical operator.

A demonstration is shown in Figure 1.2. The $f(r)$ function is defined as

$$f(r)=\begin{cases}2, & r<r_1,\\ \dfrac{r_1-r-2r_2}{r_1-r_2}, & r_1\le r\le r_2,\\ 1, & r>r_2.\end{cases} \tag{1.17}$$

This function is continuous, but not differentiable, since its first-order derivative has two singular points at r_1 and r_2. A filter is designed by a Gaussian function,

$$h(r)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{r^2}{2\sigma^2}\right), \tag{1.18}$$

where σ is the standard deviation. Convolution produces a smooth function $\tilde{f}(r)$.

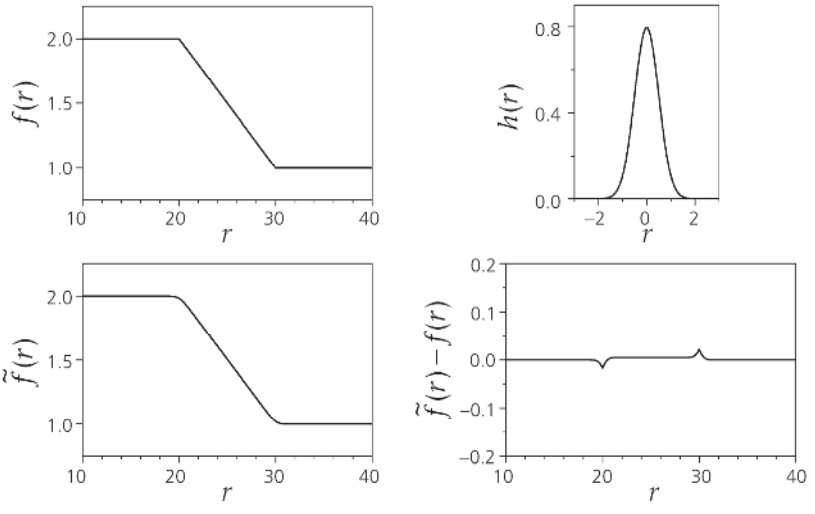

Figure 1.2 A function $f(r)$, that is not differentiable, convolved with a function $h(r)$ produces a differentiable function $\tilde{f}(r)$. The latter is differentiable without singularities, and the difference $|\tilde{f}(r)-f(r)|$ is sufficiently small.

We can verify the numerical example of Figure 1.2 against the three conditions:

1) Localisation of $h(r)$ depends upon the parameter σ. In this display, $\sigma = 0.5$.
2) The filter $h(r)$ is unimodular, because $\sum_i h_i \Delta r = 1$, where Δr is the sampling rate, and $h_i = h(i\Delta r)$.
3) The difference $|\tilde{f}(r) - f(r)|$ is sufficiently small, as the total difference $\varepsilon \le 0.00126$.

In this simple example, the operator G is the first-order differential,

$$G(r) = \frac{\mathrm{d}}{\mathrm{d}r}. \tag{1.19}$$

After regularisation, the operator becomes

$$\tilde{G}(r) = \frac{\mathrm{d}}{\mathrm{d}r} h(r) *. \tag{1.20}$$

These two operators can be understood in the Fourier transform domain as

$$G(k) = -\mathrm{i}k, \qquad \tilde{G}(k) = -\mathrm{i}kH(k), \tag{1.21}$$

where k is the wavenumber, $\mathrm{i} = \sqrt{-1}$ is the imaginary symbol, $G(k)$ and $\tilde{G}(k)$ are the Fourier transforms of $G(r)$ and $\tilde{G}(r)$, respectively, and $H(k)$ is the Fourier transform of $h(r)$.

The essence of any inverse problems is regularisation.