Chapter 1

Back-projection methods

Consider a linear set of equations relating observed data to a model:

$$\mathbf{d} = \mathbf{G}\mathbf{m} \tag{1.1}$$

using the conventional notation of $\mathbf{d}$ for the data vector, $\mathbf{m}$ for the model vector, and $\mathbf{G}$ for the linear operator that predicts the data from the model. Our goal in geophysical inverse problems is to estimate $\mathbf{m}$ from the observations, $\mathbf{d}$. Assuming there are more data points than model points, the standard way to solve this problem is to define a residual vector, $\mathbf{r} = \mathbf{d} - \mathbf{G}\mathbf{m}$, and find the $\mathbf{m}$ that minimizes $\mathbf{r} \cdot \mathbf{r}$. This is the least squares solutions and it can be shown that

$$\mathbf{m} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d}. \tag{1.2}$$

However often $\mathbf{G}^T\mathbf{G}$ is singular or ill-conditioned, or it may simply be too large to invert. What can be done is these cases? The simplest and crudest way to proceed is to make the approximation

$$(\mathbf{G}^T\mathbf{G})^{-1} \approx \mathbf{I} \tag{1.3}$$

in which case we can estimate the model as

$$\mathbf{m} \approx \mathbf{G}^T\mathbf{d}. \tag{1.4}$$

The transposed matrix $\mathbf{G}^T$ is the *adjoint* or back-projection operator. Each model point is constructed as the weighted sum of the data points that it affects. Can such a crude approximation be of any use? It's certainly easy to think of examples where (1.3) is completely invalid. However, in real geophysical problems it's

surprising how often this method works, particularly if a scaling factor is allowed to bring the data and model-predicted data into better agreement (i.e., assuming $(\mathbf{G}^T\mathbf{G})^{-1} \approx \lambda\mathbf{I}$, where λ is a constant). Indeed, it is sometimes observed that the adjoint works better than the formal inverse because it is more tolerant of imperfections in the data. Jon Claerbout discusses this in a wonderful set of notes (e.g., http://sepwww.stanford.edu/sep/prof/gee/ajt/paper_html/node1.html).

In seismology our data are typically a set of seismograms. In source inversions, we normally assume that the Earth's velocity structure is known and we solve for the locations and times of seismic wave radiators (e.g., solving for a slip model). In reflection seismology, we normally assume that the location and time of the source is known and we solve for the location of the reflector(s) that cause the observed arrivals. In each case, the model estimate at each model point is obtained by finding the times in the seismograms at which changes in the model will affect the seismogram. The model estimate from back-projection is obtained by simply summing or *stacking* the seismogram values at these points. The main thing to compute is the travel time between the model points and each recording station. These give the time shifts necessary to find the times in each seismogram that are sensitive to the model perturbations.

One way of thinking about this is that we have the computer perform a series of hypothesis tests over a time-space model grid. Is there a seismic radiator at this space-time point? If there is, we would expect it to show up in seismograms at these times. If we sum over the seismogram values at these times, we should get a large amplitude. Of course, it is possible that inference from radiation at other model points will cause us to have a biased estimate. But on average, we hope (expect) these other contributions to cancel out. This is the forward-time way of thinking about the problem.

But we could also think about this in reverse time. In this case we start with the seismograms and project their values backward in time through the model grid. As we do this, we accumulate the values in the model grid points. The model points that are likely sources will experience constructive interference as the time-reversed wavefields focus to these points. This is why this process is sometimes called back-

projection or reverse time migration. But the result is exactly the same as the forward modeling approach described in the previous paragraph.

Left unstated in this discussion is how the amplitudes in the seismograms should be scaled. If one wants to recover true model amplitudes, then geometrical spreading and other factors should be taken into account. Often, however, the goal is simply an image of the model and the absolute amplitude is not that important. For example, in reflection seismology automatic gain control is often used to equalize the contributions from different records and true amplitude information is lost. These amplitude normalization methods can make back-projection more robust with respect to noisy data or uncertainties in the velocity model.

1.1 Migration in reflection seismology

In reflection seismology, complicated structures will produce scattered and diffracted arrivals that cannot be modeled by simple plane-wave reflections, and accurate interpretation of data from such features requires a theory that takes these arrivals into account. Most of the analysis techniques developed for this purpose are based on the idea that velocity perturbations in the medium can be thought of as generating secondary seismic sources in response to the incident wavefield, and the reflected wavefield can be modeled as a sum of these secondary wavelets.

1.1.1 Huygens' principle

Huygens' principle, first described by Christiaan Huygens (c. 1678), is most commonly mentioned in the context of light waves and optical ray theory, but it is applicable to any wave propagation problem. If we consider a plane wavefront traveling in a homogeneous medium, we can see how the wavefront can be thought to propagate through the constructive interference of secondary wavelets (Fig. 1.1). This simple idea provides, at least in a qualitative sense, an explanation for the behavior of waves when they pass through a narrow aperture.

The bending of the ray paths at the edges of the gap is termed *diffraction*. The degree to which the waves diffract into the "shadow" of the obstacle depends upon the wavelength of the waves in relation to the size of the opening. At relatively

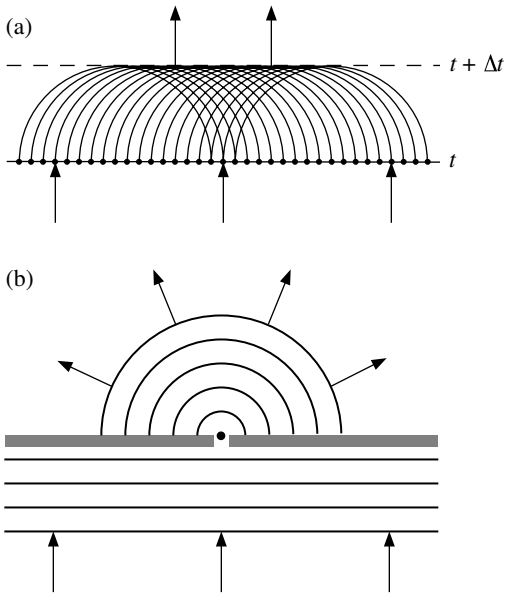

Figure 1.1: Illustrations of Huygens' principle. (a) A plane wave at time $t + \Delta t$ can be modeled as the coherent sum of the spherical wavefronts emitted by point sources on the wavefront at time t. (b) A small opening in a barrier to incident waves will produce a diffracted wavefront if the opening is small compared to the wavelength.

long wavelengths (e.g., ocean waves striking a hole in a jetty), the transmitted waves will spread out almost uniformly over 180°. However, at short wavelengths the diffraction from the edges of the slot will produce a much smaller spreading in the wavefield. For light waves, very narrow slits are required to produce noticeable diffraction. These properties can be modeled using Huygens' principle by computing the effects of constructive and destructive interference at different wavelengths.

1.1.2 Diffraction hyperbolas

We can apply Huygens' principle to reflection seismology by imagining that each point on a reflector generates a secondary source in response to the incident wavefield. This is sometimes called the "exploding reflector" model. Consider a single point scatterer in a zero-offset section (Fig. 1.2). The minimum travel time is given by

$$t_0 = \frac{2h}{v}, \tag{1.5}$$

where h is the depth of the scatterer and v is the velocity (assumed constant in this case). More generally, the travel time as a function of horizontal distance, x, is

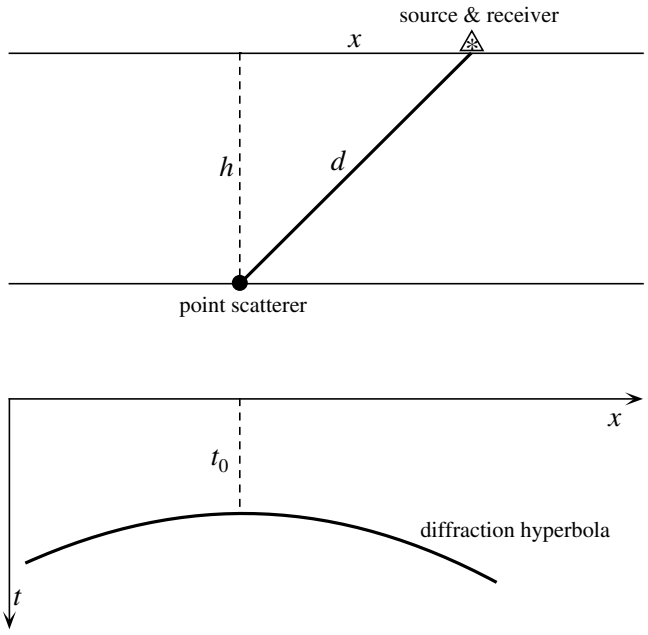

Figure 1.2: A point scatterer will produce a curved "reflector" in a zero-offset section.

given by

$$t(x) = \frac{2\sqrt{x^2 + h^2}}{v}. \tag{1.6}$$

Squaring and rearranging, this can be expressed as

$$\frac{v^2 t^2}{4h^2} - \frac{x^2}{h^2} = 1 \tag{1.7}$$

or

$$\frac{t^2}{t_0^2} - \frac{4x^2}{v^2 t_0^2} = 1 \tag{1.8}$$

after substituting $4h^2 = v^2 t_0^2$ from (1.5). The travel time curve for the scattered arrival has the form of a hyperbola with the apex directly above the scattering point. This equation describes travel time as a function of distance away from a point scatterer at depth for zero-offset data (the source and receiver are coincident).

1.1.3 Migration methods

Consider a horizontal reflector that is made up of a series of point scatterers, each of which generates a diffraction hyperbola in a zero-offset profile (Fig. 1.3). Following Huygens' principle, these hyperbolas sum coherently only at the time of the main reflection; the later contributions cancel out. However, if the reflector vanishes at

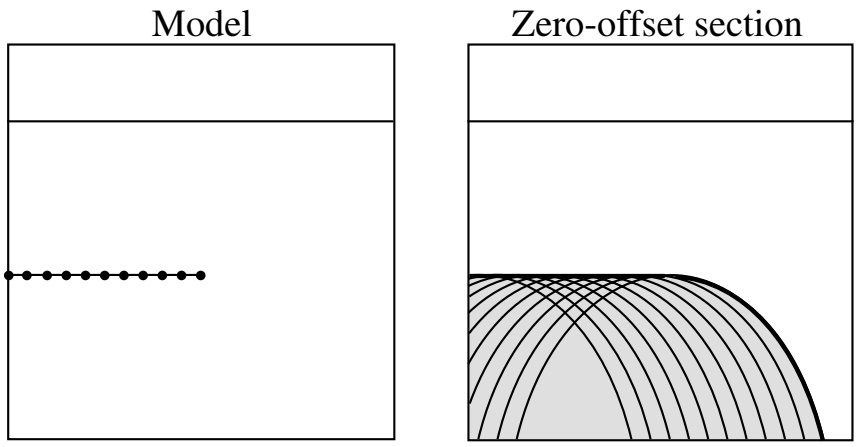

Figure 1.3: The endpoint of a horizontal reflector will produce a diffracted arrival in a zero-offset section. The reflector itself can be modeled as the coherent sum of the diffraction hyperbola from individual point scatterers. The diffracted phase, shown as the curved heavy line, occurs at the boundary of the region of scattered arrivals.

some point, then there will be a diffracted arrival from the endpoint that will show up in the zero-offset data. This creates an artifact in the section that might be falsely interpreted as a dipping, curved reflector.

Techniques for removing these artifacts from reflection data are termed *migration* and a number of different approaches have been developed. The simplest of these methods is termed *diffraction summation migration* and involves assuming that each point in a zero-offset section is the apex of a hypothetical diffraction hyperbola. The value of the time series at that point is replaced by the average of the data from adjacent traces taken at points along the hyperbola. In this way, diffraction artifacts are "collapsed" into their true locations in the migrated section. In many cases migration can produce a dramatic improvement in image quality (e.g., Fig. 1.4).

A proper implementation of diffraction summation migration requires wave propagation theory that goes beyond the simple ideas of Huygens' principle. In particular, the scattered amplitudes vary as a function of range and ray angle, and the Huygens secondary sources are given, for a three-dimensional geometry, by the time derivative of the source-time function (in the frequency domain this is described by the factor $-i\omega$, a $\pi/2$ (90 degree) phase shift with amplitude proportional to frequency). In the case of a two-dimensional geometry, the secondary sources are the "half-derivative" of the source function (a 45 degree phase shift with amplitude

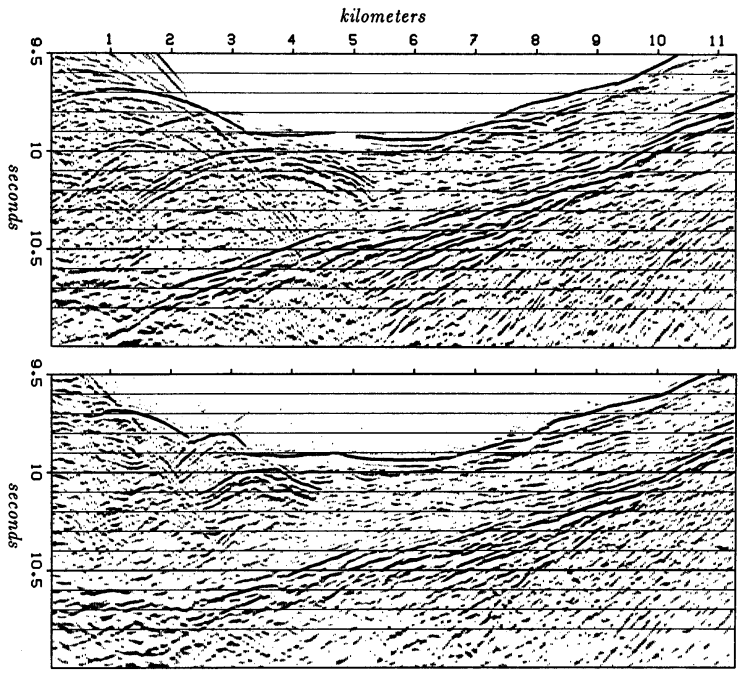

Figure 1.4: Original (top) and migrated (bottom) reflection data from a survey line across the Japan trench (figure modified from Claerbout, 1985; data from the Tokyo University Oceanographic Research Institute).

scaled by the square root of frequency). These details are provided by Kirchhoff theory, which is discussed later in this chapter. The diffraction hyperbola equation assumes a uniform velocity structure, but migration concepts can be generalized to more complicated velocity models. However, it is important to have an accurate velocity model, as use of the wrong model can result in "undermigrated" or "overmigrated" sections.

In common practice, data from seismic reflection experiments are first processed into zero-offset sections through common midpoint (CMP) stacking. The zero-offset section is then migrated to produce the final result. This is termed *poststack migration*. Because CMP stacking assumes horizontal layering and may blur some of the details of the original data, better results can be obtained if the migration is performed prior to stacking. This is called *prestack migration*. Although prestack migration is known to produce superior results, it is not implemented routinely owing to its much greater computational cost.

1.2 EXERCISES

1. (COMPUTER, get files from http://igppweb.ucsd.edu/~shearer/SCECERI/) Your 20-station seismic network recorded some high-frequency tremor. You are provided 250 s of data from each station, sampled at 20 samples/s. The ascii file, tremor_data.txt, contains the (x, y) station location in km, followed by 5000 points, etc., for the 20 stations. The station locations alone are given in the file stations.xy.txt. You have reason to believe that the tremor is located at 30 km depth and that the crust has a uniform P velocity of 6 km/s. Use P-wave backprojection to locate likely source(s) of the tremor. (HINT: Do a grid search of possible tremor source locations, using x values from 1 to 100 km and y values from 1 to 100 km (i.e., try 10,000 possible source locations, all at 30 km depth). Stack the seismograms at each source grid point using the appropriate time shifts for the computed travel times to each station. Then compute the RMS for the time series, save this value for each of the 100×100 source locations, and plot the resulting image of source amplitudes.) The Python program plotimage.py is an example of how you can plot your results, assuming your back-projection results are contained in an ascii file, out.backdata.txt, written as 100 lines of 100 values each, e.g., using the F90 format:

```
do iy = 1, 100
   write (12, '(100f6.3)') (backarray(ix, iy), ix=1,100)
enddo
```

The file out.backdata.txt can be used to test the Python program, but will then need to be replaced with the output of your back-projection program. Once you have your program working, you may find it interesting to perform some additional tests: (a) What happens if you assume an incorrect value for the P velocity? (b) Compute synthetic resolution kernels for a hypothetical source located in the center of the grid (x = 50, y = 50) as well as one located at (x = 5, y = 95). (HINT: Use one of the station time series as an example, then stack suitably time shifted versions of it over all the stations, where the

time shift is given by the difference between the P travel time at the true location and the grid point you are considering.) What explains the difference in the size and shape of the kernels?

1.3 Additional reading

Allmann, B.P., and P.M. Shearer, A high-frequency secondary event during the 2004 Parkfield earthquake, *Science*, **318**, 1279, doi: 10.1126/science.1146537, 2007.

Ishii, M., P.M. Shearer, H. Houston and J.E. Vidale, Extent, duration and speed of the 2004 Sumatra-Andaman earthquake imaged by the Hi-Net array, *Nature*, **435**, doi:10.1038/nature03675, 2005.

Ishii, M., P.M. Shearer, H. Houston, and J.E. Vidale, Teleseismic P wave imaging of the 26 December 2004 Sumatra-Andaman and 28 March 2005 Sumatra earthquake ruptures using the Hi-net array, *J. Geophys. Res.*, **112**, B11307, doi:10.1029/2006JB004700, 2007.

Walker, K.T., M. Ishii and P.M. Shearer, Rupture details of the 28 March 2005 Sumatra Mw 8.6 earthquake imaged with teleseismic P waves, *Geophys. Res. Lett.*, **32**, L24303, doi: 10.1029/2005GL024395, 2005.

Walker, K. T., and P. M. Shearer, Illuminating the near-sonic rupture velocities of the intracontinental Kokoxili Mw 7.8 and Denali fault Mw 7.9 strike-slip earthquakes with global P wave back projection imaging, *J. Geophys. Res.*, **114**, doi: 10.1029/2008JB005738, 2009.