Chapter 6

Nonuniqueness and Localized Averages

6.1 NULL VECTORS AND NONUNIQUENESS

In Chapters 3–5, we presented the basic method of finding estimates of the model parameters in a linear inverse problem. We showed that we could always obtain such estimates but that sometimes in order to do so we had to add *a priori* information to the problem. We shall now consider the meaning and consequences of nonuniqueness in linear inverse problems and show that it is possible to devise solutions that do not depend at all on *a priori* information. As we shall show, however, these solutions are not estimates of the model parameters themselves but estimates of *weighted averages* (linear combinations) of the model parameters.

When the linear inverse problem $\mathbf{Gm} = \mathbf{d}$ has nonunique solutions, there exist nontrivial solutions (that is, solutions with some nonzero m_i) to the homogeneous equation $\mathbf{Gm} = 0$. These solutions are called the *null vectors* of the inverse problem as premultiplying them by the data kernel yields zero. To see why nonuniqueness implies null vectors, suppose that the inverse problem has two distinct solutions $\mathbf{m}^{(1)}$ and $\mathbf{m}^{(2)}$ as

$$\begin{aligned} \mathbf{Gm}^{(1)} &= \mathbf{d} \\ \mathbf{Gm}^{(2)} &= \mathbf{d} \end{aligned} \tag{6.1}$$

Subtracting these two equations yields

$$\mathbf{G}(\mathbf{m}^{(1)} - \mathbf{m}^{(2)}) = 0 \tag{6.2}$$

Since the two solutions are by assumption distinct, their difference $\mathbf{m}^{\text{null}} = \mathbf{m}^{(1)} - \mathbf{m}^{(2)}$ is nonzero. The converse is also true; any linear inverse problem that has null vectors is nonunique. Note that the equation $\mathbf{Gm}^{\text{null}} = 0$ can be interpreted to mean that $\mathbf{m}^{\text{null}}$ is perpendicular to every row of $\mathbf{G}$ (as its dot product with every row is zero). Consequently, no linear combination of the rows of $\mathbf{G}$ can be a null vector. If $\mathbf{m}^{\text{par}}$ (where par stands for "particular") is any nonnull solution to $\mathbf{Gm} = \mathbf{d}$ (for instance, the minimum length solution), then $\mathbf{m}^{\text{par}} + \alpha\mathbf{m}^{\text{null}}$ is also a solution with the same error for any choice of α.

Geophysical Data Analysis: Discrete Inverse Theory. DOI: 10.1016/B978-0-12-397160-9.00006-0

Note that as $\alpha \mathbf{m}^{\text{null}}$ is a null vector for any nonzero α, null vectors are only distinct if they are linearly independent. If a given inverse problem has q distinct null solutions, then the most general solution is

$$\mathbf{m}^{\text{gen}} = \mathbf{m}^{\text{par}} + \sum_{i=1}^{q} \alpha_i \mathbf{m}^{\text{null}(i)} \tag{6.3}$$

where gen stands for "general." We shall show in Section 7.6 that $0 \leq q \leq M$, that is, that there can be no more linearly independent null vectors than there are unknowns.

6.2 NULL VECTORS OF A SIMPLE INVERSE PROBLEM

As an example, consider the following very simple equations:

$$\mathbf{Gm} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix} = [d_1] \tag{6.4}$$

This equation implies that only the mean value of a set of four model parameters has been measured. One obvious solution to this equation is $\mathbf{m} = [d_1, d_1, d_1, d_1]^{\text{T}}$ (in fact, this is the minimum length solution).

Three linearly independent null solutions can be determined by inspection as

$$\mathbf{m}^{\text{null}(1)} = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{m}^{\text{null}(2)} = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} \quad \mathbf{m}^{\text{null}(3)} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \tag{6.5}$$

The most general solution is then

$$\mathbf{m}^{\text{gen}} = \begin{bmatrix} d_1 \\ d_1 \\ d_1 \\ d_1 \end{bmatrix} + \alpha_1 \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \tag{6.6}$$

where the αs are arbitrary parameters.

Finding a particular solution to this problem now consists of choosing values for the parameters α_i. If one chooses these parameters so that $\|\mathbf{m}\|_2$ is minimized, one obtains the minimum length solution. Since the first vector is orthogonal to all the others, this minimum occurs when $\alpha_i = 0$, $i = 1, 2, 3$. We shall show in Chapter 7 that this is a general result: the minimum length solution never contains any null vectors. Note, however, that if other definitions of solution simplicity are used (e.g., flatness or smoothness), those solutions will contain null vectors.

6.3 LOCALIZED AVERAGES OF MODEL PARAMETERS

In Chapters 3–5, we have sought to estimate the elements of the solution vector $\mathbf{m}$. Another approach is to estimate some average of the model parameter $\langle m \rangle = \mathbf{a}^T\mathbf{m}$, where $\mathbf{a}$ is some averaging vector. The average is said to be localized if this averaging vector consists mostly of zeros (except for some group of nonzero elements that multiplies model parameters centered about one particular model parameter). This definition makes particular sense when the model parameters possess some natural ordering in space and time, such as acoustic velocity as a function of depth in the earth. For instance, if $M = 8$, the averaging vector $\mathbf{a} = [0,0,¼,½,¼,0,0,0]^T$ could be said to be localized about the fourth model parameter. The averaging vectors are usually normalized so that the sum of their elements is unity.

The advantage of estimating averages of the model parameters rather than the model parameters themselves is that quite often it is possible to identify unique averages even when the model parameters themselves are not unique. To examine when uniqueness can occur, we compute the average of the general solution as

$$\langle \mathrm{m} \rangle = \mathbf{a}^T\mathbf{m}^{\mathrm{gen}} = \mathbf{a}^T\mathbf{m}^{\mathrm{par}} + \sum_{i=1}^{q} \alpha_i \mathbf{a}^T\mathbf{m}^{\mathrm{null}(i)} \tag{6.7}$$

If $\mathbf{a}^T\mathbf{m}^{\mathrm{null}(i)}$ is zero for all i, then $\langle m \rangle$ is unique. The process of averaging has completely removed the nonuniqueness of the problem. Since $\mathbf{a}$ has M elements and there are $q \leq M$ constraints placed on $\mathbf{a}$, one can always find at least one vector that cancels (or "annihilates") the null vectors. One cannot, however, always guarantee that the averaging vector is localized around some particular model parameter. But, if $q < M$, one has some freedom in choosing $\mathbf{a}$ and there is some possibility of making the averaging vector at least somewhat localized. Whether this can be done depends on the structure of the null vectors, which, in turn, depends on the structure of the data kernel $\mathbf{G}$. Since the small-scale features of the model are unresolvable in many problems, unique localized averages can often be found.

6.4 RELATIONSHIP TO THE RESOLUTION MATRIX

During the discussion of the resolution matrix $\mathbf{R}$ (Section 4.3), we encountered in a somewhat different form the problem of determining averaging vectors. We showed that any estimate $\mathbf{m}^{\mathrm{est}}$ computed from a generalized inverse $\mathbf{G}^{-\mathrm{g}}$ was related to the true model parameters by

$$\mathbf{m}^{\mathrm{est}} = \mathbf{R}\mathbf{m}^{\mathrm{true}} = \mathbf{G}^{-\mathrm{g}}\mathbf{G}\mathbf{m}^{\mathrm{true}} \quad \text{or} \quad m_i^{\mathrm{est}} = \sum_{j=1}^{M} R_{ij} m_j^{\mathrm{true}} = \sum_{j=1}^{M}\sum_{k=1}^{N} G_{ik}^{-g} G_{kj}\, m_j^{\mathrm{true}} \tag{6.8}$$

The ith row of $\mathbf{R}$ (or rather its transpose) can be interpreted as a unique averaging vector $\mathbf{a}$ that is centered about m_i, with the averaging vector $\mathbf{a}$ being built up from linear combinations of the rows of the data kernel $\mathbf{G}$

$$m_i^{\text{est}} = \langle m \rangle^{(i)} = \sum_{j=1}^{M} a_j^{(i)} m_j^{\text{true}} \quad \text{with} \quad a_j^{(i)} = \sum_{k=1}^{N} c_k^{(i)} G_{kj} \quad \text{and} \quad c_k^{(i)} = G_{ik}^{-g} \tag{6.9}$$

Note that the resolution matrix in Equation (6.8) is composed of the product of the generalized inverse and the data kernel. We can interpret this product as meaning that a row of the resolution matrix is composed of a weighted sum of the rows of the data kernel $\mathbf{G}$ (where the elements of the generalized inverse are the weighting factors $c_k^{(i)} = G_{ik}^{-g}$) regardless of the generalized inverse's particular form. An averaging vector $\mathbf{a}$ produces a unique average $\langle m \rangle$ if and only if $\mathbf{a}^{\mathrm{T}}$ can be represented as a linear combination of the rows of the data kernel $\mathbf{G}$, since the rows of $\mathbf{G}$ are guaranteed to be perpendicular to every null vector.

Whether or not the average is truly localized depends on the structure of $\mathbf{R}$. The spread function discussed previously in Section 4.6 is a measure of the degree of localization.

The process of forming the generalized inverse is equivalent to "shuffling" the rows of the equation $\mathbf{Gm} = \mathbf{d}$ by forming linear combinations until the data kernel is as close as possible to an identity matrix. Each row of the data kernel can then be viewed as a localized averaging vector, and each corresponding row of the shuffled data vector is the estimated value of the average.

6.5 AVERAGES VERSUS ESTIMATES

We can, therefore, identify a type of dualism in inverse theory. Given a generalized inverse $\mathbf{G}^{-g}$ that in some sense solves $\mathbf{Gm} = \mathbf{d}$, we can speak either of estimates of model parameters $\mathbf{m}^{\text{est}} = \mathbf{G}^{-g}\mathbf{d}$ or of localized averages $\langle \mathbf{m} \rangle = \mathbf{G}^{-g}\mathbf{d}$. The numerical values are the same but the interpretation is quite different. When the solution is interpreted as a localized average, it can be viewed as a unique quantity that exists independently of any *a priori* information applied to the inverse problem. Examination of the resolution matrix may reveal that the average is not especially localized and the solution may be difficult to interpret. When the solution is viewed as an estimate of a model parameter, the location of what is being solved for is clear. The estimate can be viewed as unique only if one accepts as appropriate whatever *a priori* information was used to remove the inverse problem's underdeterminacy. In most instances, the choice of *a priori* information is somewhat *ad hoc*, so the solution may still be difficult to interpret.

In the sample problem stated above, the data kernel has only one row. There is therefore only one averaging vector that will annihilate all the null vectors: one proportional to that row

$$\mathbf{a} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}^{\mathrm{T}} \tag{6.10}$$

This averaging vector is clearly unlocalized. In this problem, the structure of $\mathbf{G}$ is just too poor to form good averages. The generalized inverse to this problem is by inspection

$$\mathbf{G}^{-g} = [\,1 \quad 1 \quad 1 \quad 1\,]^{T} \tag{6.11}$$

The resolution matrix is therefore

$$\mathbf{R} = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \tag{6.12}$$

which is very unlocalized and equivalent to Equation (6.8).

6.6 NONUNIQUE AVERAGING VECTORS AND *A PRIORI* INFORMATION

There are instances in which even nonunique averages of model parameters can be of value, especially when they are used in conjunction with other *a priori* knowledge of the nature of the solution (Wunsch and Minster, 1982). Suppose that one simply picks a localized averaging vector that does not necessarily annihilate all the null vectors and that, therefore, does not lead to a unique average. In the above problem, the vector $\mathbf{a} = (1/3)[1,1,1,0]^{T}$ might be such a vector. It is somewhat localized, being centered about the second model parameter. Note that it does not lead to a unique average, since

$$\langle m \rangle = \mathbf{a}^{T}\mathbf{m}^{gen} = d_1 + 0 + 0 + \frac{1}{3}\alpha_3 \tag{6.13}$$

is still a function of one of the arbitrary parameters α_i. Suppose, however, that there is *a priori* knowledge that every m_i must satisfy $0 \le m_i \le 2d_1$. Then from the equation for $\mathbf{m}^{gen}$, α_3 must be no greater than d_1 and no less than $-d_1$. Since $-d_1 \le \alpha_3 \le d_1$, the average has bounds $(2/3)d_1 \le \langle m \rangle \le (4/3)d_1$. These constraints are considerably tighter than the *a priori* bounds on m_i, which demonstrates that this technique has indeed produced some useful information. This approach works because even though the averaging vector does not annihilate all the null vectors, $\mathbf{a}^{T}\mathbf{m}^{null(i)}$ is small compared with the elements of the null vector. Localized averaging vectors often lead to small products since the null vectors often fluctuate rapidly about zero, indicating that small-scale features of the model are the most poorly resolved. A slightly more complicated example of this type is solved in Figure 6.1.

This approach can be generalized as follows (Oldenberg, 1983):

$$\begin{aligned} &\text{maximize/minimize } \langle m \rangle = \mathbf{a}^{T}\mathbf{m} \quad \text{with respect to } \mathbf{m} \\ &\text{with the constraints } \mathbf{Gm} = \mathbf{d}^{obs} \quad \text{and} \quad \mathbf{m}^{(l)} \le \mathbf{m} \le \mathbf{m}^{(u)} \end{aligned} \tag{6.14}$$

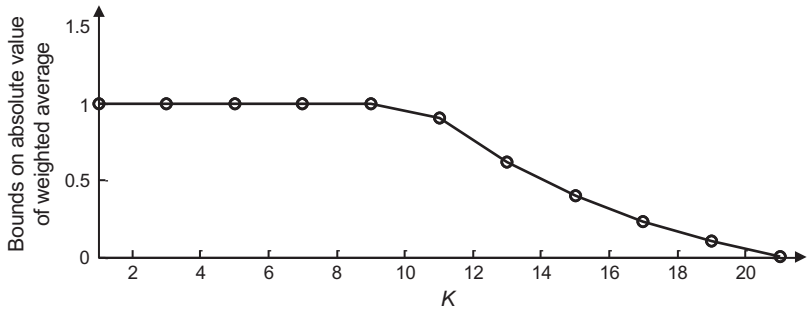

FIGURE 6.1 Bounds on weighted averages of model parameters, m_i, in a problem in which the only datum is that the sum of all model parameters is zero. When this observation is combined with the *a priori* information that each model parameter must satisfy $|m_i| \leq 1$, bounds can be placed on the weighted averages of the model parameter. The bounds shown here are for averages of K neighboring model parameters. Note that the bounds are tighter than the *a priori* bounds only when $K > 10$. *MatLab* script gda06_01.

Here, $\mathbf{m}^{\mathrm{l}}$ and $\mathbf{m}^{\mathrm{u}}$ are the *a priori* lower and upper bounds on $\mathbf{m}$, respectively. Note that this formulation does not explicitly include null vectors; the constraint $\mathbf{Gm} = \mathbf{d}^{\mathrm{obs}}$ is sufficient to ensure that the model parameters satisfy the data. This problem is a special case of the *linear programming problem*

$$\begin{aligned} &\text{find } \mathbf{x} \text{ that minimizes } z = \mathbf{f}^{\mathrm{T}}\mathbf{x} \\ &\text{with the constraints } \mathbf{Ax} \leq \mathbf{b} \quad \text{and} \quad \mathbf{Cx} = \mathbf{d} \quad \text{and} \quad \mathbf{x}^{(\mathrm{l})} \leq \mathbf{x} \leq \mathbf{x}^{(\mathrm{u})} \end{aligned} \tag{6.15}$$

Note that the minimization problem can be converted into a maximization problem by flipping the sign of $\mathbf{f}$. Furthermore, "$\geq$ type" inequality constraints can be converted to "$\leq$ type" by multiplication by -1.

The linear programming problem was first studied by economists and business operations analysts. For example, z might represent the profit realized by a factory producing a product line, where the number of each product is given by $\mathbf{x}$ and the profit on each item given by $\mathbf{f}$. The problem is to maximize the total profit $\mathbf{f}^{\mathrm{T}}\mathbf{x}$ without violating the constraint that one can produce only a positive amount of each product, or any other linear inequality constraints that might represent labor laws, union regulations, physical limitation of machines, etc.

MatLab provides a `linprog()` function for solving the linear programming problem. Equation (6.13) is solved by calling it twice, once to minimize $\langle m \rangle$ and the other to maximize it

```
[mest1, amin]=linprog(a,[],[],G,dobs,mlb,mub);
[mest2, amax]=linprog(-a,[],[],G,dobs,mlb,mub);
amax=-amax;
```

(*MatLab* script gsa06_01)

Here, `amin` and `amax` are its lower and upper bounds on $\langle m \rangle$, respectively. While the corresponding model parameter vectors mest1 and mest2 are also

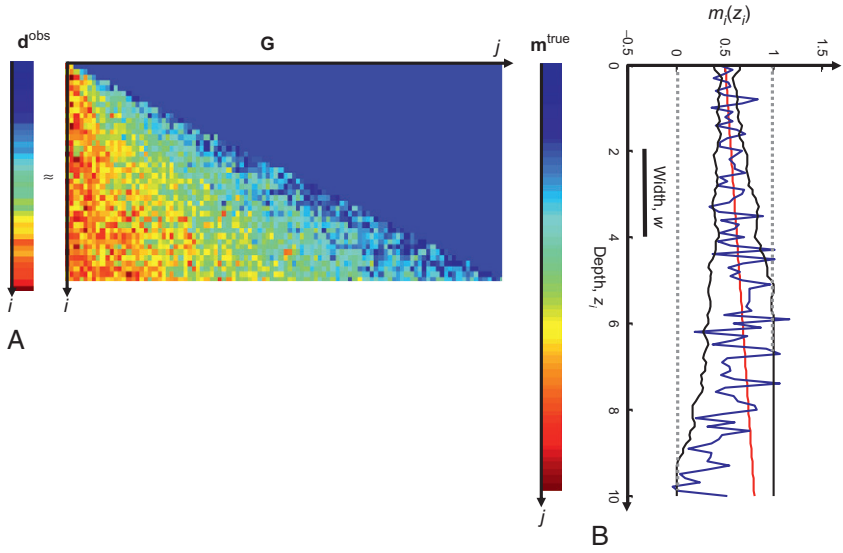

FIGURE 6.2 (A) This underdetermined inverse problem, $\mathbf{d} = \mathbf{Gm}$, has $M = 100$ model parameters m_i and $N = 40$ data d_i. The data are weighted averages of the model parameters, from the surface down to a depth, z, that increases with index, i. The observed data d_i^{est} include additive noise. (B) The true model parameters (red curve) increase linearly with depth z. The estimated model parameters (blue curve), computed using the minimum length method, scatter about the true model at shallow depths ($z < 6$) but decline toward zero at deeper depths due to poor resolution. Bounds on localized averages of the model parameters, with an averaging width, $w = 2$ (black curves), are for *a priori* information, $0 < m_i < 1$ (gray dotted lines). *MatLab* script gda06_02.

calculated, they are not usually of interest. The arguments of `linprog()` include the averaging vector `a`, the data kernel `G`, the observed data `dobs`, and the *a priori* upper and lower bounds `mlb` and `mub` on the model parameters. An example using a Laplace transform-like data kernel is shown in Figure 6.2.

6.7 PROBLEMS

6.1. What is the general solution to the problem $\mathbf{Gm} = \mathbf{d}$, with

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

6.2. Give some examples of physical problems where the model parameters can be assumed, with reasonable certainty, to fall between lower and upper bounds (and identify the bounds).

6.3. Suppose that $M = 21$, model parameters are known to have bounds $-1 \leq m_i \leq 1$. Suppose that the unweighted average of each three adjacent model parameters are observed so that the data kernel has the form

$$\mathbf{G} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ & & & & & & \cdots & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Can useful bounds be placed on the weighted average of the three adjacent model parameters, where the weights are [¼, ½, ¼]? Adapt *MatLab* script gda06_01 to explore this problem.

REFERENCES

Oldenburg, D.W., 1983. Funnel functions in linear and nonlinear appraisal. J. Geophys. Res. 88, 7387–7398.

Wunsch, C., Minster, J.F., 1982. Methods for box models and ocean circulation tracers: mathematical programming and non-linear inverse theory. J. Geophys. Res. 87, 5647–5662.