Lecture Notes on

Nonlinear Inversion and Tomography:

I. Borehole Seismic Tomography

From a Series of Lectures by

James G. Berryman
University of California
Lawrence Livermore National Laboratory
Livermore, CA 94550

Originally Presented at

Earth Resources Laboratory
Massachusetts Institute of Technology

July 9–30, 1990

Revised and Expanded

October, 1991

Chapter 3

Least-Squares Methods

We consider solutions to the inversion problem for block models. Given a set of weights $w_i > 0$, $i = 1, \ldots, m$, we define the functional $\Psi^*: \mathcal{S} \to \mathbf{R}$ by

$$\Psi^*(\mathbf{s}) = \sum_{i=1}^{m} w_i \left[\tau_i^*(\mathbf{s}) - t_i\right]^2 . \tag{3.1}$$

$\Psi^*(\mathbf{s})$ measures the degree of misfit between the observed data and traveltimes predicted by the model $\mathbf{s}$. Ψ^* is the nonlinear least-squares functional since it uses the full traveltime functional τ^* in the error calculation. The linear least-squares functional was defined previously in (2.20).

3.1 Normal Equations

The standard least-squares problem is a simplified version of (3.1) with all weights equal to unity and the traveltime functional replaced by its linear approximation $\mathbf{Ms}$ for a cell model. Then, the squared error functional is

$$\Psi(\mathbf{s}) = (\mathbf{t} - \mathbf{Ms})^T (\mathbf{t} - \mathbf{Ms}). \tag{3.2}$$

The minimum of this functional is found by differentiating with respect to the value of the slowness in each cell. At the minimum, all these derivatives must vanish, so

$$\frac{\partial \Psi}{\partial s_j} = 2 \left[\mathbf{M}^T (\mathbf{t} - \mathbf{Ms})\right]_j = 0, \tag{3.3}$$

for all $j = 1, \ldots, n$. Thus, (3.3) implies the slowness at the minimum of (3.2) satisfies

$$\sum_{i=1}^{m} \sum_{k=1}^{n} l_{ij} l_{ik} \hat{s}_k = \sum_{i=1}^{m} l_{ij} t_i \quad \text{for} \quad j = 1, \ldots, n, \tag{3.4}$$

or equivalently that

$$\mathbf{M}^T \mathbf{M} \hat{\mathbf{s}} = \mathbf{M}^T \mathbf{t}. \tag{3.5}$$

There are n equations for the n unknowns s_j, since $\mathbf{M}^T\mathbf{M}$ is an $n \times n$ square and symmetric matrix.

These equations are known as the *normal equations* for the solution $\hat{\mathbf{s}}$ of the standard least-squares problem. If the number of data m exceeds the number of cells n in the discretized model so $m > n$, we say the discretized inversion problem is *overdetermined.*[1] If the the number of cells n exceeds the number of data m so $m < n$, we say the discretized inversion problem is *underdetermined.* The normal equations may be used in either case, but the form of the resulting solution is substantially different. We generally assume that the inversion problem is overdetermined, but there may still be situations where we want to use only a small part of the available data to make corrections to the slowness model; then the resulting problem is equivalent to the underdetermined version of the normal equations. General methods for solving (3.5) will be discussed in Chapter 4.

PROBLEMS

PROBLEM 3.1.1 *Use the chain rule to show that the minimum of a least-squares functional occurs at the same model whether we use slowness or velocity as the variable.*

PROBLEM 3.1.2 *An experimental configuration has m source-receiver pairs and the region to be reconstructed is modeled using n cells, so the ray-path matrix $\mathbf{M}$ is $m \times n$. Suppose that p independent measurements of the traveltimes have been made, resulting in p traveltime m-vectors $\mathbf{t}_1, \ldots, \mathbf{t}_p$. Then, the inversion problem can be formulated as*

$$\begin{pmatrix} \mathbf{M} \\ \mathbf{M} \\ \vdots \\ \mathbf{M} \end{pmatrix} \mathbf{s} = \begin{pmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \\ \vdots \\ \mathbf{t}_p \end{pmatrix} .$$

Show that the normal equations for this problem become

$$\mathbf{M}^T\mathbf{M}\mathbf{s} = \mathbf{M}^T \langle \mathbf{t} \rangle ,$$

where $\langle \mathbf{t} \rangle = \frac{1}{p}\sum_{q=1}^{p} \mathbf{t}_q$. Explain the significance of this result.

3.2 Scaled Least-Squares Model

DEFINITION 3.2.1 (SCALED LEAST-SQUARES MODEL) *The scaled least-squares model with respect to a given model $\mathbf{s}_0$, and set of weights w_i, is the model $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0]}$ minimizing Ψ^* subject to the constraint that $\mathbf{s} = \gamma\mathbf{s}_0$ for $\gamma > 0$. Thus*

$$\Psi^*(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0]}) = \min_{\gamma} \Psi^*(\gamma\mathbf{s}_0). \tag{3.6}$$

[1] Recall that the underlying physical problem is essentially the reconstruction of a continuous function from finite data, so this continuous reconstruction problem is always grossly *underdetermined.*

The scaled least-squares model associated with $\mathbf{s}_0$ is unique.

To solve for the scaled least-squares model, we expand $\Psi^*(\gamma\mathbf{s}_0)$ as

$$\Psi^*(\gamma\mathbf{s}_0) = \sum_i w_i[\tau_i^*(\gamma\mathbf{s}_0) - t_i]^2 \tag{3.7}$$

$$= \sum_i w_i\tau_i^*(\gamma\mathbf{s}_0)^2 - 2\sum_i w_i t_i \tau_i^*(\gamma\mathbf{s}_0) + \sum_i w_i t_i^2. \tag{3.8}$$

Using the homogeneity of τ_i^*, we can write

$$\Psi^*(\gamma\mathbf{s}_0) = \gamma^2\sum_i w_i\tau_i^*(\mathbf{s}_0)^2 - 2\gamma\sum_i w_i t_i \tau_i^*(\mathbf{s}_0) + \sum_i w_i t_i^2. \tag{3.9}$$

This is simply a second-order polynomial in γ and achieves its minimum at $\gamma = \gamma_{\mathrm{LS}[\mathbf{s}_0]}$, where

$$\gamma_{\mathrm{LS}[\mathbf{s}_0]} = \frac{\sum_i w_i t_i \tau_i^*(\mathbf{s}_0)}{\sum_i w_i \tau_i^*(\mathbf{s}_0)^2}. \tag{3.10}$$

Thus

$$\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0]} = \mathbf{s}_0\frac{\sum_i w_i t_i \tau_i^*(\mathbf{s}_0)}{\sum_i w_i \tau_i^*(\mathbf{s}_0)^2}. \tag{3.11}$$

THEOREM 3.2.1 *For any* $\mathbf{s}_0$, $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0]} \notin \mathrm{Int}\,\mathcal{F}^*$.

Proof: We have from (3.10)

$$\gamma_{\mathrm{LS}[\mathbf{s}_0]}\sum_{i=1}^m w_i\tau_i^*(\mathbf{s}_0)^2 = \sum_{i=1}^m w_i t_i \tau_i^*(\mathbf{s}_0), \tag{3.12}$$

or, given the homogeneity of τ_i^*,

$$\sum_{i=1}^m w_i\tau_i^*(\mathbf{s}_0)[\tau_i^*(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0]}) - t_i] = 0. \tag{3.13}$$

Since the w_i and values of τ_i^* are positive, this can only be true if either $\tau_i^*(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0]}) = t_i$ for all i (i.e., $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0]} \in \mathrm{Bdy}\,\mathcal{F}^*$ and is an exact solution to the inversion problem) or if $\tau_i^*(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0]}) < t_i$ for at least one i (*i.e.*, $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0]} \notin \mathcal{F}^*$). Thus, the scaled least-squares model cannot be in $\mathrm{Int}\,\mathcal{F}^*$. ▮

This important result shows that

> *a scaled least-squares slowness model can never be a strictly interior point of the global feasible set.*

The only way for a scaled least-squares point to be in the feasible set is for it to be on the boundary and then only if it solves the inversion problem.

We can write the scaled least-squares model in matrix notation as follows. Let $\mathbf{W}$ be the diagonal matrix formed from the positive weights w_i:

$$\mathbf{W} = \begin{pmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_m \end{pmatrix}. \tag{3.14}$$

Further, let $\mathbf{M}_0$ be the ray-path matrix computed from $\mathbf{s}_0$. Thus, $\tau_i^*(\mathbf{s}_0) = [\mathbf{M}_0\mathbf{s}_0]_i$. In matrix notation, (3.10) becomes

$$\gamma_{\mathrm{LS}[\mathbf{s}_0]} = \frac{\mathbf{s}_0^T\mathbf{M}_0^T\mathbf{W}\mathbf{t}}{\mathbf{s}_0^T\mathbf{M}_0^T\mathbf{W}\mathbf{M}_0\mathbf{s}_0}, \tag{3.15}$$

implying

$$\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0]} = \mathbf{s}_0\frac{\mathbf{s}_0^T\mathbf{M}_0^T\mathbf{W}\mathbf{t}}{\mathbf{s}_0^T\mathbf{M}_0^T\mathbf{W}\mathbf{M}_0\mathbf{s}_0}. \tag{3.16}$$

PROBLEM

PROBLEM 3.2.1 *If $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0]}$ is defined in terms of the linear least-squares functional $\Psi^{\mathcal{P}}$ instead of Ψ^*, is there a result corresponding to Theorem 3.2.1 for this model?*

3.3 Nonlinear Least-Squares Models

DEFINITION 3.3.1 (LEAST-SQUARES MODEL) *A least-squares model, with respect to weights w_i, is a vector $\hat{\mathbf{s}}_{\mathrm{LS}}$ which minimizes Ψ^*, i.e.,*

$$\Psi^*(\hat{\mathbf{s}}_{\mathrm{LS}}) = \min_{\mathbf{s}} \Psi^*(\mathbf{s}). \tag{3.17}$$

The least-squares model may be nonunique. Nonuniqueness is expected when $m < n$, *i.e.*, there are fewer traveltime data than model cells, or when $m > n$ and the ray-path matrix has a right null space containing ghosts $\mathbf{g}$. The most common method of picking the "best" least-squares solution [Penrose, 1955b] is to choose the one of minimum Euclidean norm. This "best" solution has some nice properties as we shall see when we discuss ghosts in tomography, but it may not represent the "best" solution to the inversion problem.

THEOREM 3.3.1 $\hat{\mathbf{s}}_{\mathrm{LS}} \notin \mathrm{Int}\,\mathcal{F}^*$.

Proof: This theorem follows from the fact that $\hat{\mathbf{s}}_{\mathrm{LS}} = \hat{\mathbf{s}}_{\mathrm{LS}[\hat{\mathbf{s}}_{\mathrm{LS}}]}$, *i.e.*, a least-squares model is the scaled least-squares model with respect to itself (or otherwise there would be a model yielding smaller Ψ^*). ∎

Any nonlinear least-squares solution is infeasible unless it solves the inversion problem, in which case it lies on the boundary of the feasible set.

The preceding proof is entirely adequate to establish the infeasibility of the least-squares point. However, it may be enlightening to present a second proof based on stationarity of the ray paths.

Consider the deviation of the least-squares functional induced by a small change in the model:

$$\delta\Psi^* = \Psi^*(s+\delta s) - \Psi^*(s) = \sum_{i=1}^{m} w_i \left[\tau_i^*(s+\delta s) - t_i\right]^2 - \sum_{i=1}^{m} w_i \left[\tau_i^*(s) - t_i\right]^2. \quad (3.18)$$

This equation may be rearranged without approximation into the form

$$\delta\Psi^* = 2\sum_{i=1}^{m} w_i \left[\tau_i^*(s+\delta s) - \tau_i^*(s)\right]\left[(\tau_i^*(s+\delta s) + \tau_i^*(s))/2 - t_i\right]. \quad (3.19)$$

For small slowness perturbations δs, the first bracket in the sum of (3.19) is clearly of order δs, while any contributions of order δs in the second bracket are therefore of second order and may be neglected. If $dl_i^*[s]$ is the infinitesimal increment of the (or a) least-time ray along path i for s, then

$$\tau_i^*(s+\delta s) - \tau_i^*(s) = \int (s+\delta s)\, dl_i^*[s+\delta s] - \int s\, dl_i^*[s]. \quad (3.20)$$

Recall that stationarity of the ray paths near the one of least time implies that

$$\int s\, dl_i^*[s+\delta s] = \int s\left\{dl_i^*[s] + d\delta l_i^*\right\} \simeq \int s\, dl_i^*[s], \quad (3.21)$$

where $d\delta l_i^*$ is the perturbation in the infinitesimal increment $dl_i^*[s]$ of the ray path induced by the fact that $dl_i^*[s+\delta s]$ is the one for the perturbed model and therefore generally[2] only slightly different from that for s. Using (3.21) in (3.20), we find that

$$\tau_i^*(s+\delta s) - \tau_i^*(s) \simeq \int \delta s\, dl_i^*[s+\delta s] \simeq \int \delta s\, dl_i^*[s] \quad (3.22)$$

to lowest order in δs. Thus, (3.19) becomes

$$\delta\Psi^* = 2\sum_{i=1}^{m} w_i \left(\int \delta s\, dl_i^*[s]\right)\left[\tau_i^*(s) - t_i\right]. \quad (3.23)$$

Equation (3.23) is the expression needed to construct the functional (Frechét) derivative of Ψ^*. If s produces the minimum of $\Psi^*(s)$, then the functional derivative should vanish. We see that the weights w_i are positive, the coefficient of δs is the integral of the increment of the ray path itself in the regions of change which is strictly positive, and if the the traveltime function $\tau_i^*(s) - t_i \geq 0$ as is required for all i in order for the model s to be feasible, then the derivative cannot vanish and therefore s is not the minimum. This contradiction shows again that either the minimum of the traveltime function must be infeasible, or it must solve the inversion problem.

Problem

Problem 3.3.1 *Determine whether there is a result analogous to Theorem 3.3.1 for the linear least-squares functional* $\Psi^{\mathcal{P}}$.

[2] There are pathological cases where a small change in the model s can induce a large change in the ray path, but we will ignore this possibility for the present argument.

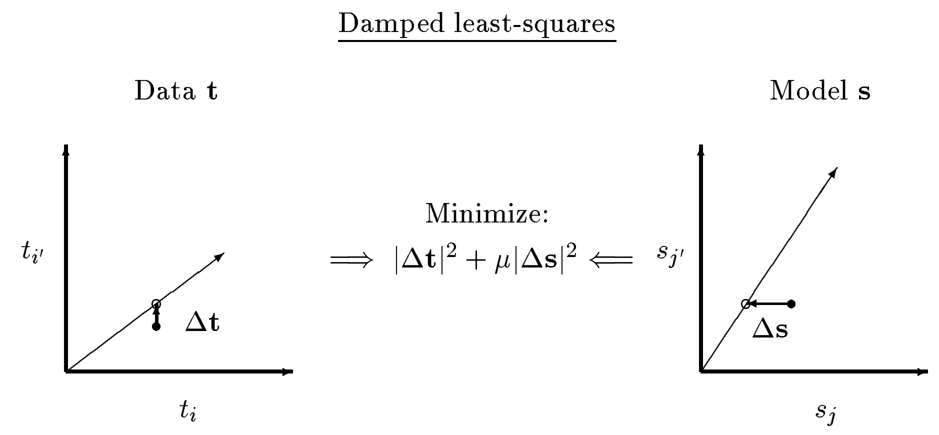

Figure 3.1: Schematic illustration of deamped least squares analysis.

3.4 Damped Least-Squares Model

Let $\mathbf{C}$ be a diagonal (coverage) matrix formed from the positive weights c_j:

$$\mathbf{C} = \begin{pmatrix} c_1 & & & \\ & c_2 & & \\ & & \ddots & \\ & & & c_n \end{pmatrix}. \tag{3.24}$$

The c_js may be treated here as arbitrary positive weights, but a definite choice of the c_js will be found later.

Definition 3.4.1 (damped least-squares model) *The damped least-squares model with respect to a given model* $\mathbf{s}_0$*, and set of weights* w_i, c_j*, is the model* $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]}$ *minimizing*

$$\Psi^*(\mathbf{s}) + \mu(\mathbf{s} - \mathbf{s}_0)^T \mathbf{C}(\mathbf{s} - \mathbf{s}_0). \tag{3.25}$$

[Levenberg, 1944]

Like the scaled least-squares model, the damped least-squares model is unique.

We can solve for the damped least-squares model based on a linear approximation to the traveltime functionals. Given the model $\mathbf{s}_0$, let P_i^0 denote the least-time ray paths through $\mathbf{s}_0$. Then, to first order in $\mathbf{s} - \mathbf{s}_0$ we have

$$\tau_i^*(\mathbf{s}) \simeq \tau_i^{P_i^0}(\mathbf{s}). \tag{3.26}$$

This approximation yields

$$\Psi^{\mathcal{P}}(\mathbf{s}) = (\mathbf{t} - \mathbf{M}_0\mathbf{s})^T \mathbf{W}(\mathbf{t} - \mathbf{M}_0\mathbf{s}), \tag{3.27}$$

where $\mathbf{M}_0$ is the ray-path matrix obtained from the ray paths P_i^0.

Using the first-order approximation, the damped least-squares model becomes

$$\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]} = \mathbf{s}_0 + (\mathbf{M}_0^T\mathbf{W}\mathbf{M}_0 + \mu\mathbf{C})^{-1}\mathbf{M}_0^T\mathbf{W}(\mathbf{t} - \mathbf{M}_0\mathbf{s}_0). \tag{3.28}$$

This equation can be rearranged to show that

$$\mu\mathbf{C}(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]} - \mathbf{s}_0) = \mathbf{M}_0^T\mathbf{W}(\mathbf{t} - \mathbf{M}_0\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]}). \tag{3.29}$$

Then, we obtain the following two theorems:

THEOREM 3.4.1 *If* $\mathbf{s}_0 \notin \mathcal{F}^{\mathcal{P}^0}$, *then the model* $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]}$ *defined in (3.28) does not solve the inversion problem for any* $\mu > 0$.

THEOREM 3.4.2 *If* $\mathbf{s}_0 \notin \mathcal{F}^{\mathcal{P}^0}$, *then* $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]} \notin \mathcal{F}^{\mathcal{P}^0}$.

Proof: The proofs are by contradiction.

First, suppose that $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]}$ solves the inversion problem so $\mathbf{M}_0\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]} \equiv \mathbf{t}$. Then, (3.29) shows that $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]} = \mathbf{s}_0$ (if $\mu > 0$) so $\mathbf{M}_0\mathbf{s}_0 \equiv \mathbf{t}$. But this result contradicts the assumption that $\mathbf{s}_0 \notin \mathcal{F}^{\mathcal{P}^0}$ so $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]}$ does not solve the inversion problem if $\mathbf{s}_0$ is infeasible.

Second, suppose that $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]}$ is feasible (*i.e.*, $\mathbf{M}_0\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]} \geq \mathbf{t}$ and, using the previous theorem, we may exclude the possibility that it solves the inversion problem so in fact $\mathbf{M}_0\hat{\mathbf{s}}_s ls\mathbf{s}_0, \mu > \mathbf{t}$), then it follows from the positivity of all the matrix elements in (3.29) that $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]} \leq \mathbf{s}_0$. But, if $\mathbf{s}_0$ is infeasible so $(\mathbf{M}_0\mathbf{s}_0)_i < t_i$ for some i, then it also follows that $(\mathbf{M}_0\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]})_i < t_i$ for the same i so $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]}$ is infeasible, which contradicts the original feasibility supposition on $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]}$. ∎

To paraphrase the results, "we cannot get there from here." If we start our computation at any local infeasible point, we cannot get to a solution or to any local feasible point using the damped least-squares method. These results are very strong because they show the infeasibility of $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]}$ holds for any value of the *damping parameter* $\mu > 0$ and also for any choice of the weight matrix $\mathbf{C}$. In fact, damped least-squares always leads to a biased estimate:

> *unless the starting point* $\mathbf{s}_0$ *already solves the inversion problem, the damped least-squares solution never solves the inversion problem for any* $\mu > 0$.

Practical application of damped least-squares requires a definite choice of the damping parameter μ. Some methods for choosing its magnitude will be explored in the problems and also in the section discussing linear inversion algorithms (Section 4). Types of damping more general than the norm damping considered in (3.24) will also be discussed.

PROBLEMS

PROBLEM 3.4.1 *Assume that* $\mathbf{s}_0$ *solves the inversion problem except for a scale factor* γ, *i.e.*,

$$\mathbf{M}\gamma\mathbf{s}_0 = \mathbf{t}.$$

Show that the damped least-squares solution with $\mathbf{s}_0$ *as the starting model does not solve the inversion problem for any* $\mu > 0$ *unless* $\gamma = 1$.

PROBLEM 3.4.2 *Assume the starting point* $\mathbf{s}_0$ *is not a stationary point of* $\Psi^*(\mathbf{s})$. *Then, show that the damped least-squares point always gives a function value less than that of the starting point, i.e., that this point satisfies*

$$\Psi^*(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]}) < \Psi^*(\mathbf{s}_0).$$

[Hint: Show that $\Psi^*(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]}) + \mu(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]} - \mathbf{s}_0)^T\mathbf{C}(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_o,\mu]} - \mathbf{s}_0) \le \Psi^*(\mathbf{s}_0)$*.] Thus, the squares of the residuals will be reduced. [Levenberg, 1944]*

PROBLEM 3.4.3 *If* $\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,0]}$ *is the standard least-squares solution, show that*

$$(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]} - \mathbf{s}_0)^T\mathbf{C}(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]} - \mathbf{s}_0) < (\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,0]} - \mathbf{s}_0)^T\mathbf{C}(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,0]} - \mathbf{s}_0).$$

[Hint: Show that $\Psi^*(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,0]}) \le \Psi^*(\hat{\mathbf{s}}_{\mathrm{LS}[\mathbf{s}_0,\mu]})$*.] Thus, the weighted sums of the squares of the model corrections is less than that for the standard least-squares problem. [Levenberg, 1944]*

PROBLEM 3.4.4 *Consider the model correction* $\Delta\mathbf{s}_d = \mathbf{s} - \hat{\mathbf{s}}$ *as a function of the damping parameter* μ, *where*

$$(\mathbf{M}^T\mathbf{M} + \mu\mathbf{I})\Delta\mathbf{s}_d = \mathbf{M}^T(\mathbf{t} - \mathbf{M}\hat{\mathbf{s}}) \equiv \Delta\mathbf{s}_g. \tag{3.30}$$

The angle θ *between the damped least-squares solution* $\Delta\mathbf{s}_d$ *and the negative of the least-squares functional gradient* $\Delta\mathbf{s}_g$ *is determined by*

$$\cos\theta = \frac{\Delta\mathbf{s}_g^T\Delta\mathbf{s}_d}{||\Delta\mathbf{s}_g||\,||\Delta\mathbf{s}_d||}.$$

Show that

1. $\cos\theta \to 1$ *as* $\mu \to \infty$;

2. $\cos\theta \to 0$ *as* $\mu \to 0$ *if* $\mathbf{M}^T\mathbf{M}$ *is singular.*

Use these results to characterize (3.30) as an interpolation formula. [Marquardt, 1963]

PROBLEM 3.4.5 *Suppose that traveltime measurements have been repeated* K *times, resulting in the set of data vectors* $\mathbf{t}^{(k)}$ *for* $k = 1, \ldots, K$. *Form the data matrix*

$$\mathbf{T} = (\,\mathbf{t}^{(1)} \quad \mathbf{t}^{(2)} \quad \cdots \quad \mathbf{t}^{(K)}\,)$$

and the associated solution matrix

$$\mathbf{S} = (\,\mathbf{s}^{(1)} \quad \mathbf{s}^{(2)} \quad \cdots \quad \mathbf{s}^{(K)}\,).$$

Also, introduce the noise matrix $\mathbf{N}$ *defined by*

$$\mathbf{N} = \mathbf{T} - \mathbf{MS} = \Delta\mathbf{T} - \mathbf{M}\Delta\mathbf{S},$$

where the correction matrices $\Delta\mathbf{T}$ *and* $\Delta\mathbf{S}$ *are defined similarly by column vectors* $\Delta\mathbf{s}^{(k)} \equiv \mathbf{s}^{(k)} - \bar{\mathbf{s}}$ *and* $\Delta\mathbf{t}^{(k)} \equiv \mathbf{t}^{(k)} - \mathbf{M}\bar{\mathbf{s}}$. *The correlation matrices are then defined by*

$$\mathcal{C}_{ss} = \Delta\mathbf{S}\Delta\mathbf{S}^T, \quad \mathcal{C}_{tt} = \Delta\mathbf{T}\Delta\mathbf{T}^T, \quad \mathcal{C}_{nn} = \mathbf{N}\mathbf{N}^T,$$

$$\mathcal{C}_{st} = \Delta\mathbf{S}\Delta\mathbf{T}^T, \quad \mathcal{C}_{nt} = \mathbf{N}\Delta\mathbf{T}^T, \quad \mathcal{C}_{ns} = \mathbf{N}\Delta\mathbf{S}^T,$$

etc. Suppose that $\mathcal{C}_{ns} \simeq 0 \simeq \mathcal{C}_{sn}$ *(i.e., these correlation matrices are essentially negligible compared with the others) so the noise is uncorrelated with the solution. Then, show the following:*

1. $\mathcal{C}_{nn} \simeq \mathcal{C}_{tn}$;
2. $\mathcal{C}_{nn} = \mathcal{C}_{tt} - \mathbf{M}\mathcal{C}_{st} - \mathcal{C}_{ts}\mathbf{M}^T + \mathbf{M}\mathcal{C}_{ss}\mathbf{M}^T$;
3. $\mathcal{C}_{ts} \simeq \mathbf{M}\mathcal{C}_{ss}$;
4. $\mathcal{C}_{tt} = \mathbf{M}\mathcal{C}_{st} + \mathcal{C}_{nt}$.

Show that the best linear unbiased estimate (BLUE) of the solution based on the data set k is

$$\hat{\mathbf{s}}^{(k)} = \mathcal{C}_{st}\mathcal{C}_{tt}^{-1}\mathbf{t}^{(k)} \tag{3.31}$$

and that

$$\mathcal{C}_{st}\mathcal{C}_{tt}^{-1} \simeq \mathcal{C}_{ss}\mathbf{M}^T\left(\mathbf{M}\mathcal{C}_{ss}\mathbf{M}^T + \mathcal{C}_{nn}\right)^{-1}.$$

Contrast this result with the damped least-squares method assuming that the correlation matrix $\mathcal{C}_{nn}$ *is diagonal. The result (3.31) is known as the "stochastic inverse" [Franklin, 1970; Jordan and Franklin, 1971].*

3.5 Physical Basis for Weighted Least-Squares

So far we have treated the weights w_i as if they are arbitrary positive constants. But are they arbitrary? If they are not arbitrary, then what physical or mathematical feature of the inversion problem determines the weights?

Our goal is ultimately to solve (if possible) the nonlinear inversion problem, so we must keep in mind that the arguments often given for determining the weights in weighted least-squares schemes in other contexts may not be relevant to our problem. In particular, these weights are very often chosen on the basis of statistical (uncorrelated) errors in the data. The assumption behind these choices may be very good indeed in some cases, but generally not in the nonlinear inversion problem. Our working hypothesis for this analysis is that the major source of error in nonlinear inversion is not the measurement error, but the error due to the erroneous choices of ray paths currently in use in the algorithm. The statistical errors in the data become a significant issue only after we have constructed a reliable set of ray paths so that the errors due to wrong ray paths are smaller than the errors in the traveltime data. In fact, for high contrast reconstructions, it may be the case that the errors in the traveltime data are only a small fraction of one percent while the errors introduced by erroneous choices of ray paths are on the order of several percent, or even more in pathological cases.

We envision a two step process. First, we solve the inversion problem iteratively to find a good set of ray paths. This step requires the weighting scheme described here. Second, once we have a reliable set of paths, the weighting scheme can be changed to take proper account of the statistical errors in the data.

Now we use physical arguments to construct a proper set of weights [Berryman, 1989]. Suppose that the traveltime data in our reconstruction actually come from a model that is homogeneous, *i.e.*, with constant slowness σ_0. What will be the characteristics of such data? Clearly, the rays will in fact be straight and the average wave slowness along each ray will be the same constant

$$\sigma_0 = \frac{t_1}{L_1} = \frac{t_2}{L_2} = \cdots = \frac{t_m}{L_m}, \tag{3.32}$$

where

$$L_i = \sum_{j=1}^{n} l_{ij}. \tag{3.33}$$

Furthermore, it follows that the constant value of slowness is also given by the formula

$$\sigma_0 = \frac{\sum_{i=1}^m t_i}{\sum_{i=1}^m L_i}. \tag{3.34}$$

This problem is an ideal use for the scaled least-squares approach presented earlier. We know the ray paths are straight, so we know the ray-path matrix $\mathbf{M}$. We also know that the slowness has the form $\mathbf{s} = \gamma\mathbf{v}$, where $\mathbf{v}^T = (1, \ldots, 1)$ is an n-vector of ones. We want to minimize the least-squares error

$$\Psi(\gamma\mathbf{v}) = (\mathbf{t} - \mathbf{M}\gamma\mathbf{v})^T\mathbf{W}(\mathbf{t} - \mathbf{M}\gamma\mathbf{v}) \tag{3.35}$$

with respect to the coefficient γ. The minimum of (3.35) occurs for

$$\mathbf{v}^T\mathbf{M}^T\mathbf{W}(\mathbf{t} - \mathbf{M}\gamma\mathbf{v}) = 0. \tag{3.36}$$

Solving for γ gives

$$\sigma_0 = \gamma = \frac{\mathbf{v}^T\mathbf{M}^T\mathbf{W}\mathbf{t}}{\mathbf{v}^T\mathbf{M}^T\mathbf{W}\mathbf{M}\mathbf{v}}. \tag{3.37}$$

For easier comparison of (3.37) and (3.34), we now introduce some more notation. Define the m-vector of ones $\mathbf{u}^T = (1, \ldots, 1)$. Then,

$$\mathbf{M}\mathbf{v} = \mathbf{L}\mathbf{u} \tag{3.38}$$

and

$$\mathbf{M}^T\mathbf{u} = \mathbf{C}\mathbf{v}, \tag{3.39}$$

where $\mathbf{L}$ is a diagonal $m \times m$ *length matrix* whose diagonal elements are the row sums of $\mathbf{M}$ given by (3.33) and $\mathbf{C}$ is a diagonal $n \times n$ matrix whose diagonal elements are the column sums of $\mathbf{M}$ given by

$$C_{jj} = \sum_{i=1}^{m} l_{ij}. \tag{3.40}$$

In our later analysis, we will see that the matrix $\mathbf{C}$ (which we call the *coverage matrix*) is a good choice for the second weight matrix in damped least-squares (3.24).

Now we see that (3.37) can be rewritten in this notation as

$$\sigma_0 = \frac{\mathbf{u}^T\mathbf{LWt}}{\mathbf{u}^T\mathbf{LWLu}}, \tag{3.41}$$

while (3.34) becomes

$$\sigma_0 = \frac{\mathbf{u}^T\mathbf{t}}{\mathbf{u}^T\mathbf{Lu}}. \tag{3.42}$$

Comparing (3.41) to (3.42), we see that these two equations would be identical if

$$\mathbf{WLu} = \mathbf{u}. \tag{3.43}$$

Equation (3.43) states that $\mathbf{u}$ is an eigenvector of the matrix $\mathbf{WL}$ with eigenvalue unity. Two choices for the product $\mathbf{WL}$ are

$$\mathbf{WL} = \mathbf{I}, \tag{3.44}$$

where $\mathbf{I}$ is the identity matrix and

$$\mathbf{WL} = \mathbf{L}^{-1}\mathbf{MC}^{-1}\mathbf{M}^T. \tag{3.45}$$

The choice (3.45) is undesirable because it leads to a weight matrix that is not positive definite which would lead to spurious zeroes of the least-squares functional. The choice (3.44) leads to

$$\mathbf{W} = \mathbf{L}^{-1}, \tag{3.46}$$

which is both positive definite and diagonal.

The full significance of the result (3.46) becomes more apparent when we consider that the traveltime data $\mathbf{t} = \bar{\mathbf{t}} + \Delta\mathbf{t}$ will generally include some experimental error $\Delta\mathbf{t}$. If we assume the data are unbiased and the number of source/receiver pairs is sufficiently large, then to a good approximation we should have $\mathbf{u}^T\Delta\mathbf{t} = 0$. The result (3.41) can be rewritten as

$$\gamma = \frac{\mathbf{a}^T\mathbf{L}^{-1}\mathbf{t}}{\mathbf{a}^T\mathbf{u}}, \tag{3.47}$$

where $\mathbf{a} = \mathbf{LWLu}$ may be treated for these purposes as an arbitrary weighting vector. For γ to be unbiased, we must have

$$\mathbf{a}^T\mathbf{L}^{-1}\Delta\mathbf{t} = \mathbf{u}^T\Delta\mathbf{t} = 0. \tag{3.48}$$

Since the $\Delta\mathbf{t}$s are otherwise arbitrary, we must have

$$\mathbf{u} = \mathbf{L}^{-1}\mathbf{a} = \mathbf{WLu}, \tag{3.49}$$

which is the same condition as that found in (3.43). Thus, the choice (3.46) produces the simplest weight matrix giving a linear unbiased estimator of the scale factor for a constant slowness model. In Section 7.2, we derive weights producing unbiased estimates for arbitrary slowness.

Weighting inversely with respect to the lengths of the ray paths can be justified on physical grounds using several different arguments [Frank and Balanis, 1989]. Signal-to-noise ratio is expected to be better on shorter paths than longer ones, since the overall attenuation will typically be smaller and the likelihood of missing the true first arrival therefore smaller. Shorter trial paths are more likely to correspond to real paths that remain completely in the image plane for two-dimensional reconstruction problems.

A disadvantage of using this weighting scheme is that sometimes the ray path is long because the source and receiver are far apart (*e.g.*, from the top of one borehole to the bottom of the other). Yet the information contained in the ray is important because such diagonal rays may help to determine the horizontal extent of some feature of interest, especially when the experimental view angles are severely limited as in crosshole tomography. Weighting inversely with respect to the ray-path length tends to reduce the possibly significant improvement in horizontal resolution that can come from inclusion of these rays. This disadvantage can be circumvented to some extent by using more of these diagonal rays, *i.e.*, using more closely spaced sources and receivers for the diagonal rays. Then, the weights of the individual rays are smaller, but their overall influence on the reconstruction can still be significant.

In Section 4.3, we show that an argument based on stability and regularization leads to the same choice of weight matrices.

3.6 Partial Corrections Using Backprojection

Suppose we have found a solution $\hat{\mathbf{s}}$ of the overdetermined ($m > n$) normal equations

$$\mathbf{M}^T\mathbf{M}\hat{\mathbf{s}} = \mathbf{M}^T\mathbf{t}, \tag{3.50}$$

but this solution does not satisfy the data exactly so

$$\mathbf{M}\hat{\mathbf{s}} \neq \mathbf{t}. \tag{3.51}$$

Then, we argue that a correction $\Delta\mathbf{s}$ could be added to $\hat{\mathbf{s}}$ and the correction should satisfy

$$\mathbf{M}\Delta\mathbf{s} = \Delta\mathbf{t} \equiv \mathbf{t} - \mathbf{M}\hat{\mathbf{s}}. \tag{3.52}$$

Now suppose further that for some subset of the ray paths either $\Delta t_i = 0$, or we are satisfied for some other reason with the agreement between the predicted and measured data (*e.g.*, $|\Delta t_i| \leq \epsilon$ for some small threshold ϵ, or ray path i corresponds to a feasible ray path with $\Delta t_i \leq 0$). Then, we may want to make corrections using only the ray paths that are considered unsatisfactory. We renumber the ray paths so the unsatisfactory ones are the first m' of the m total paths and suppose $m' < n$. Next we rewrite (3.52) as

$$\mathbf{M}'\Delta\mathbf{s} = \Delta\mathbf{t}', \tag{3.53}$$

where $\mathbf{M}'$ is an $m' \times n$ matrix and $\Delta\mathbf{t}'$ is the corresponding m'-vector of unsatisfactory traveltime errors. The problem of solving for the Δs_js is underdetermined as stated.

We can solve (3.53) using a type of backprojection. We argue that the correction vector component $\Delta\mathbf{s}_j$ should be a sum whose terms are proportional to l_{ij} (so that rays not passing through cell j make no contribution) and it should be a linear combination of the traveltime errors Δt_i. However, these corrections should also be made in a way that minimizes the overall effect on the agreement already attained in (3.50). One way to do this approximately is to weight inversely with respect to the cell coverage C_{jj}; then, the cells with the most coverage will change the least and therefore the result should have the smallest effect on (3.50). This argument results in a general form for the correction

$$\Delta s_j = C_{jj}^{-1} \sum_{k'k}^{m'} l_{k'j} w_{k'k} \Delta t_k, \tag{3.54}$$

where $w_{k'k}$ is some weight matrix to be determined by substituting (3.54) into (3.53). On making the substitution, we find that

$$\sum_{j=1}^{n} \sum_{kk'}^{m'} \left(l_{ij} C_{jj}^{-1} l_{k'j} \right) w_{k'k} \Delta t_k = \Delta t_i, \tag{3.55}$$

which implies that

$$\sum_{i=1}^{m'} \left(\sum_{j=1}^{n} l_{ij} C_{jj}^{-1} l_{k'j} \right) w_{k'k} = \delta_{ik}, \tag{3.56}$$

showing that $w_{k'k}$ is the inverse of the matrix

$$\left[\mathbf{M}'\mathbf{C}^{-1}(\mathbf{M}')^T \right]_{kk'} = \sum_{j=1}^{n} l_{kj} C_{jj}^{-1} l_{k'j}. \tag{3.57}$$

Thus, we have

$$\Delta\mathbf{s} = \mathbf{C}^{-1}(\mathbf{M}')^T \left[\mathbf{M}'\mathbf{C}^{-1}(\mathbf{M}')^T \right]^{-1} \Delta\mathbf{t}'. \tag{3.58}$$

It is straightforward to verify that (3.58) is a formal solution by substituting it into (3.53).

To use (3.58) in general requires that the inverse of the $m' \times m'$ matrix $\mathbf{M}'\mathbf{C}^{-1}(\mathbf{M}')^T$ must always exist. This may not always be true, but there is one particularly simple case where the formula can be evaluated: $m' = 1$. Then,

$$\Delta s_j = \frac{l_{1j} \Delta t_1 / C_{jj}}{\sum_{k=1}^{n} l_{1k}^2 / C_{kk}}. \tag{3.59}$$

Problems

Problem 3.6.1 *Let* $\mathbf{D}$ *be an arbitrary* $n \times n$ *positive diagonal matrix. Show that another set of model corrections can be chosen to be*

$$\Delta\mathbf{s} = \mathbf{D}^{-1}(\mathbf{M}')^T \left[\mathbf{M}'\mathbf{D}^{-1}(\mathbf{M}')^T\right]^{-1} \Delta\mathbf{t}'. \tag{3.60}$$

Other than $\mathbf{C}$*, what are some physically relevant choices of the weight matrix* $\mathbf{D}$*?*

Problem 3.6.2 *Consider the stochastic inverse (3.31) when the noise correlation is negligible so that* $C_{nn} \simeq 0$*. Compare the resulting formula with (3.58) and (3.60).*

Problem 3.6.3 *Using (3.58), find an explicit expression for* Δs_j *when slownesses along only two ray paths need correction, i.e.,* $m' = 2$*. Show that the required matrix inverse exists for this problem if*

$$\left(\sum_{j=1}^{n} l_{1j}l_{2j}/C_{jj}\right)^2 < \left(\sum_{j=1}^{n} l_{1j}^2/C_{jj}\right)\left(\sum_{k=1}^{n} l_{2k}^2/C_{kk}\right). \tag{3.61}$$

Use Cauchy's inequality for sums to show that (3.61) is always satisfied unless $l_{1j} = \gamma l_{2j}$ *for all* j*, where* $\gamma > 0$ *is some scalar. Explain the physical significance of the special case when* $l_{1j} = \gamma l_{2j}$ *and suggest a method of solving the problem in this case.*

Problem 3.6.4 *Using (3.58), find an explicit expression for* Δs_j *when slownesses along three ray paths need correction* ($m' = 3$)*. Determine conditions on the matrix elements necessary to guarantee that the required matrix inverse exists.*