On anisotropy and inhomogeneity parameter estimation using traveltimes

Ayiaz Kaderali*, Theodore Stanoev†

Abstract

We consider an anisotropic inhomogeneous model to simulate measured vertical-seismic-profile traveltimes. In this model, we assume that velocity increases linearly with depth and anisotropy is the result of elliptical velocity dependence. Using a series of sources in a line to a single receiver, we minimize the least-squares residual between measured and modelled traveltimes to estimate the anisotropy and inhomogeneity parameters of the subsurface. To verify the approach, we construct synthetic traveltime data for a one- and two-layer model and perform the traveltime inversion. We justify the convergence-ensuring specifications and model-parameter restrictions that are intrinsic to the parameter estimation. To determine the approach’s practicality, we assess the reliability of results under the influence of noise. From this assessment, we discern the noise threshold for both models and determine that the noise restriction is severely restrictive for increasing model complexity. We estimate parameters, using the same methodology, for a real-data case and conclude with a discussion of results.

1 Introduction

It has been said that seismology is the most successful geophysical technique for learning about the Earth’s interior (Aki and Richards, 2002, Section 9.4). A common approach used by seismologists to gain information about the stratified subsurface of the Earth is traveltime inversion.

To perform such an inversion, seismologists require at least two mathematical formalisms. The first is the forward problem, which, in the case of this paper, is derived using ray theory. As such, the path of the ray that connects a source and receiver must obey Fermat’s principle of stationary time. The model must account for the material properties of the subsurface, which is satisfied by considering a velocity profile in the form of interval velocities. The second is the quantitative modelling, which, through a numerical procedure, alters the model parameters of the forward problem to resemble a set of physical observations. We refer to the second formalism, as the inverse problem, sensu lato. The choice of numerical procedure depends greatly on the structure of the forward-problem model. In this paper, we consider an inverse problem of the least-squares type, which, in the case of traveltime inversion, seeks to minimize the difference between modelled and measured traveltimes. The optimization technique that we consider is Newton’s method in optimization, which is discussed in detail in textbooks of numerical optimization (e.g., Nocedal and Wright (1999, Chapter 6) and Fletcher (2000, Chapter 3)).

In seismology, the combination of these mathematical formalisms forms the basis of many studies. From Dix (1955) calculating interval velocities to Stewart (1984) using least-squares optimization techniques of traveltimes to refine interval-velocity accuracy, seismology has been used routinely to estimate layer properties of the stratified subsurface of the Earth. Further, Pujol et al. (1985) formulate multilayer ray-tracing equations for traveltime calculation of surface shot gathers. Therein, they consider a linearized solution for the Taylor series expansion of the expressions of traveltimes and offsets, in terms of the speed and ray parameter, are truncated after the first order. The resulting linear system is solved using the Levenberg-Marquardt algorithm, which is an iterative approach that uses the information from the gradient to determine the next
iterate. Similarly, Mao and Stuart (1997) consider a ray-tracing multilayer model with laterally heterogeneous layers for the purpose of transmission and reflection tomography. Therein, to minimize the misfit of their traveltime expressions, they truncate the Taylor series expansion after the first order. Consequently, they use the Gauss-Newton algorithm, which approximates the Hessian using first-order derivatives, to minimize the traveltime misfit.

Another specification of ray theory is to model media whose velocity changes not only with depth, but also with direction, which we refer to as an anisotropic inhomogeneous medium. Indeed, both the anisotropy and inhomogeneity of physical material have an intertwined effect on the propagation of seismic waves. However, within the framework of mathematics, it is entirely possible to quantify their effects separately. To this end, there is a wealth of research on the effects of anisotropy and inhomogeneity, of which we highlight Červený (1989, 2001), Winterstein (1990), Helbig (1994), Slawinski (2020a,b). For the purposes of inversion, it is hopeless to attempt to invert for the model parameters that govern the symmetry class of general anisotropy. Instead, as indicated by Červený (1989, p. 91), it is advisable to consider simpler forms of anisotropy, such as elliptically anisotropic media.

One model of this type is that of Slawinski et al. (2004), wherein the defining characteristics are that the velocity of a propagating wave increases linearly with depth and exhibits an elliptical velocity dependence along the direction of propagation. They demonstrate that the statistical significance of the anisotropy parameter, $\chi$, is small, but significant, and remain consistent with the weakly anisotropic assumptions discussed by Thomsen (1986). The model is extended by Rogister and Slawinski (2005), who use Hamilton’s equations to obtain the ray-tracing equations and derive analytic solutions for the traveltime equations, which are valid for both upgoing and downgoing ray paths. This traveltime model has been the topic of subsequent studies during the past decade.

Let us call attention to a few such studies. Danek and Slawinski (2012) use the model to perform a stochastic interpretation of geophysical data to justify the inclusion of the anisotropy parameter by means of the Bayesian Information Criterion. Therein, to perform the traveltime inversion, they use the Metropolis-Hastings algorithm, which is a random walk algorithm that perturbs model parameters numerically to reduce the traveltime misfit. In applying a one-layer model to synthetic and real traveltime data, they demonstrate the necessity of the anisotropy parameter to model adequately the observed traveltimes. Zhang et al. (2017) approximate the model to calculate numerically ray paths through laterally homogeneous and heterogeneous velocity models. By imposing triangular grids on the model subsurface, they solve the equations of motion intrinsic to the problem using the Runge-Kutta method. Gierlach and Danek (2018) construct two synthetic, multilayer setups, each containing one anisotropic layer. Using the Nelder-Mead simplex algorithm, they use the traveltime model to invert for the layer properties and, using Bayesian Information Criterion, they determine the optimal number of model parameters to model adequately the traveltimes under the influence of noise.

In this paper, we consider the mathematical model developed by Slawinski et al. (2004), and extended by Rogister and Slawinski (2005), to estimate anisotropy and inhomogeneity parameters of the subsurface using vertical seismic profile traveltimes. Its structure is as follows. We begin by adapting the traveltime equations for multilayered models, providing an overview of Newton’s method and its convergence-ensuring specifications, and establishing the requirement of logarithmic barrier functions to guarantee model-parameter restrictions. Then, we construct synthetic traveltime models to assess the reliability of results of the optimization under the influence of six noise profiles. Next, we repeat the optimization for a one- and two-layer model and determine the threshold of noise that yields acceptable results for each model, which we assess using relative errors between the estimated and true model parameters. Finally, we perform the optimization on a real-data example and conclude with a discussion of the results. Also, this paper includes an appendix that exemplifies the control experiment of the optimization technique as well as the details of the convergence specifications and parameter restrictions.
2 Formulation

2.1 Mathematical model

Let us consider a model of $N$ horizontal layers each exhibiting anisotropy and inhomogeneity.\(^1\) The inhomogeneity is a linear increase of velocity, $V(z) = a + bz$, where $a$ is the vertical speed of the wavefront at the top of a layer, $b$ is the measure of inhomogeneity within that layer, and $z$ is the vertical component that corresponds to depth. The anisotropy is an elliptical velocity dependence of a wavefront (Slawinski et al., 2004),

$$\chi = \frac{v_h^2 - v_v^2}{2v_v^2}, \quad (1)$$

where $v_h$ and $v_v$ are the horizontal and vertical velocity; for $v_h = v_v$, $\chi = 0$ and, hence, the wavefront velocity is isotropic.

We consider a series of $M$ point sources aligned in the $xz$-plane along the surface of the multilayered model and a single receiver. Along the $j$th ray, the coordinates at the $\ell$th layer interface are given by the sequence

$$S_j = \{(x_{\ell,j}, z_{\ell})\}_{\ell=1}^{N+1} \quad \text{for} \quad j \in \{1, \ldots, M\}; \quad (2)$$

we denote $S$ as the $M \times 1$ vector of sequences. As such, the source coordinates are $(x_{1,j}, z_1)$, where $x_{1,j}$ are the horizontal offsets and the vertical position $z_1 = 0$ for all $M$ sources. The receiver coordinates are $(x_{N+1,j}, z_{N+1})$, where $x_{N+1,j} = 0$ by design and $z_{N+1}$ is the depth of the receiver. For the $i$th layer, $a_i$ is the vertical speed at the top of the layer, $b_i$ is the linear increase in vertical speed within the layer, and $\chi_i$ is the anisotropy of the layer. Herein, $b_i > 0$ corresponds to the increasing compaction of the subsurface with depth, which increases the speed of a propagating wave. In view of physical considerations of applying the model to shale, we require $\chi_i > 0$ due to the increase in the horizontal component of speed of a propagating wave, which is a directional property of shale.

Thus, the traveltime of a wave propagating along the $j$th ray is

$$t(S_j; \beta) = \sum_{i=1}^{N} \frac{\arctanh \left( p_{i,j} b_i (x_{i+1,j} - x_{i,j}) - \sqrt{1 - p_{i,j}^2 a_i^2 (1 + 2 \chi_i)} \right) + \arctanh \sqrt{1 - p_{i,j}^2 a_i^2 (1 + 2 \chi_i)}}{b_i}, \quad (3)$$

where $\beta = \left[\left\{a_i\right\}_{i=1}^{N}, \left\{b_i\right\}_{i=1}^{N}, \left\{\chi_i\right\}_{i=1}^{N}\right]^t$ is the $3N \times 1$ vector of model parameters and

$$p_{i,j} = \frac{2 (x_{i+1,j} - x_{i,j})}{\sqrt{(x_{i+1,j} - x_{i,j})^2 + (1 + 2 \chi_i) (z_{i+1} - z_i)^2} \left(2 a_i + b_i (z_{i+1} - z_i) \right)^2 (1 + 2 \chi_i) + b_i^2 (x_{i+1,j} - x_{i,j})^2}}, \quad (4)$$

is the ray parameter in the $i$th layer, which is a conserved quantity along the ray. Herein, we denote $t := t(S; \beta)$ as the $M \times 1$ vector of modelled traveltimes. Also, let us remark that the symbol $\beta$ should not be confused with $S$-wave speeds, which is a common use of the symbol in seismology. To calculate the traveltime along the $j$th ray, we require a complete sequence $S_j$, for which the coordinates of the $N - 1$ intermediate $x_{\ell,j}$ are yet unknown. As such, for a $N$-layer model, we form the system of equations

$$\begin{align*}
p_{1,j} - p_{2,j} &= 0 \\
p_{2,j} - p_{3,j} &= 0 \\
&\vdots \\
p_{N-1,j} - p_{N,j} &= 0
da \in \{2, \ldots, N\}.
\end{align*} \quad (5)$$

\(^{1}\)Throughout this paper, in keeping with SI units, the units for $a$ and $b$ are m s\(^{-1}\) and s\(^{-1}\), respectively, for traveltimes are s, for the ray parameter are m\(^{-1}\) s.
To determine a local solution of problem (6), we require an algorithm to generate a sequence of iterates

$$
\alpha
$$

is the Newton direction, which relies on second-order Taylor series approximation of \( f(\beta) \), of \( f(\beta) \), exist and are continuous, whose components are

$$
g(\beta)_i = \frac{\partial f(\beta)}{\partial \beta_i} \quad \text{and} \quad H(\beta)_{i,j} = \frac{\partial^2 f(\beta)}{\partial \beta_i \partial \beta_j};
$$

the former is a 3\( N \times 1 \) vector and the latter is a 3\( N \times 3N \) matrix. We use Newton’s method in optimization, which relies on second-order Taylor series approximation of \( f(\beta) \) at \( \beta = \beta^{(k)} \), to form a quadratic approximation to problem (6), such that

$$
f(\beta^{(k)} + \alpha \Delta \beta^{(k)}) = f(\beta)|_{\beta = \beta^{(k)}} + (\alpha \Delta \beta^{(k)})^T g(\beta)|_{\beta = \beta^{(k)}} + \frac{1}{2} (\alpha \Delta \beta^{(k)})^T H(\beta)|_{\beta = \beta^{(k)}} (\alpha \Delta \beta^{(k)}),
$$

where

$$
\Delta \beta^{(k)} = - \left( H(\beta)|_{\beta = \beta^{(k)}} \right)^{-1} g(\beta)|_{\beta = \beta^{(k)}},
$$

is the Newton direction, \( \alpha \in [0,1] \) is the step size, and \( k \) is an iterative superscript. To obtain the Newton direction, we set the derivative, with respect to \( \Delta \beta^{(k)} \), of expression (8) to zero and solve for \( \Delta \beta^{(k)} \), where we assume \( H(\beta) \) is positive definite. Henceforth, we adopt the notation

$$
f^{(k)} := f(\beta)|_{\beta = \beta^{(k)}} , \quad g^{(k)} := g(\beta)|_{\beta = \beta^{(k)}} , \quad H^{(k)} := H(\beta)|_{\beta = \beta^{(k)}},
$$

To determine a local solution of problem (6), we require an algorithm to generate a sequence of iterates \( \{ \beta \}^\infty_{k=0} \) that minimize the traveltime residual. The value of \( f \) at each iterate must be such that \( f^{(k+1)} < f^{(k)} \), or, in other words, the sequence of iterates must lead to a strict decrease in \( f \). The algorithm is terminated when either a sufficient residual value is achieved or the condition of strict decrease is not met. To generate the sequence of iterates, we consider a line search Algorithm 1, such that

$$
\beta^{(k+1)} = \beta^{(k)} + \alpha^{(k)} \Delta \beta^{(k)},
$$

which computes a step size \( \alpha^{(k)} \) along the direction \( \Delta \beta^{(k)} \) to minimize \( f \).

For Algorithm 1 to produce a sequence of iterates that are globally convergent, the search direction \( \Delta \beta^{(k)} \) must be a descent direction, which is true if \( H^{(k)} \) is positive definite. However, as indicated by Nocedal and Wright (1999, p. 136), if \( H^{(k)} \) is not positive definite, or near singular, the Newton direction may lead to ascending values of \( f \) or may be excessively long. To ensure positive definiteness of \( H^{(k)} \), we execute the modified Cholesky factorization algorithm on \( H^{(k)} \) prior to calculating the Newton direction. Details of this algorithm are outlined in Nocedal and Wright (1999, p. 145–149), along with its use in this paper in Appendix A.2 and its necessity in Appendix A.3.

### 2.3 Logarithmic barrier function

The solution to problem (6) is vector \( \hat{\beta} \), which is referred to as the estimator of problem (6). In unconstrained optimization, there are no restrictions on the model-parameter values that comprise \( \hat{\beta} \). However, to maintain physically meaningful results in fit-for-purpose applications of minimizing traveltime residuals, it is reasonable to impose restrictions on some, if not all, model parameters. To do so, we introduce a logarithmic barrier function to \( f \), which changes problem (6) to an inequality-constrained optimization problem. Thus, we have

$$
\min_{\beta} f(\beta) \quad \text{subject to} \quad c_i \geq 0, \quad i \in I,
$$

---

---
Algorithm 1: Backtracking line search

For $\beta^{(k)}$, calculate $f^{(k)}$, $g^{(k)}$, $H^{(k)}$, $\Delta \beta^{(k)}$. Then, set $r = f^{(k)}$, $r' = r$, $\alpha' = 0$, $\Delta \alpha = 0.1$;

while $r' = r$ do

$\beta' = \beta^{(k)} + \alpha' \Delta \beta^{(k)}$;
$r = f(\beta')_{|\beta' = \alpha'}$;
if $r < r'$ then

$r' = r$;
else

if $\alpha' - 2\Delta \alpha > 0$ then

$\alpha' = \alpha' - 2\Delta \alpha$;
else

$\alpha' = 0$;
end
$\Delta \alpha = \Delta \alpha / 10$;
$\beta' = \beta^{(k)} + \alpha' \Delta \beta^{(k)}$;
$r = f(\beta')_{|\beta' = \alpha'}$;
$r' = r$;
end
$\alpha' = \alpha' + \Delta \alpha$;
if $\Delta \alpha < 1 \times 10^{-12}$ then

break;
end
if $\alpha' \geq 1$ then

break;
end

$\alpha^{(k)} = \alpha'$;
$\beta^{(k+1)} = \beta^{(k)} + \alpha^{(k)} \Delta \beta^{(k)}$;

where $c_i$ is the set of inequality constraints and $\mathcal{I}$ is the set of integers. The objective function for problem (12) is

$$ P(\beta) = f(\beta) - \sum_{i \in \mathcal{I}} \log l_i, \quad (13) $$

where the summand of the latter term in $P$ is the logarithmic barrier function. In a similar manner to notation (10), we let $P^{(k)}$ denote value of the objective function at the $k$th iteration of the optimization—it is analogous to the $k$th value of $f^{(k)}$ for problem (6). Furthermore, in the context of problem (12), we abuse notation (7) by stating equivalently that the components of the gradient and Hessian of $P(\beta)$ are

$$ g(\beta)_i = \frac{\partial P(\beta)}{\partial \beta_i} \quad \text{and} \quad H(\beta)_{i,j} = \frac{\partial^2 P(\beta)}{\partial \beta_i \partial \beta_j}. \quad (14) $$

According to Nocedal and Wright (1999, Section 17.2), the barrier functions have the properties such that they are infinite everywhere except in a particular domain, smooth inside the domain, and their value approaches infinity at the boundary of the domain. To satisfy these properties, we consider the logistic function,

$$ l(x) = \frac{1}{1 + \exp (-r(x - x_0))}, \quad (15) $$

where $r$ is the growth rate and $x_0$ is its center. For large $r$, equation (15) resembles a step function about $x_0$ with a minimum value of zero and a maximum value of one. Since $\log(0) = -\infty$ and $\log(1) = 0$, we exploit (15) to restrict the values of model parameters. We provide further details on the use of the logistic function as a logarithmic barrier function in Appendix A.3.
3 Synthetic-data testing

Let us assess the reliability of results in using Newton’s method to solve problem (13). For consistency, we retain the same source-receiver combinations across all executions of the algorithm. We specify $M = 139$ point sources along the surface, with offsets ranging from approximately 80 m to 3300 m, at intervals of approximately 25 m, and a single receiver at a depth of approximately 1850 m. To ensure physically meaningful results, we use the logarithmic barrier function to restrict model parameters such that $b_i > 0$ and $\chi_i > 0$.

First, we set up the following control experiment. For a known set of model parameters, $\beta$, we generate sequences $S$ and calculate traveltimes $t(S; \beta)$, which we designate as the measured traveltimes $T$. Then, from a set of starting values, $\beta^{(1)}$, we perform Newton’s method to obtain the sequence of iterates $\{\beta^{(k)}\}$ that minimize $f$ and, hence, solve problem (13), whose solution is the estimator vector, $\hat{\beta}$. We quantify the “goodness of fit” by calculating the relative errors, $\delta(\hat{\beta}) := ((\hat{\beta} - \beta)/\beta) \times 100\%$ between the observed $\beta$ and expected $\beta$.

Second, we assess the reliability of results under the influence of random noise, which serves to simulate errors in the traveltimes data. The range of the random noise is set at a constant percentage of the measured range of the noise increases by a fixed percentage. Within the fixed-percentage range, we generate the noise profile using Matlab’s `rand` function, which returns uniformly distributed random numbers within a determined interval. We generate six noise profiles for which the fixed-percentage range increases by a factor of 10 from $\pm 0.000001\%$ to $\pm 0.1\%$. As the range increases for each noise profile, the magnitude of noise within the noise profile also increases. The corresponding mode of the orders of magnitude increases by a factor of 10 from nanoseconds ($10^{-9}$) to milliseconds ($10^{-3}$). Throughout the assessment, we use the six noise profiles that are plotted in Figure 1. Similarly to the control experiment, we quantify the goodness of fit of estimator $\hat{\beta}$.

Third, we perform the control and noise-assessment experiments for a one- and two-layer model. This process is repeated on eight sets of model parameters, with increasing anisotropy. We distinguish one set of model parameters from another using notation $[\beta]_s$ for $s \in \{1, \ldots, 8\}$ and their corresponding estimators by $[\hat{\beta}]_s$. For the one-layer model, we have

$$[\beta]_1 = \begin{bmatrix} 1500 \\ 0.75 \\ 0.0015 \end{bmatrix}, \quad [\beta]_2 = \begin{bmatrix} 1500 \\ 0.75 \\ 0.0408 \end{bmatrix}, \quad [\beta]_3 = \begin{bmatrix} 1500 \\ 0.75 \\ 0.0832 \end{bmatrix}, \quad [\beta]_4 = \begin{bmatrix} 1500 \\ 0.75 \\ 0.1728 \end{bmatrix}$$

for the two-layer model, we have

$$[\beta]_5 = \begin{bmatrix} 911 \\ 3285 \\ 1.5 \\ 0.5 \\ 0.0015 \\ 0.0019 \end{bmatrix}, \quad [\beta]_6 = \begin{bmatrix} 911 \\ 3285 \\ 1.5 \\ 0.5 \\ 0.0408 \\ 0.0618 \end{bmatrix}, \quad [\beta]_7 = \begin{bmatrix} 911 \\ 3285 \\ 1.5 \\ 0.5 \\ 0.0832 \\ 0.1272 \end{bmatrix}, \quad [\beta]_8 = \begin{bmatrix} 911 \\ 3285 \\ 1.5 \\ 0.5 \\ 0.1728 \\ 0.2688 \end{bmatrix}$$

As we recall from equation (1), the value of anisotropy corresponds to the difference between the horizontal and vertical velocity. To quantify the amount of anisotropy, we set four categories based on the relative percentage increase in horizontal to vertical velocity. They are: near isotropy, ranging from 0% up to 1%; low anisotropy, from 1% up to 7%; moderate anisotropy, from 7% up to 15%; and high anisotropy, from 15% onward. For model parameter sets (16) and (17), the anisotropy category to which each belongs is indicated by its respective underbraces.

We restrict model-parameter sets to such values to remain consistent with reasonable values of anisotropy exhibited in shale (e.g., Vernik and Liu, 1997, pp. 531-532). Finally, we tabulate the observed $[\beta]_s$ and their corresponding $\delta([\beta]_s)$ as well as discuss the reliability of results.
3.1 One-layer optimization

For the one-layer model, we seek to solve problem (13), for which

\[ l_1(b_1) := \frac{1}{1 + \exp(-10^6 b_1)} \quad \text{and} \quad l_2(\chi_1) := \frac{1}{1 + \exp(-10^6 \chi_1)} . \] (18a,b)

To do so, first, we ensure the convergence of the control experiment on model parameters (16), for which we require the use of Hessian modifications and logarithmic barrier functions, which are explained in Sections 2.2 and 2.3, respectively, and justified in Appendix A. Second, we repeat the optimization on model parameters (16) under the influence of the six noise profiles plotted in Figure 1. We tabulate the resulting estimating vectors and their relative errors in Table 1. Therein, we observe two prominent trends concerning the relative errors.

The first trend is the increase in relative error corresponding to the increase in the range of the noise profile. We observe that for each increase in the order of magnitude of the noise profile, there is an increase by approximately one order of magnitude of the relative errors of the model parameters. The second trend is the decrease in relative error corresponding to the increase in anisotropy. We observe that, for the nearly isotropic case, the relative-error ordering is \( \delta(a_1) < \delta(b_1) < \delta(\chi_1) \). In contrast, as the amount of anisotropy increases, the relative error in \( \chi_1 \) diminishes to approximately the same level in \( a_1 \), with the largest relative error manifesting in \( b_1 \). Thus, we state that, for the nearly isotropic case, the relative error manifests disproportionately in the anisotropy parameter whereas, for increasing levels of anisotropy, the relative error
is comprised predominantly of shale will yield reliable results. We expect this result because, for low anisotropy, small errors in $\chi_1$ greatly affect the modelled traveltimes, which is not the case for high anisotropy. In other words, small errors in $\chi_1$ greatly affect the modelled traveltimes in low-anisotropy cases, but have less of an effect for increasing levels of anisotropy.

Along with these trends, we determine that the reliability of results is satisfactory for fixed-percentage noise ranges of at least $\pm 0.1\%$. This percentage corresponds to noise-profile values on the order of milliseconds, which is the same order of magnitude associated with errors in picking traveltimes in real-data cases. Thus, in the context of optimization, applying a one-layer model to traveltime data in regions where the subsurface is comprised predominantly of shale will yield reliable results.

| +0.000001% | ±0.00001% | ±0.001% | ±0.01% | ±0.1% |
|------------|-----------|---------|--------|-------|
| $\hat{\beta}_1$ | 1500.00005 | 1499.99990 | 1499.99866 | 1499.98059 | 1500.12664 | 1498.25090 |
| 0.74999993 | 0.75000015 | 0.75000198 | 0.75002729 | 0.74983824 | 0.75254910 |
| ±0.001% | ±0.00001% | ±0.0001% | ±0.01% | ±0.1% |
| $\hat{\beta}_2$ | 1500.00005 | 1499.99988 | 1499.99842 | 1499.97821 | 1500.13668 | 1497.96923 |
| 0.74999993 | 0.75000017 | 0.7500231 | 0.75003055 | 0.74982463 | 0.75294017 |
| ±0.001% | ±0.00001% | ±0.0001% | ±0.01% | ±0.1% |
| $\hat{\beta}_3$ | 1500.00006 | 1499.99987 | 1499.99815 | 1499.97551 | 1500.14766 | 1497.64443 |
| 0.74999992 | 0.75000020 | 0.75000269 | 0.7500442 | 0.74989072 | 0.75339073 |
| ±0.001% | ±0.00001% | ±0.0001% | ±0.01% | ±0.1% |
| $\hat{\beta}_4$ | 1500.00008 | 1499.99983 | 1499.99749 | 1499.96937 | 1500.17137 | 1496.88716 |
| 0.74999989 | 0.75000025 | 0.75000306 | 0.75004267 | 0.74977744 | 0.75443983 |
| ±0.001% | ±0.00001% | ±0.0001% | ±0.01% | ±0.1% |

Table 1: Tabulated results of performing the inequality-constrained Newton’s method in a one-layer model for $[\hat{\beta}_1, \ldots, \hat{\beta}_4]$, for increasing ranges of constant-percentage noise profiles.

### 3.2 Two-layer optimization

For the two-layer model, we seek to solve problem (13), for which we require $b_i > 0$ and $\chi_i > 0$, for $i = 1, 2$. As such, we require logarithmic barrier functions (18), along with another two for the second layer. Then, we repeat the control experiment on model parameters (17), in the same manner described in Appendix A. As with the one-layer case, we repeat the optimization under the influence of the six noise profiles plotted in Figure 1 and tabulate the resulting estimating vectors, and their relative errors, in Table 2. Therein, we observe the same trends concerning the relative errors described in Section 3.1—increasing noise-profile values result in increasing relative errors of model parameters, and increasing anisotropy results in greater error for all model parameters as opposed to anisotropy parameters only.

Pertaining to the relative errors in Table 2, the threshold beyond which estimator vectors $\hat{\beta}$ cease to approximate the model parameters is between fixed-percentage ranges $\pm 0.001\%$ and $\pm 0.01\%$. The corresponding modes of the orders of magnitude are microseconds and tens of microseconds, respectively, which are several orders of magnitude less than the one-layer-optimization threshold. In contrast to the one-layer model, the increase in the number of model parameters from three to six reduces drastically the reliability of results for
increasing ranges of constant-percentage noise profiles. With reference to optimization, this is a substantial limitation in applying a two-layer model to real traveltime data.

Let us go further and consider a three-layer model; this would increase the number of model parameters to nine. The increase from a one- to two-layer model reduces the reliability-of-results threshold from greater than the order of milliseconds to the order of microseconds. We postulate that there would be—at least—a similar threshold reduction upon increasing the model complexity to three layers. Apart from the fact that nine-parameter optimization may lead to many local minima and nonphysical results, the severe noise restrictions that beset the three-layer model render its optimization nonviable, especially under the influence of the amount of noise that may be present in real datasets, which we exemplify in Section 4. Thus, for pragmatic considerations, we do not consider the application of three-layer models to real traveltime data.

| ±0.000001% | ±0.00001% | ±0.0001% | ±0.001% | ±0.1% |
|------------|-----------|----------|---------|------|
| [β₅]       |           |          |         |      |
| 0.00149763 | ±0.00001% | ±0.0001% | ±0.001% | ±0.1% |
| 0.00050928 | ±0.00001% | ±0.0001% | ±0.001% | ±0.1% |
| 0.00015943 | ±0.00001% | ±0.0001% | ±0.001% | ±0.1% |
| 0.00027780 | ±0.00001% | ±0.0001% | ±0.001% | ±0.1% |
| 0.00032869 | ±0.00001% | ±0.0001% | ±0.001% | ±0.1% |
| 0.00035276 | ±0.00001% | ±0.0001% | ±0.001% | ±0.1% |
| 0.00018697 | ±0.00001% | ±0.0001% | ±0.001% | ±0.1% |
| 0.00031064 | ±0.00001% | ±0.0001% | ±0.001% | ±0.1% |
| 0.00027860 | ±0.00001% | ±0.0001% | ±0.001% | ±0.1% |
| 0.00015094 | ±0.00001% | ±0.0001% | ±0.001% | ±0.1% |

Table 2: Tabulated results of performing the inequality-constrained Newton’s method in a two-layer model for [β₅], ..., [β₈], for increasing ranges of constant-percentage noise profiles.
4 Real-data example

Let us perform the optimization on a set of measured traveltimes, $T$, obtained in offshore Newfoundland (Kaderali, 2009). The purpose is to use traveltime model (3) to solve problem (13) and, in turn, quantify the anisotropy and inhomogeneity of the subsurface in the region.

The subsurface unit in question is comprised predominantly of shale units that can be amalgamated effectively to a single unit. We are able to corroborate this statement by considering the interval velocities corresponding to the dataset, which we plot in Figure 2. Therein, we observe a steady linear increase in velocity throughout the entirety of the 2 km subsurface. As such, for this dataset, we consider a one-layer model to be an adequate, albeit low-resolution, representation of the subsurface.

![Figure 2: Interval velocities with depth](image)

Further inspection of Figure 2 reveals that there are zones of changing interval velocities, namely at depths $z = 1212 \text{ m}$ and $z = 1585 \text{ m}$. Upon this basis, it is reasonable to apply a multilayer model that segments the subsurface, depending on the desired level of resolution. Thus, for the two-layer model, we bisect the subsurface and set the layer interface at depth $z_1 = 1212 \text{ m}$. We select this layer-interface depth instead of $z = 1585 \text{ m}$ to avoid creating a model with a single layer that dominates the effect of the modelled traveltimes.

Prior to performing the optimizations, we must assess the amount of noise present in the real data. To this end, we apply a three-point median filter on the traveltimes and calculate the difference between $T$ and the filtered traveltimes to obtain the noise profile. We plot the noise profile in Figure 3 and, through visual inspection, we conclude that the amount of noise is on the order of milliseconds. We calculate that the mode of the orders of magnitude of the noise profile is $10^{-4}$, which indicates that the most frequent noise-profile value is on the order of a fraction of a millisecond.

As we recall from Sections 3.1 and 3.2, the reliability-of-results threshold for one-layer optimization is greater than milliseconds and for two-layer optimization is on the order of microseconds. As such, we recognize that, in the context of the optimization, applying a one-layer model to traveltimes $T$ will yield reliable results, but
not so for the two-layer model. Be that as it may, we perform the optimization for both one- and two-layer models, with restrictions $0 < b_i < 2$ and $0 < \chi_i < 1$. The estimator vector and objective-function value, for the former, are

$$[\hat{\beta}]_{N=1} = \begin{bmatrix} 1293.79163 \\ 1.12823361 \\ 0.05416071 \end{bmatrix} \quad \text{and} \quad \hat{f}_{N=1} = 0.00023430 ,$$

(19a,b)

and, for the latter, are

$$[\hat{\beta}]_{N=2} = \begin{bmatrix} 897.950408 \\ 3296.42981 \\ 1.98094661 \\ 0.00003242 \\ 0.00002178 \\ 0.00002216 \end{bmatrix} \quad \text{and} \quad \hat{f}_{N=2} = 0.00009644 .$$

(20a,b)

To gain insight on the values of objective functions (19b) and (20b), let us use model parameters (19a) and (20a) to calculate the best-fit traveltimes using equation (3). Taking their difference with traveltimes $T$ for every offset, we obtain the traveltime residuals for both optimizations, which we plot in Figure 4. Therein, we observe that both traveltime residuals vary with offset. Since the general trends of both the residuals remain within a $\pm 2$ ms range, we maintain that both fits are sufficiently within the traveltime picking error. However, it is clear that both residual plots exhibit a cyclic pattern, with Figure 4(a), the one-layer case, being more pronounced than Figure 4(b), the two-layer case. We attribute the cyclic pattern to an inadequacy of the model to simulate the measured traveltimes. The inadequacy may be rooted in an insufficient number of layers to model the subsurface or an absence of elliptical velocity dependence in all layers. Within the context of optimization, we expect the residuals to be random and appear to be centred about zero, which is the case of the noise profile of the measured traveltimes in Figure 3. To this end, increasing the number of layers not only improves the representation of the medium, but also reduces the cyclic pattern, as can be observed in Figure 4(b). With that being said, the cyclic pattern persists, which could be the consequence of imposing elliptical anisotropy in all layers. Let us examine Figures 4(a) and 4(b) individually.

The traveltime residual in Figure 4(a) does not remain near zero with any offset, which could be attributed to a suboptimal model choice. Indeed, consolidating nearly two kilometres of subsurface into a single layer does not necessarily lead to an accurate representation of its anisotropy and inhomogeneity. However, since the subsurface is considered to be predominantly shale and the general trend of the traveltime residual remains within the experimental picking error limits, we claim that this model is an adequate representation of the subsurface, at least at a macroscopic level. Furthermore, in view of the one-layer-optimization threshold described in Section 3.1, the amount of noise exhibited in $T$ should not lead to poor numerical results. Thus, we contend that model parameters (19a) are reliable.

Recalling the adage that an increase in the number of model parameters results in an increase of numerical fit, the traveltime residual in Figure 4(b) remains closer to zero than in Figure 4(a), and hence, $\hat{f}_{N=2} < \hat{f}_{N=1}$.
Figure 4: Traveltime residuals with offset upon performing the (a) one- and (b) two-layer optimization on travel-times $T$ as expected. Indeed, bisecting the subsurface improves the numerical fit of traveltimes $T$. However, in view of the reliability-of-results threshold for two-layer optimization described in Section 3.2, the amount of noise exhibited in $T$, which is on the order of milliseconds, is three orders of magnitude greater than the threshold, which is on the order of microseconds. As such, the reliability of model parameters (19b) is poor; they indicate that the first layer is strongly inhomogeneous and exhibits low anisotropy, and that the second layer is nearly homogeneous and exhibits near isotropy. Such results are not consistent with the interval velocities plotted in Figure 2, nor with the knowledge that the subsurface in question is comprised predominantly of shale. Thus, despite the increase of model adequacy for physical representation of the subsurface and the numerical fit of traveltimes, there is a significant tradeoff in the reliability of results.

5 Conclusions

To estimate the values of parameters, attributed to anisotropy and inhomogeneity within the context of an analytical mathematical model, using vertical seismic profile traveltimes, we use Newton’s method to minimize the least-squares residual between measured and modelled traveltimes. To understand applications of this method to real traveltimes, we construct synthetic models to analyze the reliability of results under the influence of increasing percentage ranges of noise. We determine that there is a threshold for the acceptable level of noise, which increases in severity as the model complexity increases. In particular, the reliability-of-results threshold for a one-layer model is three orders of magnitude greater than that of a two-layer model.

Within the context of traveltime equation (3), greater complexity is tantamount to segmenting the subsurface into a greater number of layers. Insofar as constructing physically representative models of the subsurface is concerned, the greater the model complexity, the better the fit, as expected due to the larger number of parameters used to accommodate the data. However, as determined in Section 4, the amount of noise present in real traveltime data imposes a tradeoff in the reliability of results with an increase in model complexity.

Pertaining to the real-data example considered in Section 4, we find that, despite our desire to construct a multilayer model, the noise present in this dataset exceeds the reliability-of-results threshold for a two-layer model. Although we are able to minimize the traveltime residual for a model comprised of two or more layers, we are restricted to a one-layer model for reliable results. However, since the subsurface in question is comprised predominantly of shale, we contend that the one-layer model is an adequate representation and that model parameters (19) are reliable and acceptable.

Acknowledgments

We wish to acknowledge discussions with Michael A. Slawinski and the graphic support of Elena Patarini. Also, this research was partially supported by the Natural Sciences and Engineering Research Council of
Canada, grant 202259.

References

Aki, K. and Richards, P. G. (2002). *Quantitative seismology*, volume 2. University Science Books.

Červený, V. (1989). Ray tracing in factorized anisotropic inhomogeneous media. *Geophysical Journal International*, 99(1):91–100.

Červený, V. (2001). *Seismic ray theory*. Cambridge University Press.

Danek, T. and Slawinski, M. A. (2012). Bayesian inversion of VSP traveltimes for linear inhomogeneity and elliptical anisotropy. *Geophysics*, 77(6):R239–R243.

Dix, C. H. (1955). Seismic velocities from surface measurements. *Geophysics*, 20(1):68–86.

Fletcher, R. (2000). *Practical Methods of Optimization*. John Wiley & Sons, Ltd, 2 edition.

Gierlach, B. and Danek, T. (2018). Inversion of velocity parameters in multilayered elliptical anisotropy medium - synthetic data example. *E3S Web of Conferences*, 66(01017):1–8.

Helbig, K. (1994). Foundations of anisotropy for exploration seismics. In *Handbook of geophysical exploration: Seismic exploration*, volume 22. Elsevier Science Serials.

Kaderali, A. (2009). Investigating anisotropy and inhomogeneity using tomographic inversion of VSP traveltimes: Validation of analytic expressions for linearly inhomogeneous elliptically anisotropic models. Master’s thesis, Memorial University of Newfoundland.

Mao, W. and Stuart, G. W. (1997). Transmission-reflection tomography: Application to reverse VSP data. *Geophysics*, 62(3):884–894.

Nocedal, J. and Wright, S. J. (1999). *Numerical optimization*. Springer series in operations research. Springer.

Pujol, J., Burridge, R., and Smithson, S. B. (1985). Velocity determination from offset vertical seismic profiling data. *Journal of Geophysical Research*, 90(B2):1871–1880.

Rogister, Y. and Slawinski, M. A. (2005). Analytic solution of ray-tracing equations for a linearly inhomogeneous and elliptically anisotropic velocity model. *Geophysics*, 70(5):D37–D41.

Slawinski, M. A. (2020a). *Waves and rays in elastic continua*. World Scientific, 4 edition.

Slawinski, M. A. (2020b). *Waves and rays in seismology: Answers to unasked questions*. World Scientific, 3 edition.

Slawinski, M. A., Wheaton, C. J., and Powojowski, M. (2004). VSP traveltime inversion for linear inhomogeneity and elliptical anisotropy. *Geophysics*, 69(2):373–377.

Stewart, R. R. (1984). VSP interval velocities from traveltime inversion. *Geophysical Prospecting*, 32:608–628.

Thomsen, L. (1986). Weak elastic anisotropy. *Geophysics*, 51(10):1954–1966.

Vernik, L. and Liu, X. (1997). Velocity anisotropy in shales: A petrophysical study. *Geophysics*, 62(2):521–532.

Winterstein, D. F. (1990). Velocity anisotropy terminology for geophysicists. *Geophysics*, 55(8):956–1112.

Zhang, M., Xu, T., Bai, Z., Liu, Y., Hou, J., and Yu, G. (2017). Ray tracing of turning wave in elliptically anisotropic media with an irregular surface. *Earthquake Science*, 30:219–228.
A Optimization demonstration for $N = 1$

In this appendix, we exemplify the optimization technique, described in Section 2.2, used in the control experiment for the one-layer case as well as the justification of the logarithmic barrier function, described in Section 2.3.

A.1 Control experiment

Let us consider the one-layer case for which the true model parameters are $\beta = [1500, 0.75, 0.0015]^t$. To perform the optimization, we compute sequences $S$ and generate the measured traveltimes, $T = t(S; \beta)$. We begin the optimization by selecting a set of starting values, $\beta^{(1)} = [1700, 1, 0.1]^t$, after which we calculate directly the values of the objective function, gradient, and Hessian,

$$f^{(1)} = 5.55225227, \quad g^{(1)} = \begin{bmatrix} 0.02234679 \\ 18.4370840 \\ 24.9494726 \end{bmatrix}, \quad H^{(1)} = \begin{bmatrix} 0.00002688 & 0.02337750 & 0.04114517 \\ 0.02337750 & 16.9699206 & 30.3450287 \\ 0.04114517 & 30.3450287 & -13.0394538 \end{bmatrix}. \quad (A.1)$$

Since diagonal entry $H_{3,3}^{(1)}$ is negative, it is clear that $H^{(1)}$ is not positive definite, which we verify by its eigenvalues, $\text{eig}(H^{(1)}) = \{-0.00000532, 35.8173333, -31.8868343\}^t$. As stated in Section 2.2, Newton’s method requires a positive definite Hessian to produce a sequence of iterates, $\beta^{(k)}$, that are globally convergent. Hence, we perform the modified Cholesky factorization to alter $H^{(1)}$ such that it is positive definite; we distinguish the modified Hessian by $\dagger$. The result of the alteration is

$$H^{(1)} := \begin{bmatrix} 0.00009976 & 0.02337750 & 0.04114517 \\ 0.02337750 & 30.7360048 & 30.3450287 \\ 0.04114517 & 30.3450287 & 80.9191375 \end{bmatrix}, \quad \text{where } \text{eig}(H^{(1)}) = \begin{bmatrix} 0.00007558 \\ 16.4523579 \\ 95.2028074 \end{bmatrix}, \quad (A.2)$$

which satisfies the positive definite requirement. In view of equation (9), we use $H^\dagger$ to calculate the step direction,

$$\Delta \beta^{(1)} = -\begin{bmatrix} 1323.17683 & -5.43303647 & -4.69057983 \\ -5.43303647 & 0.05389310 & -0.01744760 \\ -4.69057983 & -0.01744760 & 0.02128597 \end{bmatrix}^{-1} \begin{bmatrix} 0.02234679 \\ 18.4370840 \\ 24.9494726 \end{bmatrix} = -\begin{bmatrix} 78.4907572 \\ 0.43691234 \\ 0.10457155 \end{bmatrix}. \quad (A.3)$$

To ensure a strict decrease in $f$, we must calculate an appropriate step size length, $\alpha$, along the $\Delta \beta^{(1)}$ direction. In Figure A.1, we plot the values of $f(\beta^{(1)} + \alpha \Delta \beta^{(1)})$ for $\alpha \in [0, 1]$. Therein, we observe that, for a positive definite Hessian, the resulting curve of residuals resembles a parabola. In ensuring positive definiteness, we find that this smooth and predictable behaviour is preserved. However, for nonpositive definiteness, and especially in the presence of noise, we found that we lose this smooth and predictable behaviour, which serves as a tacit justification for ensuring the Hessian remains positive definite for all calculations.

Using Algorithm 1, we find that the minimum occurs at $\alpha^{(1)} := 0.73296960$, for which we set

$$\beta^{(2)} := \beta^{(1)} + \alpha^{(1)} \Delta \beta^{(1)} = \begin{bmatrix} 1700 \\ 1 \\ 0.01 \end{bmatrix} - 0.73296960 \begin{bmatrix} 78.4907572 \\ 0.43691234 \\ 0.10457155 \end{bmatrix} = \begin{bmatrix} 1642.46866 \\ 0.67975654 \\ -0.06664777 \end{bmatrix}, \quad (A.4)$$

after which we repeat the previous steps by calculating directly the values of the objective function, gradient, and Hessian—the objective function value for $\beta^{(2)}$ is $f^{(2)} = 0.07156195$. The process is repeated until either Algorithm 1 produces $\alpha^{(k)}$ whose order of magnitude is less than some value or the decrease in residual is less than some value. The remaining iterations of the optimization are tabulated in Table A.1.
Table A.1: Iterations of unconstrained Newton’s method performed on the one-layer case using Hessian modifications from starting values \( \beta^{(1)} = [1700, 1, 0.01]^T \), converging to true values \( \beta_1 = [1500, 0.75, 0.0015]^T \).

| \( k \) | \( \{\beta^{(k)}\} \) | \( f^{(k)} \) |
|---|---|---|
| 1 | 1700 | 1 | 0.01 | 5.5225227 |
| 2 | 1642.46866 | 0.67975654 | -0.06664777 | 0.07156195 |
| 3 | 1303.54209 | 1.07977513 | -0.07656328 | 0.01186727 |
| 4 | 1346.92235 | 0.98946329 | -0.04426630 | 0.00239513 |
| 5 | 1436.89201 | 0.84021609 | -0.01334487 | 0.00051080 |
| 6 | 1473.69567 | 0.78820908 | -0.00461155 | 0.00004540 |
| 7 | 1497.03543 | 0.75408963 | 0.00091052 | 3.3154×10^-16 |
| 8 | 1499.90130 | 0.75014074 | 0.00147849 | 6.3848×10^-10 |
| 9 | 1499.99994 | 0.75000008 | 0.00149999 | 3.3154×10^-16 |
| 10 | 1500 | 0.75 | 0.0015 | 2.0268×10^-26 |

Figure A.1: Evaluations of \( f(\beta^{(1)} + \alpha \Delta \beta^{(1)}) \), where \( \Delta \beta^{(1)} = -\left( H^{(1)} \right)^{-1} g^{(1)} \) and \( \alpha \in [0, 1] \)

### A.2 Control experiment without Hessian modification

In Section A.1, we ensured a positive definite Hessian throughout the optimization using the modified Cholesky factorization. However, if the Hessian is not altered, the optimization is not guaranteed to converge to the true solution from all starting values. Indeed, considering the same model parameters, \( \beta \), and starting values, \( \beta^{(1)} \), the optimization converges to the correct solution. It is not the case from another set of starting values, say, \( \beta^{(1)} = [2400, 1, 0.2]^T \). To demonstrate this, we calculate the values of the objective function, gradient, and Hessian,

\[
f^{(1)} = 31.08341448, \quad g^{(1)} = \begin{bmatrix} 0.03018866 \\ 25.0184043 \\ 28.8129228 \end{bmatrix}, \quad H^{(1)} = \begin{bmatrix} -0.00000424 & -0.00248880 & 0.00556654 \\ -0.00248880 & -4.71975983 & 1.49180975 \\ 0.00556654 & 1.49180975 & -53.6965777 \end{bmatrix}.
\]  

(A.5)

As indicated by its main-diagonal entries and its eigenvalues,

\[
eig(H^{(1)}) = [-53.7419760, -0.00000250, -4.67436327]^T.
\]  

(A.6)
the Hessian is negative definite. Without a modification to its values, we use $H^{(1)}$ to calculate the search direction. We obtain

$$\Delta \beta^{(1)} = [8112.60299, 1.47122367, 1.41846688]^T,$$

which does not permit a decrease in $f$ for any step size length. As such, the optimization is terminated on the first iteration. For other starting values, it is not guaranteed to encounter such a scenario on the first iteration, but the risk remains at each iteration of the optimization.

$$k \{ \beta^{(k)} \} \quad \begin{array}{ccc} f^{(k)} \\ \hline 1 & 2400 & 1 \quad 0.2 & 31.0834145 \\
2 & 2204.86452 & -0.3136164 & 0.10542578 & 1.50964783 \\
3 & 2327.63552 & -0.3084871 & 0.04845627 & 0.00606885 \\
4 & 2389.84780 & -0.3137287 & 0.07165287 & 0.00606885 \\
5 & 2497.11162 & -0.40068941 & 0.04845627 & 0.00228533 \\
6 & 2678.09091 & -0.57165728 & 0.03132588 & 0.00000004 \\
7 & 2792.97026 & -0.66992183 & 0.01475213 & 0.00000004 \\
8 & 2873.14100 & -0.73880958 & 0.00362632 & 0.00000004 \\
9 & 2885.52025 & -0.74884791 & 0.00169315 & 0.00000004 \\
10 & 2886.87975 & -0.75 & 0.0015 & 1.1947 \times 10^{-22} \\
11 & 2886.87975 & -0.75 & 0.0015 & 3.0863 \times 10^{-26} \\
\end{array}$$

Table A.2: Iterations of unconstrained Newton’s method performed on the one-layer case using Hessian modifications from starting values $\beta^{(1)} = [2400, 1, 0.2]^T$, not converging to true values $[\beta]_1 = [1500, 0.75, 0.0015]^T$.

In view of this result, we insist that the Hessian remains positive definite throughout all executions of the optimization. Thus, we repeat the optimization for starting values $\beta^{(1)} = [2400, 1, 0.2]^T$, following the procedure indicated in Section A.1. The iterations of the optimization are tabulated in Table A.2. Therein, we observe that the optimization converges to

$$\tilde{\beta} = \begin{bmatrix} 2886.87975 \\ -0.75 \\ 0.0015 \end{bmatrix} \neq \begin{bmatrix} 1500 \\ 0.75 \\ 0.0015 \end{bmatrix} = \beta.$$

Under the context of the optimization, estimator (A.8) is an acceptable solution. However, in practice, we reject this solution as we expect $b_1 > 0$ and $\chi_1 > 0$ due to the physical considerations indicated in Section 2.1. Namely, model parameter $\tilde{\beta}_2 = b_1 = -0.75 < 0$ is a nonphysical result.

### A.3 Logarithmic barrier functions in control experiment

In Section A.2, we observed that the optimization used in the control experiment can result in solutions with nonphysical parameters. To ensure that solutions remain physically meaningful, we use the logarithmic barrier function, described in Section 2.3, to restrict the values of model parameters. In other words, we introduce logarithmic barrier functions to $f$ that change the unconstrained optimization problem to an inequality-constrained optimization problem. In view of equation (13), the objective function for this problem is

$$P(\beta) = f(\beta) - \sum_{i \in I} \log l_i.$$

To maintain the restrictions imposed by physical considerations, $b_1 > 0$ and $\chi_1 > 0$, we recall the logistic function for some argument $x$,

$$l(x) = \frac{1}{1 + \exp \left( -r \left( x - x_0 \right) \right)}.$$

To use logistic function (A.10) as a logarithmic barrier function for $b_1 > 0$ and $\chi_1 > 0$, we require that it resembles a step function ranging from zero to one with a midpoint of $x = 0$, which requires a large growth rate, $r$. We select the value of $r$ by increasing its magnitude and inspecting its curve in the left-hand plot.
of Figure A.2. Since we do not expect $\chi_1 < 10^{-3}$, we require that the value of $r$ that results in a step-like logistic function for $\chi_1$ on the order of $10^{-4}$—this value is $r = 10^6$. Also, since we expect the order of magnitude of $b_i$ greater than or equal to the order of magnitude of $\chi_i$, we use $r = 10^6$ for its barrier function as well.

As such, we restate equations (18) so that, for both restrictions, we have

$$l_1(b_1) := \frac{1}{1 + \exp(-10^6 b_1)} \quad \text{and} \quad l_2(\chi_1) := \frac{1}{1 + \exp(-10^6 \chi_1)}. \quad (A.11a,b)$$

Taking the logarithm of equation (A.11b) and plotting the curves on the right-hand plot of Figure A.2, we observe that its value equals zero for $\chi_1 > 0$, and rapidly increases in negative value outside that domain. We tabulate these values for $\chi_1$ in Table A.3; since the restrictions for $b_1$ and $\chi_1$ are the same, the barrier-function values in Table A.3 are the same for both $b_1$ and $\chi_1$. Subtracting the logarithms of equations (A.11) from $f(\beta)$ results in increasingly large positive values of $P$ for $b_1 < 0$ and $\chi_1 < 0$. Effectively, this restricts the optimization to positive values $b_1$ and $\chi_1$.

Using restrictions (A.11), we repeat the optimization for starting values $\beta^{(1)} = [1700, 1, 0.04]^t$ and $\beta^{(1)} = [2400, 1, 0.2]^t$. The iterations of the optimization are tabulated, for the former, in Table A.4 and, for the latter, in Table A.5.

Let us compare the sequence of iterates for the unconstrained and inequality-constrained problems. For the former, the value of $\chi_1$ in the unconstrained problem becomes negative on the second iteration and improves
the residual by nearly an order of magnitude for each of the subsequent four iterations before returning to positive values. In the inequality-constrained problem, the value of $\chi_1$ attempts to become negative on the second iteration, but is restricted by the logarithmic barrier function. For the following fourteen iterations, the improvement of the objective function is slow as a consequence of the value of $\chi_1$ being affected by the barrier. Once the value of $\chi_1$ is unaffected by the barrier, the pace of the optimization improves considerably and it converges to the correct solution within four iterations.

For the latter, the value of $b_1$ in the unconstrained problem becomes negative on the second iteration, which leads the optimization toward a nonphysical solution. Similarly, in the inequality-constrained problem, the value of $b_1$ attempts to become negative on the second iteration, but is restricted by the logarithmic barrier function. The progress of the optimization becomes extremely slow as the objective function’s order of magnitude does not change for 24 iterations as both $b_1$ and $\chi_1$ appear to undergo marginal improvements due to their proximity to the barriers. Although the number of iterations nearly triples in comparison to the unconstrained problem, the inequality-constrained optimization converges to the correct solution, which was not possible without the barriers.

The results of both optimizations demonstrate the necessity of the logarithmic barrier functions to guarantee convergence from different starting points. As such, we include such functions in all calculations.

| $\chi_1$ | $t_2$ | $\log(t_2)$ |
|----------|-------|-------------|
| −0.0010  | 0     | $-\infty$  |
| −0.0009  | 0     | $-\infty$  |
| −0.0008  | 0     | $-\infty$  |
| −0.0007  | 9.8597×10$^{-305}$ | −700 |
| −0.0006  | 2.6504×10$^{-261}$ | −600 |
| −0.0005  | 7.1246×10$^{-218}$ | −500 |
| −0.0004  | 1.9152×10$^{-174}$ | −400 |
| −0.0003  | 5.1482×10$^{-131}$ | −300 |
| −0.0002  | 1.38390×10$^{-87}$  | −200 |
| −0.0001  | 3.72007×10$^{-44}$  | −100 |
| 0        | 0.5   | −0.693      |
| 0.0001   | 1     | 0           |
| 0.0002   | 1     | 0           |
| 0.0003   | 1     | 0           |
| 0.0004   | 1     | 0           |
| 0.0005   | 1     | 0           |
| 0.0006   | 1     | 0           |
| 0.0007   | 1     | 0           |
| 0.0008   | 1     | 0           |
| 0.0009   | 1     | 0           |
| 0.0010   | 1     | 0           |

Table A.3: Increasingly negative values of logarithmic barrier function for $\chi_1 < 0$
Table A.4: Iterations of inequality-constrained Newton’s method performed on the one-layer case using Hessian modifications from starting values $\beta^{(1)} = [1700, 1, 0.01]^T$, converging to true values $[\beta]_1 = [1500, 0.75, 0.0015]^T$. 

| $k$ | $\{\beta^{(k)}\}$ | $P^{(k)}$ |
|-----|---------------------|----------|
| 1   | 1700 1 1 0.01       | 5.55225227 |
| 2   | 1692.54515 0.95850321 0.00006807 | 4.41369483 |
| 3   | 1233.20128 1.08472339 0.00007378 | 0.03674654 |
| 4   | 1426.63490 0.83058675 0.00008276 | 0.00451468 |
| 5   | 1478.73012 0.77649374 0.00009275 | 0.00020029 |
| 6   | 1494.14148 0.75855244 0.00010275 | 0.00003661 |
| 7   | 1494.76944 0.75783063 0.00011276 | 0.00001461 |
| 8   | 1494.80826 0.75777284 0.00012278 | 0.00000656 |
| 9   | 1494.84601 0.75771622 0.00013284 | 0.00000359 |
| 10  | 1494.88414 0.75765904 0.00014300 | 0.00000247 |
| 11  | 1494.92333 0.75760026 0.00015345 | 0.00000205 |
| 12  | 1494.96553 0.75753698 0.00016469 | 0.00000187 |
| 13  | 1495.01729 0.75745936 0.00017848 | 0.00000178 |
| 14  | 1495.11025 0.75731996 0.00020324 | 0.00000170 |
| 15  | 1495.72385 0.75639986 0.00036667 | 0.00000129 |
| 16  | 1499.97374 0.75003000 0.00149732 | $6.7360 \times 10^{-10}$ |
| 17  | 1499.99998 0.75000003 0.00149999 | $2.8015 \times 10^{-17}$ |
| 18  | 1500 0.74999999 0.0015 | $2.4500 \times 10^{-26}$ |
| 19  | 1500 0.75 0.0015 | $9.4772 \times 10^{-27}$ |
\[
\begin{align*}
\{\beta^{(k)}\} & \quad P^{(k)} \\
1 & 2400 & 1 & 0.2 & 31.0834145 \\
2 & 2251.46463 & 0.00008704 & 0.1280146 & 1.61680199 \\
3 & 2041.92934 & 0.00012801 & 0.06785129 & 0.04946932 \\
4 & 2091.67981 & 0.00013801 & 0.06785129 & 0.00454632 \\
5 & 2101.53157 & 0.00014801 & 0.06785129 & 0.00454442 \\
6 & 2102.01283 & 0.00015801 & 0.06785129 & 0.00454434 \\
7 & 2101.96707 & 0.00016801 & 0.06785129 & 0.00454432 \\
8 & 2101.93914 & 0.00017801 & 0.06785129 & 0.00454432 \\
9 & 2101.91910 & 0.00018801 & 0.06785129 & 0.00454432 \\
10 & 2101.89851 & 0.00019801 & 0.06785129 & 0.00454432 \\
11 & 2101.87932 & 0.00020801 & 0.06785129 & 0.00454432 \\
12 & 2100.71385 & 0.00152699 & 0.06785128 & 0.00454429 \\
13 & 2099.30273 & 0.00305433 & 0.06785037 & 0.00454416 \\
14 & 2090.82795 & 0.01223931 & 0.06779151 & 0.00454432 \\
15 & 2079.43643 & 0.02462133 & 0.06779151 & 0.00454432 \\
16 & 2055.79366 & 0.05044920 & 0.06760345 & 0.00450040 \\
17 & 1998.83778 & 0.11331286 & 0.06667920 & 0.00432868 \\
18 & 1833.50096 & 0.29898848 & 0.06191242 & 0.00359029 \\
19 & 1668.54327 & 0.50434975 & 0.04235992 & 0.00212884 \\
20 & 1593.85828 & 0.61394512 & 0.02313033 & 0.00051620 \\
21 & 1510.37809 & 0.73180548 & 0.00545920 & 0.00007704 \\
22 & 1503.25052 & 0.74518445 & 0.00229869 & 0.00000069 \\
23 & 1500.00008 & 0.74999490 & 0.00150227 & 2.0564\times10^{-10} \\
24 & 1500 & 0.74999999 & 0.0015 & 5.2284\times10^{-19} \\
25 & 1500 & 0.75 & 0.0015 & 1.3276\times10^{-26} \\
\end{align*}
\]

Table A.5: Iterations of inequality-constrained Newton’s method performed on the one-layer case using Hessian modifications from starting values \(\beta^{(1)} = [2400, 1, 0.2]^t\), converging to true values \(\beta_1 = [1500, 0.75, 0.0015]^t\).