Homework 3

David L. Olson

March 23, 2025

Contents

1 Problem 1

Exercise 2 in Section 3.6

1.1 Solution

Note: *My MATLAB code for this homework problem repeats all the steps for example 3.1 so that I can take on this problem. I will only cover the checkerboard test in this write-up.*

The checkerboard test using $\boldsymbol{m}_{true}$ can be reshaped to $\boldsymbol{m}_{true} \in \mathbb{R}^9$ such that

$$\boldsymbol{m}_{true} = \begin{bmatrix} -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}^T$$

which allows for the creation of test data $\boldsymbol{d}_{true}$ and a recovered model $\boldsymbol{m}_{\dagger}$.

$$\boldsymbol{d}_{true} = G\boldsymbol{m}_{true}$$

$$\boldsymbol{m}_{\dagger} = G^{\dagger}\boldsymbol{d}_{true}$$

Recall from example 3.1 that $G \in \mathbb{R}^{8\times 9}$ with rank 7. Therefore the generalized pseudo-inverse of G, represented as $G^{\dagger}$, was computed using the Moore-Penrose pseudo-inverse function `pinv(G)` in MATLAB®. Figure 1 shows how the recovered model compares to the true model.

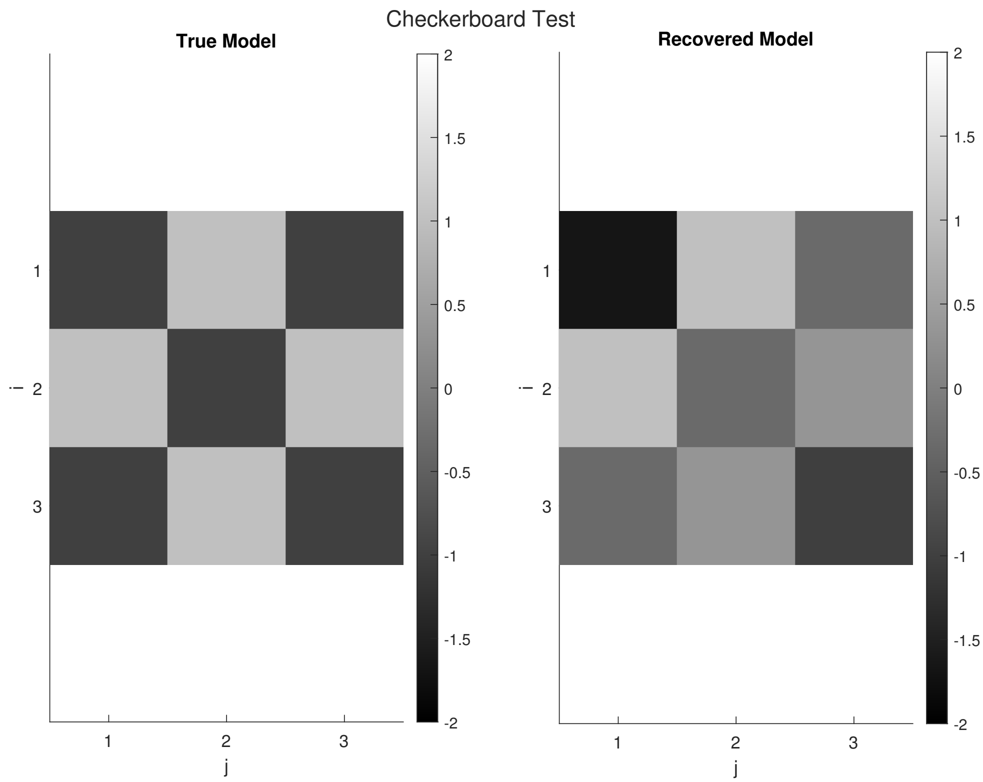

Figure 1: Checkerboard Test

Recall that this problem is of the situation "$p < m$ and $p < n$", in which both the model and data null spaces are non-trivial and $\boldsymbol{m}_\dagger$ is the minimum length solution. The resulting model null space vector, $\boldsymbol{m}_0 := \boldsymbol{m}_{true} - \boldsymbol{m}_\dagger$, is shown below both numerically and graphically in figure 2.

$$\boldsymbol{m}_0 = \begin{bmatrix} \frac{2}{3} & 0 & \frac{-2}{3} & 0 & \frac{-2}{3} & \frac{2}{3} & \frac{-2}{3} & \frac{2}{3} & 0 \end{bmatrix}^T$$

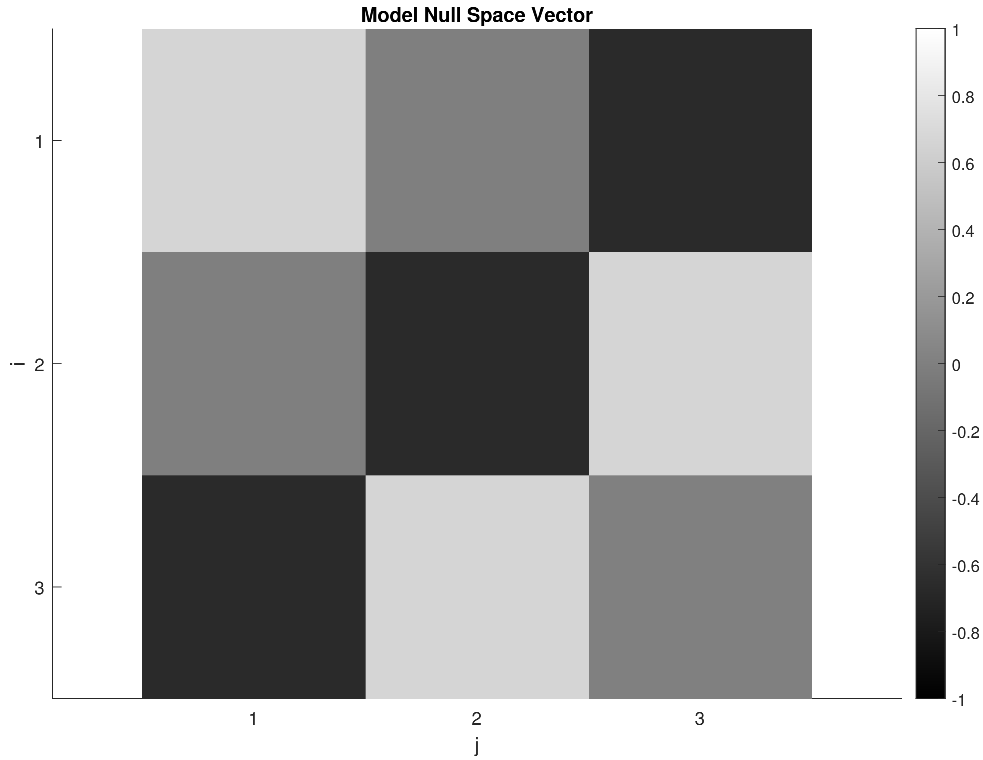

Figure 2: Model Null Space Vector

Notice that the sum of the model null space vector is zero. Also notice that all of the rows and columns depicted in figure 2 also sum to zero. The model null space vector, a.k.a. the model residual, could be restated as

$$\boldsymbol{m}_0 = \sum_{i=p+1}^{n} \alpha_i V_{.,i}$$

where non-zero coefficients for α_i could create other model null space vectors. In our case we computed the minimum length solution, implying that all α_i coefficients are zero. The addition of this vector which resides in the model null space would have no impact on the predicted data if added to another model solution which resides in the model range space.

The three perfectly recovered model parameters are m_2, m_4 and m_9. When examining the model resolution matrix R_m from the spike test, model parameter m_9 is the only parameter with perfect resolution.

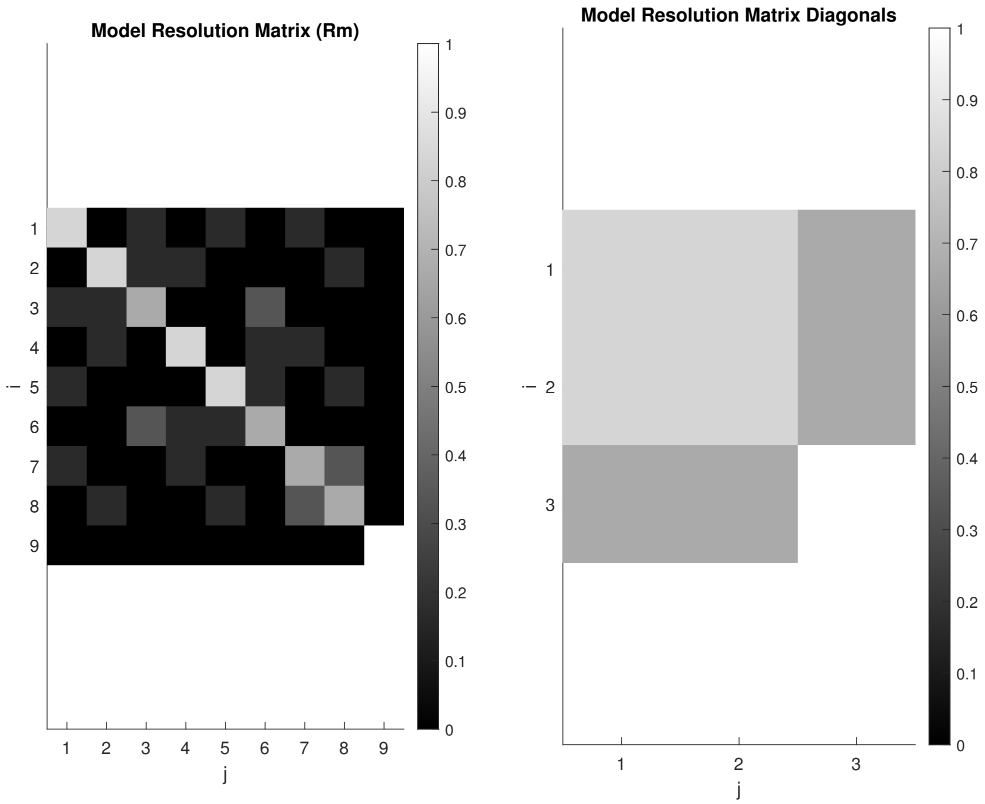

Figure 3: Checkerboard Test

Model parameters m_2, m_4 are subject to "smearing" due to the limited resolution. In this situation, this could have caused us to assume that model parameters m_2, m_4 also have perfect resolution, when instead the smearing for the parameters just happened to bring these model parameters to zero.

2 Problem 2

Exercise 4 in Section 3.6

2.1 Solution

NOTE: *I do not have any experience in seismology, please forgive any mistakes in technicalities when I try to explain this exercise in my own words. It helps me understand what is going on so I can set up the problem correctly.*

The forward problem in this exercise allows mechanical waves to propagate through a 16×16 meter grid where each square in the grid have some slowness value $s_{x,y}$ in units of $\mathrm{s\,m^{-1}}$. Stations around the grid record time of arrivals of the mechanical waves as they pass through one or multiple grid squares. The time it takes to pass through a path of squares is formulated below.

$$t = \int_l s\left(\boldsymbol{x}\right) dl$$

$$\approx \sum_{blocks} s_{block} \Delta l$$

2.1.1 Part A - Row and Column Scans Only

16 row scans and 16 column scans are utilized in this part in the exercise as shown in figure 4.

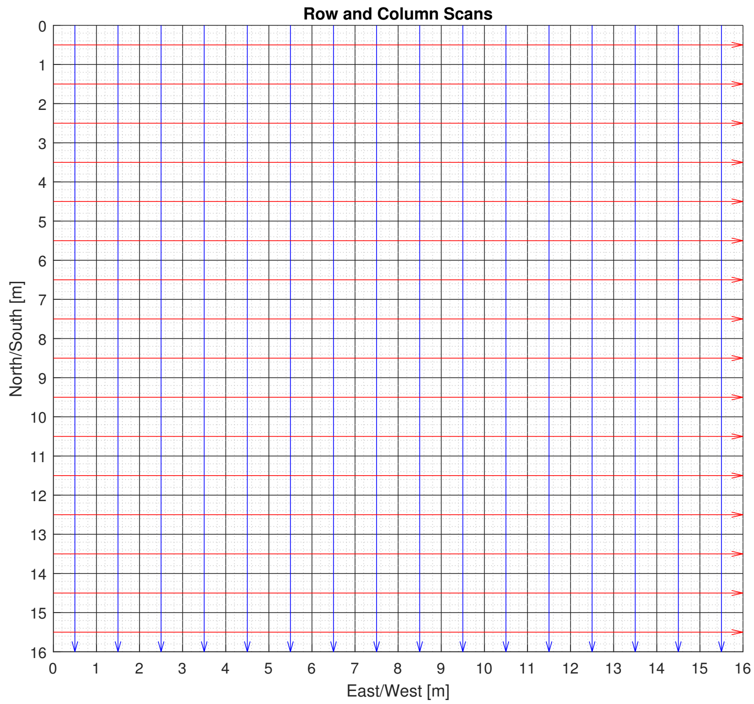

Figure 4: Row and Column Scan Visualization

This results in a total of $m = 32$ measurements. In an effort to estimate the slowness of each square in the grid, this results in a number of $n = 256$ model parameters.

$$\boldsymbol{d} \in \mathbb{R}^{32}, \quad \boldsymbol{m} \in \mathbb{R}^{256}, \quad G \in \mathbb{R}^{32\times 256}$$

The vector of measurement observations $\boldsymbol{d}$ is organized such that

$$\boldsymbol{d} = \begin{bmatrix} t_{r,1} & t_{r,2} & \dots & t_{r,16} & t_{c,1} & t_{c,2} & \dots & t_{c,16} \end{bmatrix}^T$$

where a r subscript indicates a row can and a c subscript indicates a column scan. The model parameters $\boldsymbol{m}$ are organized such that

$$\boldsymbol{m} = \begin{bmatrix} s_{1,1} & s_{1,2} & \dots & s_{1,16} & s_{2,1} & s_{2,2} & \dots & s_{16,1} & \dots & s_{16,16} \end{bmatrix}^T$$

where the first subscript indicates the row, and the second subscript indicates the column. Each row of the model operator G contain the distance traveled by the mechanical wave for each square in its path. Due to the large number of elements, a color map of the zeros and ones for this part of the problem as given instead in figure 5.

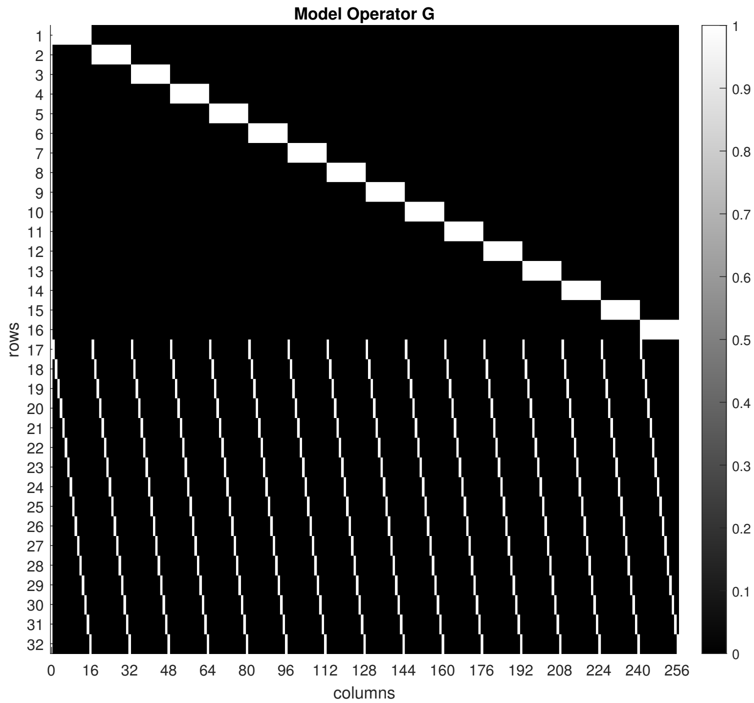

Figure 5: Model Operator G

Subpart A

Per MATLAB®, the rank of G is 31. With $m = 32$ observations and $n = 256$ model parameters, the matrix of G is not of full rank.

Subpart B

This is the environment of "$p < m$ and $p < n$", in which both the data null space and model null space are nontrivial and the solution $\boldsymbol{m}^{\dagger}$ is the minimum length least squares solution.

$$m_{\dagger} = G^{\dagger}\boldsymbol{d} = V_p S_p^{-1} U_p^T \boldsymbol{d}$$

Because G is rank-deficient, there will be no exact solution. With $p = \text{rank}\,(G) = 31$ and $m = 32$, the dimensions of the data and model range and null spaces will are

$$U_p \in \mathbb{R}^{32\times 31},\, Vp \in \mathbb{R}^{256\times 31}, \quad U_0 \in \mathbb{R}^{32\times 1},\, V_0 \in R^{256\times 225}$$

such that it fits the form below.

$$G = \begin{bmatrix} U_p & U_0 \end{bmatrix} \begin{bmatrix} S_p & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_p & V_0 \end{bmatrix}^T$$

Examples of vectors in each null space are provided in figure 6. Interpreting these patterns, the data null space vector appear to have small constant values for two groups of 16 elements. There isn't any useful data in this vector to assist in estimating the model parameters. Likewise, the example model null space vector contains mostly values close to zero which would not significantly alter a recovered if added.

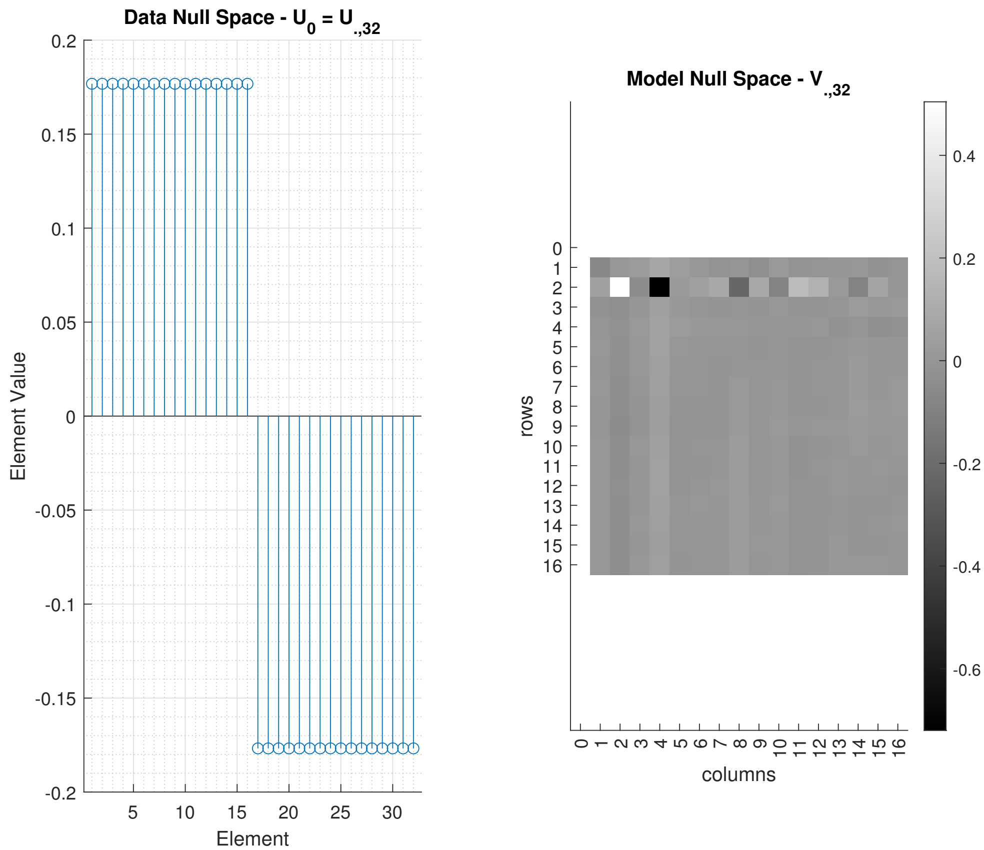

Figure 6: Null Space Examples

The model resolution matrix is provided in figure 7.

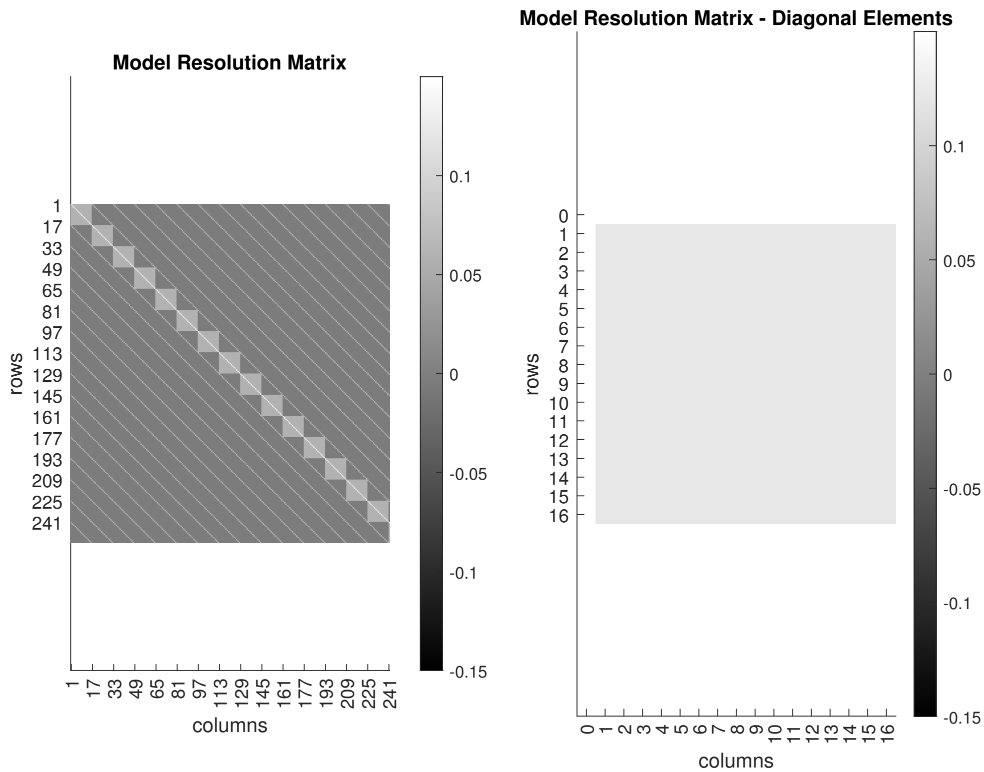

Figure 7: Model Resolution Matrix R_m

Note that there are no diagonal elements which provide perfect model resolution. Each model parameter appears to be equally smeared.

Subpart C

Model parameters are computed using the `pinv()` function in MATLAB®, and results are provided in figure 8.

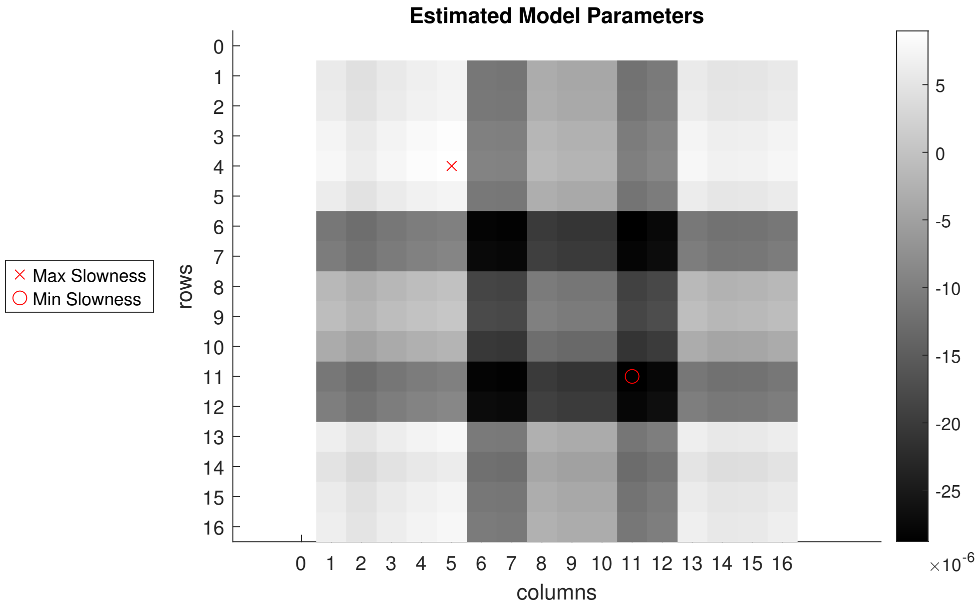

Figure 8: Estimated Model Parameters

The maximum and minimum estimated slowness values are

$$s_{max} = 8.971 \times 10^{-6}\,\mathrm{s\,m^{-1}}$$
$$s_{min} = -2.878 \times 10^{-5}\,\mathrm{s\,m^{-1}}$$

which imply velocities of

$$v_{max} = 1878457.875\,\mathrm{m\,s^{-1}}$$
$$v_{min} = -34130333.333\,\mathrm{m\,s^{-1}}$$

throughout the various square grids.

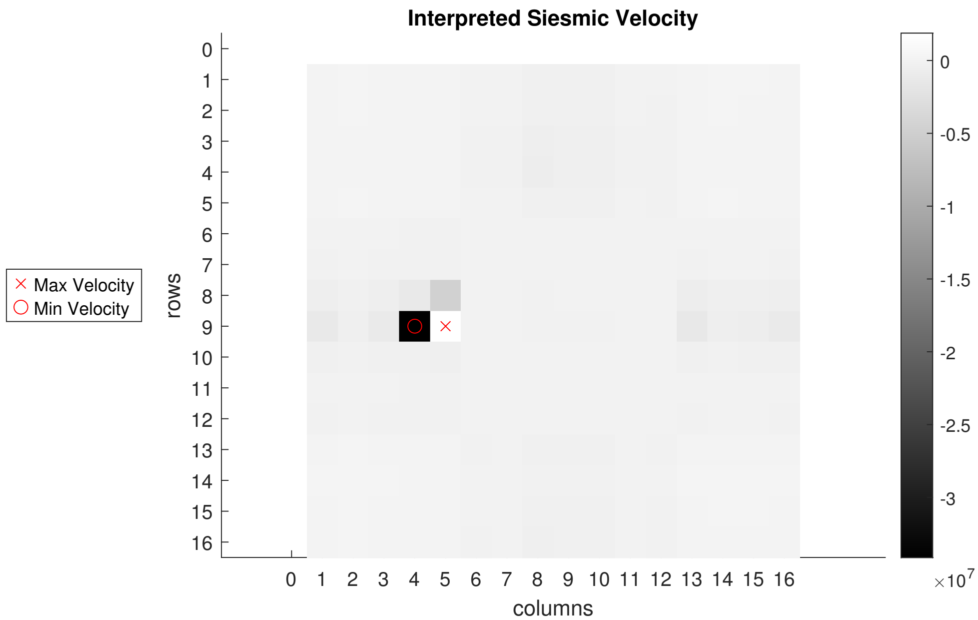

Figure 9: Interpreted Seismic Velocity

The maximum velocity square is in the 9^{th} row and 5^{th} column. *Again, I do not study seismology*, but assuming that the dinosaur bones are related to the highest velocity we expect to find in our search space, then they must be located in the square with the highest velocity value.

Obviously, these velocities reported have no physical connection to reality. The estimated slowness parameters are incredibly small which results in huge velocities far beyond the expected $3000\mathrm{m\,s^{-1}}$ velocity. This is perhaps a case when the minimum length model solution can be a hindrance to estimating meaningful parameters.

Subpart D

Given the dimensions of V_0, there are 225 example solutions that could fit the system of equations $G\boldsymbol{m} = \boldsymbol{d} = \boldsymbol{0}$. In fact, there are actually an infinite amount of example solutions which could comprise any linear combination of these 225 example model null space vectors.

To demonstrate a "wild" model that can fit this solution, let's use a linear combination of three model null space vectors such that:

$$\boldsymbol{m}_{wild} = 0.25V_{.,32} + 0.50V_{.,127} + 0.25V_{.,183}$$

Formulating this wild solution in MATLAB ® to predict the resulting data produces the following results in figure 10.

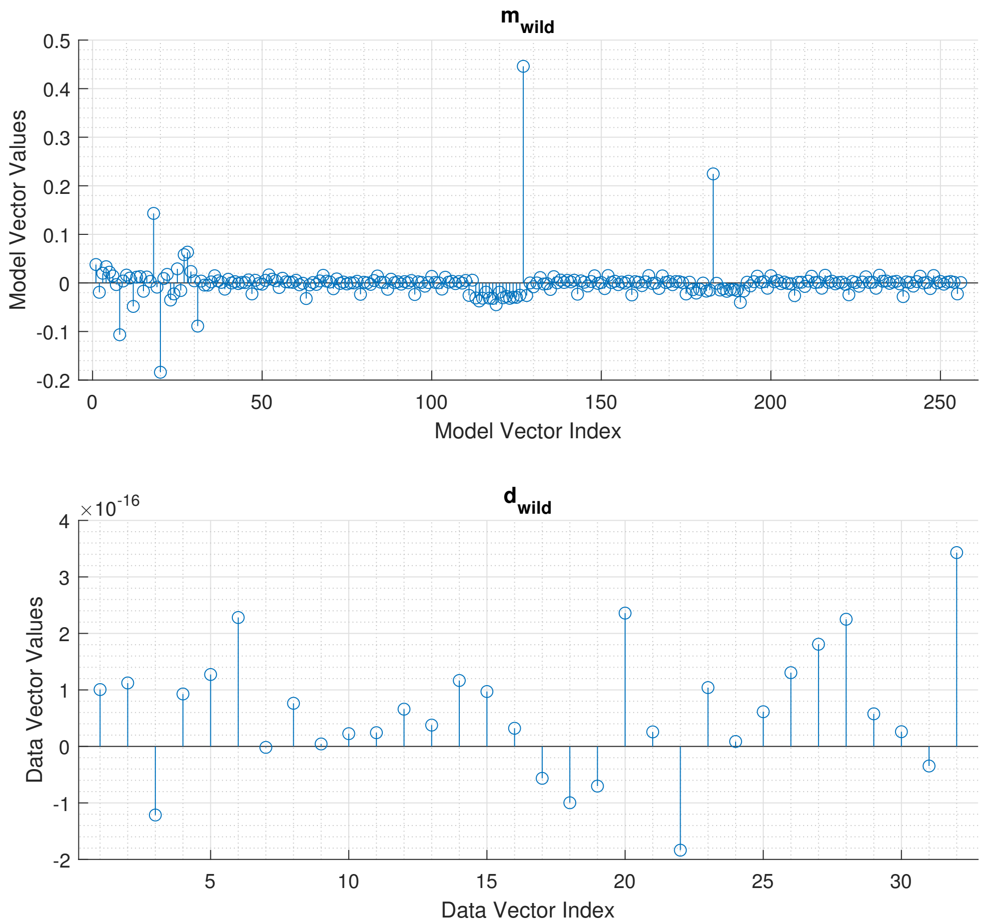

Figure 10: Wild Model and Resulting Predicted Data

Notice the scale of the values for the predicted data vector!

2.1.2 Part B - Adding in the Diagonal Scans

16 row scans, 16 column scans, and 62 diagonal scans are utilized in this part in the exercise as shown in figure 11.

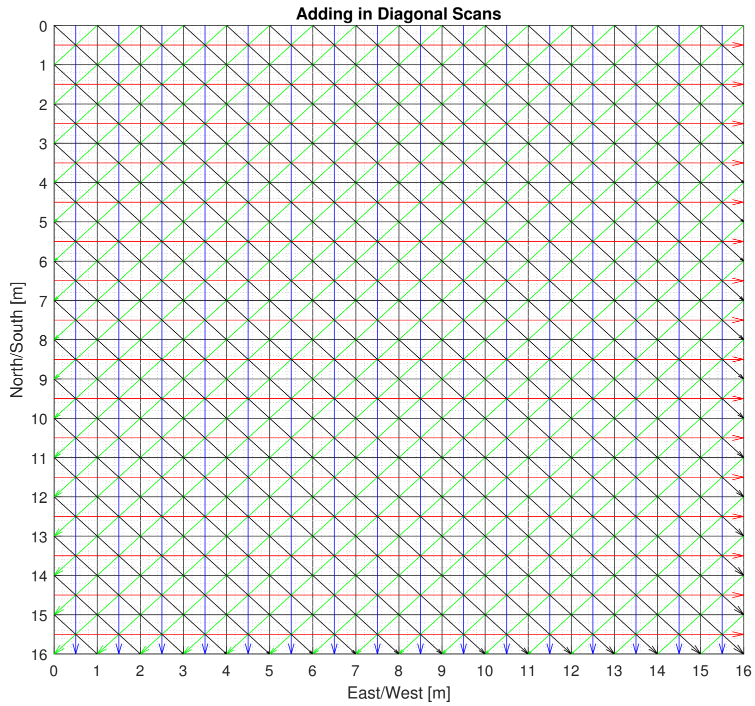

Figure 11: Adding the Diagonal Scans

This results in a total of $m = 94$ measurements. In an effort to estimate the slowness of each square in the grid, this results in a number of $n = 256$ model parameters.

$$\boldsymbol{d} \in \mathbb{R}^{94}, \quad \boldsymbol{m} \in \mathbb{R}^{256}, \quad G \in \mathbb{R}^{94\times 256}$$

Due to the large number of elements, a color map of the zeros and ones for this part of the problem as given instead in figure 12.

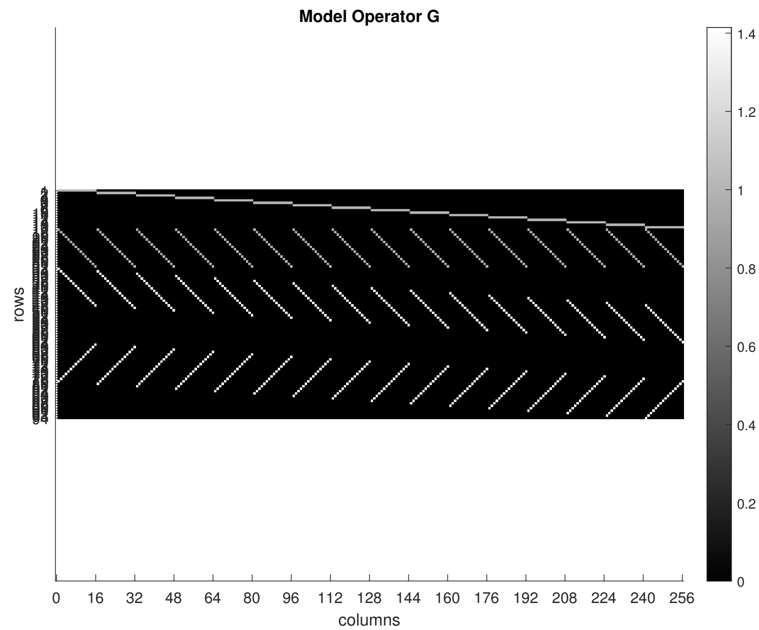

Figure 12: Model Operator G

Subpart A

Per MATLAB®, the rank of G is 87. With $m = 94$ observations and $n = 256$ model parameters, the matrix of G is not of full rank.

Subpart B

This is the environment of "$p < m$ and $p < n$", in which both the data null space and model null space are nontrivial and the solution $\boldsymbol{m}^{\dagger}$ is the minimum length least squares solution.

$$m_{\dagger} = G^{\dagger}\boldsymbol{d} = V_p S_p^{-1} U_p^T \boldsymbol{d}$$

Because G is rank-deficient, there will be no exact solution. With $p = \operatorname{rank}(G) = 97$ and $m = 94$, the dimensions of the data and model range and null spaces will are

$$U_p \in \mathbb{R}^{94\times 87},\, Vp \in \mathbb{R}^{256\times 87}, \quad U_0 \in \mathbb{R}^{94\times 7},\, V_0 \in R^{256\times 169}$$

such that it fits the form below.

$$G = \begin{bmatrix} U_p & U_0 \end{bmatrix} \begin{bmatrix} S_p & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_p & V_0 \end{bmatrix}^T$$

Examples of vectors in each null space are provided in figure 13. Interpreting these patterns, the data null space vector appears to have some sort of sinusoidal wave with a small amplitude which could perhaps be interpreted as noise filtered out of the solution. I suspect that there isn't any useful data in this vector to assist in estimating the model parameters as there is nothing about the arrival of mechanical waves that be sinusoidal. Likewise, the example model null space vector contains mostly values close to zero which would not significantly alter a recovered if added.

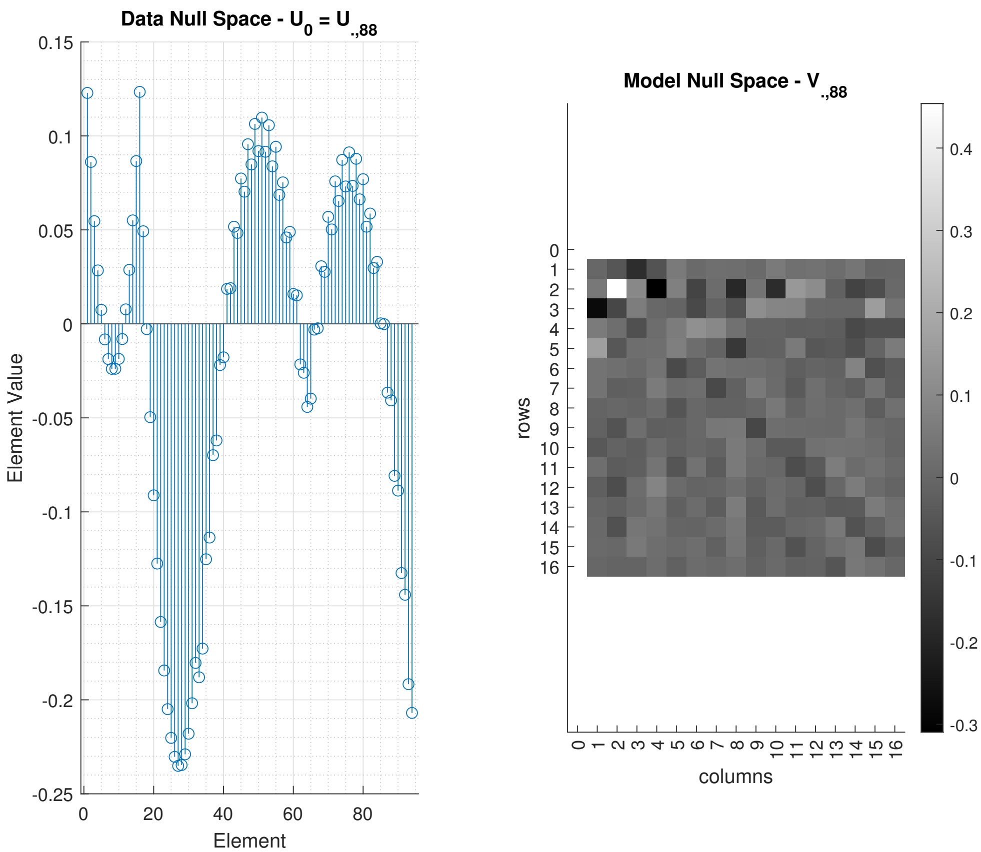

Figure 13: Null Space Examples

The model resolution matrix is provided in figure 14.

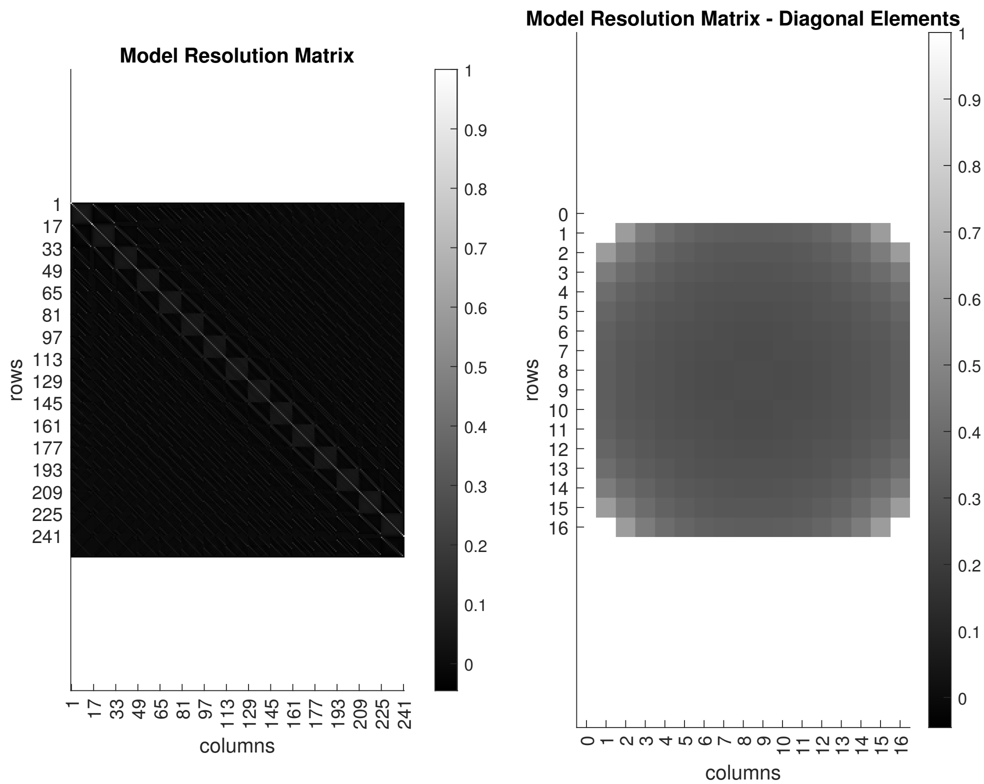

Figure 14: Model Resolution Matrix R_m

Each "corner" parameter in the 16×16 grid has perfect resolution. I suspect this is because in the diagonal scans, there are some measurements which the corner square is the only activated square for that measurement. All other parameters are subject to some amount of smearing.

Subpart C

Model parameters are computed using the `pinv()` function in MATLAB®, and results are provided in figure 15.

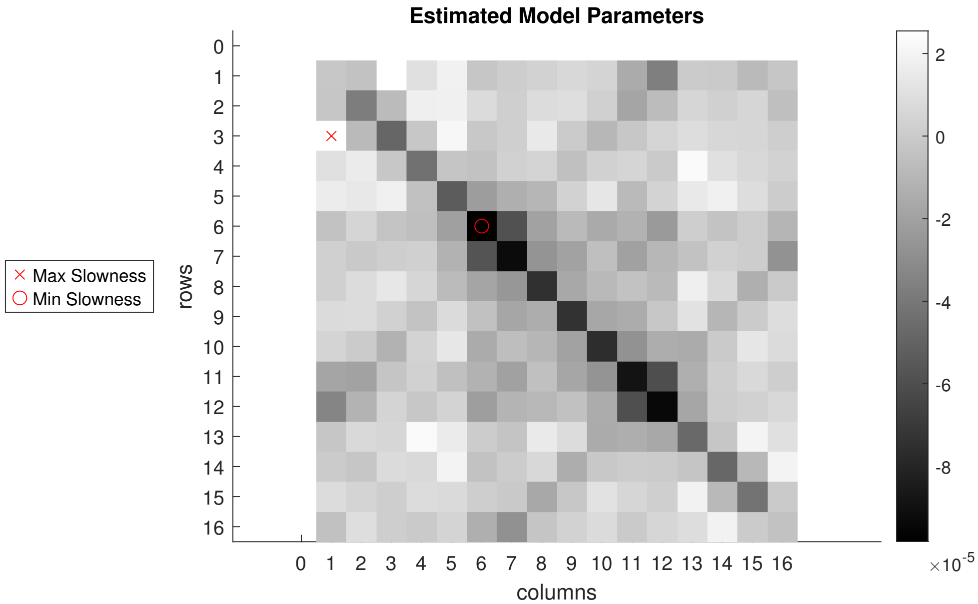

Figure 15: Estimated Model Parameters

The maximum and minimum estimated slowness values are

$$s_{max} = 2.548 \times 10^{-5}\,\mathrm{s\,m^{-1}}$$
$$s_{min} = -9.823 \times 10^{-5}\,\mathrm{s\,m^{-1}}$$

which imply velocities of

$$v_{max} = 90870796.137\,\mathrm{m\,s^{-1}}$$
$$v_{min} = -7250593.614\,\mathrm{m\,s^{-1}}$$

throughout the various square grids.

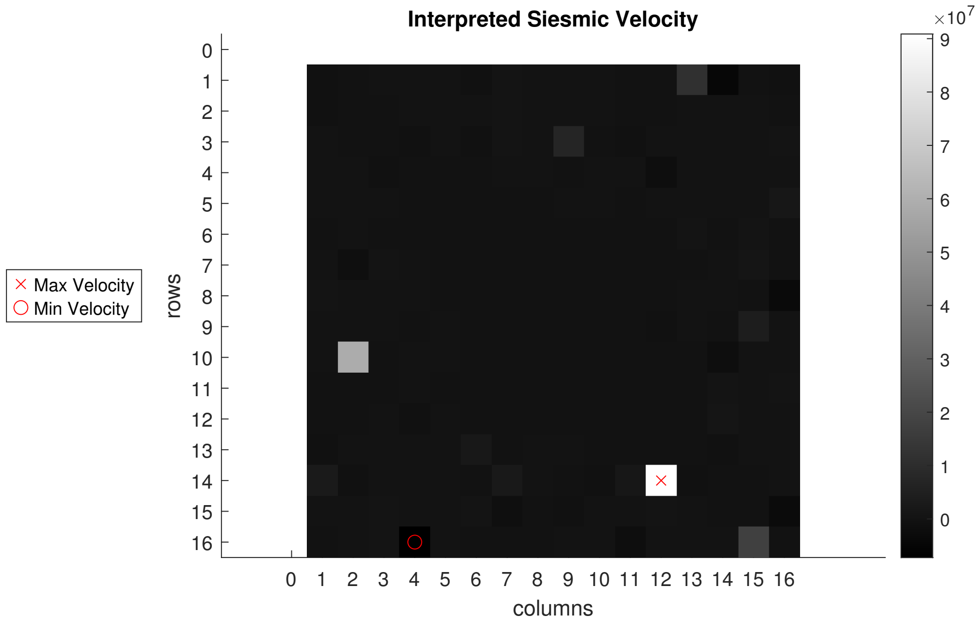

Figure 16: Interpreted Seismic Velocity

The maximum velocity square is in the 14^{th} row and 12^{th} column. *Again, I do not study seismology*, but assuming that the dinosaur bones are related to the highest velocity we expect to find in our search space, then they must be located in the square with the highest velocity value.

Obviously, these velocities reported have no physical connection to reality. The estimated slowness parameters are incredibly small which results in huge velocities far beyond the expected $3000\mathrm{m\,s^{-1}}$ velocity. This is perhaps a case when the minimum length model solution can be a hindrance to estimating meaningful parameters.

Subpart D

Given the dimensions of V_0, there are 169 example solutions that could fit the system of equations $\boldsymbol{Gm} = \boldsymbol{d} = \boldsymbol{0}$. In fact, there are actually an infinite amount of example solutions which could comprise any linear combination of these 169 example model null space vectors.

To demonstrate a "wild" model that can fit this solution, let's use a linear combination of three model null space vectors such that:

$$\boldsymbol{m}_{wild} = 0.25V_{.,88} + 0.50V_{.,132} + 0.25V_{.,201}$$

Formulating this wild solution in MATLAB ® to predict the resulting data produces the following results in figure 17.

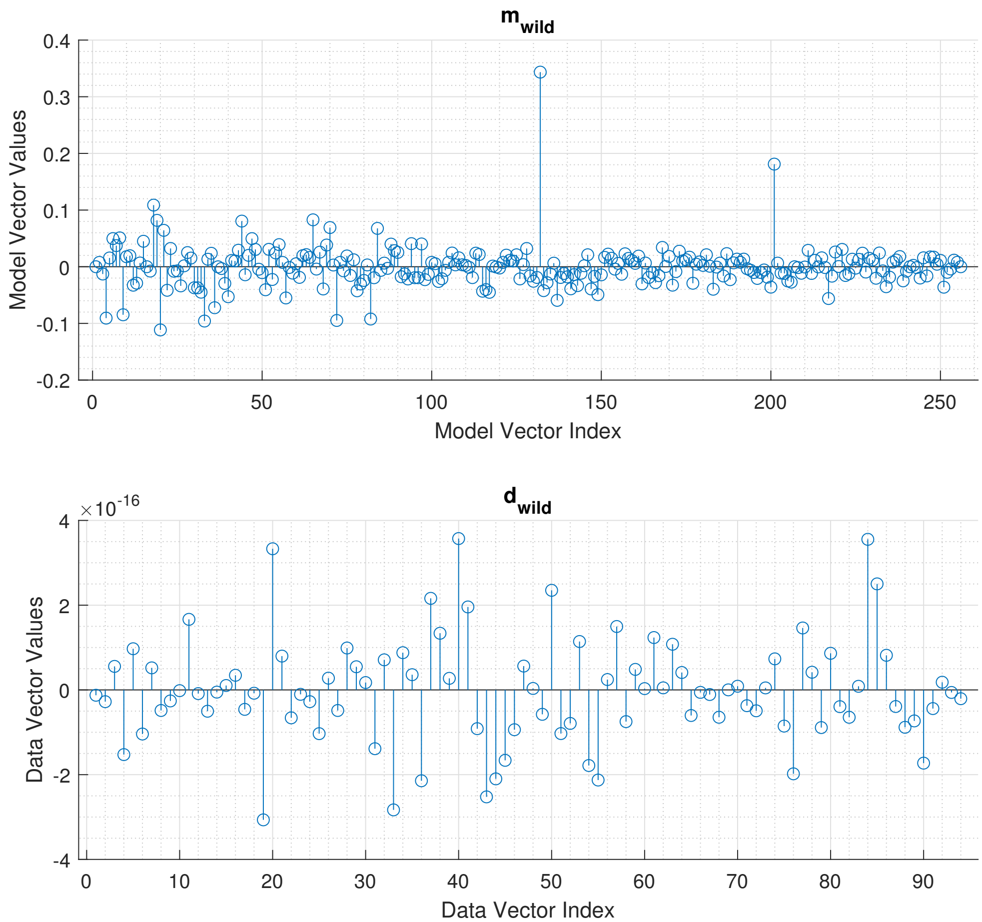

Figure 17: Wild Model and Resulting Predicted Data

Notice the scale of the values for the predicted data vector!

3 Problem 3

Exercise 3 in Section 4.10

3.1 Solution

The model operator G and noisy data vector $\boldsymbol{d}_n$ are provided in `crosswell.mat`. The problem also provides depths for the sources and receivers, and a standard deviation of $\sigma = 0.5$ms. The model operator and data are scaled by the standard deviation to form a weighted least squares solution.

$$W = \sigma I, \ \ G_w = WG, \ \ \boldsymbol{d}_w = W\boldsymbol{d}_n$$

The model operator $G \in \mathbb{R}^{256\times256}$ where both $m = n = 256$. The rank of G is 243, which implies that G_w is rank-deficient. The model operator is visualized in figure 18.

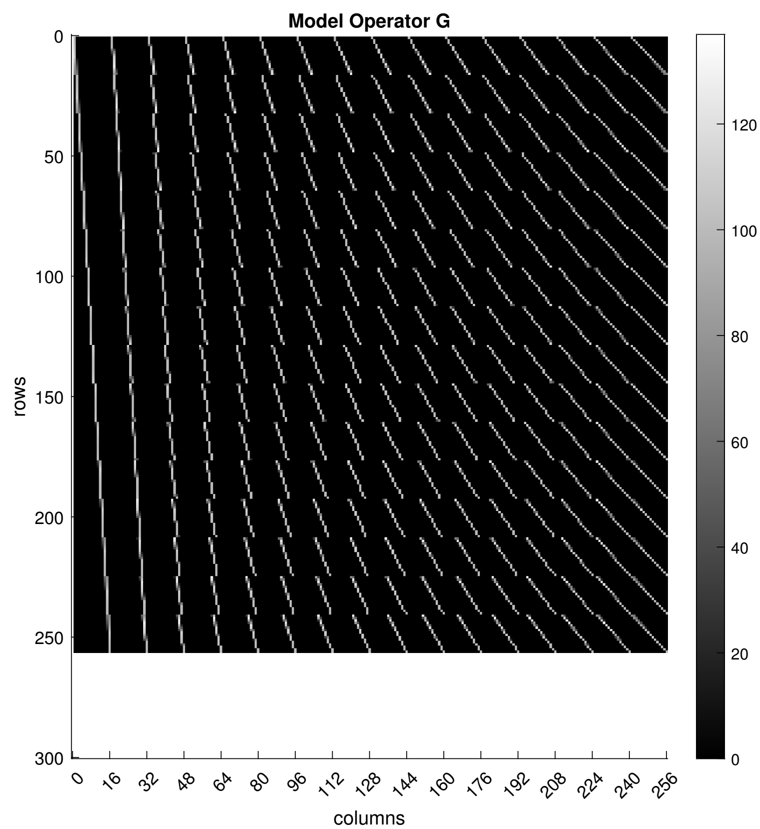

Figure 18: Model Operator G_w

3.1.1 Part A - Inverse Problem Solution via TSVD

The L-curve was created using the provided functions `l_curve_tsvd()` and `l_curve_corner()` in the `PEIP` library. The inputs to the functions used the provided values of G and $\boldsymbol{d}$, not their weighted versions. This resulted in the following plot in figure 19, where the reported truncated value $p' = 58$.

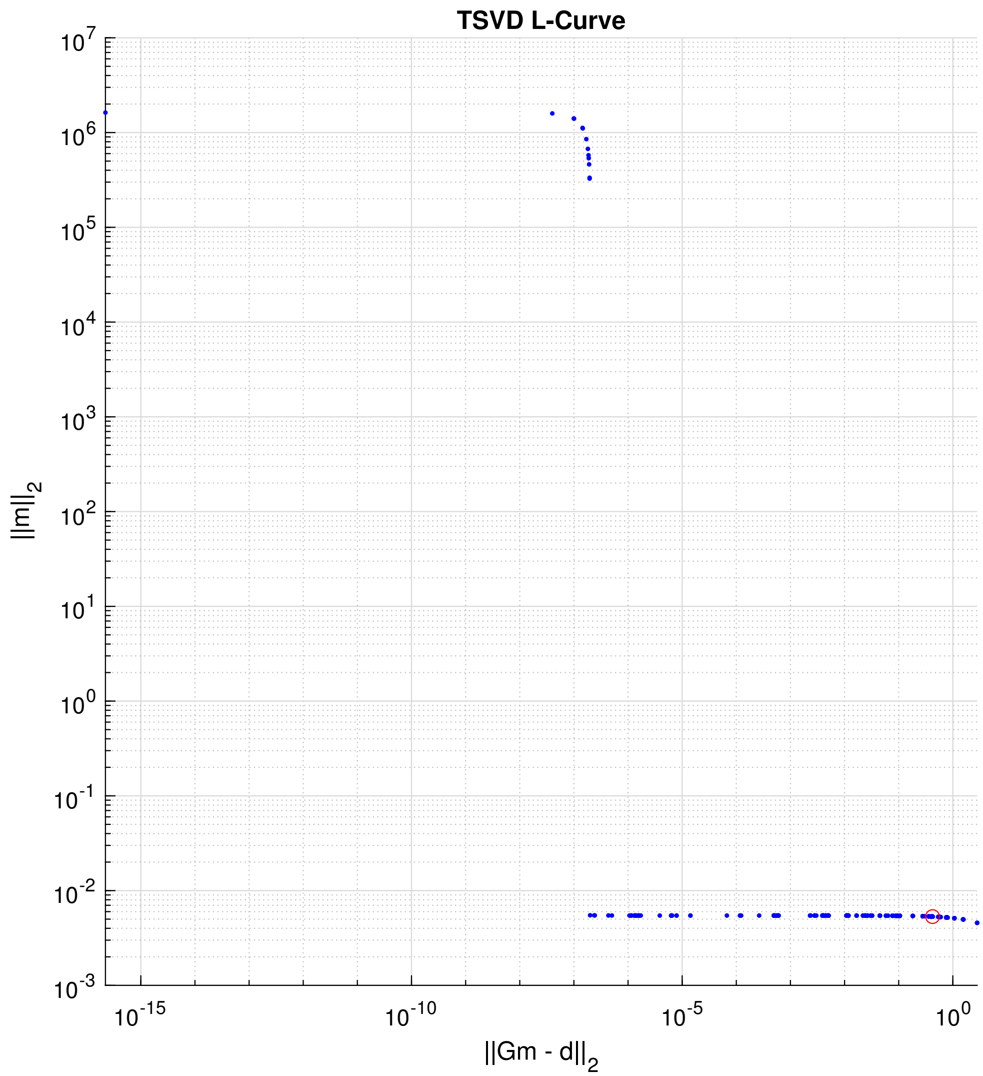

Figure 19: TSVD L-Curve

The model was solved using these generalized inverse where the SVD was re-computed using G_w. Truncating the generalized inverse solution to 58 singular values leads to the following model visualized in 20.

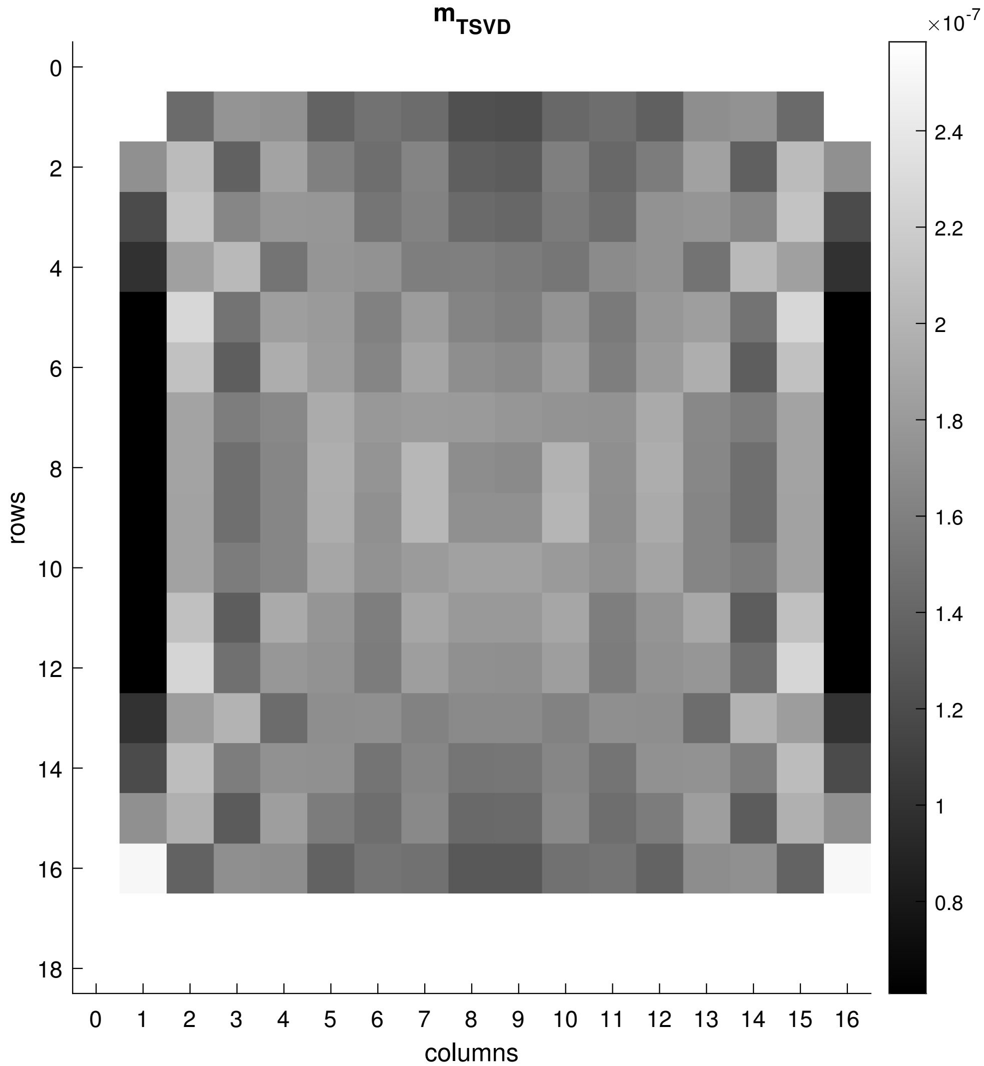

Figure 20: TSVD Solution

3.1.2 Part B - Inverse Problem Solution via Zeroth-Order Tikhonov Regularization

Again, the L-curve was created using the provided functions `l_curve_tikh_svd()` and `l_curve_corner()` in the `PEIP` library. The inputs to the functions used the provided values of G and $\boldsymbol{d}$, not their weighted versions. This resulted in the following plot in figure 21, where the reported value $\alpha = 4.117 \times 10^{-7}$.

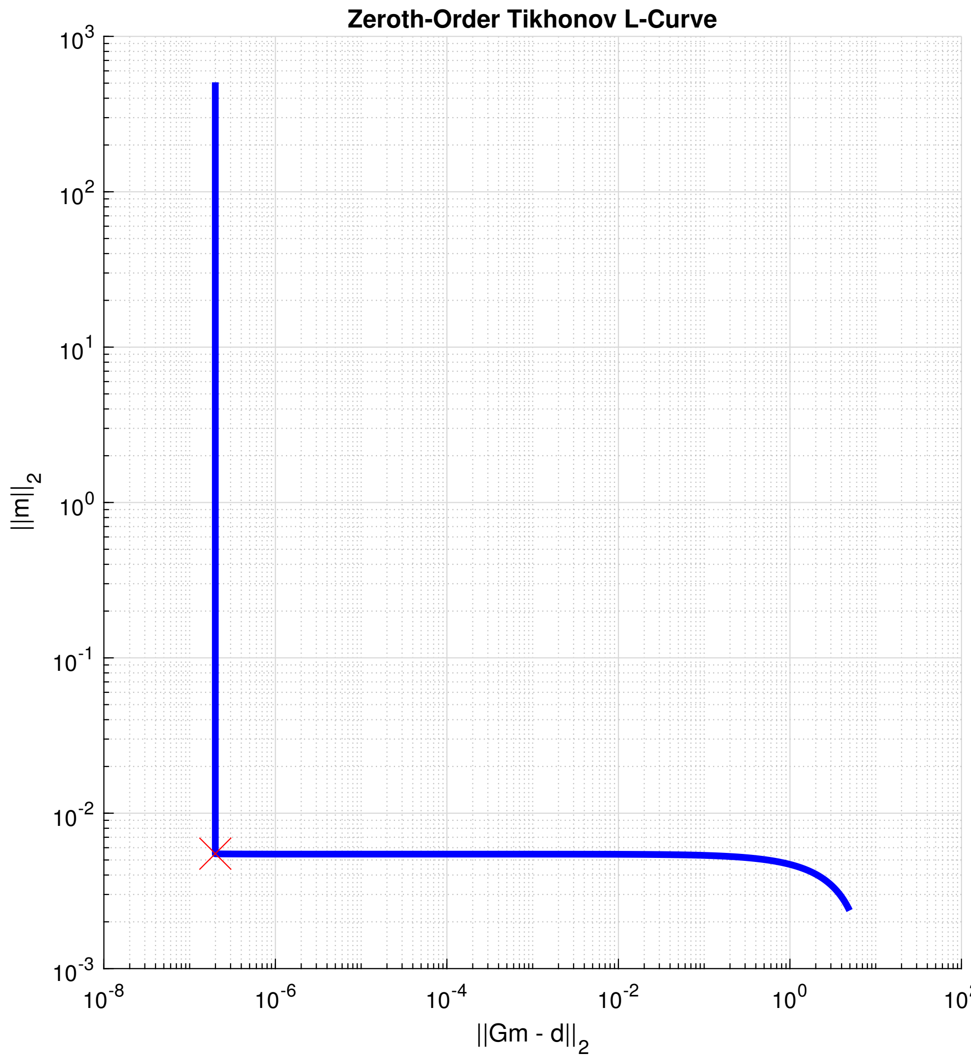

Figure 21: Zeroth-Order Tikhonov Regularization L-Curve

The model was solved with the modified normal equations below.

$$\boldsymbol{m}_{tikh} = \left(G_w^T G_w + \alpha^2 I\right)^{-1} G_w^T \boldsymbol{d}$$

The result is provided in figure 22.

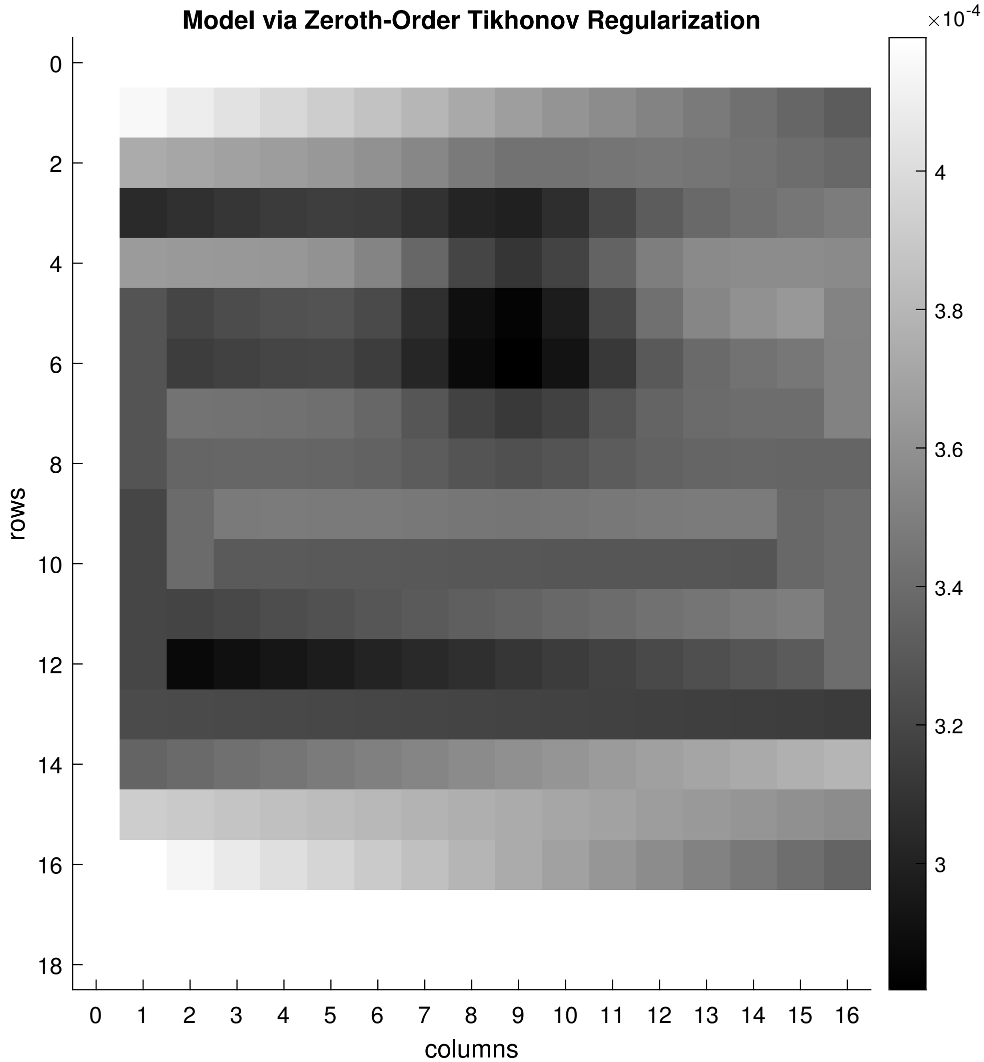

Figure 22: Zeroth-Order Tikhonov Regularization Solution

Using the discrepancy principle is difficult in this situation due to a few reasons. The first common heuristic is to bound the model residuals to the value $\delta < \sqrt{m-n}$, however in our situation $m = n$ which means bounding δ to zero is unrealistic in the presence of noise. The second heuristic is to simply bound $\delta < m$, however notice that in the L-Curve in figure 21 that all model residuals are below the value 256, which does not discriminate any potential values in our solution.

3.1.3 Part C - Inverse Problem Solution via Second-Order Tikhonov Regularization

Again, the L-curve was created using the provided functions `l_curve_tikh_gsvd()` and `l_curve_corner()` in the `PEIP` library. The inputs to the functions used the provided values of G and $\boldsymbol{d}$, not their weighted versions. This resulted in the following plot in figure 23, where the reported value $\alpha = 1.634 \times 10^{-7}$.

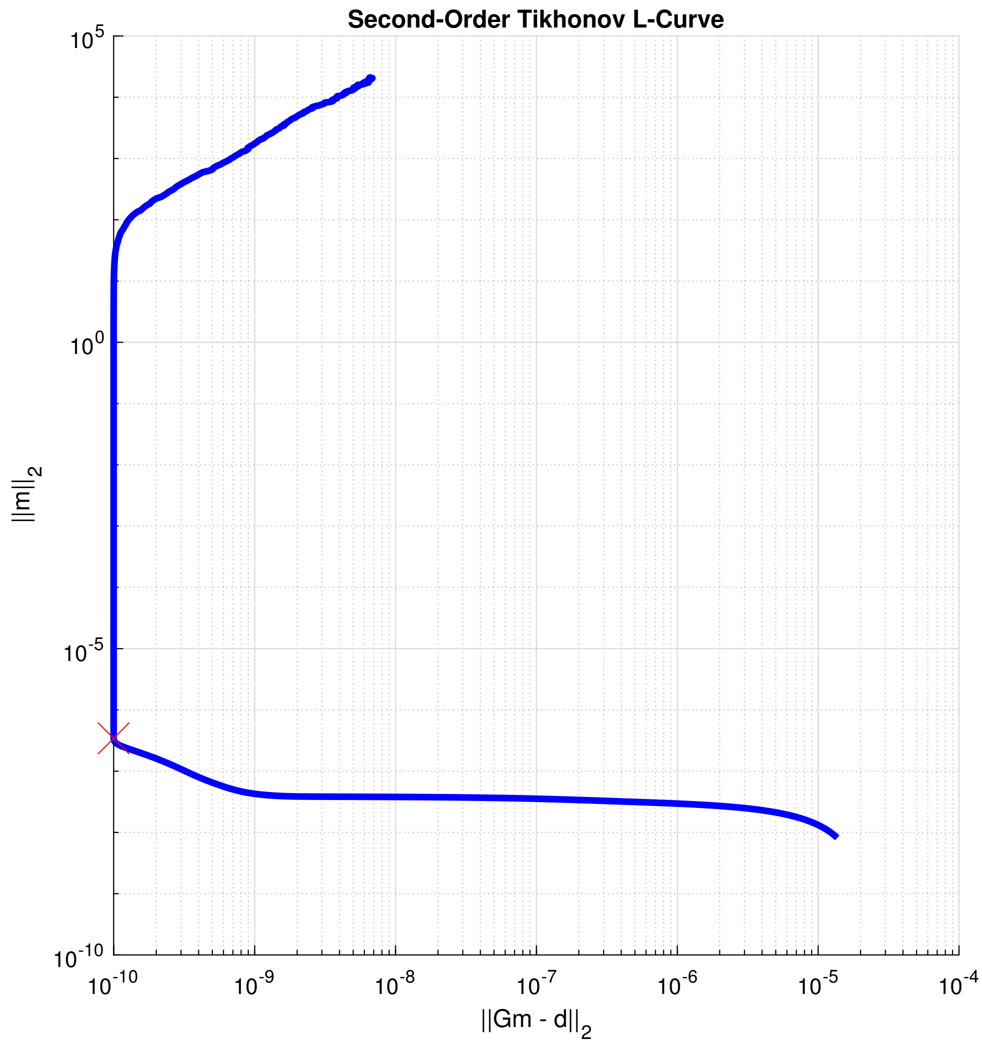

Figure 23: Second-Order Tikhonov Regularization L-Curve

Notice the curve above is not truly an L-shape as the top of the curve begins tilting into the positive model residual direction. This did not seem to have any impact in selecting the correct corner.

The model was solved with the modified normal equations below.

$$\boldsymbol{m}_{tikh,L2} = \left(G_w^T G_w + \alpha^2 L_2^T L_2\right)^{-1} G_w^T \boldsymbol{d}$$

The result is provided in figure 24.

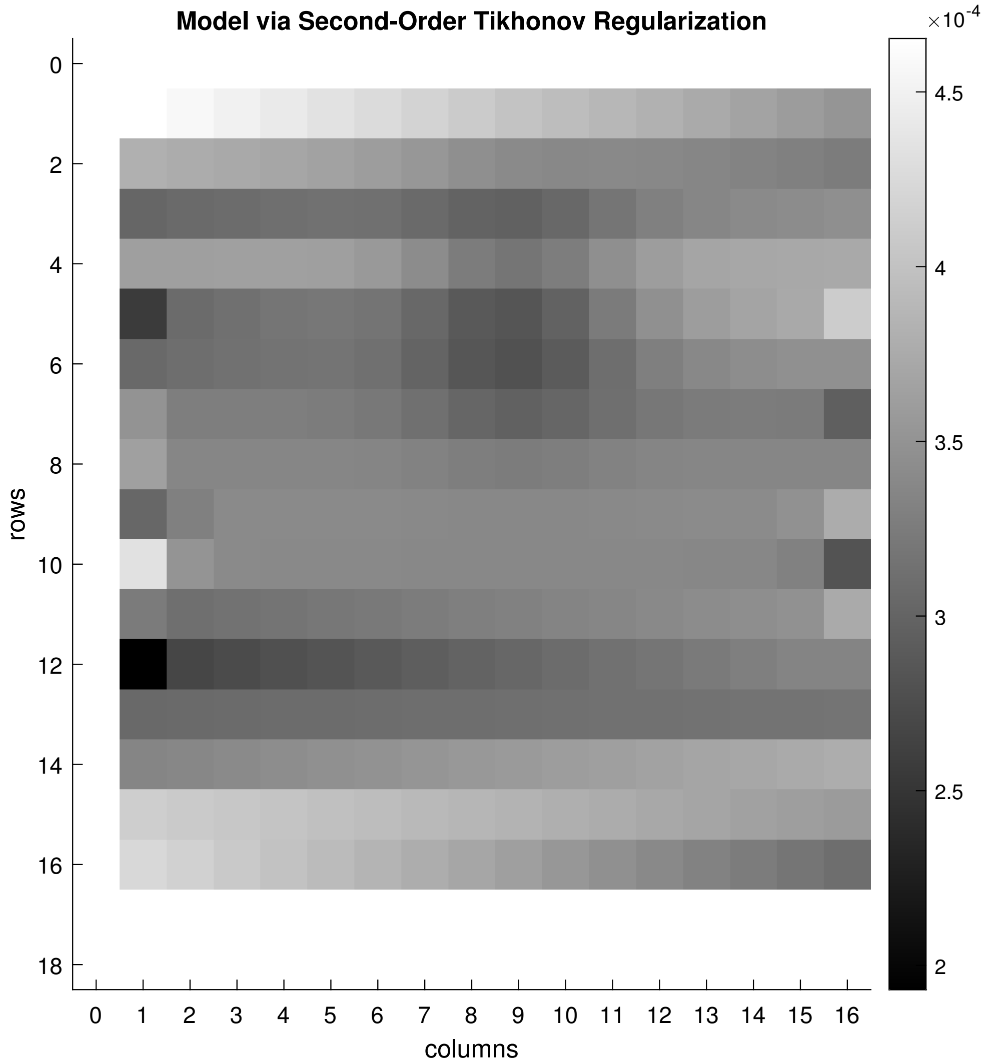

Figure 24: Second-Order Tikhonov Regularization Solution

3.1.4 Part D - Discussion

Between the three methods, TSVD, Zeroth-Order Tikhonov Regularization, and Second-Order Tikhonov Regularization, each model looks very different from the other. In the TSVD approach, vertical bands do appear within the center 8 blocks especially near the edges. However in the Zeroth-Order approach, horizontal bands appear instead. No bands in either direction appear in the 2nd order approach. The roughing matrix L_2 used in Part C consisted of the finite difference for the second derivative in both the row and column directions unlike the other L_2 matrix formulations provided in the textbook which likely explains these differences.

4 Problem 4

Exercise 5 in Section 4.10

4.1 Solution

To bias a model solution towards a particular solution $\boldsymbol{m}_0$, the objective function can be modified to be expressed accordingly below.

$$\min||G\boldsymbol{m}-\boldsymbol{d}||_2^2+\alpha^2||L\left(\boldsymbol{m}-\boldsymbol{m}_0\right)||_2^2$$

The expression within the minimization is a function of $f\left(\boldsymbol{m}\right)$ which can be expanded.

$$f\left(\boldsymbol{m}\right)=||G\boldsymbol{m}-\boldsymbol{d}||_2^2+\alpha^2||L\left(\boldsymbol{m}-\boldsymbol{m}_0\right)||_2^2$$

$$=\left(G\boldsymbol{m}-\boldsymbol{d}\right)^T\left(G\boldsymbol{m}-\boldsymbol{d}\right)+\alpha^2\left(\left(L\boldsymbol{m}-L\boldsymbol{m}_0\right)^T\left(L\boldsymbol{m}-L\boldsymbol{m}_0\right)\right)$$

$$=\left(\boldsymbol{m}^TG^T-\boldsymbol{d}^T\right)\left(G\boldsymbol{m}-\boldsymbol{d}\right)+\alpha^2\left(\left(\boldsymbol{m}^TL^T-\boldsymbol{m}_0^TL^T\right)\left(L\boldsymbol{m}-L\boldsymbol{m}_0\right)\right)$$

$$=\boldsymbol{m}^TG^TG\boldsymbol{m}-\boldsymbol{m}^TG^T\boldsymbol{d}-\boldsymbol{d}^TG\boldsymbol{m}+\boldsymbol{d}^T\boldsymbol{d}+\alpha^2\left(\boldsymbol{m}^TL^TL\boldsymbol{m}-\boldsymbol{m}^TL^TL\boldsymbol{m}_0-\boldsymbol{m}_0^TL^TL\boldsymbol{m}+\boldsymbol{m}_0^TL^TL\boldsymbol{m}_0\right)$$

Notice that terms $\boldsymbol{m}^TG^T\boldsymbol{d}$ and $\boldsymbol{d}^TG\boldsymbol{m}$ both evaluate to the same scalar value which means the can be added together to the term $2\boldsymbol{m}^TG^T\boldsymbol{d}$. Can the same be said for the other terms?... let's assume they do for now...

$$f\left(\boldsymbol{m}\right)=\boldsymbol{m}^TG^TG\boldsymbol{m}-2\boldsymbol{m}^TG^T\boldsymbol{d}\boldsymbol{d}^T\boldsymbol{d}+\alpha^2\left(\boldsymbol{m}^TL^TL\boldsymbol{m}-2\boldsymbol{m}_0^TL^TL\boldsymbol{m}+\boldsymbol{m}_0^TL^TL\boldsymbol{m}_0\right)$$

$$=\boldsymbol{m}^TG^TG\boldsymbol{m}-2\boldsymbol{m}^TG^T\boldsymbol{d}+\boldsymbol{d}^T\boldsymbol{d}+\alpha^2\boldsymbol{m}^TL^TL\boldsymbol{m}-2\alpha^2\boldsymbol{m}_0^TL^TL\boldsymbol{m}+\alpha^2\boldsymbol{m}_0^TL^TL\boldsymbol{m}_0$$

To solve for the minimum, we must set the gradient of $\nabla f\left(\boldsymbol{m}\right)=\boldsymbol{0}$. This leads to the following expression below.

$$\nabla f\left(\boldsymbol{m}\right)=2G^TG\boldsymbol{m}-2G^T\boldsymbol{d}+2\alpha^2L^TL\boldsymbol{m}-2\alpha^2L^TL\boldsymbol{m}_0=\boldsymbol{0}$$

$$G^TG\boldsymbol{m}+\alpha^2L^TL\boldsymbol{m}=G^T\boldsymbol{d}+\alpha^2L^TL\boldsymbol{m}_0$$

This expression can be expanded into matrix form resulting in a linear system of equations.

$$\begin{bmatrix}G^T & \alpha L^T\end{bmatrix}\begin{bmatrix}G\\ \alpha L\end{bmatrix}\boldsymbol{m}=\begin{bmatrix}G^T & \alpha^2L^TL\end{bmatrix}\begin{bmatrix}\boldsymbol{d}\\ \boldsymbol{m}_0\end{bmatrix}$$

The normal equations and model solution are also expressed as:

$$\left(G^T G + \alpha^2 L^T L\right) \boldsymbol{m} = G^T \boldsymbol{d} + \alpha^2 L^T L \boldsymbol{m}_0$$

$$\boldsymbol{m} = \left(G^T G + \alpha^2 L^T L\right)^{-1} \left(G^T \boldsymbol{d} + \alpha^2 L^T L \boldsymbol{m}_0\right)$$