Problem Set 4: Solutions

1a. Use Lagrange multipliers to solve the overdetermined matrix equation $\mathbf{Gm} = \mathbf{d}$, subject to the constraint that the L2 norm of $\mathbf{Hm} - \mathbf{f} = \mathbf{0}$ should be as close to zero as possible.

b. How does your solution to 1a above differ from the solution that you would obtain by augmenting the matrix $\mathbf{G}$ with the matrix $\mathbf{H}$ to create a revised matrix equation?

$$\begin{pmatrix} \mathbf{G} \\ \mathbf{H} \end{pmatrix} \mathbf{m} = \begin{pmatrix} \mathbf{d} \\ \mathbf{f} \end{pmatrix}$$

Solution:
a. In the first part of the problem, we use a Lagrange multiplier as a scalar and define a cost function:

$$\mathcal{L} = (\mathbf{Gm} - \mathbf{d})^{\mathbf{T}}(\mathbf{Gm} - \mathbf{d}) + \lambda(\mathbf{Hm} - \mathbf{f})^{\mathbf{T}}(\mathbf{Hm} - \mathbf{f})$$

Thus

$$\frac{\partial \mathcal{L}}{\partial \mathbf{m}} = 2\mathbf{G}^T(\mathbf{Gm} - \mathbf{d}) + \mathbf{2}\lambda\mathbf{H}^{\mathbf{T}}(\mathbf{Hm} - \mathbf{f}) = \mathbf{0}.$$

Solving for $\mathbf{m}$,

$$m = (\mathbf{G}^T\mathbf{G} + \lambda\mathbf{H}^T\mathbf{H})^{-1}(\mathbf{G}^T\mathbf{d} + \lambda\mathbf{H}^{\mathbf{T}}\mathbf{f})$$

b. In the second case, we define an augmented matrix:

$$\tilde{\mathbf{G}} = \begin{pmatrix} \mathbf{G} \\ \mathbf{H} \end{pmatrix} \text{ and } \tilde{\mathbf{d}} = \begin{pmatrix} \mathbf{d} \\ \mathbf{f} \end{pmatrix}.$$

Then the solution $\mathbf{m}$ can be represented as:

$$\begin{aligned} \mathbf{m} &= (\tilde{\mathbf{G}}^{\mathbf{T}}\tilde{\mathbf{G}})^{-\mathbf{1}}\tilde{\mathbf{G}}^{\mathbf{T}}\tilde{\mathbf{d}} \\ &= \left(\begin{bmatrix}\mathbf{G}^T \; \mathbf{H}^T\end{bmatrix} \begin{bmatrix} \mathbf{G} \\ \mathbf{H} \end{bmatrix}\right)^{-1} \begin{bmatrix}\mathbf{G}^T \; \mathbf{H}^T\end{bmatrix} \begin{bmatrix} \mathbf{d} \\ \mathbf{f} \end{bmatrix} \\ &= (\mathbf{G}^T\mathbf{G} + \mathbf{H}^T\mathbf{H})^{-1}(\mathbf{G}^T\mathbf{d} + \mathbf{H}^{\mathbf{T}}\mathbf{f}) \end{aligned}$$

The solutions in parts a and b are the same when $\lambda = 1$.

2. Consider the standard matrix equation $\mathbf{Gm} = \mathbf{d}$, where:

$$\mathbf{G} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0.01 \end{pmatrix},$$

and

$$\mathbf{d} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Uncertainties in the elements of $\mathbf{d}$ are identified as σ_i.

a. What is the least-squares solution for $\mathbf{m}$ if $\sigma_i = 0.1$ for all i?

b. What is the (row-weighted) least-squares solution for $\mathbf{m}$ if $\sigma_1 = \sigma_2 = 0.1$ and $\sigma_3 = 10$?

c. Comment on your results from cases a and b above? What would happen if $\sigma_1 = \sigma_3 = 0.1$ and $\sigma_2 = 10$?

Solution:
a. $\mathbf{m} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d}$, so $\mathbf{m}^T = [1 \;\; 1.01]$.
b. If the solution is row-weighted, then the 3rd row of $\mathbf{G}$ has minimal impact on the solution and $\mathbf{m}^T = [1 \;\; 1]$.
c. Alternatively, if σ is 0.1 for rows 1 and 3, and 10 for row 2, then the row weighted elements of $\mathbf{G}$ are the same in rows 2 and 3, and both rows contribute equally to the solution. In this case, $\mathbf{m}^T = [1 \;\; 50.5]$.

3. Suppose that you have temperature data at fixed depths (such as CTD bottle depths) and you would like to find a functional form to describe the vertical temperature structure in the range between 150 and 900 m depth.

a. Download the following profile data from the course web site:
http://www-mae.ucsd.edu/~sgille/sio221b/ps4_profile.dat
and least-squares fit a linear profile of the form $T = m_1 + m_2 z$ to the temperature data. In this data, column 3 contains depth, column 4 contains temperature, column 5 contains salinity, and column 6 is oxygen. The particular station was collected on 25 November 1972 at 35.32°W, 30.43°S.

b. Assume that the observational error is 0.1°C at all depths. What are the estimated errors in your parameters m_i? Is the functional misfit $\langle(\mathbf{Gm} - \mathbf{T})^2\rangle$ consistent with the assumed errors in T? You can do this by computing the variable

$$\chi^2 = \frac{(\mathbf{Gm} - \mathbf{T})^T(\mathbf{Gm} - \mathbf{T})}{\sigma^2}$$

and checking whether χ^2 is equal to $N - M$. (More formal procedure would have you compute the complete gamma function to evaluate whether the observed value of χ^2 is plausible.)

c. Verify that the formal error bars that you have derived are consistent with error bars that would be derived using a Monte Carlo simulation. To estimate alternate errors in m_i carry out a Monte Carlo simulation using the following procedure: 1. Generate 100 or more data sets of normally distributed fake perturbations with a standard deviation equivalent to the observed data (using "randn" in Matlab, for example). 2. With each set of noise, randomly perturb the temperature data, and recompute the least-squares fit solution. 3. Compute the standard deviations of your estimates of m_i. Do your error bars differ from the error bars derived in part b?

Solution:
a. Least-squares fit from rows 9 through 25 of the data. This yields $m_1 = 19.1343°\text{C}$ and $m_2 = 0.0169°\text{C/m}$.

b. The error covariance matrix is $C_{mm} = 0.1^2(\mathbf{G^TG})^{-1}$. The formal errors in the solution are the square root of the diagonals of C_{mm}. Thus $\sigma_1 = 0.065$ and $\sigma_2 = 1.19 \times 10^{-4}$

In this case χ^2 is 218 and $N - M$ is 10. This suggests that our estimated a priori error is too small compared with the typical misfit. This mismatch between χ^2 and $N - M$ can occur when the a priori error bars are erroneous or when the model used to fit the data is a poor fit to the variability.

c. Using the Monte Carlo procedure discussed in the problem, I find that the error bars for m_1 and m_2 are the same as the theoretical error bars to 2 significant digits. This demonstrates that the formalism that we used to compute error bars in part b is consistent with error bars derived using a brute strength approach, but it does not guarantee that the true error bar is the same size as our estimate.