# Problem Set 4: SIO 221B, Data Analysis

*due Friday, November 8, 2002*

**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?

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

**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 $\sigma_i$.

**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 $\sigma_2 = 10$?

**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 $\chi^2$ is equal to $N - M$. (More formal procedure would have you compute the complete gamma function to evaluate whether the observed value of $\chi^2$ 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?