**12.864 Inference from Data and Models     14 March 2005**
**Problem Set No. 3     Due: 30 March 2005**

1. Let an observation vector be given by

$$\mathbf{y} = \begin{bmatrix} -0.41043 & 0.48023 & -1.2784 & -1.0017 & -0.89310 & 0.94575 & 0.80744 & 0.065709 & -1.1805 & -0.25846 \end{bmatrix}^T$$

where $n_i$ is white noise of unit standard deviation and zero mean. $0 \le r \le 10$, sampled at points $r_j = 1, 2, ..., 10$. You have reason to believe that the data are described by a sum of sine waves,

$$y_0(t) = \sum_{n=1}^{N} a_n \sin\left(\frac{2\pi n r}{20}\right)$$

evaluated at the sample points $r_j$.

(a) Let $N = 5$. Using the singular value decomposition, find an estimate for the $a_n$ and an estimate of their standard error, assuming that you know the noise variance is 1. If you estimate the noise variance from your residuals what do you get? Are the $a_n$ fully resolved? Are the residuals consistent with the hypothesis about $n_i$?

(b) Now assume $N = 20$. How do the answers to part (a) change? Which of the equations is most important in determining the solution?

(Use the untapered SVD in (a), (b)).

(c) Now repeat (a), (b) using the tapered SVD, explaining what you are doing, and describe any differences in the solutions.

2. Solve problem 1, without using the SVD, but conventional matrix-vector least-squares. How do the solutions compare?

3. Some observations of a temperature field down a borehole (well) produce the following data, at uniform depth intervals, $z_n = n$, $n = 0, 1, 2, ...$

$$\mathbf{y} =$$
$$\begin{bmatrix} 1.57 & -0.67 & 0.12 & -0.71 & -3.14 & -1.81 & -2.81 & -5.03 & -5.67 & -6.83 & -8.19 & -8.27 & -10.59 & -8.82 & -12.14 & -12.89 \end{bmatrix}^T .$$

It is believed that the data are describable with a straightline $a + bz$, where $\mathbf{x} = [a, b]^T$ are constants. The noise in the data is thought to have zero mean, and there is no knowledge of the sign of $a$ or $b$. Using the Gauss-Markov (minimum variance estimate) find estimates of $a, b$, their standard errors, and error covariance when,

(a) $\mathbf{R} = \langle \mathbf{nn}^T \rangle = \sigma^2 \mathbf{I}$, $\sigma^2 = 1$. $\mathbf{S} = \|\langle \mathbf{xx}^T \rangle\| \to \infty$

(b) $\mathbf{R} = \langle \mathbf{nn}^T \rangle = \sigma^2 \mathbf{I}$, $\sigma^2 = 1$. $\mathbf{S} = \langle \mathbf{xx}^T \rangle = 4\mathbf{I}$.

(c) $\mathbf{S} = \left\| \langle \mathbf{xx}^T \rangle \right\| \to \infty$, and, $\mathbf{R}$ is the $16 \times 16$ matrix,

$$\mathbf{R} = \left\{ \begin{array}{cccccc} 1 & .9999 & .9998 & . & .9985 \\ .9999 & 1 & .9999 & . & .9986 \\ .9998 & .9999 & 1 & . & . \\ . & . & . & 1 & .9999 \\ .9985 & .9986 & . & .9999 & 1 \end{array} \right\}$$

(an easy way to set this up, e.g., in Matlab, is to define a vector, $\mathbf{f} = [1 : -0.0001 : 0.9985]'$ (that is a 16-component vector declining from 1, in steps of 0.0001) and writing $\mathbf{R} = $toeplitz$(\mathbf{f})$. A so-called toeplitz matrix has constant elements along each of its diagonals. Best to confirm that $\mathbf{R}$ is positive definite.)

(d) Solve (c) using the SVD.