SVD: Another Example

Shenfu Dong, a postdoctoral researcher at Scripps, has been comparing in situ and satellite measurements of sea surface temperature from Drake Passage. We would like the in situ and satellite measurements to be the same, but unfortunately, Shenfu finds that the satellite measurements disagree with the in situ measurements. The disagreement varies depending on how strongly the wind is blowing and also on the water vapor content of the atmosphere, as shown in Figure 1. Shenfu fit the difference between the temperatures to a function of the form $T = a_0 + a_1 s + a_2 w$. She found that $a_0 = 1.3$, $a_1 = -0.09$, and $a_2 = -0.03$.

Shenfu’s simple model assumes that $\Delta T$ depends linearly on wind speed $s$ and water vapor $w$. Obviously that’s not a perfect fit to these noisy observations. When we look closely at the data, we might decide that $\Delta T$ has an exponential decay with increasing wind speed, so that $\Delta T \propto \exp(s/\alpha)$, where $\alpha$ measures how quickly the exponential increases with increasing $s$. Let’s try $\alpha = 100$ as an initial test. Then:

$$A = \begin{bmatrix} 1 & s_1 & w_1 & \exp(s_1/\alpha) \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

We can plot the functions that make up the columns of $A$, as shown in Figure 2. By eye we can tell that...
exp(s/100) is nearly constant. (This isn’t surprising, since for small \( x \) we know that \( \exp(x) \approx 1 + x \).) This yields the fit \( \Delta T = 29.62 + 0.22s - 0.03w - 28.44\exp(s/100) \). This looks a little suspicious. More telling, the singular values of \( A \) are 340, 68, 6.5, and 0.004. The largest and smallest differ by 5 orders of magnitude, and that suggests to us that we may have redundant information in the columns of \( A \). (If the wind speed data spanned a larger parameter range, then 1 and \( \exp(s/100) \) would diverge and this problem would disappear.) The singular value decomposition shows solution based on three singular values is \( \Delta T = 0.66 - 0.098s - 0.027w + 0.66\exp(s/100) \).

**Error Bar Examples**

To think about error bars for least-squares fits, let’s start with a limiting case in which we solve for one unknown parameter, using a weighted fit, in which the weights \( \sigma \) are constant for all rows of \( A \). If our observations are \( x \), then their mean is \( \bar{x} = y \), so that

\[
\mathbf{x} = \mathbf{A}^{-1} y.
\]

Thus our formal solution is:

\[
y = (A_{wt}^T A_{wt})^{-1} A_{wt}^T x_{wt} = \left( \frac{N}{\sigma^2} \right)^{-1} \sum_{i=1}^{N} \frac{y_i}{\sigma^2} = \bar{x}.
\]

We know from earlier discussions, that the uncertainty in \( \bar{x} \) should be equivalent to \( \sigma/\sqrt{N} \), the uncertainty in each \( x \) divided by \( \sqrt{N} \). This is \( (A_{wt}^T A_{wt})^{-1} \), as we expected.

Importantly, this rule for finding uncertainties in \( x \) will work, even if \( \sigma \) varies. In the wind speed versus wave height example, the square-root of the diagonals gives 0.06 and 0.01. Thus, formally we can say \( s = 0.17 \pm 0.06 + (0.81 \pm 0.01)h \). If we use only the “high-quality” data, we have less information, and the error bars increase slightly, so that \( s = 0.21 \pm 0.07 + (0.080 \pm 0.01)h \). (This tells us that we can reduce our formal uncertainties by including all available data, even if some are of poorer quality than others.)

In the case represented in Figure 1 from lecture 15, error bars for the fit to all observations and the weighted fit to the monthly mean are effectively the same: \( 0.02, 0.05, 0.04 \), even though they are computed from matrices \( A \) that have vastly different numbers of elements.

**Interpreting Misfit in a Weighted Fit: Chi-Squared Criteria**

When we say that the uncertainty in our data \( b_i \) is \( \sigma_i \), we’re assuming that \( b_i \) should differ from our fit by about \( \sigma_i \). Thus we might predict:

\[
\epsilon = A_{wt} x - b_{wt} = \sum_{i=1}^{N} \sum_{j=1}^{M} (a_{ij} x_j - b_{ij})^2 \approx \sum_{i=1}^{N} 1 = N.
\]

Actually, we have to adjust this a little, since the \( x_j \) values are determined from the data, we lose one degree of freedom for each \( x \) for which we solve. For example, if \( N \) and \( M \) were both two, so that we were fitting two equations with two unknowns, the problem would be exactly determined, and \( \epsilon \) would be zero. So in reality, if our uncertainties are correct, we expect that \( \epsilon \approx N - M \). There’s quite a range to this, of course. In the wave height/wind speed example, \( \epsilon \) is predicted to be \( 30 - 2 = 28 \) and is actually 29.5. In the annual cycle of temperature example, \( \epsilon \) is predicted to be \( 12 - 3 = 9 \) for the monthly mean case and is actually 8.6.

We can use the weighted value of \( \epsilon \) to evaluate the success of our least-squares fit. For this purpose, we usually refer to the weighted \( \epsilon \) as \( \chi^2 \) or “chi-squared”:

\[
\chi^2 = \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{(a_{ij} x_j - b_{ij})^2}{\sigma_i^2} = N - M.
\]
The variable $\chi^2$ can vary quite a bit, but if $\chi^2$ differs substantially from $N - M$, then that might tell us that something is wrong with our fit. There are three major possibilities:

- Our estimates of our uncertainty are misguided. Perhaps we’ve estimated $\sigma$ based on some theoretical concept about instrument performance, when in reality errors are significantly larger. Perhaps we’ve been too conservative in estimating $\sigma$, and errors are actually much smaller. If $\chi^2$ is orders of magnitude different from $N - M$, then it’s quite possible that $\sigma$ is a poor guess.

- Alternatively, we may have chosen a poor model with which to fit our observations. Perhaps the observations vary sinusoidally, and we’re trying to fit them to a straight line. Or vice versa.

- Errors in our data are not normally distributed.