1. The Warmup Assume $\mathbf{n}$ is Gaussian with zero mean and variance $< n(t)n(t') >= \sigma\delta_{t,t'}$ in all of the problems.

   (a) For the matrix
   $$E = \begin{pmatrix} 2 & -1 & 1 \\ 3 & 2 & 1 \end{pmatrix}$$

      i. calculate the data and resolution matrices

      ii. Let there be observations $\mathbf{y} = [1, 2]^T$ such that
      $$E\mathbf{x} + \mathbf{n} = \mathbf{y}.$$
      This is clearly formally underdetermined. Find the solution which minimizes the cost function
      $$J = \mathbf{x}^T\mathbf{x}$$
      using least squares and compare the solution to that obtained via SVD (With the null space set to zero).

      iii. What is the uncertainty in the solution?

      iv. Now consider instead
      $$E = \begin{pmatrix} 2 & 3 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$$
      and the formally overdetermined problem
      $$E\mathbf{x} + \mathbf{n} = \mathbf{y},$$
      where $\mathbf{y} = [1, 2, -1]^T$. Find the least squares solution which minimizes $\mathbf{n}^T\mathbf{n}$. Again, compare this solution to the SVD.

   (b) Let A be an arbitrary $M \times N$ matrix of reals. Solve the least squares problem which minimizes
   $$J = \mathbf{x}^T\mathbf{x} + \alpha^{-2}\mathbf{n}^T\mathbf{n}$$
   and rewrite the problem in terms of SVD. Discuss what happens to the small singular value contributions.

(c) There is one observation

$$x + n_1 = 1$$

and a-priori statistics $< n_1 >=< x >= 0$, $< n_1^2 >=< x^2 >= 1/2$.

i. Waht is the best estimate of x and n?

ii. A second measurement becomes available,

$$x + n_2 = 3$$

with $< n_2 >= 0$, and $< n_2^2 >= 4$. What is the best estimate of x and what is its estimated uncertainty? Are the various a-priori statistics consistent with the final result?

(d) Two observations of unknown x produce the apparent results $x = 1$ and $x = 3$. Produce a reasonable value for x under the assumption that

- Both observations are equally reliable
- The second one is much more reliable (but not infinitely so) than the first one.

Make some reasonable numerical assumption about what "reliable" means and justify as much as possible. Can you write the two equations in a more sensible form?

(e) Two observations of 3 unknowns x, y, z produce apparent result,

$$\begin{aligned} x - y - z &= 1 \\ x - y - z &= 3 \end{aligned}$$

Discuss what if anything might be inferred from such a peculiar result. You can make sensible assumptions about what is going on, but state what these are.

(f) The temperature along coordinate x is *believed* to satisfy the linear rule $\psi_1 = a + bx$, with a and b constants. Measurements of the temperature were $y(r = 0) = 10$, $y(1) = 9.5$, $y(2) = 11.1$, $y(3) = 12$.

i. Using least squares, find an estimate of a and b and the noise in each measurement, and their standard errors.

ii. Solve again using SVD and discuss via the resolution matrices which of the observations proved to be most important. Is the solution fully resolved?

2. An application of SVD: download dna1.dat and dna2.dat from the assigment page. These are actual DNA-ligand spectroscopic results. Use the load command in matlab to input either of these files. The first row and column of the data files should be discarded. The matrix of interest is the remaining set of data values, call it D_1 and D_2. Each column represents a different measurement, each row contains light scatter spectral components.

   Plot several j of $D_i(:, j)$, $i = 1, 2$. Use the hold command in the plot so that you can superimpose them.

   The samples are made of a small number of fundamental chemicals. This means that it should be possible to write the k^{th} column of either matrix D_i of the mixture as

$$d_k = \sum_{i=1}^{r} c_{ik} \mathbf{v}_k,$$

   where r is the (small) number of different compounds, c is the concentration, and $\mathbf{v}$ is the spectral response of each of these compounds. Hence, we can write down the data of either experiment as

$$D = V \Lambda C$$

   where Λ is $r \times r$, C is $r \times n$, and V is $m \times r$. The aim of this exercise is to estimate r.

   - Use matlab SVD to obtain U, S, and V, where S is the matrix of singular values.
   - Noise in the data means that you will not see all of the singular values neatly separated into null and non-zero; but you will see their magnitude drop significantly. Examine S to determine a good estimate of r.
   - Also look at U. What should the elements of this matrix look like if there are singular values whose entries are nearly zero?
   - Using the first r singular values reconstruct an approximation of D.
   - How good is your estimate? propose some sort of a metric for this. Perhaps some sort of relative energy norm?
   - Find the relative compression of the approximation, i.e. define the number of memory units of D as n and the number of memory units required to store $\tilde{D}$ as $\tilde{n}$, then the compression is $C = \tilde{n}/n$.