ECE-486 Homework 8, Spring 2006

1. This problem involves investigating the required numeric precision to implement a particular filter.

   Floating point numbers are usually represented using a sign bit, a fractional part (usually between 0.5 and 1), and an exponent. For example, an IEEE single-precision number is represented using one sign bit, a 23-bit fractional part, and an 8-bit exponent. Double precision uses a 52-bit fractional part. In the following, assume that a similar floating-point representation is to be used, where the number of bits required for the fractional part is to be determined.

   A 6th order equiripple lowpass IIR filter for sample rate of 48 ksps and cutoff frequency 500 Hz may be designed using the MATLAB "ellip" function as follows:

   ```
   [b,a] = ellip(6,1,60,500/(48000/2));
   ```

   The frequency response of the filter (for the full-precision 'b' and 'a' coefficients) is shown below. As the numeric precision of the coefficients is reduced, the frequency response of the filter will eventually degrade.

   A simple Matlab function which rounds the fractional part to a given number of bits has been provided on the course www site.

   (a) Find the minimum resolution (in bits) for the "fractional part" of the coefficient representation to preserve the magnitude response illustrated below. That is, for a minimum resolution of n_0 bits, the frequency response calculated using $n_0 - 1$ or fewer bits will be noticeably different than that illustrated below. Make a single plot showing the magnitude responses of the filter when the filter coefficients are rounded to $n_0 - 4, n_0 - 3, \ldots, n_0$ bits. Identify the resolution at which the filter becomes unstable.

   (b) Alternatively, the filter could be implemented as a cascade of three second-order systems. You can find the required filter coefficients using the `tf2sos()` Matlab function. Repeat problem 1a assuming that this implementation is used (this time rounding the 2^{nd}-order filter coefficients). Your plots should show the *cascaded* 6^{th} order responses.