1. Introduction

The data obtained by measurements of biological systems is often noisy. To try to distinguish signal from noise, a good working assumption is that “noise frequencies are much greater than signal frequencies”.

To explain this idea more precisely, let’s assume that the data consists of a function \( f(t) \), where \( t \) denotes time, on some interval of time, \([0, T]\). We can expand \( f \) in a Fourier cosine series, obtaining some Fourier coefficients \( a_n \). What we mean by our assumption is that Fourier coefficients with large index (that is, \( a_n \) where \( n \) is large) represent noise, while Fourier coefficients with small index represent the signal. (We think of Fourier coefficients with large index \( n \) as “high frequency” because they correspond to functions \( \cos(n\pi/T)t \) which oscillate very fast.)

This means that you could remove the noise by computing the Fourier series expansion of \( f \), then taking the \( n \)th Fourier approximation \( f_n \) of \( f \), where

\[
f_n(t) = \frac{a_0}{T} + \sum_{i=1}^{n} a_i \cos \frac{i\pi t}{T}.
\]

Of course, if \( n \) is very large, \( f_n \) will approximate \( f \) very closely. This is undesirable—if you approximate \( f \) too well, you’ll approximate the noise as well as the data!

The solution to the problem is to decide (either by theory, or by measuring a system you already know), that some fraction of the signal energy, say 20%, is noise. Since we’ve already assumed that the noise is “high frequency” (that is, that the Fourier coefficients of the noise function have large index), this assumption leads to the following algorithm.

1. Compute the total energy of \( f \) (that is, the norm of \( f \)).
2. Find \( f_1 \), and find the norm of \( f_1 \). If the norm of \( f_1 \) is greater than or equal to 80% of the norm of \( f \), you’re done. If not, find \( f_2 \), and test the norms again. Continue the process until you find some \( f_n \) such that the norm of \( f_n \) is greater than 80% of the norm of \( f \).
3. This \( f_n \) is (assumed to be) the signal.

Often, the Fourier approximation \( f_n \) obtained by this procedure is referred to as an “80% smoothing” of the original data. We’ll call this the “low–frequency” smoothing of the signal.

When the fraction of the signal that’s noise is unknown, another assumption, sometimes called “thermalization” is called into play. Roughly, this assumes that the Fourier coefficients of signals tend to be large, and to vary greatly in size, while the Fourier coefficients of noise tend to be small, and roughly the same size.\(^1\)

In practice, this means that you compute the Fourier coefficients of the function, then throw out all the little ones, and take the remaining Fourier terms to be the signal. This is a different

\(^1\)For instance, “white noise” is, by definition, a signal where all the audible Fourier coefficients are equal. I’m stressing audible because the sum of the squares of the Fourier coefficients has to be finite for the Fourier series to converge—that is, if all the Fourier coefficients were equal, the signal would be infinite everywhere! “Pink noise” has Fourier coefficients which obey the relation \( a_{2n} \sim (1/2)^n a_n \) and \( b_{2n} \sim (1/2)b_n \), a decrease of about 3 decibels per octave. This decrease is characteristic of signals produced by musical instruments.
kind of smoothing—we’ll call it the “big–coefficient” smoothing of the signal. For instance, if \( f \) had only 11 Fourier coefficients, given by

\[
(a_0, \ldots, a_{10}) = (0.25, 0.117, 5, 0.06, 2.5, 0.1, 0.05, 0.04, 1.25, 0.13, 0.072), a_{50} = 0.1
\]

then it would be reasonable to decide that \( a_2 (5) \), \( a_4 (2.5) \), and \( a_8 (1.25) \) were “big”, while the remaining coefficients (averaging \( \sim 0.1 \)) were “small”. The resulting big coefficient smoothing \( f_s \) on the interval \([0, \pi]\) would look like

\[
f_s = 5 \cos 2t + 2.5 \cos 4t + 1.25 \cos 8t.
\]

Since the norm of \( f \) is about 51.837 and the norm of \( f_s \) is about 51.541, this smoothing contains more than 99% of the original signal energy. You can see the “noisy” and “smoothed” functions in the graphs below.

These kind of smoothings are harder to do. After all, the signal function usually has infinitely many nonzero Fourier coefficients. The procedure is to start from the beginning. Compute a reasonable number of Fourier coefficients, say 10 or 20. Then decide which are “big”, and assemble these into a function \( f_s \). If the norm of \( f_s \) is close to the norm of \( f \), you’re done.

If not, compute another 20 Fourier coefficients, pick the big ones, and add them to the ones you already picked to assemble a new \( f_s \). Keep doing this until your test function \( f_s \) has a reasonable fraction of the total energy of the data function \( f \) (at least 60–70%).

You can assume that if your smoothed function has most of the energy of \( f \), then it captures most of the signal, and you’re done. Be careful! If you only have 60% of the total energy of \( f \), you are throwing 40% of the data away—you might be missing important components of the signal that you’re trying to measure!

2. Problems

Please do the following problems. There is a Maple file on the Math 241 web page with some real ECG data (courtesy of the MIT ECG database) containing a fair amount of noise. The data measures electrical potentials in the human heart during normal heart function (ECG stands for electrocardiogram). In all these problems, use a Fourier cosine series.

- Find what you think are the 3 largest Fourier coefficients of this data. Can you propose biological explanations for each of these components of the signal?
- Find the total energy (norm) of the data function.
- Compute 60%, 70%, and 80% low–frequency smoothings of this data using Maple. Use a Fourier cosine series, and include graphs of your smoothings plotted along with the original data.
- Find what you think is the best big–coefficient smoothing of this data. Your smoothing \textit{must} have at least 60% of the energy of the original signal, but it is up to you to decide how much further to go.

Write one paragraph on why you took the approximation that you did (i.e. why you believe that the remaining Fourier coefficients represent noise, rather than signal). Include graphs of this smoothing plotted with the original signal, and with the high–frequency smoothing.