**Problem 11.1**

Suppose you’re given two spectra, \(X[k]\) and \(Y[k]\), each of which is the 1024-point FFT of a frame of speech with a sampling frequency of \(F_s = 16000\)Hz. You want to find out how similar these two spectra are.

In order to do that, you will compute filterbank coefficients,

\[
C_x[m] = \ln \sum_{k=0}^{1023} H_m[k] |X[k]|, \quad C_y[m] = \ln \sum_{k=0}^{1023} H_m[k] |Y[k]|
\]

where the filters, \(H_m[k]\), are given by

\[
H_m[k] = \begin{cases} 
\frac{k-k_m-1}{k_m-k_{m-1}} & k_m \leq k \leq k_{m-1} \\
\frac{k_{m+1}-k}{k_{m+1}-k} & k_{m+1} \leq k \leq k_m \\
0 & \text{otherwise}
\end{cases}
\]

The band edges, \(k_m\), should be uniformly spaced on a mel-scale, meaning that 

\[
m(f) = 2595 \log_{10} \left(1 + \frac{f}{700}\right)
\]

Assume that \(c[0] = 0\). Under that assumption, the liftering operation, Eq. 11.2-1, is equivalent to smoothing \(\ln |X[k]|\) by convolution with a digital-sinc function. What is the bandwidth, in Hertz, of the smoothing function (measure “bandwidth” as the frequency of the first null)?

**Problem 11.2**

Suppose you’re liftering a linear-frequency spectrum. The low-pass liftered spectrum is constructed from the low-pass liftered cepstrum as

\[
C_{LP}[k] = 2 \sum_{q=1}^{\frac{N}{2}-1} c_{LP}[q] \cos \left(\frac{2\pi k q}{N}\right)
\]

The low-pass liftered cepstrum is computed from the input cepstrum as

\[
c_{LP}[q] = \begin{cases} 
c[q] & 1 \leq q \leq 12 \\
0 & \text{otherwise}
\end{cases}
\]

The input cepstrum is computed from the input spectrum as

\[
c[q] = \frac{2}{N} \sum_{k=1}^{N-1} \ln |X[k]| \cos \left(\frac{2\pi k q}{N}\right)
\]

and the input spectrum, \(X[k]\), is the 1024-point FFT of a signal sampled at \(F_s = 16000\)Hz.

Assume that \(c[0] = 0\). Under that assumption, the liftering operation, Eq. 11.2-1, is equivalent to smoothing \(\ln |X[k]|\) by convolution with a digital-sinc function. What is the bandwidth, in Hertz, of the smoothing function (measure “bandwidth” as the frequency of the first null)?
Problem 11.3

Computing the MFCC involves the following steps:

1. Take the magnitude DFT, \(|X[k]|\), of one frame of audio, \(x[n]\).
2. Compute the weighted summation of \(|X[k]|\) within each mel-frequency band.
3. Take the logarithm.
4. Compute the DCT.

In these days of neural networks, it is stylish to represent every operation as a sequence of matrix multiplications followed by scalar nonlinearities. For example, suppose that \(x[n]\) is a time-domain sample of the original audio signal, and consider the following sequence of operations:

\[
\vec{a} = \begin{bmatrix} x[1] \\ \vdots \\ x[N] \end{bmatrix}
\]

\[
\vec{b} = W\vec{a}
\]

\[
\vec{c} = \begin{bmatrix} |b[1]| \\ \vdots \\ |b[M]| \end{bmatrix}
\]

\[
\vec{d} = V\vec{c}
\]

\[
\vec{e} = \begin{bmatrix} \ln(d[1]) \\ \vdots \\ \ln(d[L]) \end{bmatrix}
\]

\[
\vec{f} = U\vec{e}
\]

...where \(W\) is an \(M \times N\) matrix whose \((m, n)\)th element is \(w_{mn}\), \(V\) is an \(L \times M\) matrix whose \((l, m)\)th element is \(v_{lm}\), and \(U\) is a \(K \times L\) matrix whose \((k, l)\)th element is \(u_{kl}\).

Find formulas for \(u_{kl}\), \(v_{lm}\), and \(w_{mn}\), as functions of \(k, l, m, n\), so that the vector \(\vec{f}\) contains the MFCC.