Transform coding (continued..)

Last lecture I argued that we would like to transform code blocks of samples, $\mathbf{Y} = \mathbf{U}^T\mathbf{X}$, such that the columns of the matrix $\mathbf{U}$ are the eigenvectors of a matrix constructed from the autocorrelation function $R(d)$, namely

$$\mathbf{U\Lambda} = \mathbf{RU}$$

The eigenvalues correspond to $\mathcal{E}Y_i^2$.

Let the block sizes be of size m. For the case $m = 2$, the correlation matrix $\mathbf{R}$ is of the form:

$$\mathbf{R} = \begin{bmatrix} R(0) & R(1) \\ R(1) & R(0) \end{bmatrix}$$

In this case, it is easy to verify that (1,1) and (1,-1) are eigenvectors, and the respective eigenvalues are $\mathcal{E}Y_1^2 = R(0) + R(1)$ and $\mathcal{E}Y_2^2 = R(0) - R(1)$.

Let's consider larger values of k. If we examine samples from an image or audio file, then we find that the autocorrelation $R(d)$ is typically a decreasing function of the distance d between samples. The reason is that we expect the average squared difference between samples separated by a distance of $d+1$ to be greater than that of samples separated by a distance of d. So,

$$R(0) > R(1) > R(2) > R(3) > \ldots$$

For example, let $k = 3$, so that the blocks are triplets. To keep the numbers simple, suppose $R(0) = 1,\ R(1) = .9,\ R(2) = .8$. Then,

$$\mathbf{R} = \begin{bmatrix} R(0) & R(1) & R(2) \\ R(1) & R(0) & R(1) \\ R(2) & R(1) & R(0) \end{bmatrix} = \begin{bmatrix} 1 & .9 & .8 \\ .9 & 1 & .9 \\ .8 & .9 & 1 \end{bmatrix}$$

What are the eigenvectors of this matrix? You might guess that, like in the case of $k = 2$, the eigenvector with the largest eigenvalue would be the vector in direction $(1,1,1)^T$. This guess is close but not quite correct, however. If you calculate the eigenvectors of $\mathbf{R}$ (the columns of $\mathbf{U}$), and normalize them, you get:

$$\mathbf{U} = \begin{bmatrix} .57 & .71 & .42 \\ .59 & 0 & -.8 \\ .57 & -.71 & .42 \end{bmatrix}$$

up to two decimal points of precision. In particular, the first column is very close to $(1,1,1)$.

What are the eigenvalues? The eigenvalue of the first column of $\mathbf{U}$ is *approximately* $\mathcal{E}Y_1^2 = R(0) + R(1) + R(2) \approx 2.7$. The eigenvalue of the second column of $\mathbf{U}$ is $\mathcal{E}Y_2^2 = R(0) - R(2) \approx .2$. The eigenvalue of the third column of $\mathbf{U}$ is approximately $\mathcal{E}Y_3^2 = R(0) + R(2) - 2R(1) \approx .04$.

Note that, since the Y_1 values have a much greater range on average that Y_2 and Y_3, we should spend most of our bits encoding Y_1 and if we are trying to save bits, we should spend them on Y_1 and not on Y_2 and certainly not on Y_3.

[ASIDE (March 19): In the next few lectures, I will let the block size be m. The variable k will represent something else. I will try to go back and change the notation from the previous lecture, so that this doesn't confuse you when you study for the final exam...]

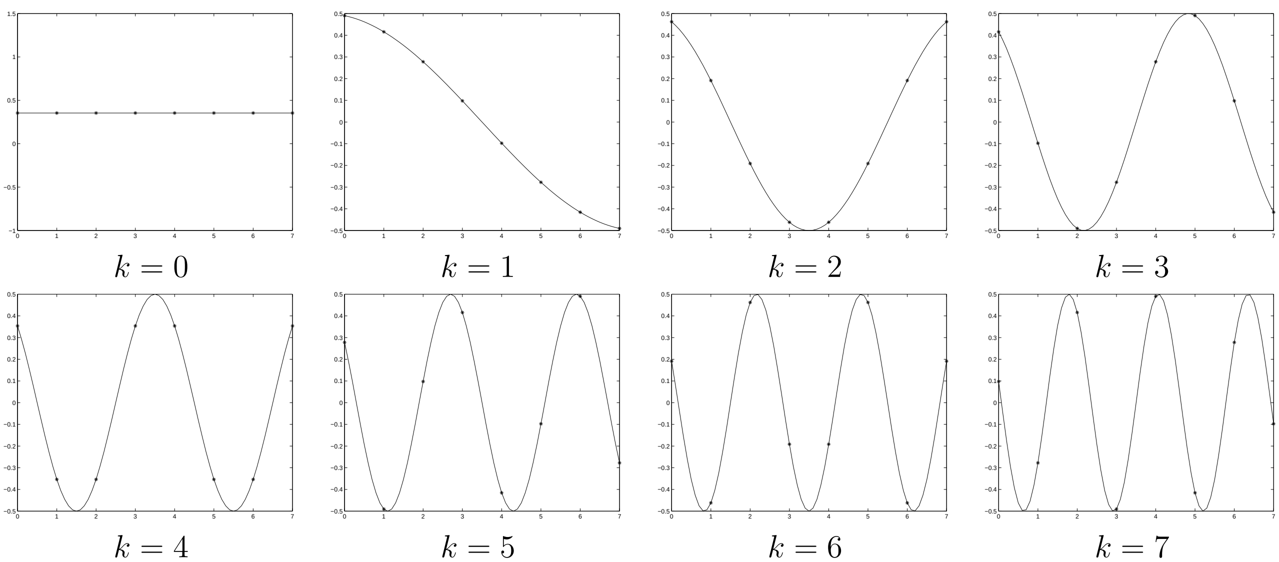
Discrete Cosine Transform (DCT)

For any data sequence $X_1, \ldots, X_n$ that one wishes to encode, one could compute the autocorrelations $R()$ and then solve for the matrix $\mathbf{U}$, and use this matrix to transform the sequence. It turns out that when the autocorrelation is a slowly decreasing function $R(d)$ of distance d between samples, a good approximation to the matrix $\mathbf{U}$ is often the following matrix $\mathbf{C}$, defined:

$$\mathbf{C}_{j,k} = \begin{cases} \frac{1}{\sqrt{m}}, & k = 0 \\ \sqrt{\frac{2}{m}} \cos(\ \frac{\pi}{m} k \ (j + \frac{1}{2}) \ ), & k = 1, \ldots, m-1 \end{cases}$$

where m is the block size, j is the index of an element within a block, k is a *frequency* of a cosine and $j, k \in \{0, 1, \ldots, m-1\}$.

The following plots show the row vectors of $\mathbf{C}^T$, – that is, the column vectors of $\mathbf{C}$ – for the case of $m = 8$. The columns of $\mathbf{C}$ are indexed by row number j. Thus, the horizontal axis (abscissa) in these plots is the variable j.



The transformation

$$\vec{Y} = \mathbf{C}^T \vec{X}$$

is called the *discrete cosine transform*, or *DCT*. The columns of $\mathbf{C}$ are orthogonal to each other and are of unit length, i.e.

$$\mathbf{C}^T \mathbf{C} = I$$

This orthogonality property is not obvious, but it can be shown to be true.

The above formula is very mysterious, at first glance. Where does it come from?

First note that the case $k = 0$ just gives a constant vector. This corresponds to the constant eigenvector case we talked about earlier. The factor $\frac{1}{\sqrt{m}}$ is there just to normalize this vector. Let's now turn to the case $k > 0$.

Let's first ignore the "shift by $\frac{1}{2}$" and the factors $\sqrt{2}m$. Let's just look at $\cos(\frac{\pi}{m} jk)$. Observe that

$$\cos(\frac{\pi}{m} jk) = \cos(\frac{2\pi}{2m} jk).$$

Since j and k are integers, this cosine function will have the value 1 whenever jk is a multiple of $2m$. In particular, this function will have k cycles over a distance $2m$ samples where j and k are both in $0, \ldots, m-1$. Thus, this cosine function will have $\frac{k}{2}$ cycles over m samples. That is what we see in the plots above.

Let's next consider the function $\cos(\frac{\pi}{m}k(j+\frac{1}{2}))$. This shifts the function $\cos(\frac{\pi}{m}kj)$ to the left by a distance $\frac{1}{2}$. The purpose of this shift is to make the cosine functions either even or odd symmetric over the points $j = 0, \ldots, 7$ – namely if you flip each function left-to-right then you get the same function back again, or you get the negative of the function. (You can convince yourself of this by examining the equations, or you can look at the above plots and see it with your eyes.)

Notice that if we don't shift by $\frac{1}{2}$, then the functions have even or odd symmetry over $j = 0, 1, \ldots, m$. But this is $m+1$ points, whereas our blocks are only of size m, i.e. our variables j and k are defined on $0, \ldots, m-1$. So, we need to shift by $\frac{1}{2}$ to get the nice symmetry property. [ASIDE: admittedly, it should not be obvious to you why this symmetry works best. Please just take my word for it for now.]

Finally, what about the factors $\sqrt{2}m$ for the case $k > 0$? As in the case $k = 0$, these factors are only there to normalize, i.e. ensure the columns of $\mathbf{C}$ are orthonormal.