Quantitative Understanding in Biology Module III: Linear Difference Equations Computer Laboratory

Scaling and Mean Centering in PCA

Generate a synthetic dataset comprised of 1000 (x,y) pairs by using the following commands in MATLAB:

```
>> u = random('norm', 5, 1, 1000, 1);
>> v = random('norm', 5, 1, 1000, 1);
>> x = 0.25 * u + 0.75 * v;
>> y = 25 * u - 75 * v + 700;
```

Note that all subsequent analysis the is to be on the (x,y) data; u and v are only used to synthesize this data set.

Make a scatter plot of the (x, y) data. Note that, unlike the example we worked through in class, this data is not even close to being mean-centered. Also, note that the scales of x and y differ by two orders or magnitude. Be sure to pay close attention to the scaling of all graphs you produce (in this lab and forever after); you may want to take manual control.

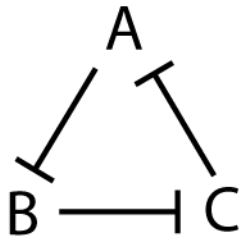
Perform a PCA analysis on the x,y data. How do you interpret the eigenvalues and the loadings? Are your results consistent with what your intuition suggested from the plot?

Now transform your (x,y) data so that each of x and y are mean centered and scaled according to their standard deviations. **Plot and perform a second PCA analysis on the transformed data, and again give your interpretation of the eigenvalues and the loadings.**

Comment on the effects of the transformation. Do you think you should always transform data in the manner above before you use PCA? Why or why not?

A Network of Interacting Genes

Consider a simple network of interacting genes:



This seemingly simple picture describes a somewhat elaborate model for the interactions of genes and their products. It says that the product of gene A inhibits the expression of gene B. In turn, the product of gene B inhibits the expression of C. Similarly, the product of gene C inhibits the expression of gene A.

A little thought reveals that this system may have some interesting dynamic characteristics. As the expression of gene A increases, we expect the expression of B to decrease. This is turn will lead to an increase in the expression of gene C (because its inhibitor, B, is present in decreasing quantities). But that, in turn, should lead to a decrease in the expression of A. Our thought experiment indicates that this system has the potential to exhibit oscillatory behavior.

We will proceed to build a simple, linear mathematical model of this system. In fact, such a system was expressed in bacteria using GFP as a reporter (Elowitz and Leibler, *Nature*, 2000). You can visually observe the intensity of GFP vary over time in the bacteria.

In order to explore this system more formally, we would begin by writing the differential equations corresponding to each biochemical reaction implied in the diagram. It is important to realize that there are two such species implied for each gene: its transcribed mRNA product, and its translated protein product. We'll denote the concentrations of mRNAs as m_A, m_B, and m_C; similarly the protein products will be denoted as p_A, p_B, and p_C.

A differential equation for the concentration of mRNA of A can be written as:

$$\frac{dm_A}{dt} = \alpha_0 - m_A - \alpha \cdot p_C$$

In the first term, α_0 represents a basal level of expression of mRNA. The second term represents degradation of the mRNA of A over time; this is simply proportional to the amount of mRNA present (you might include a coefficient here to capture the rate of degradation, but it is not necessary for our purposes).

The third term represents the inhibition of the expression of mRNA of gene A by the protein product of gene C. Protein C is a transcription factor that binds to the promoter region of the DNA encoding for A and prevents transcription. The degree of inhibition is given by the coefficient α. A key assumption in our model is that this relationship is linear. Does this seem reasonable?

We can turn this differential equation into a difference equation by considering a short time interval, τ. Our difference equation is:

$$m_A^{'} - m_A = \tau(\alpha_0 - m_A - \alpha \cdot p_C)$$

We can write similar equations for m_B and m_C.

A second set of differential equations comes from consideration of the protein products. We write:

$$\frac{dp_A}{dt} = \beta(m_A - p_A)$$

The term containing m_A indicates that the production of the protein is proportional to the amount of mRNA present. The negative p_A term represents degradation of the protein.

This equation can be written in terms of differences using the same reasoning as we followed for mRNA, and similar logic can be applied to the protein products of genes B and C.

Note that we use the same parameters, α_0, α and β for all three genes. In reality, they are probably different, but this simplified model will still be sufficient to capture the essential dynamic character of the system.

Show that the system as modeled above can be represented by the matrix equation:

$$\begin{pmatrix} m_A \\ p_A \\ m_B \\ p_B \\ m_C \\ p_C \\ 1 \end{pmatrix}_{n+1} = \begin{pmatrix} (1-\tau) & 0 & 0 & 0 & 0 & -\tau\alpha & \tau\alpha_0 \\ \tau\beta & (1-\tau\beta) & 0 & 0 & 0 & 0 & 0 \\ 0 & -\tau\alpha & (1-\tau) & 0 & 0 & 0 & \tau\alpha_0 \\ 0 & 0 & \tau\beta & (1-\tau\beta) & 0 & 0 & 0 \\ 0 & 0 & 0 & -\tau\alpha & (1-\tau) & 0 & \tau\alpha_0 \\ 0 & 0 & 0 & 0 & \tau\beta & (1-\tau\beta) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} m_A \\ p_A \\ m_B \\ p_B \\ m_C \\ p_C \\ 1 \end{pmatrix}_n$$

Using the following values for the model parameters, simulate the system: τ = 0.005; α_0 = 4; α = 1; β = 0.01. For your initial condition, use a unit amount of protein A only. Note that the time-step, τ, is fairly small. You don't need to collect and plot data for every single time-point. How can you do this efficiently?

Compute the eigenvalues and eigenvectors for this model. Comment.

Repeat both a simulation and the computation and interpretation of eigenvalues and eigenvectors for α=2.5.

Without running additional simulations, can you determine a value for α that would give a sustained, periodic oscillation of mRNAs and proteins? How sensitive is this model to this parameter? Do you think a system such as this can be used as a biological clock?

