Matrix Analysis EE312

LTS2 - EPFL

Spring 2024 - week 7

Submit your answers **individually** on Moodle before **May 3rd noon**.

1 Eigenvalues and population genetics

Genes control traits (physical characteristics, behaviors, etc) of an individual and are passed on from its parents. Most of them come in pairs with one inherited from the mother and the other from the father. Genes can have different "versions" or alleles.

We are considering a scenario where the gene for a specific trait has only two alleles, represented by the letters A and a. Individuals of a population present a pair of alleles that can be identical or different. Hence, *only three genotypes are possible*: AA, Aa, aa.

There is an **equal probability** to inherit an allele from each parent. The order of alleles does not matter. Only individuals with the same genotype can breed together to have offspring. In other words, the two parents of an individual can only have the following genotypes together: "$AA - AA$", "$Aa - Aa$" and "$aa - aa$".

We want to study the relative proportion of each possible genotypes in a population, i.e. the genotype distribution. One can ask how it evolves in a population after n generations from only the initial population. Simple probabilities can describe the changes from generation to generation.

The vector

$$\boldsymbol{x}_n = \begin{pmatrix} \alpha_n \\ \beta_n \\ \gamma_n \end{pmatrix},$$

where α_n, β_n and γ_n denote the proportion of the population with genotype AA, Aa, aa, in the n-th generation, can be used to describe the genotype distribution.

1. What is the matrix $\boldsymbol{M}$ s.t. $\boldsymbol{x}_n = \boldsymbol{M}\boldsymbol{x}_{n-1}$?

2. Express $\boldsymbol{x}_n$ as a function of $\boldsymbol{x}_0$ and M

3. What are the eigenvalues and the eigenvectors of $\boldsymbol{M}$?

4. Compute the genotype distribution of the population after $n = 10$ generations with $\alpha_0 = \frac{1}{8}$, $\beta_0 = \frac{3}{4}$ and $\gamma_0 = \frac{1}{8}$.
   <u>*Hint:*</u> $2^{10} = 1024$.

5. Compute the genotype distribution of the population when the number of generations n grows to infinity, with $\alpha_0 = \frac{1}{8}$, $\beta_0 = \frac{3}{4}$ and $\gamma_0 = \frac{1}{8}$.

6. Let us now consider the problem as a continuous one, depending on the time t. Let us denote by $\boldsymbol{x}(t)$ the continuous equivalent of $\boldsymbol{x}_n$. Rewrite the equation from the first question as

   $$\frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{Q}\boldsymbol{x}(t),$$

   using the approximation

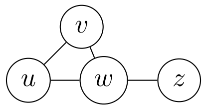
   $$\frac{d\boldsymbol{x}(t)}{dt} \approx \boldsymbol{x}_n - \boldsymbol{x}_{n-1}.$$

   - What is the relationship between $\boldsymbol{M}$ and the matrix $\boldsymbol{Q}$ involved in this differential equation ?
   - What is the relationship between the eigenvalues and eigenvectors of $\boldsymbol{M}$ and those of $\boldsymbol{Q}$?

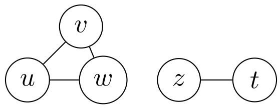
2 Graphs and eigenvalues

In this exercise, you will study some properties of graphs and their relationship with eigenvalues.

A graph G consists in a set of vertices V (or nodes) and a set E of edges. An edge connects two vertices, e.g. if $u, v \in V$, and $(u, v) \in E$ the vertices u and v are connected.



In the above picture, $V = \{u, v, w, z\}$ and $E = \{(u, v), (u, w), (v, w), (w, z)\}$. Note that in this case, the edges are *undirected*, i.e. we do not make a distinction between (u, v) and (v, u). The graph depicted above is said to be *connected*, as there exists a set of edges that connects every pair of vertices in the graph.



As a comparison, the above graph contains two *connected components*, the one formed by vertices u, v and w, and the one formed by the vertex z.

It is possible to represent the connectivity of a graph using its *adjacency matrix*, denoted by A. If $V = \{v_1, v_2, ..., v_n\}$, the adjacency matrix A is a $n \times n$ matrix, and each element a_{ij} is 1 if an edge exists between v_i and v_j and 0 otherwise. For example the adjacency matrix of the graph in the first picture is

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The adjacency matrix of the second graph would be

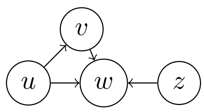
$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

There is no unique way of defining the adjacency matrix. The one in the example assumed the vertices were ordered in the following way u, v, w, z, t, but it is possible to order the vertices differently.

The *degree* of a vertex is the number of other vertices connected via an edge, e.g. in the first graph, the degree of u and v is 2, the degree of w is 3 and the degree of z is 1. The *degree matrix* of a graph is the diagonal matrix whose diagonal elements are the degree of each vertex of the graph. E.g. for the first graph

$$D = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Pick an arbitrary orientation for the edges of the graph, i.e., decide arbitrarily on a positive and negative end for each edge. Then, the *vertex-edge incidence matrix* B of a graph G is the $n \times m$ matrix whose (i,j)-th entry is 1 if the node v_i is the positive end of the edge e_j, -1 if v_i is at the negative end of the edge e_j, for $i = 1, \dots, n$ and $j = 1, \dots, m$, and 0 if the node does not appear in the edge. Let us pick an orientation for our first graph:



If using the "pointy" end of the arrow as positive end of the edge, and ordering the edges as $\{(u, v), (u, w), (v, w), (w, z)\}$, then we have

$$B = \begin{pmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

1. The graph Laplacian L is the operator defined by the matrix $L = D - A$. Show that $L = BB^\top$.
2. Show that 0 is a singular value of B and an eigenvalue of L. What is the eigenvector associated with the 0 eigenvalue of L ?
3. In the case of a fully connected graph (i.e. each vertex is connected to all other vertices of the graph and has a degree of $n-1$), what are the eigenvalues of L (Hint: consider the vectors orthogonal to the eigenvector associated with the 0 eigenvalue) ?
4. Let $x \in \mathbb{R}^n$. Prove that $x^T L x = \sum_{(i,j) \in E} (x_i - x_j)^2$
5. Let S be any subset of nodes in the graph G. The function $\text{cut}(S)$, counts the number of edges in the graph G with one endpoint in S and one endpoint outside of it. Express $\text{cut}(S)$ in closed form as a function of the matrix B and the set S. (hint: a set of nodes S can be represented by an indicator vector 1_S, where the i-th entry is 1 if $v_i \in S$ and zero otherwise).
6. If $\text{cut}(S) = 0$ for S non-empty and $S \neq V$, what is the implication for the connectivity of the graph ?
7. Prove that L is positive semi-definite. (hint: a matrix is positive semi-definite if $x^\top L x \geq 0$ for all $x \in \mathbb{R}^n$)
8. If a graph contains k connected components, what is the minimal multiplicity of the 0 eigenvalue ? (Remember that the determinant of a block diagonal matrix is the product of the determinants of each block.)
9. If a graph is connected, can the 0 eigenvalue have a multiplicity greater than 1 ?