Supplementary Discussion

Benjamin Doran et al.

2023-06-14

Table of contents

1 Relating the statistical structure of co-variation with taxonomic relatedness in a simple example of serial diversifications

2 Incorporating an example of convergent evolution into the statistical structure of co-variation

3 Defining a Spectral Tree of newly sequenced strains considering the co-evolutionary pattern of previously sequenced bacteria

4 How similarity and dissimilarity are encoded into spectral components

We sought to understand the mathematics of why spectral factorization reveals hierarchical scales of relatedness. In answer, we found it is because the similarity and differences between sub-populations are split across separate spectral components. Specifically, we show that (i) an ensemble of systems with two sub-populations will have exactly 2 irreducible spectral components up to the exact point at which those populations become identical; (ii) the change in magnitude for these spectral components — their eigenvalues — are equal and opposite to each other as we increase the degree of relatedness between the sub-populations; and (iii) the major eigenvector encodes the similarity of the sub-populations and the lesser eigenvector encodes the sub-population's dissimilarity for all points between the extrema where the sub-populations are identical or completely independent.

As it is not guaranteed that the reader has a background in the required mathematics, we will split this section into two parts. Section §4.1 will delve into the necessary detail regarding eigenvalue decomposition, the determinant, and the characteristic polynomial for readers to understand the connections between these concepts. Section §4.2 will detail a specific case of an ensemble of 3 systems, and show how changing the degree of relatedness between these systems changes specific aspects of the eigenspectrum.

4.1 Linear Algebra Background

4.1.1 What are eigenvectors?

For readers unfamiliar with linear algebra, some of the early applications for eigen decomposition in the 1700s were developed to describe the linear transformations of physical systems (rotations, shifting, scaling, and shearing of rigid bodies) [1]. Of particular interest in these descriptions are the principle axes or "eigenvectors" of the transformation which are the only vectors that do not change direction during the transform. The eigenspectrum describes the complete set of axes "eigenvectors", and is defined as the non-zero solutions to this equation

$$C\vec{v} = \lambda\vec{v} \tag{1}$$

In Equation 1, C is the linear transformation, represented as a matrix of real numbers; $\vec{v}$ is the eigenvector, represented as a list of real numbers; and λ is the eigenvalue, a real number that shortens or lengthens the eigen vector.

The eigenvectors and eigenvalues are useful precisely because they are the only stable descriptors of the transformation and can be used to consistently describe positions both before and after the transform.

To mathematically solve for these eigenvectors, one common technique is to first find the eigenvalues which are the solutions to Equation 2, and then substitute these eigenvalues into Equation 1 and solve for each of the eigenvectors ($\vec{v}$). All applications of spectral factorization (i.e. SVD and PCA) use this fundamental equation to define their spectral components ('eigenvectors').

$$\det(C - \lambda I) = 0 \tag{2}$$

In Equation 2, C is again the linear transformation; I is the identity matrix of the same size as C, defined as having 1s along the diagonal and 0 for all other entries; $det()$ is the determinant

[1] Hawkins, T. Cauchy and the spectral theory of matrices. Historia Mathematica 2, 1–29 (1975).