ABSTRACT

RIMA IZEM: Analyzing of Nonlinear Variation in Functional Data.
(Under the direction of J. Stephen Marron and Joel G. Kingsolver.)

Scientists in an increasing number of fields, including biology, medicine, and physics, collect samples of curves or images of common shape. Although these data are discrete, the processes generating them are continuous. Analyzing variation in these samples, with the aim of making inferences about the general population from which the sample is drawn, is often the main statistical interest. Usual statistical methods, such as Principal Components Analysis, are effective in analyzing linear variations, but do not always produce interpretable results. Moreover, some directions of variation in the data, of keen interest to biologists, are non-linear, and these methods fail to properly characterize them.

In this dissertation, we formulate a general model for curves or images of common shape. We use the Functional Data Analysis framework, which considers the curves as the statistical entities of interest, to exploit, in the statistical analysis, the continuity of the underlying processes generating the data. We present a new Functional Data Analysis method for analyzing variation under this model. Our method achieves two important goals. The first goal is to decompose the variation in the data into predetermined and interpretable directions of interest, and these could be linear or non-linear. The second goal is to quantify each direction by a newly defined ratio of sums of squares, to allow for a comparison of the contributions to the total variation. The new ratio of sums of squares quantifies a non-linear direction by taking into account the curvature of the space of variation. We discuss, in the general case,

consistency of our estimates of variation, using mathematical tools from differential geometry and shape statistics.

We successfully applied our method to two different examples of biological data. Our analysis shows that non-linear components are dominant. Moreover, our decomposition allows biologists to compare the prevalence of different genetic tradeoffs in a population and to quantify the effect of selection on evolution.