ENGSCI 331 – Computational Techniques 2

Module 4: Eigenproblems Assignment

Andreas W. Kempa-Liehr

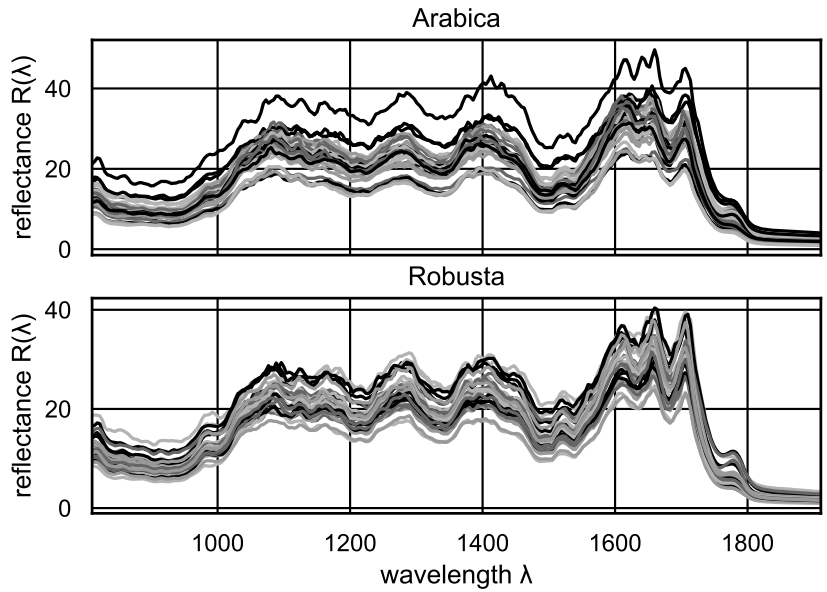
Department of Engineering Science

Due date **Wednesday 25th Sept. 10:00pm**
EP-CT2-assignment.pdf, Rev. (None)

Welcome to CoffeeAI

You have decided to join the Auckland based technology startup CoffeeAI, which develops an *intelligent* coffee machine named *eBarista*. This coffee machine is supposed to automatically tune its grinding and brewing process according to the coffee beans provided. The heart of each *eBarista* is a Fourier-Transform Infrared Spectrometer, which measures the infrared spectrum of absorption of the coffee beans.

Your task is to support the feature engineering of your data science team. For this purpose you are going to compute the principle components (eigenvectors) of the cross-correlation matrix of the reflectance spectra,



Data source: Romain Briandet, E. Katherine Kemsley, and Reginald H. Wilson. "Discrimination of Arabica and Robusta in Instant Coffee by Fourier Transform Infrared Spectroscopy and Chemometrics". In: *Journal of Agriculture and Food Chemistry* 44.1 (1996), pp. 170–174. DOI: 10.1021/jf950305a

project the spectra into their eigenvector space, and identify a pair of eigenvectors for which the infrared spectra of Arabica and Robusta coffee beans are separated as much as possible.

Fourier Transform Infrared Spectrometer

FTIR spectrometers (Fourier Transform Infrared Spectrometer) are widely used in organic synthesis, polymer science, petrochemical engineering, pharmaceutical industry and food analysis. In addition, since FTIR spectrometers can be hyphenated to chromatography, the mechanism of chemical reactions and the detection of unstable substances can be investigated with such instruments.
Source: `https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Spectroscopy/Vibrational_Spectroscopy/Infrared_Spectroscopy/How_an_FTIR_Spectrometer_Operates`

Context

- Your code will run on an embedded device, therefore you are required to write your code in C++.
- Luckily, you are part of a team and you'll be able to continue working on an existinc C++ code bases, which provides you with
  - basic IO routines including the loading of the dataset `DS19hH2_dk0_FTIR_Spectra_instant_coffee.csv`,
  - definition of an `Eigenpair` object featuring a normalization method,
  - functions for computing the dot products of two matrices, two vecors, as well as the dot product of a matrix and a vector.

Assignment - Programming tasks

- Read these instructions carefully from top to bottom.
- Familiarize yourself with the `*.cpp` and `*.h` files provided in the ZIP archive uploaded to Canvas.
- Setup our development environment and compile the code without changing anything.
- Use the provided templates to complete the implementation of the following functions in `myEigenfunctions.cpp`:
  - `power_method`
  - `deflate`
  - `CenterMatrix`
  - `CovarianceMatrix`
- Test the implementation of your functions in `myEigenMain.cpp`.

Principle Component Analysis

For the following analysis, you should compute the eigenpairs using your C++ code. It is your decision, if you compute the required projections within your C++ program or in a programming language of your choice (e.g. Matlab or Python). Create the visualizations with the tools of your choice.

- Compute the first 56 principle components of the coffee data dataset. They are given as eigenvectors of the covariance matrix.
- Export the principle components as well as the corresponding eigenvalues as two separate CV files.
- Plot the explained variance ratio of this eigenvectors. This ratio is given as cumulative sum of all eigenvalues, divided by the sum over all eigenvalues.
- How many principle components should be used for representing the spectra dataset?

Dimensionality reduction

- Project the spectra on the most important principle components.
- Identify a pair of two principle components, which provide a good separation of the different coffee types if the spectra are projected into this subspace. Hint: Use scatter plots.
- Use the six most important principle components for reconstructing one Arabica and one Robusta spectrum. Create plots, which compare the original spectra with the reconstructed spectra.

Report - Due date Wednesday 25th Sept. 10:00pm

- Upload a ZIP archive to Canvas, which includes
  - your complete sourcecode,
  - your results as CSV files, as well as
  - a short report, which discusses the testing of your implementation, as well as all points from the principle component analysis and the dimensionality reduction.
- The C++ code needs to be compilable.