M156 – Lab assignment #3: Application of PCA and Least Squares to real data: hyperspectral images

1 Introduction

Hyperspectral imaging is a an imaging technique in which a sensor (usually mounted on an aircraft or a satellite for earth observation) acquires the light intensity for many contiguous and narrow wavelengths of the electromagnetic spectrum (usually in the visible and near infrared domains). Therefore, the information contained in such images is much more important than in gray-level or color images (which only consider red, green and blue), since each pixel is a full spectrum, as shown in Fig. 1. Alternatively, the data can be seen as a collection of gray level images, each accounting for light intensity in a given wavelength.

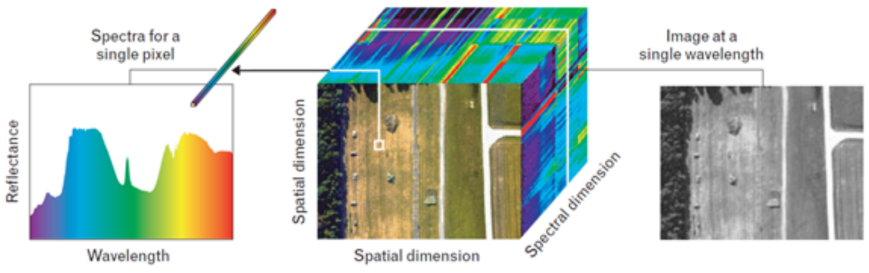

Figure 1: Hyperspectral imaging concept.

The resolution in wavelength is so fine that each material within the image is associated with a distinct spectrum, allowing automatic identification. For instance, with such images, it is easy to make the distinction between true and artificial grass, which is very difficult with color images. This is because the physical compositions of both objects are very different, and this can be seen in the way they reflect light in the infrared domain. The applications of these images are numerous: earth observation, environmental monitoring, agriculture, planetary exploration, food and pharmaceutical industries,...

However, the spatial resolution of these images is lower than regular color or gray-level images, which makes the data interpretation harder: it is frequent that the surface on the ground corresponding to a given pixel actually incorporates several materials. The corresponding pixel will then be a mixture of the contributions of the different materials.

A classical problem is then to try to find automatically in the image the spectra of the pure materials in the scene (called *endmembers*), and then to infer their proportions (called *abundances*) in each pixel. This problem is called spectral unmixing. It is illustrated for a given pixel in Fig. 2. More information on this problem can be found for instance in [1].

A hyperspectral image can be represented in a matrix form: $\mathbf{X} \in \mathbb{R}^{L \times N}$, where N is the number of pixels in the image, and L is the number of wavelengths (also called spectral bands). The spectra of the pure materials are gathered in the columns of a matrix $\mathbf{S} \in \mathbb{R}^{L \times P}$, where P is the number of materials. The proportions are gathered in a matrix $\mathbf{A} \in \mathbb{R}^{P \times N}$, where a_{ij} is the proportion of material j in pixel i.

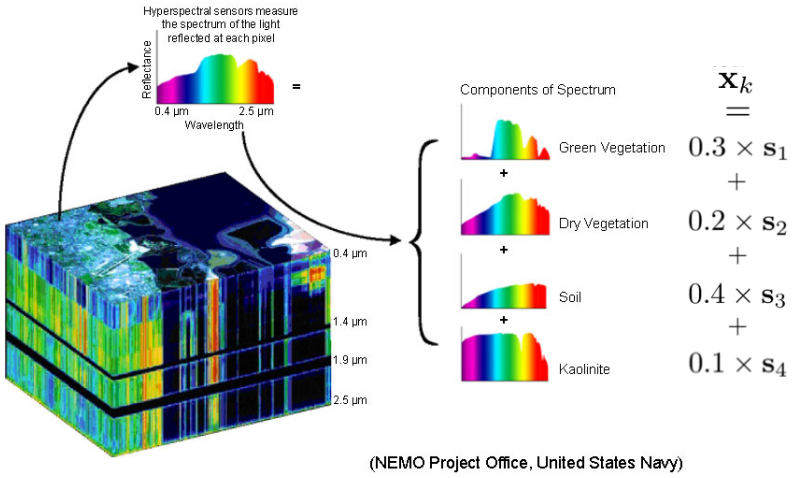

Figure 2: Spectral unmixing.

A common model for this type of data is to assume the observed pixels are a linear combination of the spectra of the pure materials, polluted by some noise:

$$\mathbf{x}_k = \sum_{i=1}^{P} a_{ik}\mathbf{s}_i + \mathbf{n}_k = \mathbf{S}\mathbf{a}_k + \mathbf{n}_k \tag{1}$$

where $\mathbf{x}_k$ represents pixel k (k^{th} column of $\mathbf{X}$), $\mathbf{s}_i$ is the i^{th} column of $\mathbf{S}$, $\mathbf{a}_k$ is the k^{th} column of $\mathbf{A}$ and $\mathbf{n}_k$ is a realization of Gaussian distributed noise.

This model can be reformulated in a matrix form:

$$\mathbf{X} = \mathbf{S}\mathbf{A} + \mathbf{N} \tag{2}$$

where $\mathbf{N}$ is a matrix containing N realizations of an L-dimensional Gaussian isotropic random variable $n \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$.

2 Theory

1. Write down the likelihood for $\mathbf{X}$, assuming that the the pixels are i.i.d., and that the matrix $\mathbf{S}$ is known.
2. Show that estimating $\mathbf{A}$ by maximum likelihood amounts to solving:

$$\arg\min_{\mathbf{A}} \frac{1}{2}||\mathbf{X} - \mathbf{S}\mathbf{A}||_F^2 \tag{3}$$

   where $||\cdot||_F$ is the Frobenius norm of a matrix: $||\mathbf{Y}||_F = \sqrt{\text{tr}(\mathbf{Y}^\top\mathbf{Y})} = \sqrt{\sum_{i=1}^{L}\sum_{j=1}^{N} y_{ij}^2}$ if $\mathbf{Y} \in \mathbb{R}^{L\times N}$. Give the solution to this least squares problem.
3. Since the coefficients of the matrix $\mathbf{A}$ are proportions, it makes sense that the sum of all the coefficients of the P materials in a given pixel equals 1:

$$\forall k = 1, ..., N, \ \sum_{i=1}^{P} \mathbf{a}_{ik} = 1 \tag{4}$$

   Another natural constraint is to force the coefficients to be positive (since a material cannot contribute negatively to an observed pixel). However, here, out of simplicity, we only consider the sum-to-one constraint (and we will see that the results do not suffer too much from not enforcing the nonnegativity constraint)

   Our problem is now equivalent to solving:

$$\arg\min_{\mathbf{A}} \frac{1}{2}||\mathbf{X} - \mathbf{S}\mathbf{A}||_F^2$$
$$\text{s.t. } \mathbf{A}^\top \mathbb{1}_P = \mathbb{1}_N \tag{5}$$

where $\mathbb{1}_.$ is a column vector of ones, whose size is given in index.

Write down the Lagrangian associated to this equality constrained problem. Hint: you have to consider as many Lagrange multipliers as there are constraint equations, and you can arrange them in a column vector, to obtain a matricial expression for the Lagrangian, using an inner product.

4. By differentiating the Lagrangian w.r.t. $\mathbf{A}$ and the vector of Lagrange Multipliers, show that looking for stationary points of the Lagrangian amounts to solving a linear system of the form:

$$\begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{bmatrix} \begin{bmatrix} \mathbf{A} \\ \boldsymbol{\lambda}^\top \end{bmatrix} = \begin{bmatrix} \mathbf{F} \\ \mathbf{G} \end{bmatrix} \tag{6}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^N$ is a column vector of Lagrange multipliers, and $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{E}$, $\mathbf{F}$ and $\mathbf{G}$ are matrices whose sizes and expressions are to be determined.

3 Experimental part

3.1 Data

The data we are interested in is a $101 \times 101 \times 195$ image acquired over the mountains near Palo Alto, CA. The wavelength range is from 400 to 2500nm, in 195 spectral bands. The spatial resolution is 17m. The image comprises mostly soil, vegetation and water.

Here, we provide the matrix $\mathbf{S}$, whose columns were extracted from the data. Each column represents the signature of one material found in the data. The spectral bands approximately corresponding to the wavelengths of red, green and blue are 30, 20 and 7, respectively. This means that you can obtain a color image by only considering those three bands: `imshow(im(x,y,[band1,band2,band3]))`.

You can plot any pixel of the image by the command `plot(squeeze(im(x,y,:)))`, where the command `squeeze` is used to shrink the $1 \times 1 \times L$ array into a vector, and `x` adn `y` are the coordinates of the considered pixel. If you have previously reshaped the image into a $L \times N$ matrix $\mathbf{X}$, you can plot a pixel by `plot(Y(:,n))`, where `n` is the index of the pixel.

3.2 Analysis

3.2.1 PCA

1. Perform a PCA on the data, seeing each pixel as an observation, and each spectral band as a variable. Plot the eigenvalues in descending order, and show a few of the corresponding PCA scores (the coordinates in the new basis). Since you have coordinates for each pixel, you can display them as images. Comment on the obtained images (salient features in the principal components). Compare with the color composition you can make using the RGB bands. Display the PCA scores for the last few components. How do you interpret them?

2. Using all the plots and images you have made, what can you say about the subspace actually occupied (approximately) by the data, w.r.t. the dimensionality of the data?

3. Display a 2D and a 3D scatterplots using the first principal components. Does this confirm your answer to the previous question?

4. Plot the first few eigenvectors. Are they physically interpretable? To guide your answer, compare with the columns of matrix $\mathbf{S}$ (in terms of the shape, but also on the properties of these two vector families).

3.2.2 Geometry of the Problem

5. The matrix $\mathbf{S}$ corresponds to the basis we are going to use in our regression. If the linear model of Eq. (3) holds, what is the actual subspace in which the data may theoretically lie (discarding the noise)?

6. Same question when the sum-to-one constraint is added. Hint: in that case, the decomposition of Eq. (1) becomes, by definition, an *affine combination* of the endmembers (i.e. of the columns of $\mathbf{S}$).

7. Same question when the sum-to-one *and* the positivity constraints are added. Hint: in this case the data points are *convex combinations* of the endmembers .

8. Conclude on the theoretical intrinsic dimensionality of the data (discarding potential model errors and noise).

9. Compute the coordinates of each column of $\mathbf{S}$ in the basis defined by PCA, and add the corresponding points to the previously drawn scatterplots. What can you say about their locations in the feature space? From this and the previous questions, is it an appropriate basis to use?

3.2.3 Estimation of the abundances

10. Code the solution of unconstrained least squares (Eq. (2)), and compute the abundances. Display the abundance coefficients you recover as spatial maps (by reshaping $\mathbf{A}$ as an image, each channel corresponding to a material). How do you interpret these maps visually (look in particular at the edges of the structures in the images, and pay attention to the scales of the abundance values).

11. As mentioned earlier, the computed values do not take into account the fact that the abundances are proportions. Code the solution to the sum-to-one constrained version of the problem, and display the new abundances. Compare to the previous question.

12. Perform the same computations as in the last question, but this time using only as inputs the projected data and endmembers on the first K principal components, where K was identified in question 8. Compare the results and running times of your algorithm in both cases (use the commands `tic` and `toc`).

13. Compute the Root Mean Squared Error in each pixel, both for the unconstrained and constrained case:

$$RMSE_k = \frac{1}{\sqrt{L}} ||\mathbf{x}_k - \hat{\mathbf{S}}\hat{\mathbf{a}}_k||_2 \tag{7}$$

    You can display them as images (be sure to display them using the same scales for a meaningful visualization) and compute their means over the whole image. Which approach gives the best values? Interpret. How do you explain the highest values of the RMSE?

Reference

[1] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, pp. 354–379, April 2012.