SC505 Linear Algebra and Optimization
Autumn 2020

Project Title: Hyperspectral Unmixing **Date:** 22-12-2020
Authors: Kishan Vaishnani-202011004 | Shivangi Gajjar-202011023 | Dhyanil Mehta-202011032

1 Introduction

1.1 What are Hyperspectral Images?

Different colors namely red, blue and green are seen due to the reflection of the light from the object. The reflected light falls under separate wavelengths in the visible spectrum of the electromagnetic radiation. These are the wavelengths that the human eye can see. However, there are a lots of wavelengths that the human eye cannot perceive. These features are often visible to other animals. Hence, we can consider an image as an Hyperspectral image while working with more than RGB bands in it. *Figure 1* shows the difference between a RGB image and a Hyperspectral image.

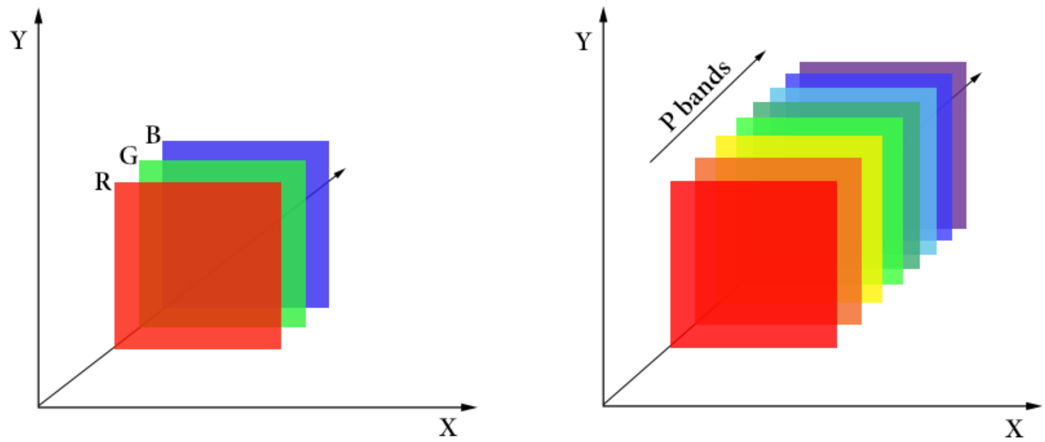

Figure 1: *RGB Image VS Hyperspectral Image*

1.2 What is Hyperspectral Unmixing?

After defining the hyperspectral images, our next task is to know where is such imagery used. Hyperspectral images are used in Hyperspectral Imaging. Hyperspectral imaging is a satellite imaging modality that senses an area of the earth across several frequency bands, instead of only R-G-B bands for usual images. This information is then used for identifying minerals, forest covers, urbanisation. For such an interpretation, we first need to extract information about the minerals/contents in the area whose image has been captured and their abundances, as there can be multiple components in an area. The problem is compounded due to low spatial resolution of the capturing device. For capturing such images, HSI camera is used.

The images captured by the HSI camera, will give us the image data in the form of a hypercube. In the hypercube, the pixels may not be a single pixel, it can be a combination of multiple spectra. Such pixels are known as **mixed-pixels** or **hyper-pixels**. We need to decompose these hyper-pixels into endmember signatures(pure pixel signatures) and corresponding abundances. The process of decomposition of hyper-pixels is known as **Hyperspectral Unmixing**.

2 Related work

Hyperspectral Unmixing has a numerous amount of application, out of which is the mineral abundance estimation, the one we discussed earlier also. Being an important preprocessing step in any type of application on Hyperspectral Imagery, there has been a significant amount of contribution in optimizing the unmixing problem. We observed that the most used technique for unmixing is Non-Negative Matrix Factroization, abbreviated as NMF.

In NMF, the hyperspectral data is factorized into an abundance matrix and a spectral matrix. NMf can be implemented using number of techniques: Alternating Least Squares(ALS), Hierarchical Alternating Least Squares(HALS), Vertex Component Analysis(VCA), Multiplicative Update(MU) and many more. Out of these, HALS and MU are most commonly used for hyperspectral unmixing. One reason can be that their time complexities are equivalent.

From above mentioned list, the most used method for implementing NMF, was provided by Lee and Seung, known as the Multiplicative Update NMF technique[4]. In this technique, they derived their own multiplicative update rules, which are applied to the Abundance matrix and the Spectral matrix iteratively. They stated that it is not possible to achieve a local minimum for the factorization problem because, here the factors are not both convex in nature. Hence, this method gives the global minimum for the approximate factorization[4]. The multplicative update rules provided are applied to the factors until they reach a convergence threshold. They also provided proofs of convergence for the method.

Before getting into the computations, it is necessary to identify the model that should be used. S. Jia and Y. Qian[2] mentions that the unsupervised linear unmixing model is best for our purpose. There are a few advantages of selecting such model. We know that the hyperspectral data comprises of multiple endmembers. So, one advantage is that it is difficult to get a priori knowledge of the endmembers present in the dataset[2]. Another advantage is that the linear spectral model is powerful tool for dealing with the mixed pixels[3]. S. Jia and Y. Qian, gave a modified version of NMF, by providing various constraints on the problem (CNMF), namely piecewise smoothness constraint and sparseness constraint[2]. They constrasted the results of CNMF with VCA, PSnsNMF and PSNMFSC. It was observed that the noise bands were removed efficiently through CNMF however, there were bit abrupt changes in the abundance values.

One more variation to the NMF method for hyperspectral unmixing is provided by N. Wang and L. Zhang. They proposed an endmember dissimilarity based method(ED), in which they defined a penalty function on the hyperspectral data[5]. According to their study, there is very less dissimilarity between the endmembers. Hence, they introduced the constraint $g(A)$, which as mentioned earlier, is called the penalty function. They contrasted ED NMF method with MVC and PSnsNMF. In conclusion, they mentioned that ED NMF method provides spectral variability robustness and purity robustness[5].

Out of these mentioned works, we have focused on the Multiplicative Update method of NMF proposed by Lee and Seung, for hyperspectral unmixing. Hence, the rest of the document discusses about MU NMF.

3 Hyperspectral Unmixing using NMF

As mentioned earlier, an HSI camera gives us the image in the form of a hypercube. We will here refer the hypercube as an Image Cube and the pure-pixel as endmember. Let us consider the hypercube X of dimensions length m, breadth n and depth of p bands as shown in the *Figure 2*. As shown in *Figure 2*, our task is to unmix the image cube X into an **Abundance Cube** W and **Spectral Matrix** H.

W represents the abundance(fraction) of each endmember, which comprises to form the whole image. Let us consider there total r endmembers known, then we can specify the dimensions of W to be $m \times n \times r$. Abundance matrix will provide us with the information of how much of fraction of endmembers are present in each pixel. H matrix contains the band-vectors of each endmember. These band-vectors are also known as the spectral signatures of each endmember. Hence, its dimensions are $r \times p$.

According to Lee and Seung [4], factorizing X, i.e. $X = WH$ is an NP Hard problem. It is convex in either H or W, but not both. Hence, we won't be getting perfect factors of X, however we can assume $X \approx WH$. We here will use the **Non-Negative Matrix Factorization (NMF)** technique to

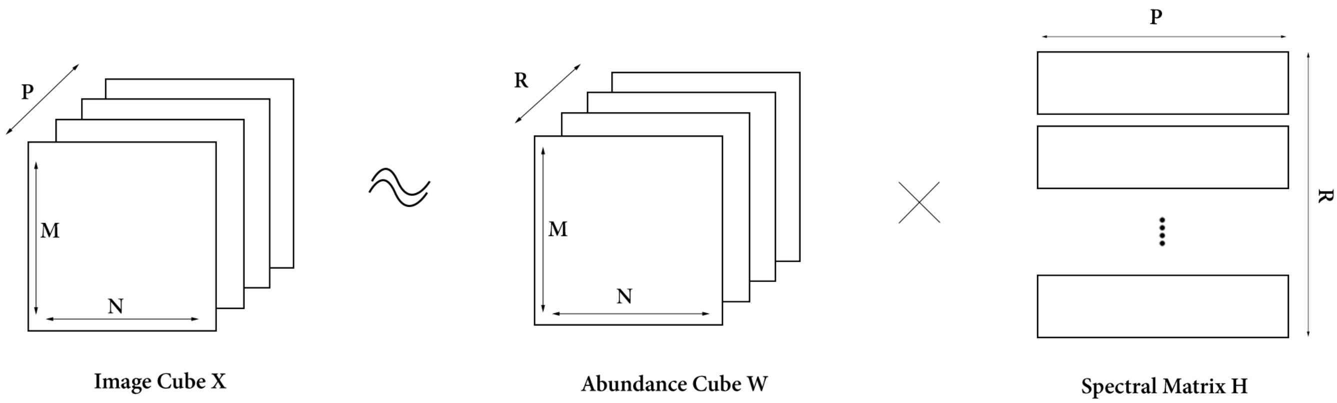

Figure 2: *Hypercube Decomposition*

solve our problem. Our target is to minimize the error in the mentioned factroization.

$$min\|X - WH\|_F^2$$

NMF includes multiple techniques to solve the above problem. Out of which, we here will use **Multiplicative Update (MU)** technique proposed by Lee and Seung.

In multiplicative update, a block coordinate descent concept is used, in which first W is updated while keeping H fixed and then H is updated keeping W fixed using the following update rules:

$$W \leftarrow W \cdot \frac{(XH^T)}{(WHH^T)}$$
$$H \leftarrow H \cdot \frac{(W^TX)}{(W^TWH)}$$

The above update rules are applied until the problem is converged. Lee and Seung has provided multiple proofs for the convergence for this problem in the paper. The following section discusses from where the above update rules are derived.

3.1 Derivation of Update Rules

For any matrix A, the Frobenius norm is defined as below, where Tr represents trace of matrix or the sum of diagonal elements,

$$\|A\|^2 = Tr(A^T A)$$

Now, let us calculate the norm of $(X - WH)$,

$$\begin{aligned}
&\|X - WH\|^2 \\
&= Tr((X - WH)^T (X - WH)) \\
&= Tr((X^T - H^T W^T)(X - WH)) \\
&= Tr(X^T X - X^T WH - H^T W^T X + H^T W^T WH) \\
&= Tr(X^T X) - Tr(X^T WH) - Tr(H^T W^T X) + Tr(H^T W^T WH) \qquad (i)
\end{aligned}$$

The traditional update rules for NMF are:

$$\begin{aligned}
&W \leftarrow W - \eta_W \cdot \nabla_W f(W, H) \qquad (ii) \\
&H \leftarrow H - \eta_H \cdot \nabla_H f(W, H) \qquad (iii)
\end{aligned}$$

where η is the learning rate, $\nabla_W f(W, H)$ is gradient of $f(W, H)$ with respect to W and $\nabla_H f(W, H)$ is gradient of $f(W, H)$ with respect to H.

Computing the derivatives of equation (i) with respect to W and H we get the gradients as,

(RULE: $\nabla_B Tr(AB) = A^T$ and $\nabla_B Tr(BA^T) = A$)

$$\begin{aligned}
&\nabla_W f(W, H) \\
&= \nabla_W Tr(X^T X) - \nabla_W Tr(X^T WH) - \nabla_W Tr(H^T W^T X) + \nabla_W Tr(H^T W^T WH) \\
&= \nabla_W Tr(X^T X) - \nabla_W Tr(HX^T W) - \nabla_W Tr(XH^T W^T) + \nabla_W Tr(WHH^T W^T) \\
&= 0 - XH^T - XH^T + 2WHH^T \\
&= -2XH^T + 2WHH^T
\end{aligned}$$

And,

$$\begin{aligned}
&\nabla_H f(W, H) \\
&= \nabla_H Tr(X^T X) - \nabla_H Tr(X^T WH) - \nabla_H Tr(H^T W^T X) + \nabla_H Tr(H^T W^T WH) \\
&= \nabla_H Tr(X^T X) - \nabla_H Tr(HX^T W) - \nabla_H Tr(H^T W^T X) + \nabla_H Tr(H^T W^T WH) \\
&= 0 - W^T X - W^T X + 2W^T WH \\
&= -2W^T X + 2W^T WH
\end{aligned}$$

Now, after putting the gradients in (ii) and (iii), ignoring the constant, we get

$$\begin{aligned}
&W \leftarrow W + \eta_W \cdot (XH^T - WHH^T) \\
&H \leftarrow H + \eta_H \cdot (W^T X - W^T WH)
\end{aligned}$$

To remove the additions, Lee and Seung set the learning rates as $\eta_W = \frac{W}{(WHH^T)}$ and $\eta_H = \frac{H}{(WW^T H)}$.

Hence, putting the values of learning rates, we get the final multiplicative update rules:

$$\begin{aligned}
&\boldsymbol{W \leftarrow W \cdot \frac{(XH^T)}{(WHH^T)}} \\
&\boldsymbol{H \leftarrow H \cdot \frac{(W^T X)}{(W^T WH)}}
\end{aligned}$$

After the problem has converged, we will get a close approximation of the Abundance matrix W and Spectral matrix H, which can further be used in numerous applications like mineral abundance estimation. This proves that Hyperspectral Unmixing is a crucial pre-processing step while working with HSI data.

4 Implementation & Results

The practical implementation of the NMF procedure for hyperspectral unmixing is based on the previously derived multiplicative update(MU) rules. We implemented the **Supervised MU NMF** in Python3. By supervised, we have fixed the number of endmembers for sake of simplicity. The algorithmic form of the code is as follows. The resultant matrices W and H will converge when

Algorithm 1 Hyperspectral Unmixing - MUNMF

procedure HSU(X)
  Initialize W and H
  Define $ConvergenceTolerance$
  $StopIter \leftarrow maxIterations$
  $counter \leftarrow 0$
  while !$Converged$ **do**
    $W_{old} \leftarrow W$
    $W \leftarrow W \cdot \frac{(XH^T)}{(WHH^T)}$
    $H \leftarrow H \cdot \frac{(W_{old}^T X)}{(W_{old}^T W_{old} H)}$
    $counter \leftarrow counter + 1$
    Check $Convergence$
    if $counter == StopIter$ **then**
      break;
    end if
  end while
  return W,H
end procedure

the error of the current iteration and previous iteration reaches the set convergence tolerance. We here set the convergence tolerance to be 10^{-4}. If, in any case, the matrices don't converge, then the execution will stop after initially defined $StopIter$ number of iterations.

4.1 Description of Datasets

As an input, we took three standard datasets of hyperspectral data: **1. Indian Pines, 2. Salinas and 3. Pavia** [1]. *Figure 3* shows the sample band images of the three datasets. As mentioned earlier, the implementation follows a supervised model, hence the number of endmembers are fixed. The endmember information is taken from the ground-truth data provided along with the datasets. *Table 1* shows the total endmembers for each dataset.

Dataset	No. of Endmembers
Indian Pines	16
Salinas	6
Pavia	9

Table 1: *Endmembers[1]*

These datasets are 220 band hyperspectral images. Hence, now we can clearly identify the dimen-

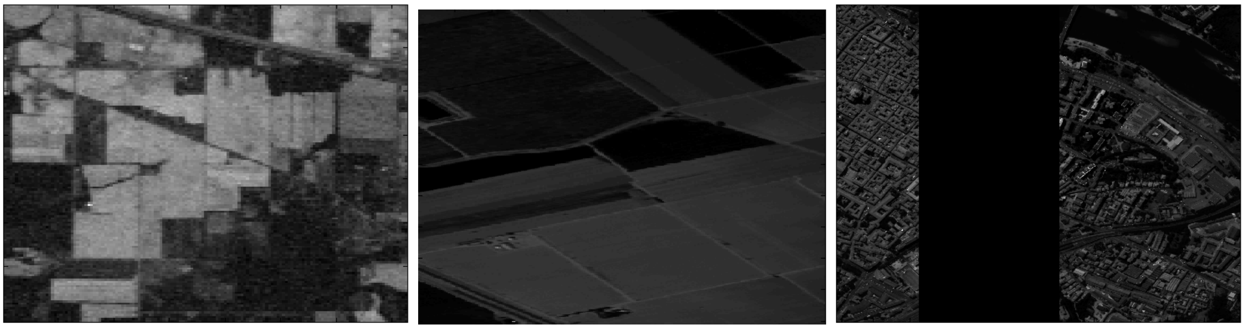

Figure 3: *Indian Pines, Salinas, Pavia Sample bands[1]*

sions of X, W and H as $145 \times 145 \times 220$, $145 \times 145 \times 16$ (in case of Indian Pines) and 16×220(in case of Indian Pines) respectively.

4.2 Results

The procedure mentioned in the algorithm provides us with an Abundance Matrix W and a spectral matrix H. We plotted the Abundance graph based on the obtained values. *Figure 4* shows the Abundance graph for each dataset, where the legend of graph shows the endmembers. The unmixing can be seen here. *Figure 5* shows the error in each iteration and the minimization of error is quite visible here.

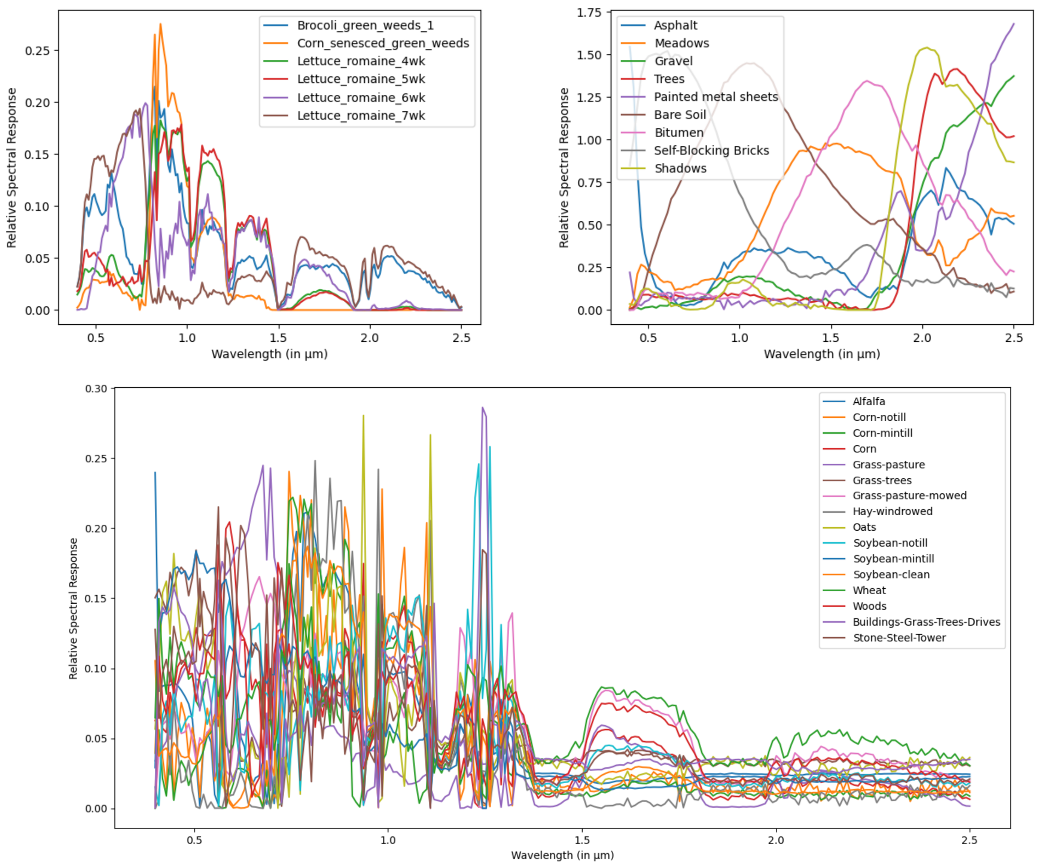

Figure 4: *Salinas, Pavia and Indian Pines Unmixed*

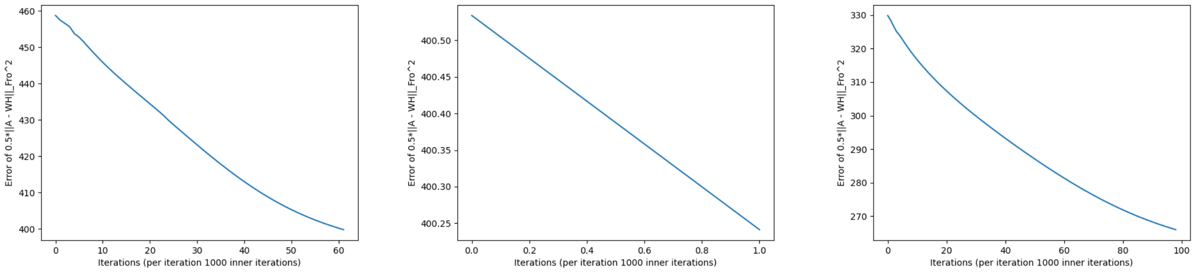

Figure 5: *Salinas, Pavia and Indian Pines Error graph*

5 Summary & Conclusion

To summarize, Hyperspectral Unmixing is a process to unmix the images containing more than standard RGB bands. By unmixing, we target to obtain two important information about the hyperspectral data: 1. Abundances of the pure pixels for each hyper-pixel and 2. Composition of spectral bands for each pure pixel(endmember). This information can be available by factorizing the given Hyperspectral data into Abundance matrix and Spectral Matrix. Numerous amount of methods are available for NMF, out of which we saw the Multiplicative Update(MU) NMF method proposed by Lee and Seung.

Apart from MU, the other method that can be explored for Unmixing is Hierarchical Alternate Least Squares(HALS) NMF method. It is similar to MU and has same time complexity as that of MU. In addition to that, we here implemented the Supervised MU NMF, where the number of endmembers were fixed in the beginning itself. To emphasize the MU NMF process, instead of focusing on estimating endmembers, we used the Supervised method. However, an Unsupervised approach can also be attempted for the same. It is a more complex task to estimate the number of endmembers in every iteration, which needs significant machine learning techniques. However, Unsupervised MU NMF is an efficient approach for Hyperspectral Unmixing.

In conclusion, HSI data is different as compared to RGB data, thus, Hyperspectral Unmixing can be considered as a significant pre-processing step to process the HSI data in any remote sensing and mineral abundance estimation applications.

References

[1] Marion F. Baumgardner, Larry L. Biehl, and David A. Landgrebe. 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3, Sep 2015.

[2] S. Jia and Y. Qian. Constrained nonnegative matrix factorization for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1):161–173, 2009.

[3] N. Keshava and J. F. Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1):44–57, 2002.

[4] Daniel Lee and Hyunjune Seung. Algorithms for non-negative matrix factorization. *Adv. Neural Inform. Process. Syst.*, 13, 02 2001.

[5] N. Wang, L. Zhang, and B. Du. An endmember dissimilarity based non-negative matrix factorization method for hyperspectral unmixing. In *2012 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–5, 2012.