Semi-NMF Regularized Autoencoders for HSU

Divyam Goel | ECE, IIT Roorkee

March 2022

Linear Mixture Model

The Linear Mixture Model assumes that the observed reflectance spectrum of a given pixel is a linear combination of unique constituent spectral signatures known as end-members. The unmixing problem then presents itself as a simple matrix factorization problem, with an additional matrix element to estimate the noise and other errors of the low-resolution hyperspectral sensors.
The mathematical formulation for the LMM is presented below:

$$\begin{gathered} \mathbf{X} = \mathbf{W}\mathbf{H} + \mathbf{E}, \\ s.t.\, \mathbf{H} \succeq 0,\, \mathbf{1}_R^T\mathbf{H} = \mathbf{1}_P^T \end{gathered} \tag{1}$$

where $\mathbf{E} \in \mathbb{R}^{B\times P}$ is the error matrix that encompasses all the errors that may have affected the data collection process. $\mathbf{H} \succeq 0$ represents the abundance non-negativity constraint, while $\mathbf{1}_R^T\mathbf{H} = \mathbf{1}_P^T$ presents the abundance sum-to-one constraint. With no constraints on $\mathbf{W}_d$ and a non-negativity constraint on $\mathbf{H}$, we find that the LMM unmixing problem is part of a special class of problems in multi-variate analysis known as the Semi-NMF. Hence, the optimization objective is:

$$(\mathbf{W}, \mathbf{H}) = \arg\min_{\mathbf{W}, \mathbf{H}} \frac{1}{2}||\mathbf{X} - \mathbf{W}\mathbf{H}||_F^2 + \mu f_1(\mathbf{W}_d) + \lambda f_2(\mathbf{H}) \tag{2}$$

where $||\cdot||_F^2$ is the Frobenius Norm, $f_1(\mathbf{W}_d)$ and $f_2(\mathbf{H})$ are regularization functions over $\mathbf{W}_d$ and $\mathbf{H}$ with positive biases μ and λ respectively. While many optimization techniques have been suggested in classic literature, the most recent advancements in the field of hyperspectral unmixing have come in the form of a deep neural network architecture known as the Autoencoder (AE).

0.1 Theorem 1

We first prove that there exists a closed form solution for the non-convex Semi-NMF optimization problem given in Equation 2. By using an iterative procedure for the optimization and assuming the knowledge of $\mathbf{W}_d$ from the previous step, Equation 2 can be written as a convex optimization problem as given below:

$$(\mathbf{H}^k) = \arg\min_{\mathbf{H}^k} \frac{1}{2}||\mathbf{X} - \mathbf{W}_d^{k-1}\mathbf{H}^k||_F^2 + \mu f_1(\mathbf{W}_d^{k-1}) + \lambda f_2(\mathbf{H}^k) \tag{3}$$

For the rest of this document, we use $\mathbf{W}_d$ in place of $\mathbf{W}_d^{k-1}$ and $\mathbf{H}$ in place of $\mathbf{H}^k$ (unless stated otherwise) for brevity. This optimization problem can also be written in the following variable separable form:

$$\begin{aligned} &min\ f_o\ :\ ||X_i - \mathbf{W}_d H_i - E_i||_F^2, \\ &s.t.\ f_i : H_i\ \succeq\ 0, g_i : \mathbf{1}_R^T H_i\ =\ 1 \end{aligned} \tag{4}$$

The solution for the system of equations in 4 can be found by invoking convex analysis and writing the Karush-Kuhn-Tucker (KKT) conditions. These are presented below:

$$\begin{aligned} \textbf{Stationarity} : \nabla f_o + \nabla f_i^T \mu_i + \nabla g_i^T \lambda_i\ =\ 0 \\ \textbf{Dual Feasibility} : \mu_i, \lambda_i\ \succeq\ 0 \\ \textbf{Complementary Slackness} : \mu_i^T H_i = 0 \end{aligned} \tag{5}$$

Substituting Equation 4 in the stationarity condition we get:

$$-2\mathbf{W}_d^T X_i + 2\mathbf{W}_d^T \mathbf{W}_d H_i + 2\mathbf{W}_d^T E_i + \mu_i + \mathbf{1}_R \lambda_i = 0 \tag{6}$$

A look at the complementary slackness condition reveals the existence of two cases. Case 1 (active constraint) is when the Lagrangian $\mu_i \succ 0$. This gives a trivial solution (H_i = 0). The other case is more complex, it's analysis is presented below:

$$\begin{aligned} &\textbf{Case 2: inactive constraint} \rightarrow \mu_i = 0, \\ &\implies 2\mathbf{W}_d^T \mathbf{W}_d H_i = 2\mathbf{W}_d^T X_i - 2\mathbf{W}_d^T E_i - \mathbf{1}_R \lambda_i, \\ &\implies H_i = \mathbf{W}_d^{\dagger} X_i - \mathbf{W}_d^{\dagger} E_i - \frac{1}{2}(\mathbf{W}_d^T \mathbf{W}_d)^{-1} \mathbf{1}_R \lambda_i \end{aligned} \tag{7}$$

$\mathbf{W}_d^{\dagger}$ is the Moore-Penrose pseudo inverse of the matrix $\mathbf{W}_d$. It is important to note that while μ_i was a vector quantity, λ_i is scalar in nature and may be unique for each unique H_i. Equation **??** can be simplified further by making use of the ASC constraint over the abundance matrix.

$$\begin{aligned} &\mathbf{1}_R^T H_i = 1, \\ &\implies \mathbf{1}_R^T \mathbf{W}_d^{\dagger} X_i - \mathbf{1}_R^T \mathbf{W}_d^{\dagger} E_i - \frac{1}{2}\mathbf{1}_R^T (\mathbf{W}_d^T \mathbf{W}_d)^{-1} \mathbf{1}_R \lambda_i\ =\ 1, \\ &\implies \frac{1}{2}\mathbf{1}_R^T (\mathbf{W}_d^T \mathbf{W}_d)^{-1} \mathbf{1}_R \lambda_i\ =\ \mathbf{1}_R^T \mathbf{W}_d^{\dagger} X_i - \mathbf{1}_R^T \mathbf{W}_d^{\dagger} E_i - 1 \end{aligned} \tag{8}$$

Rearranging Equation **??** yields a closed form solution over the Lagrangian scalar λ_i which we denote by the symbol α.

$$\lambda_i\ =\ \frac{\mathbf{1}_R^T \mathbf{W}_d^{\dagger} X_i - \mathbf{1}_R^T E_i - 1}{\frac{1}{2}\mathbf{1}_R^T (\mathbf{W}_d^T \mathbf{W}_d)^{-1} \mathbf{1}_R}\ =\ \boldsymbol{\alpha} \tag{9}$$

Substituting this solution into Equation **??**, we get a closed form solution for the variable separated optimization problem for LMM unmixing.

$$H_i = \mathbf{W}_d^{\dagger} X_i - \mathbf{W}_d^{\dagger} E_i - \frac{1}{2}(\mathbf{W}_d^T \mathbf{W}_d)^{-1} \mathbf{1}_R \boldsymbol{\alpha} \tag{10}$$

0.2 Designing the Regularization Terms

Recently, deep neural network autoencoders have been used for the spectral unmixing of Linear Mixture Models. Autoencoders learn a representation of a given input data sequence, typically for dimensionality reduction. The training objective naturally lends itself to the problem of hyperspectral unmixing. The mathematical formulation of this objective is given by:

$$(\mathbf{D}^*, \mathbf{E}^*) = \underset{\mathbf{D}, \mathbf{E}}{\arg\min} \sum_{i=1}^{P} \mathcal{L}(X_i, \mathbf{D}(\mathbf{E}(X_i))). \tag{11}$$

where the objective has been represented in the familiar variable separable form. $\mathcal{L}$ represents the choice over a loss function, while **D** and **E** represent the decoder and encoder of the model respectively. For the problem of spectral unmixing, the endmember matrix **W** is replaced by the weights of the decoder $\mathbf{W}_d$. The abundance matrix **H** is the set of hidden representations learned by the model:

$$H_i = \mathbf{W}_e X_i + \mathbf{b} \tag{12}$$

where $\mathbf{W}_e$ and **b** are the weights and biases of the encoder. The regularization terms arise from the comparative analysis of Equations 10 and 12. By introducing proximal measures between the relevant components on the R.H.S of the two equations, and adding those terms to the overall training loss we can expect to see a smoother convergence to the most optimal solution for the unmixing problem. The final training objective along with the regularization terms is as follows:

$$\begin{aligned} \mathcal{L} : MSE(X_i, \mathbf{D}(\mathbf{E}(X_i))) + ||\mathbf{W}_e - \mathbf{W}^{\dagger}||_F^2 + ||\mathbf{b} - \mathbf{b}^{`}||_F^2, \\ where\, \mathbf{b}^{`} \ = \ -\mathbf{W}^{\dagger} N_i - \frac{1}{2}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{1}_R\boldsymbol{\alpha} \end{aligned} \tag{13}$$

Note that the choice for the loss function can be varied. The most common choices for this loss are Mean Squared Error (MSE), Spectral Angle Distance (SAD) and Spectral Information Distance (SID). Recent studies have shown that SAD performs better than its counterparts and hence is more widely adopted.