Non-Negative Matrix Factorization.

Rohit Kumar, M.Tech Signal Processing, 16100
AR Shaarad, PhD student, CSA, 16056

Defination

Non-Negative Matrix Factorization(NMF) is an Linear Dimensionality Reduction(LDR) technique where given a **non-negative matrix**, we find the **non-negative factors** of it i.e

$$X \approx WH \tag{1}$$

with focus on the following optimization problem

$$\min_{W \in \mathcal{R}^{p \times r}, H \in \mathcal{R}^{r \times n}} \|\mathbf{X} - \mathbf{WH}\|_F^2 \ni \mathbf{W} \geq 0, \mathbf{H} \geq 0 \tag{2}$$

Application of NMF in Speech

- Decomposing the audio which is a mixture of more than one musical instrument into its building blocks [1].
- NMF is also used for denoising of audio.

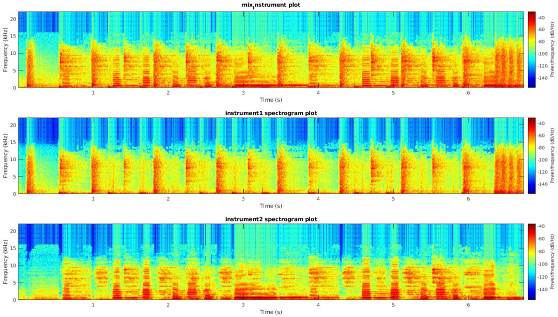

Figure: Decomposition of audio into its constituent instruments

Hyper spectral Unmixing

- The goal of blind hyperspectral unmixing is first to **identify** the ***endmembers*** i.e constitutive materials and which pixel contain which endmember and in which proportion.
- **NMF** is able to compute the **spectral signature** of the **endmembers** and simultaneously the abundance of each endmember in each pixel.

$$\underbrace{X(:,j)}_{\substack{\text{spectral signature}\\ \text{of } j\text{th pixel}}} \approx \sum_{k=1}^{r} \underbrace{W(:,k)}_{\substack{\text{spectral signature}\\ \text{of } k\text{th endmember}}} \underbrace{H(k,j)}_{\substack{\text{abundance of } k\text{th endmember}\\ \text{in } j\text{th pixel}}} .$$

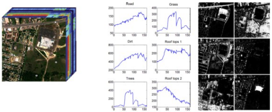

Figure: Decomposition of urban hyperspectral image

Two Block Coordinate Descent-Framework of most NMF

Update alternatively over one of the two factors, W or H while keeping the other fixed. The reason is that the **subproblem** in one factor is **convex** or precisely **nonnegative least square problem(NNLS)**.

Algorithm 1: Two Block Coordinate Descent

Input: Input nonnegative matrix X $\in \mathcal{R}^{p \times r}$ and factorization rank r
Output: (W,H)$\geq$ 0: A rank-r NMF of $X \approx WH$
Initialization: Generate some initial matrices $W^{(0)} \geq 0$ and $H^{(0)} \geq 0$
for *t=1,2,...* **do**

end
W$^{(t)} = update(X, H^{(t-1)}, W^{(t-1)})$
$H^{(t)} = update(X, W^{(t)^T}, W^{(t-1)^T})$

Alternating Least Squares(ALS)

Solving unconstrained least square problem $||\mathbf{X} - \mathbf{WH}||_F^2$ and then project the solution onto nonnegative orthant:

$$W \longleftarrow \max(\mathrm{argmin}_{Z \in \mathcal{R}^{pxr}}(||X - ZH||_F, 0) \tag{3}$$

Drawback
ALS does not converge and might oscillate under ALS updates.

Alternating Nonnegative Least Squares(ANLS)

Solve the subproblem exactly, i.e

$$W \longleftarrow (\mathrm{argmin}_{W \geq 0}(||X - WH||_F) \tag{4}$$

Drawback
Each iteration of CD is computationally expensive.

Multiplicative updates

For minimising $||V - WH||_F$, at each iteration, update H and W as [2]

$$H_{\alpha\mu} \leftarrow H_{\alpha\mu} \frac{(W^T V)_{\alpha\mu}}{(W^T WH)_{\alpha\mu}}$$

$$W_{ia} \leftarrow W_{ia} \frac{(VH^T)_{ia}}{(WHH^T)_{ia}}$$

Multiplicative updates

For minimising $||V - WH||_F$, at each iteration, update H and W as [2]

$$H_{\alpha\mu} \leftarrow H_{\alpha\mu} \frac{(W^T V)_{\alpha\mu}}{(W^T WH)_{\alpha\mu}}$$

$$W_{ia} \leftarrow W_{ia} \frac{(VH^T)_{ia}}{(WHH^T)_{ia}}$$

- Clearly, when $V = WH$, the updates are stationary.

Interpretation/Intuition

Consider a simple additive update for minimising $\frac{1}{2}||V - WH||_F^2$ w.r.t H, using gradient descent

$$H_{\alpha\mu} \leftarrow H_{\alpha\mu} + \eta_{\alpha\mu}[(W^T V)_{\alpha\mu} - (W^T WH)_{\alpha\mu}].$$

Interpretation/Intuition

Consider a simple additive update for minimising $\frac{1}{2}||V - WH||_F^2$ w.r.t H, using gradient descent

$$H_{\alpha\mu} \leftarrow H_{\alpha\mu} + \eta_{\alpha\mu}[(W^T V)_{\alpha\mu} - (W^T WH)_{\alpha\mu}].$$

This becomes equivalent to the multiplicative updates for

$$\eta_{\alpha\mu} = \frac{H_{\alpha\mu}}{(W^T WH)_{\alpha\mu}}.$$

Proof of Convergence

- Auxiliary function
  - $G(h, h')$ is an auxiliary function for $f(h)$ if $G(h, h') \geq F(h)$ and $G(h, h) = F(h)$.
- If G is an auxiliary function for F, then F is non increasing under the update $h^{t+1} = \arg\min_h G(h, h^t)$.

Proof of Convergence

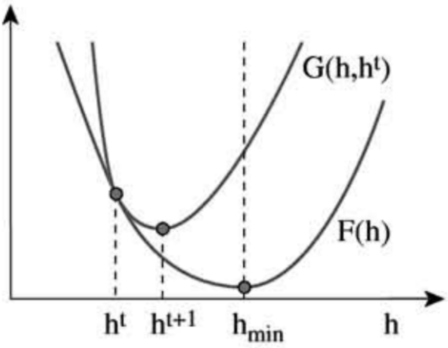

Figure: Auxiliary function

Proof of Convergence

Since, $||V - WH||_F^2 = \sum_j ||V_j - WH_j||_2^2$, for a fixed column v of V and corresponding column h of H, we consider the objective function

$$F(h) = \frac{1}{2}\sum_i (v_i - \sum_a W_{ia}h_a)^2.$$

It can be shown that the following is an auxiliary function

$$G(h, h^t) = F(h^t) + (h - h^t)^T \nabla F(h^t) + \frac{1}{2}(h - h^t)^T K(h^t)(h - h^t),$$

where $K_{ab}(h^t) = \delta_{ab}(W^T W h^t)_a / h_a^t$.

Hierarchical Alternating Least Squares(HALS)

Exact coordinate descent method, updating one column of W at a time.

$$W(:,l) \longleftarrow \operatorname{argmin}_{W(:,l)\geq 0} ||V - \sum_{k\neq l} W(:,k)H(k,:) - W(:,l)H(l,:)||_F \quad (5)$$

$$= \max\left(0, \frac{VH(l,:)^T - \sum_{k\neq l} W(:,k)(H(k,:)H(l,:))^T}{||H(l,:)||_2^2}\right) \quad (6)$$

Key Points

- Converges faster than Multiplicative Updates.
- Almost the same computational cost.

Comparison

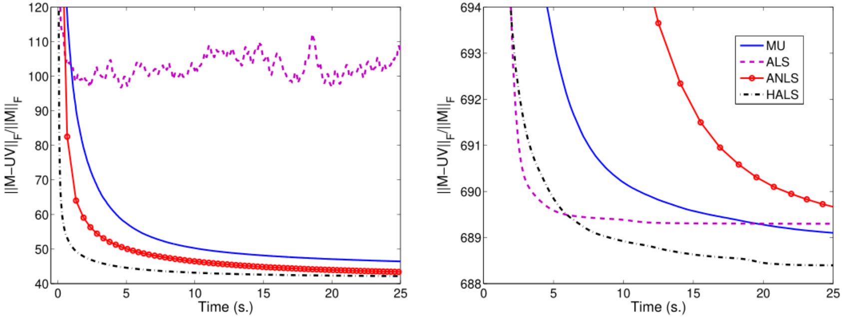

Figure: [3] Comparison

Bibliography I

1. Christian Dittmar, Patricio López-Serrano, and Meinard Müller.
   Unifying local and global methods for harmonic-percussive source separation.
   2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 176–180, 2018.

2. Daniel D. Lee and H. Sebastian Seung.
   Algorithms for non-negative matrix factorization.
   In *Proceedings of the 13th International Conference on Neural Information Processing Systems*, NIPS'00, pages 535–541, Cambridge, MA, USA, 2000. MIT Press.

3. Nicolas Gillis.
   The why and how of nonnegative matrix factorization.
   In *Regularization, Optimization, Kernels, and Support Vector Machines*. Chapman and Hall/CRC, 2014.