Physics-Aware Deep Nonnegative Matrix Factorization (PAD-NMF)

Erik Chinellato
Fabio Marcuzzi
February 22, 2024

Università degli Studi di Padova

Deep-NMF: Introduction

Let us consider a regularized NMF problem in the form

$$\begin{array}{ll} \text{minimize} & D_1(X, WH) + \mu\|H\|_1 \\ \text{subject to} & W \in \mathcal{M}_{M\times R}(\mathbb{R}^+) \quad , \quad H \in \mathcal{M}_{R\times N}(\mathbb{R}^+) \end{array}$$

which is tackled by alternate optimization of the factors

$$H^{(k+1)} = f(X; W^{(k)}, H^{(k)}) \qquad W^{(k+1)} = g(X; W^{(k)}, H^{(k+1)})$$

Deep-NMF: Introduction

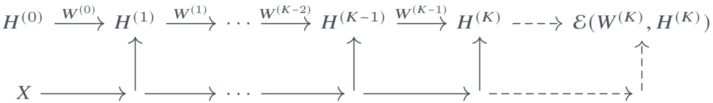
Let us consider a regularized NMF problem in the form

$$\begin{array}{ll} \text{minimize} & D_1(X, WH) + \mu\|H\|_1 \\ \text{subject to} & W \in \mathcal{M}_{M\times R}(\mathbb{R}^+) \quad , \quad H \in \mathcal{M}_{R\times N}(\mathbb{R}^+) \end{array}$$

which is tackled by alternate optimization of the factors

$$H^{(k+1)} = f(X; W^{(k)}, H^{(k)}) \qquad W^{(k+1)} = g(X; W^{(k)}, H^{(k+1)})$$

By **interpreting the iterative update scheme as a neural network**, where $H^{(k+1)}$ is the output of the k-th layer given the input $H^{(k)}$ and activation function f, Deep-NMF *unfolds* the iterations and *unties* the bases across layers: the result is a trainable neural network with parameters $\{W^{(k)}\}_{k=0,\ldots,K}$.



Deep-NMF: Introduction

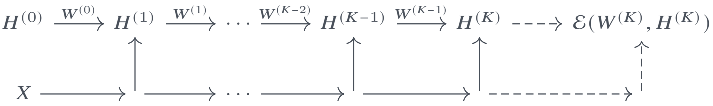
Let us consider a regularized NMF problem in the form

$$\begin{aligned} &\text{minimize} && D_1(X, WH) + \mu \|H\|_1 \\ &\text{subject to} && W \in \mathcal{M}_{M\times R}(\mathbb{R}^+) \quad , \quad H \in \mathcal{M}_{R\times N}(\mathbb{R}^+) \end{aligned}$$

which is tackled by alternate optimization of the factors

$$H^{(k+1)} = f(X; W^{(k)}, H^{(k)}) \qquad W^{(k+1)} = g(X; W^{(k)}, H^{(k+1)})$$

By **interpreting the iterative update scheme as a neural network**, where $H^{(k+1)}$ is the output of the k-th layer given the input $H^{(k)}$ and activation function f, Deep-NMF *unfolds* the iterations and *unties* the bases across layers: the result is a trainable neural network with parameters $\{W^{(k)}\}_{k=0,\dots,K}$.



Q: Why do we care about Deep-NMF? Why is it useful?

Deep-NMF: Introduction

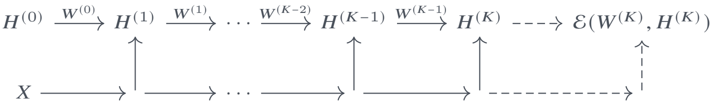
Let us consider a regularized NMF problem in the form

$$\begin{array}{ll} \text{minimize} & D_1(X, WH) + \mu \|H\|_1 \\ \text{subject to} & W \in \mathcal{M}_{M\times R}(\mathbb{R}^+) \quad , \quad H \in \mathcal{M}_{R\times N}(\mathbb{R}^+) \end{array}$$

which is tackled by alternate optimization of the factors

$$H^{(k+1)} = f(X; W^{(k)}, H^{(k)}) \qquad W^{(k+1)} = g(X; W^{(k)}, H^{(k+1)})$$

By **interpreting the iterative update scheme as a neural network**, where $H^{(k+1)}$ is the output of the k-th layer given the input $H^{(k)}$ and activation function f, Deep-NMF *unfolds* the iterations and *unties* the bases across layers: the result is a trainable neural network with parameters $\{W^{(k)}\}_{k=0,\dots,K}$.



Q: Why do we care about Deep-NMF? Why is it useful?

A: It provides a nonnegative, *additive* decomposition of X

$$X \approx W^{(K)} H^{(K)} = W_S^{(K)} H_S^{(K)} + W_N^{(K)} H_N^{(K)} = S + N$$

where

$$W^{(K)} = \begin{bmatrix} W_S^{(K)} & W_N^{(K)} \end{bmatrix} \qquad H^{(K)} = \begin{bmatrix} H_S^{(K)\top} & H_N^{(K)\top} \end{bmatrix}^\top$$

Since the weights $W^{(k)}$ **must remain nonnegative** to retain interpretability, the backpropagation algorithm is non-conventional.

Indeed, the weights are updated with:

$$W^{(k)} \Leftarrow W^{(k)} \circ \frac{\left[\nabla_{W^{(k)}} \mathcal{E}\right]_-}{\left[\nabla_{W^{(k)}} \mathcal{E}\right]_+}$$

Since the weights $W^{(k)}$ **must remain nonnegative** to retain interpretability, the backpropagation algorithm is non-conventional.

Indeed, the weights are updated with:

$$W^{(k)} \Leftarrow W^{(k)} \circ \frac{\left[\nabla_{W^{(k)}} \mathcal{E}\right]_-}{\left[\nabla_{W^{(k)}} \mathcal{E}\right]_+}$$

Hence the need to split the gradients into positive and negative parts and back-propagate both quantities:

$$\left[\frac{\partial \mathcal{E}}{\partial W^{(k)}_{\bar{n},\bar{r}}}\right]_+ = \sum_{m,r} \left(\left[\frac{\partial \mathcal{E}}{\partial H^{(k+1)}_{r,m}}\right]_+ \left[\frac{\partial H^{(k+1)}_{r,m}}{\partial W^{(k)}_{\bar{n},\bar{r}}}\right]_+ + \left[\frac{\partial \mathcal{E}}{\partial H^{(k+1)}_{r,m}}\right]_- \left[\frac{\partial H^{(k+1)}_{r,m}}{\partial W^{(k)}_{\bar{n},\bar{r}}}\right]_- \right)$$

$$\left[\frac{\partial \mathcal{E}}{\partial W^{(k)}_{\bar{n},\bar{r}}}\right]_- = \sum_{m,r} \left(\left[\frac{\partial \mathcal{E}}{\partial H^{(k+1)}_{r,m}}\right]_+ \left[\frac{\partial H^{(k+1)}_{r,m}}{\partial W^{(k)}_{\bar{n},\bar{r}}}\right]_- + \left[\frac{\partial \mathcal{E}}{\partial H^{(k+1)}_{r,m}}\right]_- \left[\frac{\partial H^{(k+1)}_{r,m}}{\partial W^{(k)}_{\bar{n},\bar{r}}}\right]_+ \right)$$

$$\left[\frac{\partial \mathcal{E}}{\partial H^{(k)}_{\bar{r},\bar{m}}}\right]_+ = \sum_{r} \left(\left[\frac{\partial \mathcal{E}}{\partial H^{(k+1)}_{r,\bar{m}}}\right]_+ \left[\frac{\partial H^{(k+1)}_{r,\bar{m}}}{\partial H^{(k)}_{\bar{r},\bar{m}}}\right]_+ + \left[\frac{\partial \mathcal{E}}{\partial H^{(k+1)}_{r,\bar{m}}}\right]_- \left[\frac{\partial H^{(k+1)}_{r,\bar{m}}}{\partial H^{(k)}_{\bar{r},\bar{m}}}\right]_- \right)$$

$$\left[\frac{\partial \mathcal{E}}{\partial H^{(k)}_{\bar{r},\bar{m}}}\right]_- = \sum_{r} \left(\left[\frac{\partial \mathcal{E}}{\partial H^{(k+1)}_{r,\bar{m}}}\right]_+ \left[\frac{\partial H^{(k+1)}_{r,\bar{m}}}{\partial H^{(k)}_{\bar{r},\bar{m}}}\right]_- + \left[\frac{\partial \mathcal{E}}{\partial H^{(k+1)}_{r,\bar{m}}}\right]_- \left[\frac{\partial H^{(k+1)}_{r,\bar{m}}}{\partial H^{(k)}_{\bar{r},\bar{m}}}\right]_+ \right)$$

Physics-Aware Deep-NMF

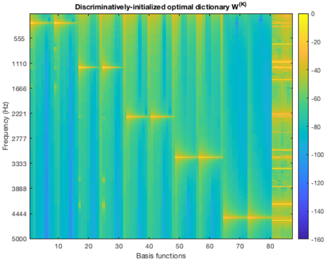
The Deep-NMF algorithm and architecture was originally designed for speech enhancement, which effectively makes it ill-suited to deal with datasets stemming from *physico-mathematical models*, where the clean components S have strong **intrinsic structure** that should be preserved by the network.

We proposed several **physics-aware enhancements**:

The Deep-NMF algorithm and architecture was originally designed for speech enhancement, which effectively makes it ill-suited to deal with datasets stemming from *physico-mathematical models*, where the clean components S have strong **intrinsic structure** that should be preserved by the network.

We proposed several **physics-aware enhancements**:

- Creation of an **optimal discriminative dictionary** $W^{(K)} = \hat{W}$ (not modified by back-propagation) which enables the recognition of physically-characterized clean components;

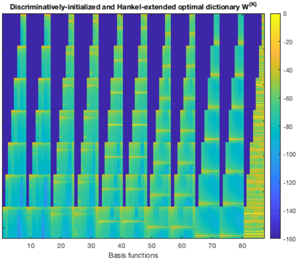


Università degli Studi di Padova

The Deep-NMF algorithm and architecture was originally designed for speech enhancement, which effectively makes it ill-suited to deal with datasets stemming from *physico-mathematical models*, where the clean components S have strong **intrinsic structure** that should be preserved by the network.

We proposed several **physics-aware enhancements**:

- Creation of an **optimal discriminative dictionary** $W^{(K)} = \hat{W}$ (not modified by back-propagation) which enables the recognition of physically-characterized clean components;
- Embedding of the (time-)correlation between consecutive columns of X by employing **block-Hankel** weights $W^{(k)}$;

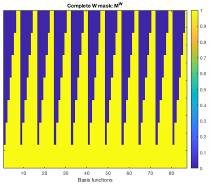


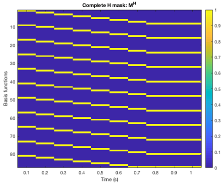
Physics-Aware Deep-NMF

The Deep-NMF algorithm and architecture was originally designed for speech enhancement, which effectively makes it ill-suited to deal with datasets stemming from *physico-mathematical models*, where the clean components S have strong **intrinsic structure** that should be preserved by the network.

We proposed several **physics-aware enhancements**:

- Creation of an **optimal discriminative dictionary** $W^{(K)} = \hat{W}$ (not modified by back-propagation) which enables the recognition of physically-characterized clean components;
- Embedding of the (time-)correlation between consecutive columns of X by employing **block-Hankel** weights $W^{(k)}$;
- Preservation of the block-Hankel structure by **projection**, thus modifying both the forward- and back-propagation;





Physics-Aware Deep-NMF

The Deep-NMF algorithm and architecture was originally designed for speech enhancement, which effectively makes it ill-suited to deal with datasets stemming from *physico-mathematical models*, where the clean components S have strong **intrinsic structure** that should be preserved by the network.

We proposed several **physics-aware enhancements**:

- Creation of an **optimal discriminative dictionary** $W^{(K)} = \hat{W}$ (not modified by back-propagation) which enables the recognition of physically-characterized clean components;
- Embedding of the (time-)correlation between consecutive columns of X by employing **block-Hankel** weights $W^{(k)}$;
- Preservation of the block-Hankel structure by **projection**, thus modifying both the forward- and back-propagation;
- Construction of a **suitable loss function** for the training process, enforcing an accurate reconstruction of the clean component.

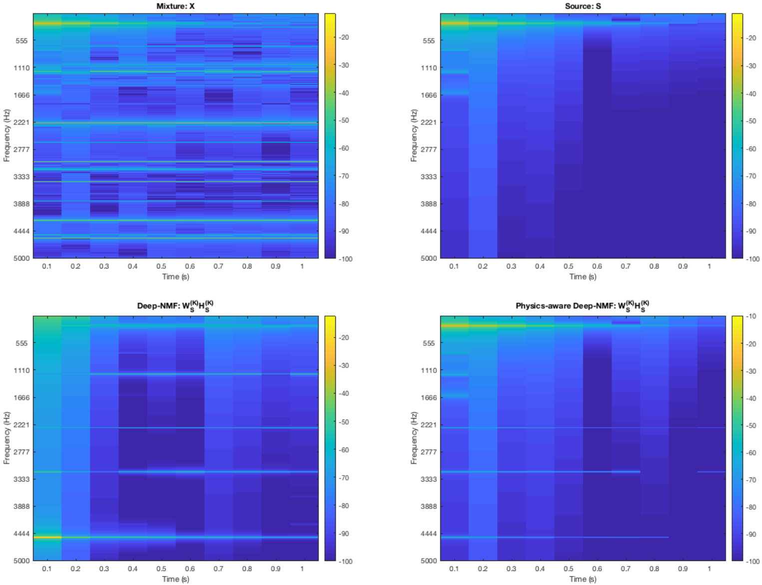
Penalty hyperparameter

Penalty term

$$\mathcal{E} = \|S - F \circ X\|_2^2 + \gamma \|S - W_S^{(K)} H_S^{(K)}\|_2^2$$

Problem-specific term
(Wiener filter)

Example 1: Hits detection on synthetic data



Example 2: Hits detection on real data