NON-NEGATIVE SPARSE CODING

by P. O. Hoyer,2002

Yunfei Wang

Sep 14,2012

Outline

- Non-negative sparse coding
- Estimating the hidden components
- Learning the basis
- Experiments
- Proof of monotonic convergence

NMF

$$X \approx AS(X \in \mathbb{R}^{n\times m}) \Rightarrow \boldsymbol{x}_i \approx [\boldsymbol{a}_1, \boldsymbol{a}_2 \ldots \boldsymbol{a}_r] \begin{bmatrix} \boldsymbol{s}_{i1} \\ \boldsymbol{s}_{i2} \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{s}_{ir} \end{bmatrix}$$

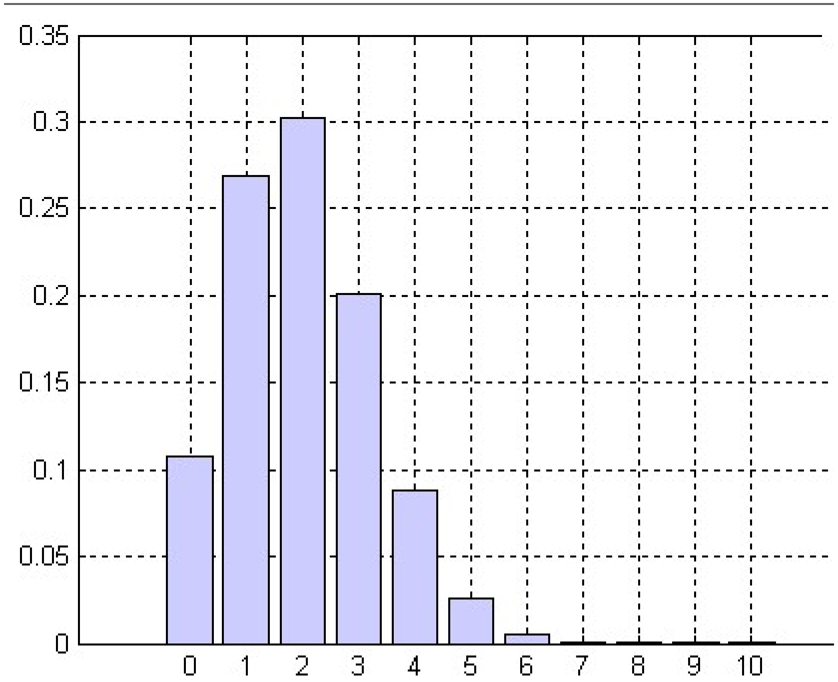
- Linear Decomposition

  1.columns of A contain the basis vector

  2.rows of S contain the corresponding hidden component

Linear Sparse Coding

- Goal-Representing any given input vector using only a few significantly non-zero hidden coefficients.



Non-negative Sparse Coding

- Both NMF and sparseness are important for learning parts-based representations.
- Combining the goal of small reconstruction error of NMF with that of sparseness.

$$\begin{cases} \min\limits_{\boldsymbol{A},\boldsymbol{S}} C(\boldsymbol{A},\boldsymbol{S}) = \dfrac{1}{2}\left\|\boldsymbol{X}-\boldsymbol{A}\boldsymbol{S}\right\|^2 + \lambda \sum\limits_{ij} f\left(\boldsymbol{S}_{ij}\right) \\ \boldsymbol{A} \geq 0, \boldsymbol{S} \geq 0, \forall i: \left\|\boldsymbol{a}_i\right\| = 1, \lambda \geq 0 \end{cases}$$

Estimating the hidden components

- For a given basis A, the objective is convex with respect to S, so the global minimum exists.
- Update rule for S:

$$\boldsymbol{S}^{t+1} = \boldsymbol{S}^{t} .* \left(\boldsymbol{A}^{T} \boldsymbol{X} \right) ./ \left(\boldsymbol{A}^{T} \boldsymbol{A} \boldsymbol{S} + \lambda \right)$$

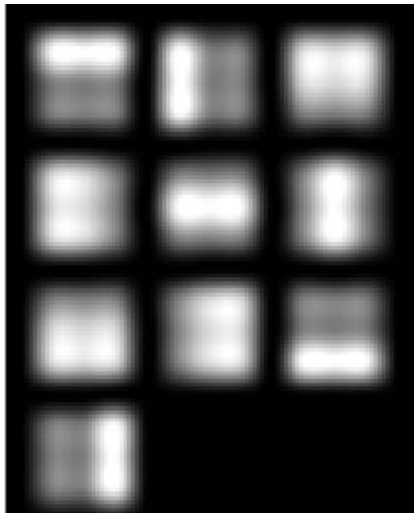
  the proof of monotonous convergence is similar to that of NMF

Learning the basis

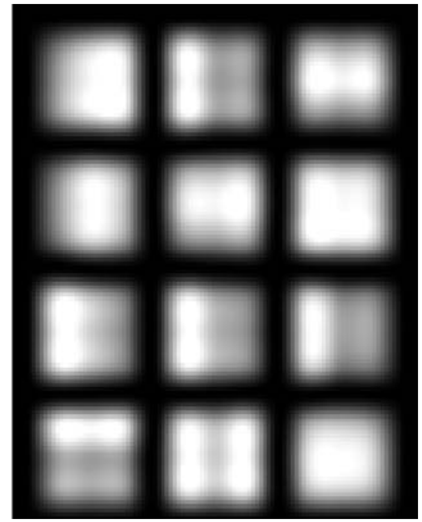
- Consider the optimization of A, holding S fixed
- Updating steps of A:
  - 1. $\boldsymbol{A}^{t+1} = \boldsymbol{A}^{t} + \mu\left(\boldsymbol{X} - \boldsymbol{A}^{t}\boldsymbol{S}\right)\boldsymbol{S}^{T}$
  - 2. Set all the negative values in A^{t+1} to zero.
  - 3. Rescale each column of A^{t+1} to unit norm.

Experiments-learning basis

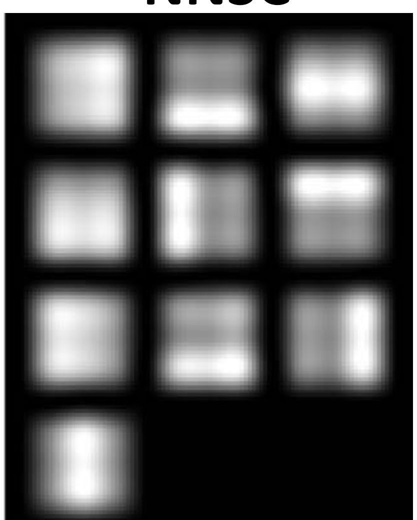
Features



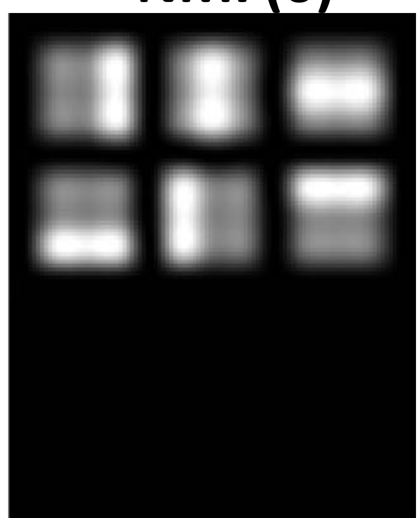
Samples



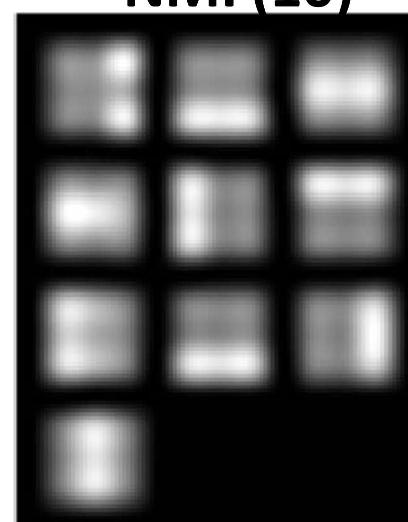
NNSC



NMF(6)



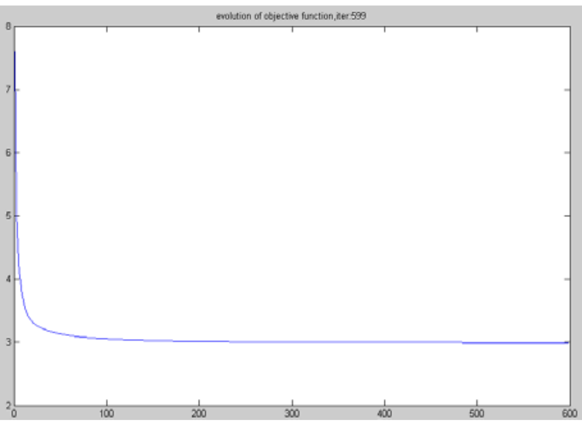
NMF(10)

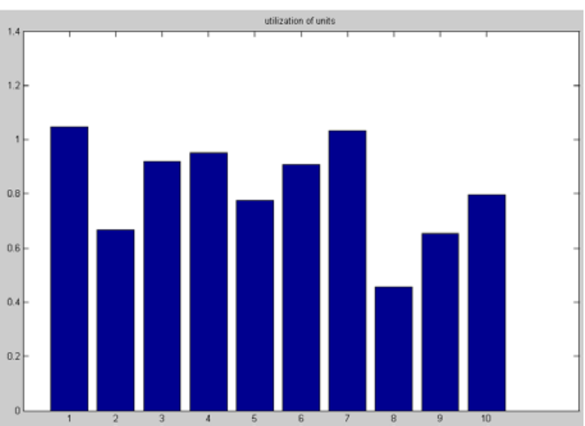


Experiments
convergence and hidden components

NNSC

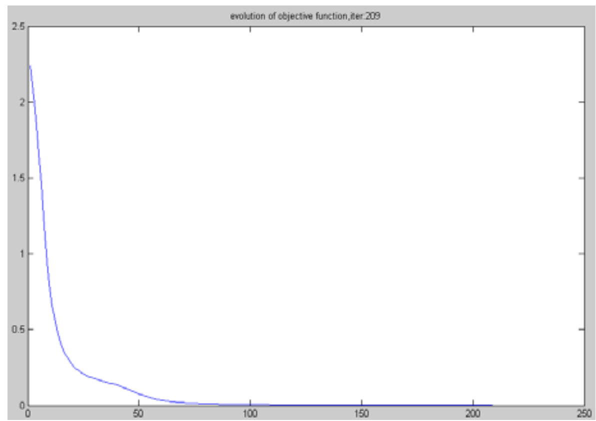
[583]:2.94223/4.11878

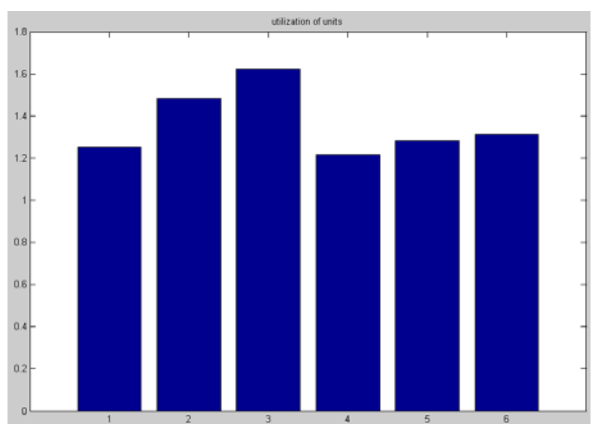




NMF(6)

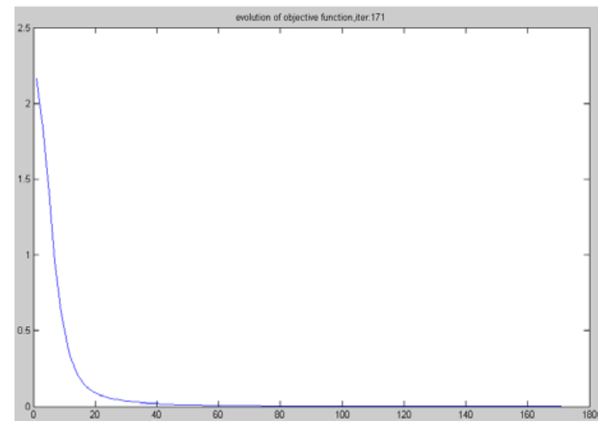
[209]: 0.00129 / 0.00000

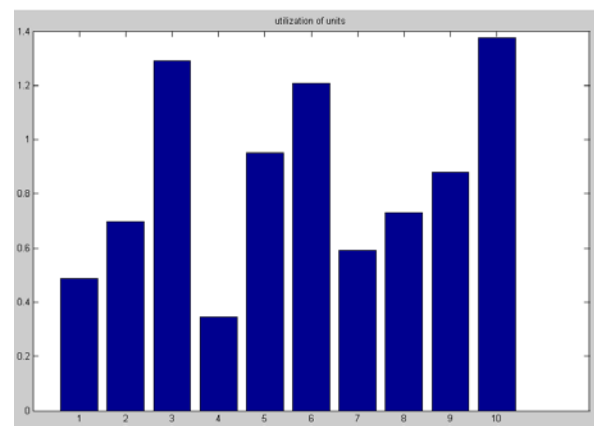




NMF(10)

[171]: 0.00097 / 0.00000

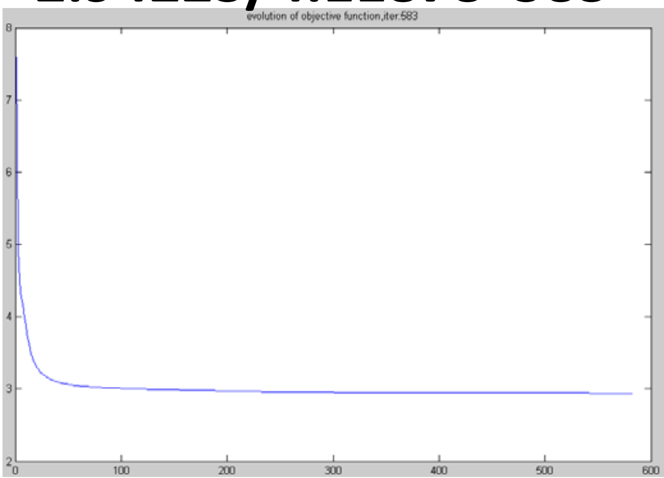


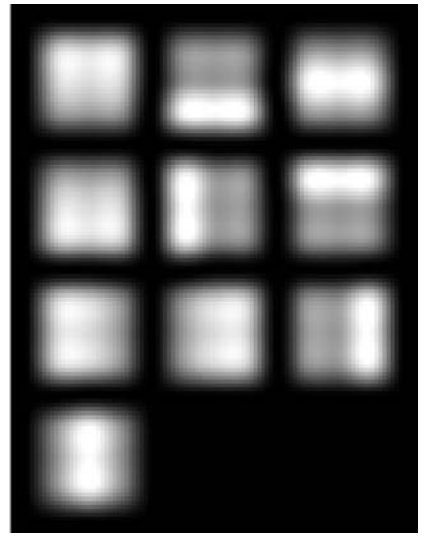


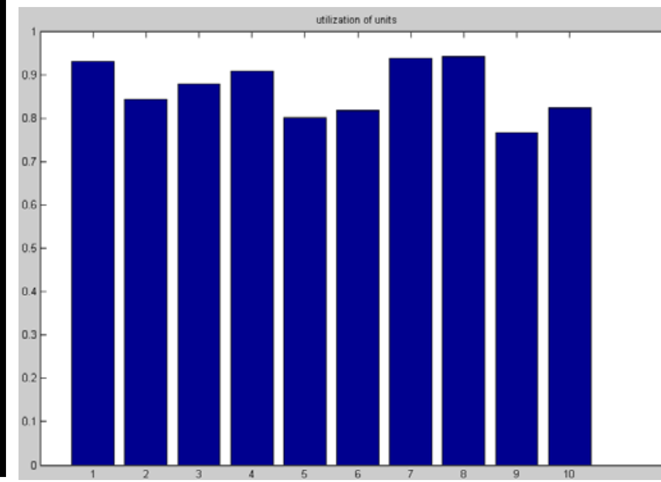
Experiments-Change NNSC a little

NNSC

2.94223/4.11878 583

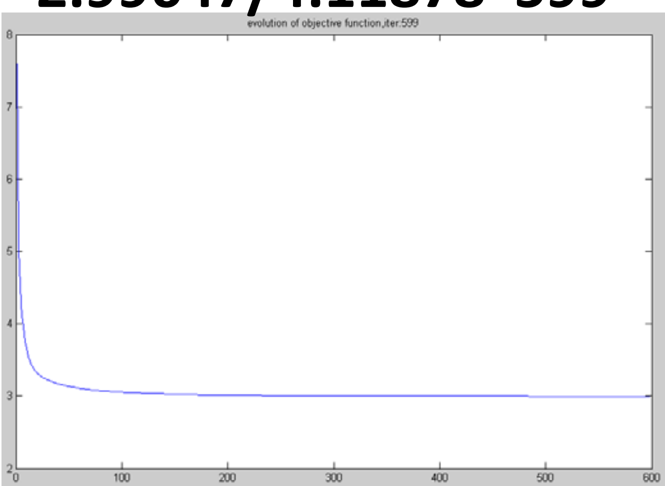


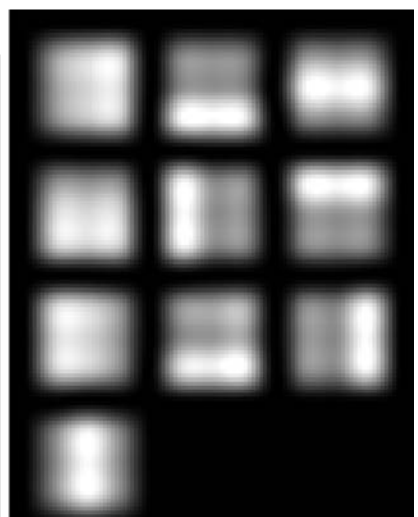


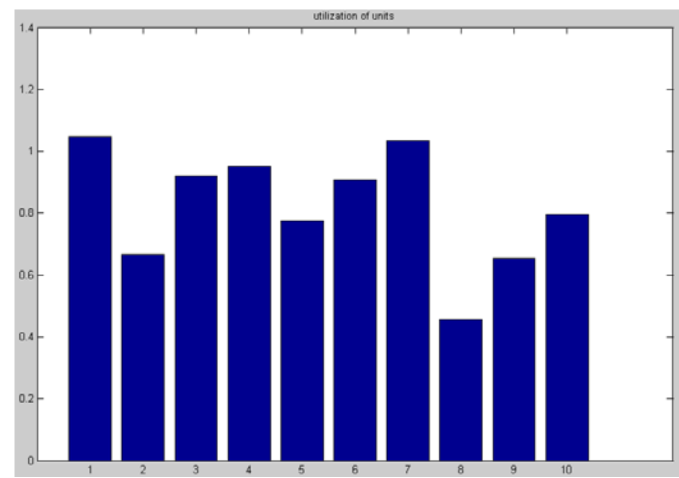


Changed NNSC

2.99647/4.11878 599







Three problems remained of NMF

- How to get the step size(numerator and denominator)?
- How to handle the concave case?
- How to construct auxiliary function?

Proof of monotonic convergence(1)

- Objective function:

$$F(\boldsymbol{h})=\frac{1}{2}\sum_i\left(\boldsymbol{v}_i-\sum_a \boldsymbol{W}_{ia}\boldsymbol{h}_a\right)^2=\frac{1}{2}\left\|\boldsymbol{v}-\boldsymbol{W}\boldsymbol{h}\right\|^2=\frac{1}{2}Tr\left[\left(\boldsymbol{v}-\boldsymbol{W}\boldsymbol{h}\right)\left(\boldsymbol{v}-\boldsymbol{W}\boldsymbol{h}\right)^T\right]$$

- Auxiliary function:

$$G(\boldsymbol{h},\boldsymbol{h}^t):\begin{cases}G(\boldsymbol{h},\boldsymbol{h}^t)\geq F(\boldsymbol{h})\\ G(\boldsymbol{h},\boldsymbol{h})=F(\boldsymbol{h})\end{cases}$$

$$\boldsymbol{h}^{t+1}=\arg\min_{\boldsymbol{h}} G(\boldsymbol{h},\boldsymbol{h}^t)$$

- F(***h***) is convergent: $F(\boldsymbol{h}^{t+1})\leq G(\boldsymbol{h}^{t+1},\boldsymbol{h}^t)\leq G(\boldsymbol{h}^t,\boldsymbol{h}^t)=F(\boldsymbol{h}^t)$
- Does the auxiliary function exists?

Proof of monotonic convergence(2)

- Constructing the auxiliary function for F(***h***)

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

$$\mathrm{K}_{ab}(\boldsymbol{h}^t) = \delta_{ab}\frac{\left(\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{h}\right)_a}{\boldsymbol{h}_a^t} = \begin{pmatrix} \frac{\left(\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{h}\right)_1}{\boldsymbol{h}_1^t} & & 0 \\ & \ddots & \\ 0 & & \frac{\left(\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{h}\right)_r}{\boldsymbol{h}_r^t} \end{pmatrix}$$

$$\frac{\partial F(\boldsymbol{h})}{\partial \boldsymbol{h}} = \boldsymbol{W}^T\boldsymbol{W}\boldsymbol{h}, \frac{\partial^2 F(\boldsymbol{h})}{\partial \boldsymbol{h}^2} = \boldsymbol{W}\boldsymbol{W}^T$$

- Second-order Taylor expansion of F(***h***) at ***h***^{t}

$$F(\boldsymbol{h}) = F(\boldsymbol{h}^t) + (\boldsymbol{h}\text{-}\boldsymbol{h}^t)^T\nabla F(\boldsymbol{h}^t) + \frac{1}{2}(\boldsymbol{h}\text{-}\boldsymbol{h}^t)^T\boldsymbol{W}^T\boldsymbol{W}(\boldsymbol{h}\text{-}\boldsymbol{h}^t)$$

Proof of monotonic convergence(3)

- Proving G($\boldsymbol{h}$,$\boldsymbol{h}^t$) is the auxiliary function of F($\boldsymbol{h}$)

1.$G(\boldsymbol{h},\boldsymbol{h})=F(\boldsymbol{h})$ is obvious.

$$2.\mathrm{G}(\boldsymbol{h},\boldsymbol{h}^t)\geq F(\boldsymbol{h})\Leftarrow(\boldsymbol{h}\text{-}\boldsymbol{h}^t)^T\left[K(\boldsymbol{h}^t)\text{-}\boldsymbol{W}^T\boldsymbol{W}\right](\boldsymbol{h}\text{-}\boldsymbol{h}^t)\geq 0$$

$$M_{ab}(\boldsymbol{h}^t)=\boldsymbol{h}_a^t\left[K(\boldsymbol{h}^t)\text{-}\boldsymbol{W}^T\boldsymbol{W}\right]_{ab}\boldsymbol{h}_b^t$$

$$\boldsymbol{v}^T M\boldsymbol{v}=\sum_{ab}\boldsymbol{v}_a\boldsymbol{h}_a^t K(\boldsymbol{h}^t)_{ab}\boldsymbol{h}_b^t\boldsymbol{v}_b-\sum_{ab}\boldsymbol{v}_a\boldsymbol{h}_a^t(\boldsymbol{W}^T\boldsymbol{W})_{ab}\boldsymbol{h}_b^t\boldsymbol{v}_b$$

$$=\sum_{a}\boldsymbol{v}_a\boldsymbol{h}_a^t K(\boldsymbol{h}^t)_{aa}\boldsymbol{h}_a^t\boldsymbol{v}_a-\sum_{ab}\boldsymbol{v}_a\boldsymbol{h}_a^t(\boldsymbol{W}^T\boldsymbol{W})_{ab}\boldsymbol{h}_b^t\boldsymbol{v}_b$$

$$=\sum_{a}\boldsymbol{v}_a^2(\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{h}^t)_a\boldsymbol{h}_a^t-\sum_{ab}\boldsymbol{v}_a\boldsymbol{h}_a^t(\boldsymbol{W}^T\boldsymbol{W})_{ab}\boldsymbol{h}_b^t\boldsymbol{v}_b$$

$$=\sum_{a}\boldsymbol{v}_a^2\boldsymbol{h}_a^t(\sum_{b}(\boldsymbol{W}^T\boldsymbol{W})_{ab}\boldsymbol{h}_b^t)-\sum_{ab}\boldsymbol{v}_a\boldsymbol{h}_a^t(\boldsymbol{W}^T\boldsymbol{W})_{ab}\boldsymbol{h}_b^t\boldsymbol{v}_b$$

$$=\sum_{ab}\boldsymbol{v}_a^2\boldsymbol{h}_a^t(\boldsymbol{W}^T\boldsymbol{W})_{ab}\boldsymbol{h}_b^t-\sum_{ab}\boldsymbol{v}_a\boldsymbol{h}_a^t(\boldsymbol{W}^T\boldsymbol{W})_{ab}\boldsymbol{h}_b^t\boldsymbol{v}_b$$

$$=\sum_{ab}\boldsymbol{v}_a^2\boldsymbol{h}_a^t(\boldsymbol{W}^T\boldsymbol{W})_{ab}\boldsymbol{h}_b^t-\sum_{ab}\boldsymbol{v}_a\boldsymbol{h}_a^t(\boldsymbol{W}^T\boldsymbol{W})_{ab}\boldsymbol{h}_b^t\boldsymbol{v}_b$$

$$=\frac{1}{2}\sum_{ab}\boldsymbol{h}_a^t(\boldsymbol{W}^T\boldsymbol{W})_{ab}\boldsymbol{h}_b^t\left(\boldsymbol{v}_a^2+\boldsymbol{v}_b^2\right)-\sum_{ab}\boldsymbol{v}_a\boldsymbol{h}_a^t(\boldsymbol{W}^T\boldsymbol{W})_{ab}\boldsymbol{h}_b^t\boldsymbol{v}_b$$

$$=\frac{1}{2}\sum_{ab}\boldsymbol{h}_a^t(\boldsymbol{W}^T\boldsymbol{W})_{ab}\boldsymbol{h}_b^t\left(\boldsymbol{v}_a-\boldsymbol{v}_b\right)^2\geq 0$$

K($\boldsymbol{h}^t$)-$\boldsymbol{W}^T\boldsymbol{W}$ is semi-positive definite if and only if $\boldsymbol{M}$ is semi-positive definite.

$$\boldsymbol{v}^T\left[K(\boldsymbol{h}^t)-\boldsymbol{W}^T\boldsymbol{W}\right]\boldsymbol{v}=\frac{1}{2}\sum_{ab}\left(\boldsymbol{W}^T\boldsymbol{W}\right)_{ab}\left(\boldsymbol{v}_a-\boldsymbol{v}_b\right)^2\geq 0$$

The update rule of h

$$\boldsymbol{h}^{t+1} = \arg\min_{\boldsymbol{h}} G(\boldsymbol{h}, \boldsymbol{h}^t)$$

$$\frac{\partial G(\boldsymbol{h}, \boldsymbol{h}^t)}{\partial \boldsymbol{h}} = \nabla F(\boldsymbol{h}^t) + K(\boldsymbol{h}^t)\left(\boldsymbol{h} - \boldsymbol{h}^t\right) = 0$$

$$\boldsymbol{h}^{t+1} = \boldsymbol{h}^t - K(\boldsymbol{h}^t)^{-1} \nabla F(\boldsymbol{h}^t)$$

$$\boldsymbol{h}_a^{t+1} = \boldsymbol{h}_a^t \frac{\left(\boldsymbol{W}^T \boldsymbol{v}\right)_a}{\left(\boldsymbol{W}^T \boldsymbol{W} \boldsymbol{h}^t\right)}$$

Handle the concave case

- Solution one:

$$G(\boldsymbol{h},\boldsymbol{h}^t):\begin{cases} G(\boldsymbol{h},\boldsymbol{h}^t) \le F(\boldsymbol{h}) \\ G(\boldsymbol{h},\boldsymbol{h}) = F(\boldsymbol{h}) \end{cases}$$

$$\boldsymbol{h}^{t+1} = \arg\max_{\boldsymbol{h}} G(\boldsymbol{h},\boldsymbol{h}^t)$$

$$F(\boldsymbol{h}^{t+1}) \ge G(\boldsymbol{h}^{t+1},\boldsymbol{h}^t) \ge G(\boldsymbol{h}^t,\boldsymbol{h}^t) = F(\boldsymbol{h}^t)$$

- Solution two:
  - Converting the objective function to convex by multiple it by a negative number.