$$\lambda_{ia} = \frac{\left(\mathbf{W}^t_{\mathbf{m+k+1}}\right)_{ia}}{\left(\mathbf{W}^t_{\mathbf{m+k+1}}\mathbf{H}^{t+1}_{\mathbf{m+k+1}}\mathbf{T}^{t+1^T}\right)_{ia}} \quad \text{where } T_{ij} = S_j(\alpha)\left(\mathbf{H}_{\mathbf{m+k+1}}\right)_{ij} \tag{A0}$$

The aim here is to show that selecting the step size given in Eq. (A0) will make the cost function given in Eq.(A1), nonincreasing and will guarantee the convergence of gradient descent algorithm. In Eq.(A1), **v** and **w** are row vectors of matrices **V** and **W**, respectively.

$$F(\mathbf{w}) = \frac{1}{2}\sum_j S_j(\alpha)\left(\bar{v}_j - \sum_a w_a H_{aj}\right)^2 \tag{A1}$$

Definition 1 $G(\mathbf{w}, \mathbf{w}^t)$ can be selected as an auxiliary function for $F(\mathbf{w})$ if it satisfies the conditions given in Eq.(A2).

$$G(\mathbf{w}, \mathbf{w}^t) \geq F(\mathbf{w}), \;\; G(\mathbf{w}, \mathbf{w}) = F(\mathbf{w}). \tag{A2}$$

Lemma 1 If *G* is an auxiliary function (definition 1) for *F*, then *F* will be nonincreasing if the update iterations are done as follows:

$$w^{t+1} = \arg\min_{\mathbf{w}} G(\mathbf{w}, \mathbf{w}^t) \tag{A3}$$

Proof: $F(\mathbf{w}^{t+1}) \leq G(\mathbf{w}^{t+1}, \mathbf{w}^t) \leq G(\mathbf{w}^t, \mathbf{w}^t) = F(\mathbf{w}^t)$

Since $G(\mathbf{w}, \mathbf{w}^t)$ is an auxiliary function *F*, according to the first part of Eq.(A2), $F(\mathbf{w}^{t+1}) \leq G(\mathbf{w}^{t+1}, \mathbf{w}^t)$ holds. In addition to this, $G(\mathbf{w}^{t+1}, \mathbf{w}^t) \leq G(\mathbf{w}^t, \mathbf{w}^t)$ (definition of *argmin*). Finally, $G(\mathbf{w}^t, \mathbf{w}^t) = F(\mathbf{w}^t)$ is valid due to definition 1 for auxiliary function $G(\mathbf{w}, \mathbf{w}^t)$.

Lemma 2 For the diagonal matrix **K** in Eq.(A4), and the cost function in Eq.(A1) , function $G(\mathbf{w},\mathbf{w}^t)$ in Eq.(A5) can be selected as an auxiliary fuction.

$$K\left(\mathbf{w}^t\right)_{aa} = \frac{\left(\mathbf{w}^t\mathbf{H}\mathbf{T}^T\right)_a}{w^t_a} \quad \text{, a=1,2,….} r \tag{A4}$$

where **w** represents a row of matrix **W** and subscript *a* is scanned through the interval [1,*r*]. The recently introduced matrix in Eq.(A4), **T,** is the scaled version of the encoding matrix, as it can be seen in Eq.(A5).

$$T_{ij} = S_j(\alpha)H_{ij}, \quad \text{where} \;\; i = 1,2,...r \;\&\; j = 1,2,..m \tag{A5}$$

$$G\left(\mathbf{w}, \mathbf{w}^t\right) = F\left(\mathbf{w}^t\right) + \left(\mathbf{w} - \mathbf{w}^t\right)\nabla F\left(\mathbf{w}^t\right) + \frac{1}{2}\left(\mathbf{w} - \mathbf{w}^t\right)K\left(\mathbf{w}^t\right)\left(\mathbf{w} - \mathbf{w}^t\right)^T \tag{A6}$$

First of all, availability of $G(\mathbf{w},\mathbf{w}^t)$ on being an auxiliary function for *F(***w***)* should be shown. To achieve this, the two conditions given in definition 1 should be verified. For $\mathbf{w}=\mathbf{w}^t$, it is clear in Eq.(A6) that $G(\mathbf{w}^t,\mathbf{w}^t)=F(\mathbf{w}^t)$ since the second and third components of the right-hand side will be zero. Next, the existence of the other condition of definition 1 should be proved.

In the task of comparing $G(\mathbf{w},\mathbf{w}^t)$ and $F(\mathbf{w}^t)$, a new expression for $F(\mathbf{w}^t)$, which is given in Eq.(A7) will be very useful.

$$F(\mathbf{w})=F(\mathbf{w}^t)+(\mathbf{w}-\mathbf{w}^t)\nabla F(\mathbf{w}^t)+\frac{1}{2}(\mathbf{w}-\mathbf{w}^t)(\mathbf{H}\mathbf{T}^T)(\mathbf{w}-\mathbf{w}^t)^T \tag{A7}$$

In order to show the equality between Eq.(A1) and Eq.(A7), firstly gradient of $F(\mathbf{w}^t)$ should be found. Eq.(A8) to Eq.(A13) lead to an open expression for $\nabla F(\mathbf{w}^t)$.

$$\nabla F(\mathbf{w}^t)=\left(\frac{\partial F(\mathbf{w}^t)}{\partial w_1^t},\frac{\partial F(\mathbf{w}^t)}{\partial w_2^t},\frac{\partial F(\mathbf{w}^t)}{\partial w_3^t},\ldots,\frac{\partial F(\mathbf{w}^t)}{\partial w_r^t}\right) \tag{A8}$$

$$\frac{\partial F(\mathbf{w}^t)}{\partial w_b^t}=\frac{\partial}{\partial w_b^t}\left(\frac{1}{2}\sum_j S_j(\alpha)\left(v_j-\sum_a w_a^t H_{aj}\right)^2\right) \tag{A9}$$

$$\frac{\partial F(\mathbf{w}^t)}{\partial w_b^t}=\frac{\partial}{\partial w_b^t}\left(\frac{1}{2}\sum_j S_j(\alpha)v_j^2-\sum_j S_j(\alpha)v_j\sum_a w_a^t H_{aj}+\frac{1}{2}\sum_j S_j(\alpha)\left(\sum_a w_a^t H_{aj}\right)^2\right) \tag{A10}$$

$$\frac{\partial F(\mathbf{w}^t)}{\partial w_b^t}=0-\sum_j v_j T_{bj}+\frac{1}{2}.2.\sum_j\sum_a H_{aj}w_a^t(T_{bj}) \tag{A11}$$

$$\frac{\partial F(\mathbf{w}^t)}{\partial w_b^t}=-\sum_j v_j T_{jb}^T+\sum_j\sum_a H_{aj}w_a^t(T_{jb}^T) \tag{A12}$$

$$\frac{\partial F(\mathbf{w}^t)}{\partial w_b^t}=-(\mathbf{v}\mathbf{T}^T)_b+(\mathbf{w}^t\mathbf{H}\mathbf{T}^T)_b \tag{A13}$$

Multiplying both sides of Eq.(A13) with $(\mathbf{w}-\mathbf{w}^t)$ yields Eq.(A14) and Eq.(A15).

$$(\mathbf{w}-\mathbf{w}^t)\nabla F(\mathbf{w}^t)=-\sum_b(\mathbf{v}\mathbf{T}^T)_b(w_b-w_b^t)+\sum_b(\mathbf{w}^t\mathbf{H}\mathbf{T}^T)_b w_b-\sum_b(\mathbf{w}^t\mathbf{H}\mathbf{T}^T)_b w_b^t \tag{A14}$$

$$(\mathbf{w}-\mathbf{w}^t)\nabla F(\mathbf{w}^t)=-\mathbf{v}\mathbf{T}^T\mathbf{w}^T+\mathbf{v}\mathbf{T}^T\mathbf{w}^{tT}+\mathbf{w}^t\mathbf{H}\mathbf{T}^T\mathbf{w}^T-\mathbf{w}^t\mathbf{H}\mathbf{T}^T\mathbf{w}^{tT} \tag{A15}$$

Eq.(A16) is yielded by subtracting $F(\mathbf{w}^t)$ from $F(\mathbf{w})$ of Eq.(A1).

$$F(\mathbf{w})-F(\mathbf{w}^t)=-\sum_j s_j v_j \sum_a H_{aj}\left(w_a-w_a^t\right)+\frac{1}{2}\sum_j s_j\left(\sum_a H_{aj}w_a\right)^2-\frac{1}{2}\sum_j s_j\left(\sum_a H_{aj}w_a^t\right)^2 \tag{A16}$$

Using Eq.(A5) and Eq.(A16) leads to Eq.(A17).

$$F(\mathbf{w})-F(\mathbf{w}^t)=-\mathbf{v}\mathbf{T}^T\mathbf{w}^T+\mathbf{v}\mathbf{T}^T\mathbf{w}^{tT}+\frac{1}{2}\mathbf{w}\mathbf{H}\mathbf{T}^T\mathbf{w}^T-\frac{1}{2}\mathbf{w}^t\mathbf{H}\mathbf{T}^T\mathbf{w}^{tT} \tag{A17}$$

The difference between Eq.(A17) and Eq.(A15) is given in Eq.(A18).

$$F(\mathbf{w})-F(\mathbf{w}^t)-(\mathbf{w}-\mathbf{w}^t)\nabla F(\mathbf{w}^t)=\frac{1}{2}\mathbf{w}\mathbf{H}\mathbf{T}^T\mathbf{w}^T+\frac{1}{2}\mathbf{w}^t\mathbf{H}\mathbf{T}^T\mathbf{w}^{tT}-\mathbf{w}^t\mathbf{H}\mathbf{T}^T\mathbf{w}^T \tag{A18}$$

Eq.(A18) can equivalently be rewritten as follows:

$$F(\mathbf{w})-(\mathbf{w}-\mathbf{w}^t)\nabla F(\mathbf{w}^t)=F(\mathbf{w}^t)+\frac{1}{2}\left[\sum_j s_j\left(\sum_a H_{aj}w_a\right)^2-2\sum_j\sum_a s_j w_b^t\left(HH^T\right)_{ja}w_a^T+\sum_j s_j\left(\sum_a H_{aj}w_a^t\right)^2\right] \tag{A19}$$

$$F(\mathbf{w})=F(\mathbf{w}^t)+(\mathbf{w}-\mathbf{w}^t)^T\nabla F(\mathbf{w}^t)+\frac{1}{2}(\mathbf{w}-\mathbf{w}^t)(\mathbf{H}\mathbf{T}^T)(\mathbf{w}-\mathbf{w}^t)^T \tag{A20}$$

After deriving a suitable expression for $F(\mathbf{w})$, it's time to check whether $G(\mathbf{w},\mathbf{w}^t)$ satisfies the second condition of definition 1, which claims that $G(\mathbf{w},\mathbf{w}^t)$ is bigger than $F(\mathbf{w})$. It is clear that the inequality in Eq.(20) is equivalent to this statement.

$$0\le(\mathbf{w}-\mathbf{w}^t)(\mathbf{K}(\mathbf{w}^t)-\mathbf{H}\mathbf{T}^T)(\mathbf{w}-\mathbf{w}^t)^T \tag{A21}$$

A quick glance will be sufficient to realize that Eq.(A21) is the definition of positive semidefiniteness. Therefore, availability of $G(\mathbf{w},\mathbf{w}^t)$ as an auxiliary function will be shown by proving positive semidefiniteness of $(\mathbf{K}(\mathbf{w}^t)-\mathbf{H}\mathbf{T}^T)$. In addition to this, proving positive semidefiniteness of the expression in Eq.(A22) is another way to achieve our goal as it is only a scaled version of $(\mathbf{K}(\mathbf{w}^t)-\mathbf{H}\mathbf{T}^T)$.

$$\mathbf{M}_{ab}(\mathbf{w}^t)=w_a^t\left(\mathbf{K}(\mathbf{w}^t)-\mathbf{H}\mathbf{T}^T\right)_{ab}w_b^t\ ,\ a=1,2,\ldots r\ ;\ b=1,2,\ldots r \tag{A22}$$

So, $(\mathbf{K}(\mathbf{w}^t)-\mathbf{H}\mathbf{T}^T)$ is positive semidefinite if and only if $\mathbf{M}$ is positive semidefinite. Following equations verify that $\mathbf{M}$ is positive semidefinite. Here, p is a r-dimensional vector:

$$\mathbf{p}^T\mathbf{M}\mathbf{p}=\sum_{ab}p_a M_{ab}p_b \tag{A23}$$

$$\mathbf{p}^T\mathbf{M}\mathbf{p}=\sum_{ab}p_a w_a^t K(\mathbf{w}^t)_{ab}w_b^t p_b-\sum_{ab}p_a w_a^t\left(\mathbf{H}\mathbf{T}^T\right)_{ab}w_b^t p_b \tag{A24}$$

Diagonality of matrix $\mathbf{K}$ makes it possible to convert Eq.(A24) to Eq.(A25).

$$\mathbf{p}^T\mathbf{M}\mathbf{p} = \sum_b p_b^2 w_b^t K\left(\mathbf{w}^t\right)_{bb} w_b^t - \sum_{ab} p_a w_a^t \left(\mathbf{H}\mathbf{T}^T\right)_{ab} w_b^t p_b \tag{A25}$$

By using Eq.(A4),

$$\mathbf{p}^T\mathbf{M}\mathbf{p} = \sum_b p_b^2 w_b^t \left(\sum_a w_a^t \left(\mathbf{H}\mathbf{T}^T\right)_{ab}\right) - \sum_{ab} p_a w_a^t \left(\mathbf{H}\mathbf{T}^T\right)_{ab} w_b^t p_b \tag{A26}$$

$$\mathbf{p}^T\mathbf{M}\mathbf{p} = \sum_{ab} p_a^2 w_a^t \left(\mathbf{H}\mathbf{T}^T\right)_{ab} w_b^t - \sum_{ab} p_a w_a^t \left(\mathbf{H}\mathbf{T}^T\right)_{ab} w_b^t p_b \tag{A27}$$

$$\mathbf{p}^T\mathbf{M}\mathbf{p} = \sum_{ab}\left(\frac{p_a^2}{2} + \frac{p_b^2}{2}\right) w_a^t \left(\mathbf{H}\mathbf{T}^T\right)_{ab} w_b^t - \sum_{ab}\left(p_a p_b\right) w_a^t \left(\mathbf{H}\mathbf{T}^T\right) w_b^t \tag{A28}$$

$$\mathbf{p}^T\mathbf{M}\mathbf{p} = \sum_{ab}\left(\mathbf{H}\mathbf{T}^T\right)_{ab} w_a^t w_b^t \left[\frac{1}{2}p_a^2 + \frac{1}{2}p_b^2 - p_a p_b\right] \tag{A29}$$

$$\mathbf{p}^T\mathbf{M}\mathbf{p} = \frac{1}{2}\sum_{ab}\left(\mathbf{H}\mathbf{T}^T\right)_{ab} w_a^t w_b^t \left(p_a - p_b\right)^2 \tag{A30}$$

$$\mathbf{p}^T\mathbf{M}\mathbf{p} \geq 0 \tag{A31}$$

As a conclusion for this step, $G\left(\mathbf{w}, \mathbf{w}^t\right)$ is a suitable auxiliary function for $F\left(\mathbf{w}\right)$. The following step is to find the update rule which is derived from Eq.(2).

$$\frac{\partial}{\partial \mathbf{w}} G\left(\mathbf{w}, \mathbf{w}^t\right) = \nabla F\left(\mathbf{w}^t\right) + \frac{1}{2}.2.\left(\mathbf{w} - \mathbf{w}^t\right)\mathbf{K}\left(\mathbf{w}^t\right) = 0 \tag{A32}$$

$$\left(\mathbf{w} - \mathbf{w}^t\right)\mathbf{K}\left(\mathbf{w}^t\right) = -\nabla F\left(\mathbf{w}^t\right) \tag{A33}$$

$$\left(\mathbf{w} - \mathbf{w}^t\right) = -\nabla F\left(\mathbf{w}^t\right)\mathbf{K}^{-1}\left(\mathbf{w}^t\right) \tag{A34}$$

$$\mathbf{w}^{t+1} \stackrel{\Delta}{=} \mathbf{w} = \mathbf{w}^t - \mathbf{K}\left(\mathbf{w}^t\right)^{-1} \nabla F\left(\mathbf{w}^t\right) \tag{A35}$$

$$\mathbf{w}_a^{t+1} = \mathbf{w}_a^t - \mathbf{K}\left(\mathbf{w}^t\right)_{aa}^{-1}\left(\frac{\partial F\left(\mathbf{w}^t\right)}{\partial w_a^t}\right) \tag{A36}$$

So, it is proven that by selecting $K\left(\mathbf{w}^t\right)_{aa}^{-1}$ as the step size for the gradient descent formulation in Eq.(A36) convergence of the algorithm is guaranteed. The open expression for $K\left(\mathbf{w}^t\right)_{aa}^{-1}$, which is given in Eq.(A37), is equivalent to the step size in Eq.(24). This completes the proof.

$$K\left(\mathbf{w}^t\right)_{aa}^{-1} = \frac{w_a^t}{\left(\mathbf{w}^t\mathbf{H}\mathbf{T}^T\right)_a} \tag{A37}$$