1 Introduction

1. Suppose we have some sequence T, a combination with repetition of M tokens taken from a set of $W = \{w_k, 0 \leq k < N\}$.

2. Now, we define the one hot encoding $\mathbf{e}_k \in B^N$, where $\mathbf{e}_k$ is the $k-$esim column of $\mathbf{I}_N$, the identity matrix in $\Re^{N\times N}$. Thus, we assign the vector $\mathbf{e}_k$ to each token w_k.

3. Then, we define a matrix $\mathbf{D} \in B^{N\times M}$, where $T_m = w_k$, $0 \leq m < M$ implies $\mathbf{D}_m \doteq \mathbf{e}_k$.

Note that

1. $\mathbf{D}\cdot\mathbf{D}^t = \mathbf{F}$, where $\mathbf{F} \in \Re^{N\times N}\}$ is a diagonal matrix such that $D_{i,k} \times D_{i,k} = 1$ or $D_{i,k} \times D_{i,k} = 0$, $F_{i,i} = \sum_{k=1}^{M} D_{i,k} \times D_{k,i} = F_{i,i} \neq 0$, $0 < F_{i,i} < M$, $F_{i,j} = \sum_{k=1}^{M} D_{i,k} \times D_{k,j} = F_{i,j} = 0$.

2. $\mathbf{D}\cdot\mathbf{u} = \mathbf{f}$, where $\mathbf{f} \in \Re^{N}\}$ is a vector such that $f_i = \sum_k D_{i,k} = F_{i,i} \neq 0$.

We want to evaluate

$$\mathbf{D}\cdot\left(\mathbf{I}_m - \frac{\mathbf{u}\cdot\mathbf{u}^t}{m}\right)\cdot\mathbf{x} = \lambda\,\mathbf{y}, \tag{1}$$

$$\left(\mathbf{I}_m - \frac{\mathbf{u}\cdot\mathbf{u}^t}{m}\right)\cdot\mathbf{D}^t\cdot\mathbf{y} = \lambda\,\mathbf{x}. \tag{2}$$