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 $\mathbf{D}_m \doteq \mathbf{e}_k$.

We want to evaluate

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

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