Supplemental material for Feature Sequence Representation Via Slow Feature Analysis For Action Classification

Takumi Kobayashi
takumi.kobayashi@aist.go.jp
National Institute of Advanced Industrial Science and Technology
Tsukuba, Japan

In this material, we explain the slow feature analysis of small-sized samples from the more practical (discrete) viewpoint than [1].

Let $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_T] \in \mathbb{R}^{d\times T}$ denote a feature sequence of T feature vectors $\boldsymbol{x}_t \in \mathbb{R}^d$. For simplicity, the feature sequence $\boldsymbol{X}$ is centered, $\sum_t \boldsymbol{x}_t = \boldsymbol{X}\mathbf{1} = \mathbf{0}$. The differential vector is computed as

$$\dot{\boldsymbol{x}}_t = \boldsymbol{x}_{t+1} - \boldsymbol{x}_t,\ \forall t \in \{1, \cdots, T-1\}, \tag{1}$$

$$\dot{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{D} \in \mathbb{R}^{d\times T-1}, \quad \text{where } \boldsymbol{D} = \begin{pmatrix} -1 & & & \\ 1 & -1 & & \\ & 1 & & \\ & & \ddots & \\ & & & -1 \\ & & & 1 \end{pmatrix} \in \mathbb{R}^{T\times T-1}. \tag{2}$$

By using these notations, the slow feature analysis is formulated as

$$(\dot{\boldsymbol{X}}\dot{\boldsymbol{X}}^\top)\boldsymbol{w} = \lambda(\boldsymbol{X}\boldsymbol{X}^\top)\boldsymbol{w} \ \Leftrightarrow\ \boldsymbol{X}\boldsymbol{D}\boldsymbol{D}^\top\boldsymbol{X}^\top\boldsymbol{w} = \lambda(\boldsymbol{X}\boldsymbol{X}^\top)\boldsymbol{w} \ \Leftrightarrow\ \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^\top\boldsymbol{w} = \lambda(\boldsymbol{X}\boldsymbol{X}^\top)\boldsymbol{w}, \tag{3}$$

where λ and $\boldsymbol{w}$ are a eigenvalue and its eigenvector, respectively, and

$$\boldsymbol{G} = \boldsymbol{D}\boldsymbol{D}^\top = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{T\times T}. \tag{4}$$

By applying SVD to $\boldsymbol{X}$, we obtain the form of $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ where $\boldsymbol{U} \in \mathbb{R}^{d\times r}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$ and $\boldsymbol{V} \in \mathbb{R}^{T\times r}$ and r indicates the rank of $\boldsymbol{X}$. In the case of small-sized samples, $T < d$, the rank r is less than T due to the centering of $\boldsymbol{X}$, *i.e.*, $\boldsymbol{X}\mathbf{1} = \mathbf{0}$, and usually $r = T-1$ due to a small number of samples in large dimensional feature space. Note that $\boldsymbol{V}^\top\mathbf{1} = \mathbf{0}$ since $\boldsymbol{X}\mathbf{1} = \mathbf{0}$. We reparameterize $\boldsymbol{w}$ by $\boldsymbol{w} = \boldsymbol{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}$ where $\boldsymbol{\alpha} \in \mathbb{R}^r$, thereby reformulating (3) into

$$\boldsymbol{V}^\top\boldsymbol{G}\boldsymbol{V}\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha}. \tag{5}$$

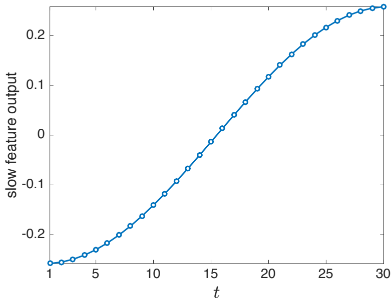

Figure 1: The slowest feature outputs $\{\boldsymbol{w}^\top \boldsymbol{x}_t\}_{t=1}^T$ of $T = 30$.

Here, $\boldsymbol{G}$ (4) is a fundamental matrix to produce discrete cosine transform (DCT) [1] and the eigenvalues/eigenvectors of $\boldsymbol{G}$ are analytically given by

$$\boldsymbol{G} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^\top, \tag{6}$$

$$\text{where } P_{tk} = -\sqrt{\frac{2}{T}}\cos\left((t-\frac{1}{2})\frac{k\pi}{T}\right),\ \Lambda_{kk} = 2\left\{1-\cos\left(\frac{k\pi}{T}\right)\right\}, \tag{7}$$

$$\forall t \in \{1,\cdots,T\},\ \forall k \in \{1,\cdots,T-1\}. \tag{8}$$

Note that the eigenvectors $\boldsymbol{P} \in \mathbb{R}^{T\times T-1}$ are DCT bases[1] of $\boldsymbol{P}^\top\boldsymbol{1} = \boldsymbol{0}$, and the eigenvalues $\{\Lambda_{kk}\}_{k=1}^{T-1}$ are increasing order. Thus, (5) is rewritten to

$$\boldsymbol{V}^\top\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^\top\boldsymbol{V}\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha}. \tag{9}$$

Both $\boldsymbol{V}$ and $\boldsymbol{P}$ are the orthonormal bases spanning the $T-1$ dimensional subspace that is perpendicular to $\boldsymbol{1}$ in the T-dimensional Euclidean space, and thus $\text{span}(\boldsymbol{V}) = \text{span}(\boldsymbol{P})$ leading to $\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{P} = \boldsymbol{P}$.

As a result, the eigenvectors $\boldsymbol{\alpha}$ and the eigenvalues λ are

$$\boldsymbol{\alpha}_k = \boldsymbol{V}^\top\boldsymbol{p}_k,\ \lambda_k = \Lambda_{kk},\ \forall k \in \{1,\cdots,T-1\}, \tag{10}$$

where $\boldsymbol{p}_k$ is the k-th column vector of the matrix $\boldsymbol{P}$. Finally, we obtain $\boldsymbol{w}_k = \boldsymbol{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}^\top\boldsymbol{p}_k$. The extracted slow features are given by $\boldsymbol{X}^\top\boldsymbol{w}_k = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}^\top\boldsymbol{p}_k = \boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{p}_k = \boldsymbol{p}_k$. Specifically, the slowest feature of the minimum eigenvalue $\lambda_1 = \Lambda_{11} = 2\{1-\cos(\frac{\pi}{T})\}$ is $\boldsymbol{x}_t^\top\boldsymbol{w}_1 = -\sqrt{\frac{2}{T}}\cos((t-\frac{1}{2})\frac{\pi}{T})$ as shown in Fig. 1. It should be noted that any feature sequences produce the identical form (Fig. 1) of the slowest features in the case of small-sized samples ($T < d$) no matter how the frame features $\boldsymbol{x}_t$ are distributed in the d-dimensional space.

References

[1] G. Strang. The discrete cosine transform. *SIAM review*, 41(1):135–147, 1999.

[2] L. Wiskott. Slow feature analysis: A theoreticalanalysis of optimal free responses. *Neural Computation*, 15(9):2147–2177, 2003.

[1] We assign negative signs to the standard DCT bases so that the extracted slowest feature is positively correlated with t (see Fig. 1).