Supplementary: Inducing Neural Collapse to a Fixed Hierarchy-Aware Frame for Reducing Mistake Severity

Anonymous ICCV submission

Paper ID 10135

1. Prove the proposed hierarchy-aware frame satisfies the frame condition

First, we give the definition of the frame condition in Euclidean space:

Definition 1 (Frame)**.** A set of vectors $\{\boldsymbol{\varphi}_k\}_{k=1}^{M}$ in $\mathbb{R}^N$ is called a frame for $\mathbb{R}^N$, if there exist constants $0 < A \leq B < \infty$ such that

$$A||\boldsymbol{x}||^2 \leq \sum_{k=1}^{M} |\langle \boldsymbol{x}, \boldsymbol{\varphi}_k \rangle|^2 \leq B||\boldsymbol{x}||^2, \, \forall \boldsymbol{x} \in \mathbb{R}^N \tag{1}$$

where $\langle .,. \rangle$ and $||.||$ denote the dot product on $\mathbb{R}^N$ and its corresponding norm.

Then we repeat the definition of our Hierarchy-Aware Frame (HAFrame):

Definition 2 (Hierarchy-Aware Frame)**.** Let a set of vectors $\{\boldsymbol{w}_i\}_{i=1}^{K}$ in $\mathbb{R}^K$ with $||\boldsymbol{w}_i||_2 = 1$, for $i = 1, ..., K$, and their pair-wise cosine similarities satisfy the following equation:

$$cos\angle(\boldsymbol{w}_i, \boldsymbol{w}_j) = \boldsymbol{w}_i^T \boldsymbol{w}_j = S_{ij}, \forall 1 \leq i \leq j \leq K \tag{2}$$

where S_{ij} is the cosine similarity between classes i and j given by our exponential mapping function. Next, let $\boldsymbol{W} = [\boldsymbol{w}_1 \, \boldsymbol{w}_2 \, ... \, \boldsymbol{w}_K]$ be the proposed hierarchy-aware frame, $\boldsymbol{W}$ is in $\mathbb{R}^{K \times K}$ and satisfies

$$\boldsymbol{S} = \boldsymbol{W}^T \boldsymbol{W} \tag{3}$$

Since $\boldsymbol{S}$ is guaranteed to be positive definite by our search for the proper s_{min} in the mapping function, we can construct $\boldsymbol{W}$ with the following matrix factorization:

$$\boldsymbol{S} = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^T = (\boldsymbol{Q}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{U}^T)(\boldsymbol{U}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{Q}^T) = \boldsymbol{W}^T\boldsymbol{W} \tag{4}$$

where $\boldsymbol{Q} \in \mathbb{R}^{K \times K}$ and $\boldsymbol{D} \in \mathbb{R}^{K \times K}$ are acquired from eigenvalue decomposition of $\boldsymbol{S}$, and $\boldsymbol{U} \in \mathbb{R}^{K \times K}$ is an orthonormal matrix obtained from QR-decomposition [2] of a random matrix in $\mathbb{R}^{K \times K}$ that allows an arbitrary rotation and satisfies $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{I}_K$. Therefore, the proposed hierarchy-aware frame is given by:

$$\boldsymbol{W} = \boldsymbol{U}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{Q}^T \tag{5}$$

Theorem 1. The proposed HAFrame $\boldsymbol{W}$ satisfies the frame condition (Eqn. 1) if the associated cosine similarity matrix $\boldsymbol{S}$ is positive definite.

Proof. To prove the proposed HAFrame $\boldsymbol{W}$ satisfies the frame condition [1] given in Eqn. 1, we substitute the set of vectors $\{\boldsymbol{\varphi}_k\}_{k=1}^{M}$ in the Eqn. 1 with the set of vectors $\{\boldsymbol{w}_i\}_{i=1}^{K}$ consisting the HAFrame $\boldsymbol{W}$. It's equivalent to proving the following condition is satisfied when $\boldsymbol{S}$ is positive definite:

$$A||\boldsymbol{x}||^2 \leq \sum_{i=1}^{K} |\langle \boldsymbol{x}, \boldsymbol{w}_i \rangle|^2 \leq B||\boldsymbol{x}||^2, \, \forall \boldsymbol{x} \in \mathbb{R}^K \tag{6}$$

Therefore, we need to prove the following:

$$A\boldsymbol{x}^T\boldsymbol{x} \leq \sum_{i=1}^{K} (\boldsymbol{x}^T\boldsymbol{w}_i)(\boldsymbol{w}_i^T\boldsymbol{x}) \leq B\boldsymbol{x}^T\boldsymbol{x} \tag{7}$$

Since $\boldsymbol{W} = [\boldsymbol{w}_1 \; \boldsymbol{w}_2 \; ... \; \boldsymbol{w}_K]$, we can further write $\sum_{i=1}^{K}(\boldsymbol{x}^T\boldsymbol{w}_i)(\boldsymbol{w}_i^T\boldsymbol{x})$ as:

$$\sum_{i=1}^{K} (\boldsymbol{x}^T\boldsymbol{w}_i)(\boldsymbol{w}_i^T\boldsymbol{x}) = \boldsymbol{x}^T \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_1^T \\ \boldsymbol{w}_2^T \\ \vdots \\ \boldsymbol{w}_K^T \end{bmatrix} \boldsymbol{x} \tag{8}$$

Therefore, we have:

$$\sum_{i=1}^{K} (\boldsymbol{x}^T\boldsymbol{w}_i)(\boldsymbol{w}_i^T\boldsymbol{x}) = (\boldsymbol{x}^T\boldsymbol{W})(\boldsymbol{W}^T\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{x} \tag{9}$$

Combining inequalities in Eqn. 7 and Eqn. 9, we have:

$$A\boldsymbol{x}^T\boldsymbol{x} \leq \boldsymbol{x}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{x} \leq B\boldsymbol{x}^T\boldsymbol{x} \tag{10}$$

If $||\boldsymbol{x}|| = \mathbf{0}$, the *frame condition* in Eqn. 6 is met. Otherwise, we have:

$$A \leq \frac{\boldsymbol{x}^T \boldsymbol{W}\boldsymbol{W}^T \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \leq B \tag{11}$$

where $\frac{\boldsymbol{x}^T \boldsymbol{W}\boldsymbol{W}^T \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}$ is the Rayleigh quotient $R(\boldsymbol{W}\boldsymbol{W}^T, x)$ [3] for a real symmetric matrix $\boldsymbol{W}\boldsymbol{W}^T$, and it is bounded by the minimum and maximum eigenvalues of $\boldsymbol{W}\boldsymbol{W}^T$. Therefore, we have:

$$\lambda_{min} \leq \frac{\boldsymbol{x}^T \boldsymbol{W}\boldsymbol{W}^T \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \leq \lambda_{max} \tag{12}$$

where λ_{min} and λ_{max} are the minimum and maximum eigenvalues of $\boldsymbol{W}\boldsymbol{W}^T$. Hence, we only need to prove that $\boldsymbol{W}\boldsymbol{W}^T$ is a positive definite matrix when $\boldsymbol{S}$ is positive definite, i.e., for any $\boldsymbol{x} \neq \mathbf{0}$, we need to prove:

$$\boldsymbol{x}^T \boldsymbol{W}\boldsymbol{W}^T \boldsymbol{x} = (\boldsymbol{W}^T \boldsymbol{x})^T (\boldsymbol{W}^T \boldsymbol{x}) = ||\boldsymbol{W}^T \boldsymbol{x}||^2 > 0 \tag{13}$$

It is equivalent to prove $\boldsymbol{W}^T = \boldsymbol{Q}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{U}^T$ is invertible (i.e., $\boldsymbol{W}^T \boldsymbol{x} \neq \mathbf{0}$ when $\boldsymbol{x} \neq \mathbf{0}$). We can find the inverse of $\boldsymbol{W}^T$:

$$(\boldsymbol{W}^T)^{-1} = \boldsymbol{U}\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{Q}^T \tag{14}$$

where $\boldsymbol{U}$ is orthonormal, $\boldsymbol{Q}$ and $\boldsymbol{D}$ are acquired via eigendecomposition of real symmetric positive definite similarity matrix $\boldsymbol{S} = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^T$, therefore $\boldsymbol{Q}$ is also orthonormal, and $\boldsymbol{D}^{-\frac{1}{2}}$ is a real matrix (eigenvalues of $\boldsymbol{S}$ are all positive real numbers). Since $\boldsymbol{W}^T$ is invertible, $\boldsymbol{W}^T \boldsymbol{x} \neq \mathbf{0}$ when $\boldsymbol{x} \neq \mathbf{0}$, this finish proving of Eqn. 13. Therefore, the matrix $\boldsymbol{W}\boldsymbol{W}^T$ is guaranteed to be positive definite when $\boldsymbol{S}$ is positive definite, and our HAFrame $\boldsymbol{W}$ satisfies the *frame condition* in Eqn. 1. The proof is completed. □

References

[1] Peter G. Casazza, Gitta Kutyniok, and Friedrich Philipp. *Introduction to Finite Frame Theory*, pages 1–53. Birkhäuser Boston, Boston, 2013. 1

[2] J. G. F. Francis. The QR Transformation A Unitary Analogue to the LR Transformation - Part 1. *Comput. J.*, 4:265–271, 1961. 1

[3] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990. 2