Indeed, minimizing $||\frac{1}{N}\mathbf{H}^\top\mathbf{H}-\mathbf{I}||_F^2$ can alleviate collapse issues and enhance representation diversity, as proposed in CCA-SSG [1], known as *decorrelation-based* methods. Our IC is **parallel** to it.

[1] From Canonical Correlation Analysis to Self-supervised Graph Neural Networks.

In our study, we establish the connection between our Isotropic Constraints (IC) and this method. Proposition 2 states that decorrelation loss can be regarded as a special case of our isotropic constraints under the conditions of ***1) only considering 2-nd central moment*** and ***2) choosing eigenvectors of covariance matrix as projection vectors***.

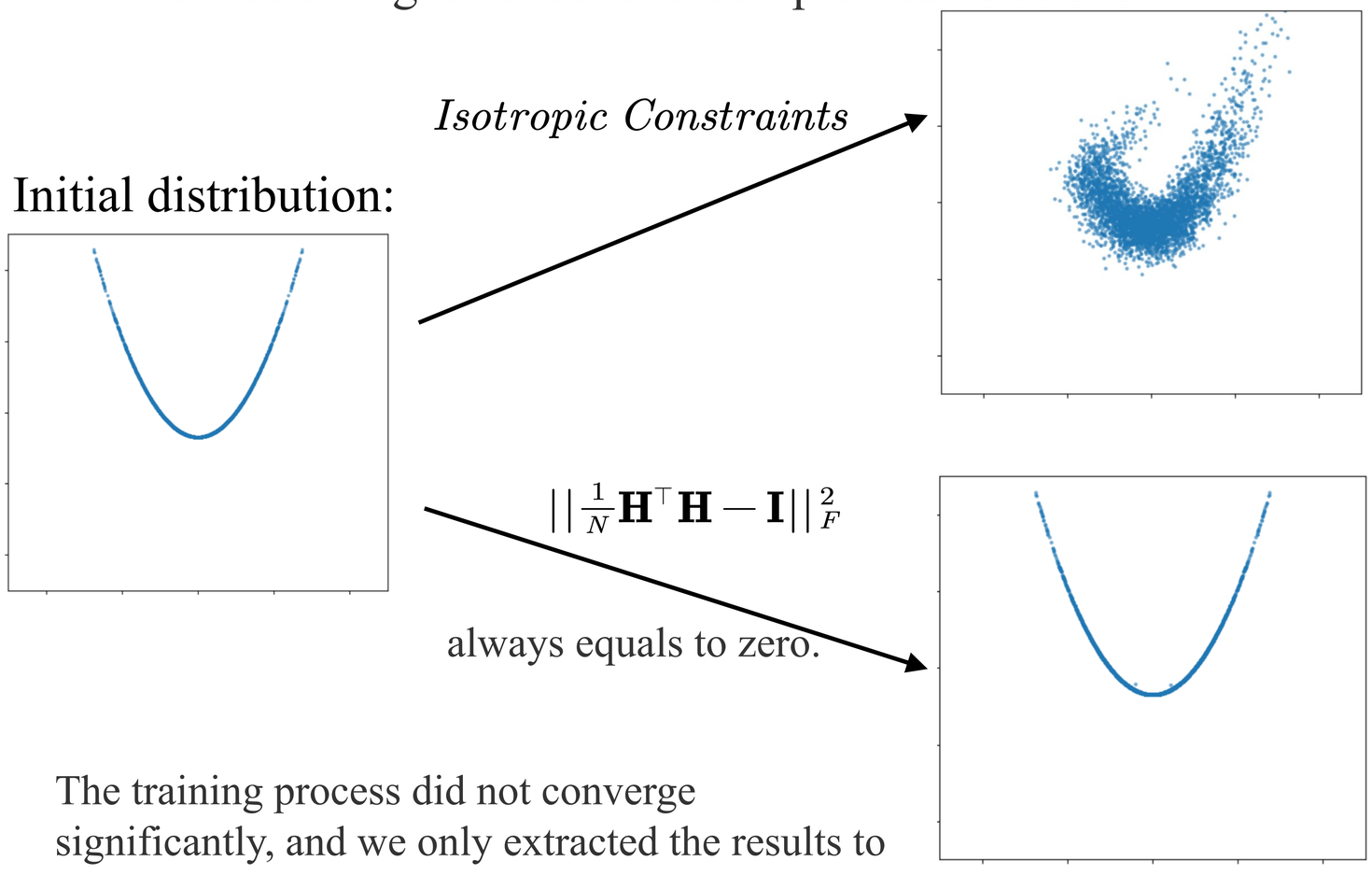
Compared to decorrelation-based strategies, our method has the following advantages:

(1) Minimizing $||\frac{1}{N}\mathbf{H}^\top\mathbf{H}-\mathbf{I}||_F^2$ minimizes Pearson correlation coefficients between each pair of dimensions, potentially requiring BatchNorm, whereas our method is not limited to specific normalization techniques.

Table 7: Performance of isotropic constraints under different normalization schemes. (These experimental results have been added to our revision.)

Dataset	Cora	Citeseer	Pubmed	Photo
Batch Norm	84.4 ± 0.5	73.5 ± 0.6	82.6 ± 0.3	93.19 ± 0.18
L2 Norm	84.3 ± 0.4	73.6 ± 0.5	82.4 ± 0.4	93.15 ± 0.20

(2) From a statistical perspective, $||\frac{1}{N}\mathbf{H}^\top\mathbf{H}-\mathbf{I}||_F^2$ can only remove Pearson linear correlations. It is powerless in the face of higher-order and complex correlations :



The training process did not converge significantly, and we only extracted the results to **highlight the differences**.

- For initial distribution, the covariance matrix of data is identity matrix $\mathbf{I}$.
- We set the sample points with 'requires_grad=True' and directly optimize them.
- From the results, it can be observed that our method exhibits a stronger capability in promoting representation diversity, because our method expects node representations to obey the same distribution along any space direction.
- This example also illustrates that eigenvalue distribution is not the optimal metric for measuring the utilization of the representation space.

In summary, our method and decorrelation-based approach investigate dimensional collapse issues and representation diversity from different perspectives. Researching the same problem from multiple angles and aspects undoubtedly contributes to the development of the self-supervised learning community.