Since several reviewers (1Kg3, bHxX, hxef) are concerned about the explanation for Observation 3.2 and Figure 2, we present a more detailed interpretation for them. The key to our interpretation is to consider the quality of the solution sets resulting from the push-away objectives (for negative pairs) in BCL and FCL: using only BCL only or only FCL would lead to a large number of redundant solutions, while combining BCL and FCL would effectively reduce this redundancy and never miss an optimal solution. We start with the special case of one negative pair (the example shown in Figure 2), followed by a general case of an arbitrary number of negative samples.

Suppose that $\mathbf{z}_1 \in \mathbb{R}^D$ and $\mathbf{z}_2 \in \mathbb{R}^D$ are representations of a negative pair residing on a hyper-sphere surface $\mathcal{S}^{D-1} \triangleq \{\mathbf{z} \in \mathbb{R}^D \mid ||\mathbf{z}||_2 = 1\}$. Then the push-away objective in BCL and FCL can be respectively formulated as

$$\min_{\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{S}^{D-1}} \underbrace{\exp(\mathbf{z}_1^\top \mathbf{z}_2)}_{\text{push-away term in BCL}} \triangleq \min_{\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{S}^{D-1}} \mathbf{z}_1^\top \mathbf{z}_2, \tag{1}$$

and

$$\min_{\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{S}^{D-1}} \underbrace{\sum_{i=1}^{D} \sum_{j \neq i} (C_{ij})^2}_{\text{push-away term in FCL}} \triangleq \min_{\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{S}^{D-1}} ||\text{off-diag}(\mathbf{z}_1 \mathbf{z}_1^\top + \mathbf{z}_2 \mathbf{z}_2^\top)||_F^2, \tag{2}$$

where $\text{off-diag}(\cdot)$ denotes a projection operator which preserves the off-diagonal elements of a matrix. It can be verified that the optimal solution to (1) and (2) is given by

$$\mathcal{Z}_{\text{B}}^\star = \{(\mathbf{z}_1, \mathbf{z}_2) \mid \mathbf{z}_1, \mathbf{z}_2 \in \mathcal{S}^{D-1}, \mathbf{z}_1 = -\mathbf{z}_2\} \tag{3}$$

and

$$\mathcal{Z}_{\text{F}}^\star = \{(\mathbf{z}_1, \mathbf{z}_2) \mid \mathbf{z}_1, \mathbf{z}_2 \in \mathcal{S}^{D-1}, \text{off-diag}(\mathbf{z}_1 \mathbf{z}_1^\top + \mathbf{z}_2 \mathbf{z}_2^\top) = \mathbf{0}\}, \tag{4}$$

respectively. In general, a solution in $\mathcal{Z}_{\text{B}}^\star$ (cf. the top right of Figure 2) does not admit a solution in $\mathcal{Z}_{\text{F}}^\star$ (cf. the bottom left of Figure 2). However, if $\mathbf{z}_1$ is a one-hot vector and $\mathbf{z}_2 = -\mathbf{z}_1$ (cf. the bottom right of Figure 2), subsituting them into (1) and (2) respectively yields the optimal value -1 and 0, implying that they exactly fall into the joint solution set $\mathcal{Z}_{\text{B}}^\star \cap \mathcal{Z}_{\text{F}}^\star \neq \emptyset$. The above analysis explains the derivation of Figure 2 in our paper.

Similarly, for the general case where N negative samples constitute a representation matrix $\mathbf{Z} \in \mathbb{R}^{N \times D}$, the push-away objective in BCL and FCL can be respectively formulated as

$$f_{\mathrm{B}}(\mathbf{Z}) \triangleq \mathbf{1}^{\top}\text{off-diag}(\mathbf{Z}\mathbf{Z}^{\top})\mathbf{1}, \tag{5}$$

and

$$f_{\mathrm{F}}(\mathbf{Z}) \triangleq ||\text{off-diag}(\mathbf{Z}^{\top}\mathbf{Z})||_F^2, \tag{6}$$

where $\mathbf{1} \in \mathbb{R}^N$ denotes a all-one vector. It is easy to verify that f_{B} and f_{F} are respectively invariant under the right rotation and the left rotation, i.e.,

$$f_{\mathrm{B}}(\mathbf{Z}\mathbf{R}_{\mathrm{B}}) = f_{\mathrm{B}}(\mathbf{Z}) \text{ and } f_{\mathrm{F}}(\mathbf{R}_{\mathrm{F}}\mathbf{Z}) = f_{\mathrm{F}}(\mathbf{Z}), \tag{7}$$

where $\mathbf{R}_{\mathrm{B}} \in \mathbb{R}^{D\times D}$ and $\mathbf{R}_{\mathrm{F}} \in \mathbb{R}^{N\times N}$ denote any rotation matrices. These rotation-invariances induce redundancy when using BCL only or FCL only. In contrast, combing BCL and FCL eliminates this redundancy since $f_{\mathrm{B+F}}(\mathbf{Z}) \triangleq f_{\mathrm{B}}(\mathbf{Z}) + f_{\mathrm{F}}(\mathbf{Z})$ is neither left-rotation invariant nor right-rotation invariant in general. On the other hand, one can construct an optimal solution to $f_{\mathrm{B+F}}$ that also admits the optimality of both f_{B} and f_{F}, so using $f_{\mathrm{B+F}}$ never miss the optimal solution.

In summary, combing the push-away objectives in BCL and FCL would reduce reduntand solutions but never miss an optimal solution, thus qualifying as a more reasonable regularization. These analyzes support our claims in Observation 3.2.