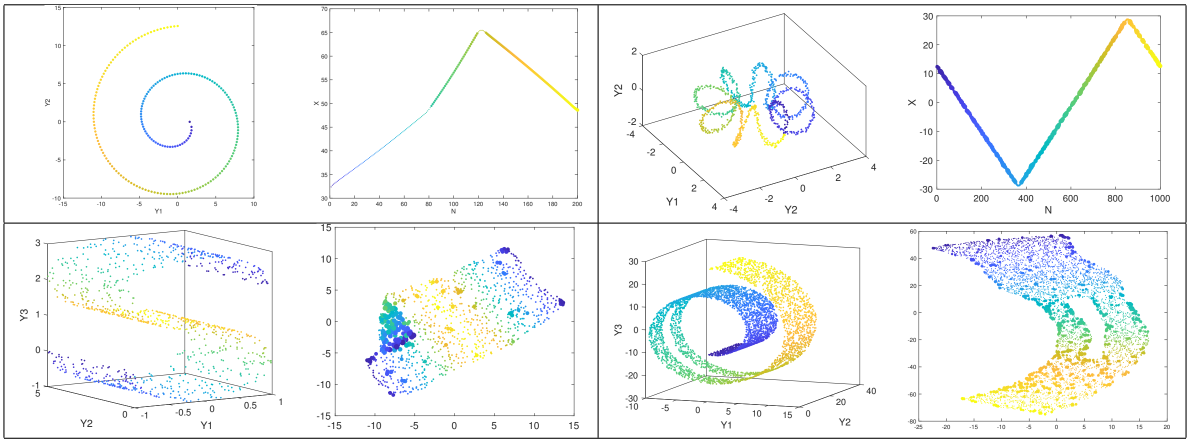

Figure 1: Pressured points for different datasets. For each of four datasets, left plot is the ground truth and the right plot is an embedding with marker size indicating pressure value. Color corresponds to the ground truth.

We thank the reviewers for the time. We are really glad that the reviewers have found that the paper provides a novel idea, is timely, is well written and motivated, and has extensive results.

Reviewer #1 *Results for UMAP*. Indeed, the objective function of UMAP is similar to t-SNE and can be written as

$$E_{\mathrm{UMAP}}(\mathbf{X}) = \sum_{i,j} (\log(1 + a \left\| \mathbf{x}_i - \mathbf{x}_j \right\|^{2b})) + \sum_{i,j} \log(1 - (1 + a \left\| \mathbf{x}_i - \mathbf{x}_j \right\|^{2b})^{-1}). \quad (1)$$

Similarly to other methods, this function also has the property of a single global minimum along a new dimension $\mathbf{Z}$ that could be found with a few iterations of Newton's method. We will make sure to update the paper with this information.

Robustness. We certainly hope that our approach would reduce the amount of "tsne engineering". We were hesitant to include any claims of robustness, since after all we are dealing with highly non-convex obj. fun. with many local minima. One would only hope to find a global solution. Our intuition does suggest that *all* of the local minima are produced by some points being pressured, however we were not yet able to prove it. In fig. 7 of the main paper, one can see that the variance of the final obj. fun. values of PP is smaller than the one from SD, however it is not exactly zero.

More aux dimensions. Mathematically, nothing prevents us from computing pressure points recursively one after another, up until all the points become non-pressured. Practically however, we would have to optimize the embedding separately for each dimension, which is costly. Our goal was to create a practical algorithm that could improve the results of existing methods, thus we have settled on increasing the dimensionality only by one.

Reviewer #4 *Comparison to other methods.* We do not propose a novel dimensionality reduction technique, but rather give insights and offer a novel optimization to the *existing* methods. Thus, the baseline should be given by the state-of-the-art optimization method (Spectral Direction), comparison to which we provide.

Results for Figure 6. We highlighted categories that differ the most from one embedding to another according to the Procrustes alignment error. The embedding for all these categories got improved (theoretically they could have gotten worse, but they did not). This is an important point and we will clarity it better in the paper.

Global minimum. The obj. fun. of the embedding methods is highly non-convex and finding a global minimum exactly is a very hard problem (see note on *Robustness* above). In fig. 5 we show the best possible results that we were able to get with a very careful and slow optimization. PP was able to get to a similar solution much faster.

Using higher-dimensional embeddings. The number of dimensions are often given as a hard constraint by the user. For example, one of the most typical application for the dimensionality reduction methods is the data visualization where the embedding dimensionality has to be equal two or three. For these cases, the goal is to find (potentially very lossy) embedding that would best represent the structure of the data. Finding the latent dimensionality is out of the scope of this paper (see also *More aux dimensions* above).

Interpretability. We discussed a typical scenario of the way pressured points arise in fig. 2 of the main paper and in the beginning of section 3. In addition, in fig. 3 we provided some examples of the pressure points for some synthetic dataset. As per reviewer suggestion, in fig. 1 above we include some additional examples of pressure points on synthetic data. Notice that the points become pressured when they are far from ground truth and are located "on top" of other points. In all the cases shown (except for the swiss roll with a hole), the original method (SNE) got stuck in a local minima. Our method was able to get out of it and achieve results that are almost identical to the ground truth.