Reviewer # **1**: Q1: The RSWL framework (7) does not seem to be well justified. Why squared loss for the regularization term instead of l1-norm or l2-norm?

A1: We would like to point out that if l1-norm or l2-norm is applied to regularization term without square loss, a trivial solution will be obtained due to the linearity of the weight, i.e., linear constraint (l1-ball) and linear index. Although nonlinear entropy term could also be applied to the regularization term, it would not bring the steerable sparsity as the proposed one does to filter out the ill ones.

Q2. The authors claim that RSWL is a general framework, yet they only applied it to PCA.

A2: It could also be applied to the graph-based $f = w_{ij}||W^T x_i - W^T x_j||_2^2$ to adaptively construct a sparse Laplacian and fuzzy k-means clustering $f = w_{ij}||x_i - c_j||_2^2$ to learn the sparse fuzzy membership, and etc.

Q3. Can the parameter k be chosen automatically, perhaps via some cross-validation methods?

A3: k serves as the key parameter here to filer out the ill ones, i.e., k active and N-k vanishing samples. In this paper, it is manually tuned to different ratio of the noised samples. It would be an interesting future direction to automatically learn hyper-parameter k.

Q4. How to choose the reduced dimension d?

A4: Reduced dimension, i.e., low rank is usually chosen between 3%-10% of raw dimension to represent the reconstruction quality, since the reconstruction approximates to the original one when reduced dimension increases.

Q5. In line 178, the authors mentioned that the error is quantified via a weighted loss where the weights p_i are learned by the algorithm. Is this fair? Because if it is fair, I imagine a naive method that assigns all weights (i.e. 1) to the data point with the smallest error would easily beat the proposed method.

A5: The comparison is conducted by normalized the weight for all the comparative methods. As for conventional PCA, weight for each term would be 1/N not 1. In other words, the sum of all the weights for the comparative methods is 1 for the fair comparison. Under this setting, smaller error does represent a better reconstruction.

Q6. For Figure 1, what is the baseline? **A**6: Baseline is directly applying k-means to the raw noised data.

Q7. If I understand it correctly, the y-value at 0.85 in figure 2a should match corresponding reconstruction error in Table 2 (which is 6.65). Have I missed something? **A**7: Thanks for the question. Figure 2 and table 2 are under different reduced dimensions.

Additionally, a more detailed literature review of robust PCA is added. We will release the code upon the final version.

Reviewer #**2**: * Was the the title of the paper meant to say "adaptive weights" instead of "adaptive neighbors"?

A: Our original meaning is that the adaptive weight vector has only k neighbors to represent the sparsity. We will fix that by developing a reasonable name.

* The spacing of the Conclusions sections seems to be smaller than 1; please fix for the final version.

A: Thanks so much for pointing out this. We will fix it as suggested.

* Lines 161-163: a very brief description of these other methods and how they differ from the proposed method would be appropriate to include here. **A**: A brief description with how they differ from the proposed method is added in lines 161-163.

Reviewer #**3**: Thus, it would be effective. However, the method would be a moderate extension from related paper [20].

A: We would like to clarify that this paper bears significant difference from reference [20] in the following sense. This paper proposes a general robust framework, whereas reference [20] promotes the robustness by exploiting the robust measure, i.e., capped l_1. This paper focuses on the general framework, whose core strategy could be applied to a variety of reconstruction functions to filter out the extremely noised samples. Novelties of methodology are also added to differentiate the proposed method from [20].

Reviewer #**4**: Although it is concise it seems that the idea is interesting and well treated and analyzed. On line 222 - Maybe the pivot point has to do with the polluted ratio being 20%? If so, doesn't that mean that the choice of the noise and the parameter k gave you algorithm a better result then it might have for different k or noise ratio?

A: Since k is the activated samples, it only relates to the polluted ratio when all noises seriously pollutes the 20% samples, which should be totally eliminated. However, as for the diverse polluted samples, i.e., some seriously polluted and rest are not, the performance and adaptivity of the proposed method lead to the better results besides the choice of the noise and the parameter k.

- How is this idea/framework relates to Core-Sets for PCA? **A**: The framework could filter out and eliminate the polluted samples, and preserve the well reconstructed ones via adaptive weight, while the conventional PCA is sensitive to the outliers. The framework is specifically proposed to enhance the robustness instead of utilizing the robust measures.

- On line 146 - p_i should be r_i? **A**: Thanks for the shrewd observation. We fix it to r_i.

- In the empirical reconstruction error test won?t the plain PCA be the best? **A**: Even in unpolluted data, robust PCA still leads to the better reconstruction, since it deals with data reconstruction pointwisely, i.e., assign each data point with different weight to evaluate the importance of each term. Diverse noises are added, such that sensitivity to outliers for each method can be separately evaluated.

- On line 216 - What do you mean by k_{rate}? Didn't you say (line 164) that $k = 0.85N$? **A**: k_{rate} is the ratio of choosing value of k from data number N. Varying parameter k_{rate} is shown in Figure 2 to illustrate the impact of different k to the reconstruction.