Supplementary material: Neuron matching in *C. elegans* with robust approximate linear regression without correspondence

Amin Nejatbakhsh, Erdem Varol
Department of Neuroscience, Department of Statistics, Zuckerman Institute
Grossman Center for the Statistics of Mind, Columbia University, New York, NY USA
mn2822@columbia.edu,ev2430@columbia.edu

1. Introduction and outline

In this supplementary we provide the proofs for the theorems in the main text and provide additional experiments to demonstrate the utility of our proposed algorithm in the presence of two types of outliers. The two types of outliers we consider are unstructured uniformly distributed outliers and structured adversarial outliers. In the former case, we demonstrate that the proposed algorithms as well as compared algorithms can handle higher than 50% outlier ratios. In the case of adversarial outliers, ie. outliers that structurally resemble the target point cloud, we demonstrate that there indeed is a ceiling of 50% ratio of outliers that no method can identifibility perform beyond.

The outline of the supplementary material is as follows. In section 2 we provide proofs for proposition 1 in the main text. In section 3 we provide qualitative and quantitative results on data with unstructured outliers. In section 4, we empirically demonstrate the limits of point set matching algorithms in the presence of adversarial outliers. In section 5 we demonstrate the phenomenon of adversarial outliers and the limits it poses to the regression without correspondence problem in a toy one-dimensional example.

2. Proofs of proposition 1

2.1. Noiseless case

Among the mn hypothetical regression coefficients obtained through all possible pairs of x_i and y_j, if a correct correspondence is encountered (i.e. $j = \pi^*(i)$), we have $y_{\pi(i)} = x_i\beta^*$ where β^* is the true coefficient. Therefore if we let $\beta^i = \frac{y_{\pi(i)}}{x_i}$ then $\beta^i = \beta^*$. Using this estimate, the distances of the remaining covariates regressed to their corresponding responses is

$$x_l\beta^i - y_{\pi(l)} = x_l\beta^* - \mathbf{y}_{\pi(l)} = 0$$

Therefore, when computing $\min_{\mathbf{\Pi}\in\mathcal{P}} \|\mathbf{x}\beta^i - \mathbf{\Pi y}\|_2^2$ via the Hungarian algorithm [2], each column of the distance matrix $[D]_{p,q} = |x_p\beta^i - y_q|$ corresponding to inlier points in $\mathbf{y}$ (i.e. $q \in \mathcal{I}^*$) will have at least one zero element. Thus, the optimal assignment $\mathbf{\Pi}^i$ will include all of the permutations $\pi^i(l) = l$ since they incur zero cost. Since there are $m-k$ of them by assumption 1, then $\sum_l \mathbf{1}(|x_l - y_{\pi^i(l)}| \leq \epsilon/2)) \geq m-k$. This is inequality because there might be additional outlier points that are by chance close to the regressed points.

Conversely, for a pair $(x_i, y_{\pi(k)})$ where $k \neq i$, we have the estimated coefficient $\beta^{i,k} = \frac{y_{\pi(k)}}{x_i} = \frac{x_k\beta^*}{x_i}$. The distances of the remaining covariates regressed with this estimate to their corresponding responses are

$$x_l\beta^{i,k} - y_{\pi(l)} = \frac{x_l x_k \beta^*}{x_i} - y_{\pi(l)} = x_l\beta^*\left(\frac{x_k}{x_i} - 1\right)$$

Therefore, without loss of generality, assuming $x_l \neq 0$ (if $x_l = 0$ the correspondence $(x_l, y_{\pi(l)})$ can be automatically inferred by choosing any $y_{\pi(l)} = 0$. If there aren't any $y_j = 0$, then this implies x_l is a point without correspondence in $\mathbf{y}$), we have

$$|x_l\beta^{i,k} - y_{\pi(l)}| \geq \epsilon$$

for some $\epsilon > 0$. ϵ can be explicitly stated as

$$\epsilon = \min_{i,l,k,\ i\neq k} x_l\beta^*\left(\frac{x_k}{x_i} - 1\right)$$

On the other hand,

$$x_i\beta^{i,k} - y_{\pi(k)} = 0$$

by construction.

Therefore, when computing $\min_{\mathbf{\Pi}\in\mathcal{P}} \|\mathbf{x}\beta^{i,k} - \mathbf{\Pi y}\|_2^2$ via Hungarian algorithm, there will less than $m-k$ assignments in the optimal assignment $\mathbf{\Pi}^{i,k}$ such that $|x_l - y_{\pi^{i,k}(l)}| \leq \epsilon/2$. Otherwise, this would imply the coefficient $\beta^{i,k}$ is a coefficient that explains the inliers, which by assumption 1 cannot be the case. Thus, $\sum_l \mathbf{1}(|x_l - y_{\pi^{i,k}(l)}| \leq \epsilon/2)) < m-k$.

This shows that the maximal cardinality of a hypothetical inlier set is at least $m - k$, and it is only achieved for a coefficient that is obtained by a correct correspondence pair. This is sufficient to show that algorithm 1 recovers the true coefficient B^* under the noiseless regime.

2.2. Noisy case

Let the noise model of the inlier regression be $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Therefore, if a correct correspondence is encountered, we have $y_{\pi(i)} = x_i\beta^* + \epsilon$ where β^* is the true coefficient. The coefficient estimated from this pairing is $\beta^i = \frac{y_\pi(i)}{x_i} = \beta^* + \frac{\epsilon}{x_i}$. When this coefficient is applied to $\mathbf{x}$ we see that

$$\mathrm{E}(x_l\beta^i - y_{\pi(l)}) = 0,\ \mathrm{Var}(x_l\beta^i - y_{\pi(l)}) = \left(\frac{x_l^2}{x_i^2} + 1\right)\sigma^2$$
$$\mathrm{E}(x_l\beta^i - y_{\pi(k)}) = (x_l - x_k)\beta^*$$
$$\mathrm{Var}(x_l\beta^i - y_{\pi(k)}) = \left(\frac{x_l^2}{x_i^2} + 1\right)\sigma^2$$

Therefore, if σ^2 is small (i.e. in the SNR regime of [4]), we have $|x_l\beta^i - y_{\pi(l)}| < |x_l\beta^i - y_{\pi(k)}|$ for $l \neq k$ with high probability. Thus the row-wise minimal cost assignment in the Hungarian algorithm will be $\pi^i(l) = l$ with high probability. However, even if $\pi^i(l) \neq l$, if we set margin ν such that $\nu = \frac{1}{2}\min_{l,k\ l\neq k}|(x_l - x_k)\beta^*|$, with high probability we will have that

$$\sum_l \mathbf{1}(|x_l\beta^i - y_{\pi^i(l)}| \leq \nu) \geq \sum_l \mathbf{1}(|x_l\beta^{i,k} - y_{\pi^i(l)}| \leq \nu) \tag{1}$$

where $\beta^{i,k}$ denotes the regression coefficient obtained via incorrect correspondence $\beta^{i,k} = \frac{y_{pi(k)}}{x_i}$. Therefore, if σ^2 is sufficiently small, with high probability, algorithm 1 recovers the coefficient $\beta^i = \beta^* + \frac{\epsilon}{x_i}$ for some $i \in \mathcal{I}$ where $\mathcal{I}$ denotes the set of inliers.

3. Unstructured outlier experiments

In this section, we describe experiments done on the `fish` point cloud dataset to supplement the results in the main text. For this experiment, the 3D `fish` point cloud was sampled at varying levels from 40 upto 100 points as the moving point cloud ($\boldsymbol{x}$). The fixed point cloud ($\boldsymbol{y}$) was sampled at the similar level but was missing 10% of correspondences to the moving point cloud. Furthermore, 30% uniformly distributed outliers was introduced to the fixed point cloud. For all experiments, we introduced three different levels of difficulty in terms of initialization. For the simplest case, the moving point cloud was randomly rotated and scaled 1:1 as the size of the fixed point cloud. We then increased the scale of the moving point cloud to be 5:1 and 10:1 relative to the fixed point cloud. Lastly, 10% noise was introduced relative to the scale of the point clouds. The qualitative and quantitative results of these experiments can be found in figure 1.

We compared the accuracy and convergence time of our proposed algorithm, rRWOC with the other algorithms ICP[1], CPD[3], and homomorphic sensing (HS) [5], the latter of which is considered to be the state of the art algorithm for regression without correspondence.

In this experiment, rRWOC performed similarly as HS in terms of accuracy and convergence time. Neither method was sensitive to the scale of initialization since they are both global methods. rWOC seem to converge faster with higher number of points since with unstructured outliers, once a good inlier set is reached, it is unlikely to confuse it with another inlier set comprising of outliers.

4. Adversarial outlier experiments

In this section, we introduced 30% adversarial outliers instead of unstructured or uniform outliers to the data. The adversarial outliers were drawn as a duplicate subset of the `fish` dataset, partially resembling the intended target of the matching problem. The remaining set up and initalization was same as what was done in section 3. The qualitative and quantitative results for this experiment can be seen in figure 2.

In this experiment, we can see that while rRWOC and HS are not sensitive to the number of points or the initialization scale in terms of accuracy, rRWOC has a higher overall accuracy than HS. In terms of convergence speed, rRWOC tends to converge slower than HS at higher number of points but faster at lower number of points. The reason why rRWOC has higher accuracy in this scenario than HS is that the search space for HS is in the space of transformations whereas rRWOC looks for maximal inlier sets. The former search space is prone to error modes where the target point cloud and adversarial point cloud span structured subspaces. rWOC is not prone to such an error model, although computational cost might be higher.

5. One-dimensional algorithm toy example

In this section we demonstrate the adversarial outlier phenomenon in the one-dimensional case using algorithm 2. We show that even in a simple one-dimensional case without noise, if the outlier set has structure and exceeds the cardinality of the inlier set (thereby constituting a more than 50% outlier ratio), we cannot recover the true inliers under any circumstance. The results of these toy examples can be seen in figures 3 and 4.

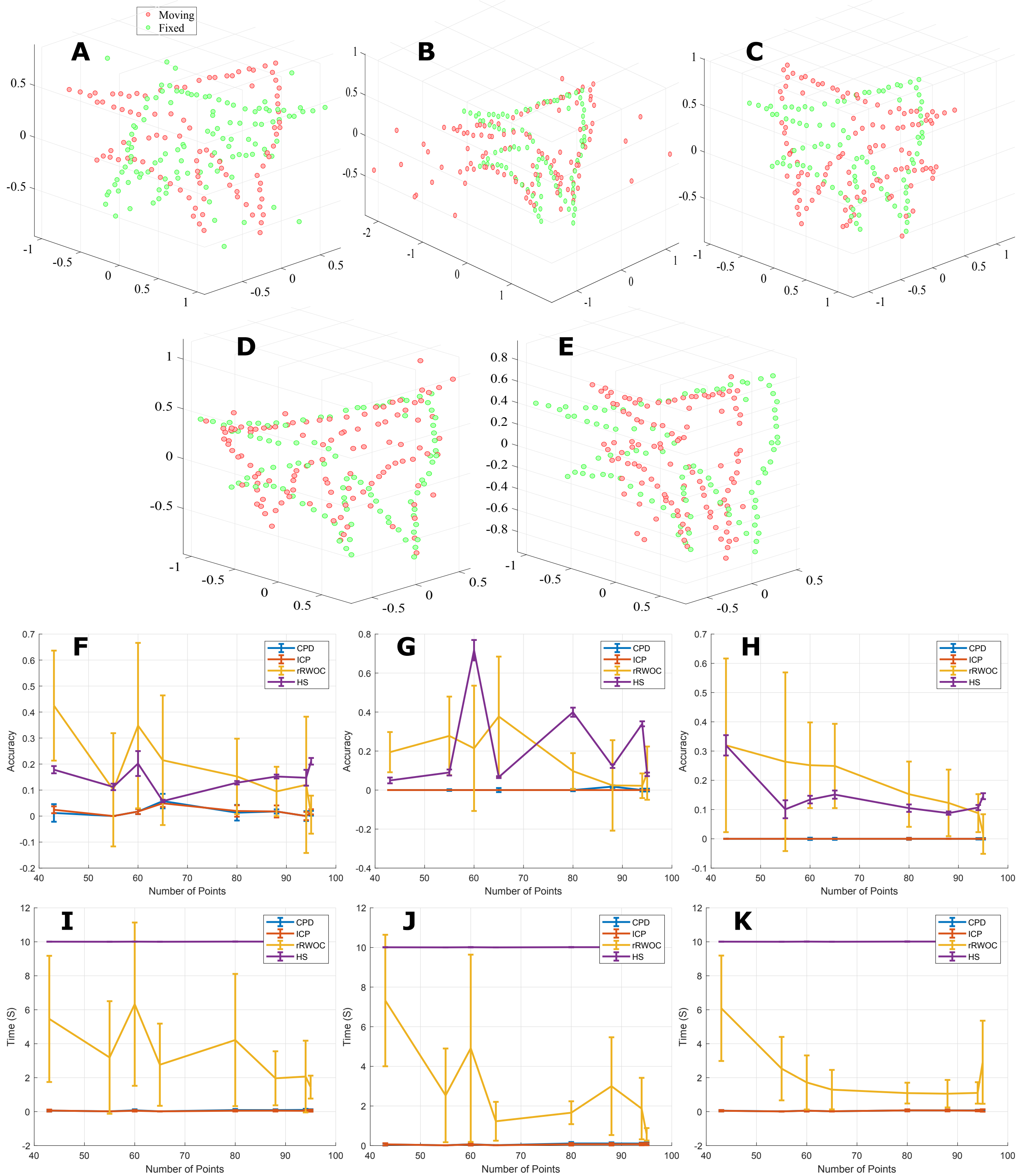

Figure 1. **Unstructured outlier scenario results**. **A:** An instance of the moving ($\boldsymbol{x}$) and fixed ($\boldsymbol{y}$) point clouds such that the fixed point cloud is corrupted by adversarial outliers, i.e. outliers that mimic the structure of the intended target point cloud. **B:** rRWOC results, **C:** ICP[1] results, **D:** CPD[3] results, **E:** Homomorphic sensing (HS) [5] results. Note that these are single instances of qualitative results, the quantitative results were computed using multiple random instantiations. **F,G,H**: The point cloud matching accuracies of compared methods for increasing scale number of points sampled for each point cloud.**F:** The ratio of the scale of the fixed point cloud to the target was 1:1. **G:** Ratio 1:5, **H:** Ratio 1:10. **I,J,K:** Convergence timing of compared methods at with initialization scales of 1:1 (**I**), 1:5 (**J**), 1:10 (**K**). Note the rRWOC and HS exhibit similar point cloud matching accuracies and computational costs in this scenario.

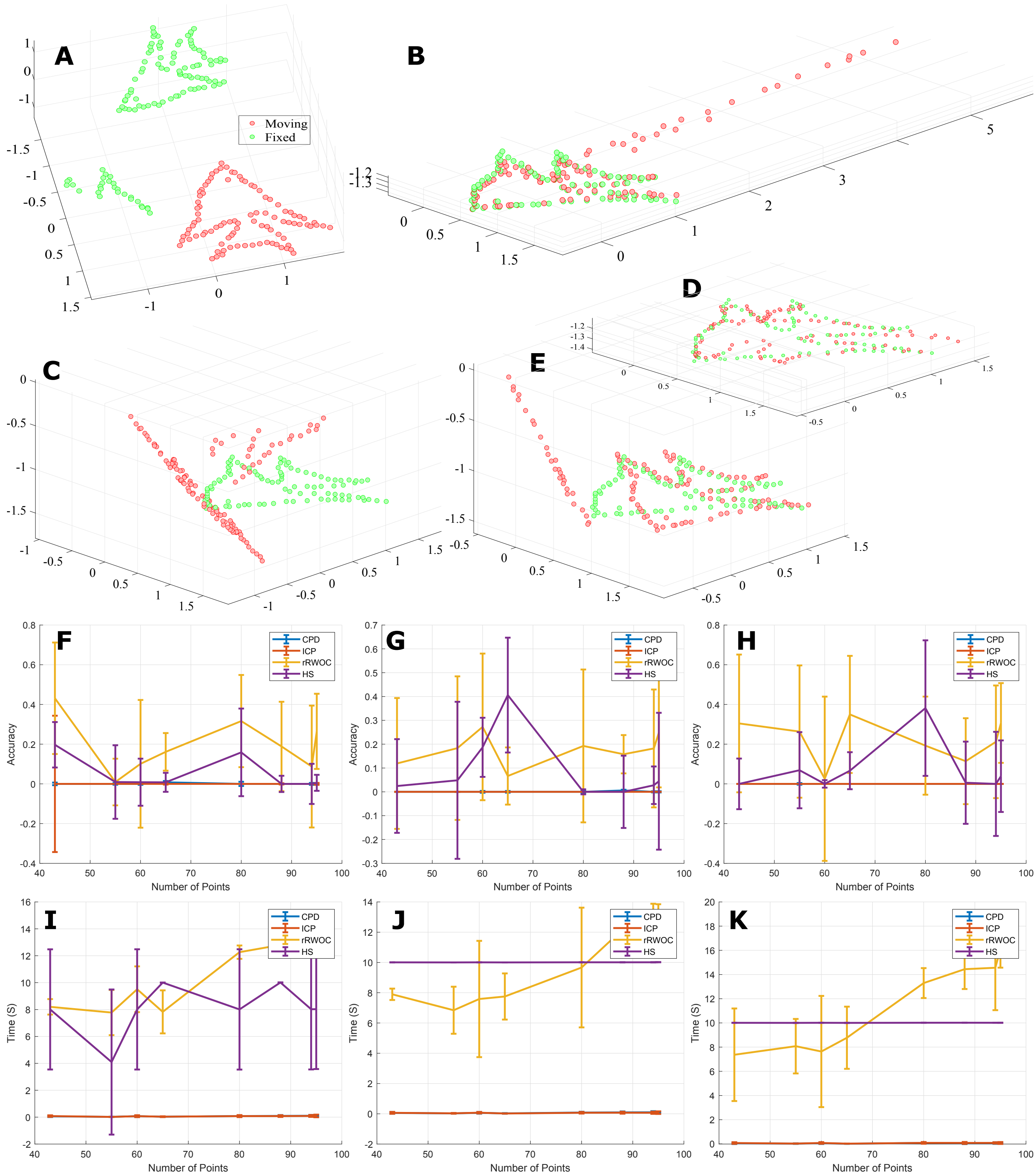

Figure 2. **Adversarial outlier scenario results**. **A:** An instance of the moving ($\boldsymbol{x}$) and fixed ($\boldsymbol{y}$) point clouds such that the fixed point cloud is corrupted by adversarial outliers, i.e. outliers that mimic the structure of the intended target point cloud. **B:** rRWOC results, **C:** ICP[1] results, **D:** CPD[3] results, **E:** Homomorphic sensing (HS) [5] results. Note that these are single instances of qualitative results, the quantitative results were computed using multiple random instantiations. **F,G,H**: The point cloud matching accuracies of compared methods for increasing scale number of points sampled for each point cloud.**F:** The ratio of the scale of the fixed point cloud to the target was 1:1. **G:** Ratio 1:5, **H:** Ratio 1:10. **I,J,K:** Convergence timing of compared methods at with initialization scales of 1:1 (**I**), 1:5 (**J**), 1:10 (**K**). Note the rRWOC maintains exhibits slighthy higher acccuracy than HS while retaining similar computation cost.

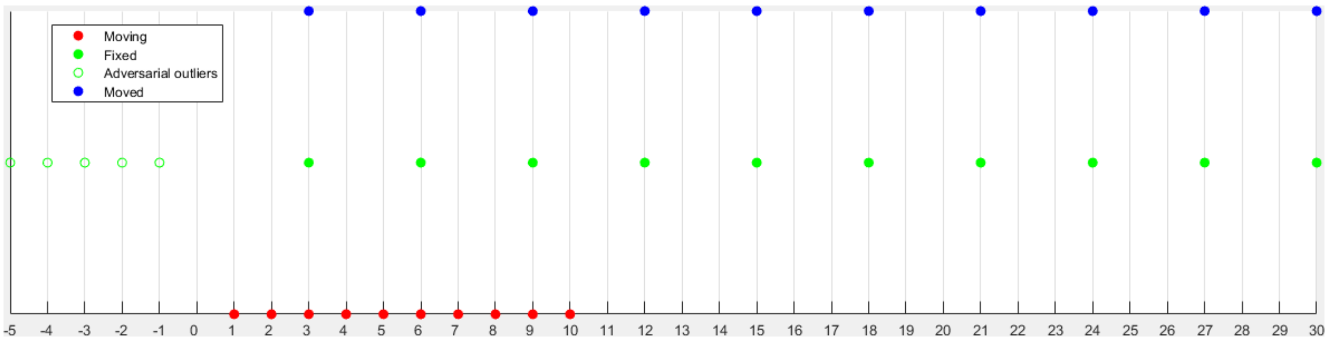

Figure 3. **One dimensional toy example of adversarial outliers**. Red points indicate the moving point cloud ($\boldsymbol{x}$) of the elements $1, 2, \ldots, 10$. The solid green points indicate the fixed point cloud that is a transformation of the moving point cloud: $\boldsymbol{y}_i = 3\boldsymbol{x}$. The hollow green points indicate adversarial points such that $\boldsymbol{y}_o = (-1) \times \{1, 2, \ldots, 5\}$. Using the proposed algorithm recovers the true $\boldsymbol{\beta} = 3$ and yields the moved point cloud as the blue dots.

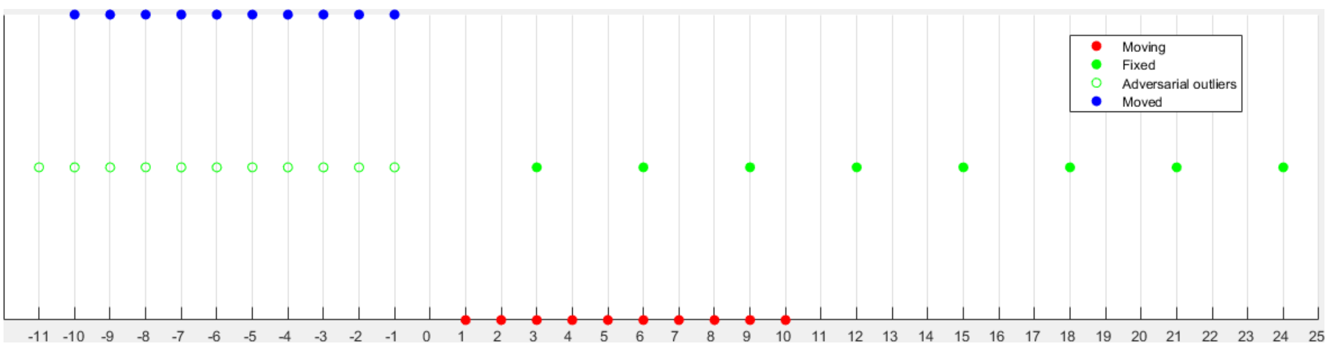

Figure 4. **One dimensional toy example of adversarial outliers — failure mode when there is more than** 50% **outliers**. In this example, the number of adversarial outliers exceed the intended number of inliers and no model can theoretically recover the intended inliers due to lack of identifiability. Here the inlier set comprising of 8 points is generated by the rule $\boldsymbol{y}_i = 3\boldsymbol{x}$ but the outlier set is generated as $\boldsymbol{y}_o = (-1) \times \{1, 2, \ldots, 11\}$. Since the putative $\beta = -1$ yields a higher tentative inlier set, the model returns the wrong regression coefficient and the wrong point cloud transformation.

References

[1] Dmitry Chetverikov, Dmitry Svirko, Dmitry Stepanov, and Pavel Krsek. The trimmed iterative closest point algorithm. In *Object recognition supported by user interaction for service robots*, volume 3, pages 545–548. IEEE, 2002.

[2] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[3] Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2262–2275, 2010.

[4] Ashwin Pananjady, Martin J Wainwright, and Thomas A Courtade. Linear regression with an unknown permutation: Statistical and computational limits. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 417–424. IEEE, 2016.

[5] Manolis Tsakiris and Liangzu Peng. Homomorphic sensing. In *International Conference on Machine Learning*, pages 6335–6344, 2019.