# A Appendix

## A.1 Result on Real world datasets

### A.1.1 Results on Shuttle dataset

Shuttle training data set from UCI Machine Learning Repository [22] contains 43,500 points. We run robust $k$-means++ on the Shuttle dataset with $k \in \{5, 10, 15\}$, and $\delta = \{0.05, 0.1\}$. We compare its performance with vanilla $k$-means++ and random initialization. We summarise our empirical findings in Tables 12, 13, 14, 15.

**Insight.** We notice that in almost every scenario robust $k$-means++ outperforms random initialization. On the comparison with $k$-means++ our algorithm gives comparable/better cost for all values of $k$. Specifically, our average cost is better than $k$-means++ on most of the instances. Our running time is slower albeit comparable to that of $k$-means++.

## A.2 Result on Synthetic datasets

We recall the dataset generation step as follows. We pick $k + z$ uniformly random points from a large $d$-dimensional hypercube of side length $s = 100$. We use $k$ points from them as means and pick $n/k$ points around each of them from a random Gaussian of unit variance. This gives a data set of $n + z$ points with $n$ points clustered into $k$ clusters and the remaining $z$ as outliers. We first run robust $k$-means++ with the values of $\alpha \in \{0, 0.25, 0.5, 1\}$, $\delta \in \{0.05, 0.1\}$ on the synthetic datasets with values $n = 1000, d = 2, k = 20$, and the number of outliers $\{25, 50, 120\}$. We summarised our results in Tables 16,17,18.

**Insight.** We notice that in almost every scenario robust $k$-means++ outperforms random initialization. On the comparison with $k$-means++ our algorithm offers better performance in all the metric – min, mean, med cost – on all possible values of $k, \alpha, \delta$. Our running time is slower albeit comparable to that of $k$-means++ and TKM + +.

| Result | $\alpha$ | Cost Min | Cost Mean | Cost Med. | Time(s) |
|---|---|---|---|---|---|
| This work $\delta$=0.05 | 0 | 1.86 | 2.1 | 2.02 | 9.5 |
| | 0.25 | 1.90 | 2.08 | 2.02 | 9.5 |
| | 0.5 | 1.85 | 2.06 | 2.02 | 9.5 |
| | 0.75 | 1.69 | **2.02** | **2.00** | 9.5 |
| | 1 | 2.46 | 2.86 | 2.60 | 9.5 |
| This work $\delta$=0.1 | 0 | 1.87 | 2.08 | 2.02 | 5.3 |
| | 0.25 | 1.85 | 2.04 | 2.01 | 5.3 |
| | 0.5 | 1.85 | 2.07 | 2.01 | 5.3 |
| | 0.75 | **1.66** | 2.08 | 2.03 | 5.3 |
| | 1 | 2.24 | 2.67 | 2.60 | 5.3 |
| TKM + + | | 1.7 | 2.26 | 2.23 | 3.42 |
| KM + + | | 1.8 | 2.4 | 2.36 | 1.6 |
| RAND | | 2.6 | 2.9 | 2.61 | 0.74 |

Table 12: Robust $k$-means++ on Shuttle dataset with $k = 5$. We delete the farthest 21 points. All cost are multiplicative of $10^8$.

| Result | $\alpha$ | Cost Min | Cost Mean | Cost Med. | Time(s) |
|---|---|---|---|---|---|
| This work $\delta$=0.05 | 0 | 8.1 | 8.87 | 8.92 | 22.4 |
| | 0.25 | 7.89 | 8.79 | 8.67 | 22.4 |
| | 0.5 | 6.99 | 8.58 | 8.76 | 22.4 |
| | 0.75 | **6.67** | 8.56 | 8.56 | 22.4 |
| | 1 | 9.47 | 12.4 | 12.6 | 22.4 |
| This work $\delta$=0.1 | 0 | 7.73 | 8.89 | 8.69 | 11.5 |
| | 0.25 | 6.8 | 8.86 | 9.11 | 11.5 |
| | 0.5 | 6.98 | 8.54 | 8.58 | 11.5 |
| | 0.75 | 6.95 | **8.31** | **8.11** | 11.5 |
| | 1 | 9.72 | 12 | 12.6 | 11.5 |
| TKM + + | | 7.0 | 8.8 | 8.9 | 7.35 |
| KM + + | | 7.12 | 8.94 | 9.08 | 3.5 |
| RAND | | 9.66 | 11.6 | 11.3 | 1.37 |

Table 13: Robust $k$-means++ on Shuttle dataset with $k = 10$. We delete the farthest 34 as outliers. All cost are multiplicative of $10^7$.

| Result | $\alpha$ | Cost | | | Time(s) |
|---|---|---|---|---|---|
| | | Min | Mean | Med. | |
| This work | 0 | 6.2 | **7.2** | **7.15** | 37.5 |
| $\delta$=0.05 | 0.25 | 6.6 | 7.43 | 7.19 | 37.5 |
| | 0.5 | 6.05 | 7.26 | 7.18 | 37.5 |
| | 0.75 | 6.16 | 7.74 | 7.71 | 37.5 |
| | 1 | 12.6 | 17.9 | 18.3 | 37.5 |
| This work | 0 | 6.24 | 7.54 | 7.37 | 19.5 |
| $\delta$=0.1 | 0.25 | 6.34 | 7.22 | 7.14 | 19.5 |
| | 0.5 | 6.37 | 7.68 | 7.69 | 19.5 |
| | 0.75 | **5.82** | 7.27 | 7.32 | 19.5 |
| | 1 | 12.2 | 19.4 | 19.2 | 19.5 |
| TKM + + | | 6.1 | 7.7 | 7.9 | 12.2 |
| KM + + | | 6.78 | 8.22 | 8.01 | 5.4 |
| RAND | | 14.1 | 16.3 | 15.1 | 1.98 |

Table 14: Robust $k$-means++ on `Shuttle` dataset with $k = 15$. We delete the farthest 17 as outliers. All cost are multiplicative of $10^7$.

| Result | $\alpha$ | Cost | | | Time(s) |
|---|---|---|---|---|---|
| | | Min | Mean | Med. | |
| This work | 0 | 4.2 | 4.94 | 4.91 | 37.5 |
| $\delta$=0.05 | 0.25 | 4.2 | 5 | 5.03 | 37.5 |
| | 0.5 | 4.41 | 5.04 | 4.84 | 37.5 |
| | 0.75 | 4.21 | 4.99 | 4.94 | 37.5 |
| | 1 | 3.94 | 4.98 | 4.93 | 37.5 |
| This work | 0 | 4.13 | 4.96 | 5.01 | 19.5 |
| $\delta$=0.1 | 0.25 | 4.54 | 5.09 | 5.07 | 19.5 |
| | 0.5 | 4.29 | 4.92 | 4.90 | 19.5 |
| | 0.75 | 4.13 | **4.56** | **4.57** | 19.5 |
| | 1 | 4.22 | 5.28 | 5.37 | 19.5 |
| TKM + + | | 3.6 | 4.8 | 4.8 | 11.2 |
| KM + + | | 3.89 | 4.58 | **4.57** | 5.4 |
| RAND | | **3.10** | 4.91 | 5.34 | 2.02 |

Table 15: Robust $k$-means++ on `Shuttle` dataset with $k = 15$. We delete the farthest 51 as outliers. All cost are multiplicative of $10^7$.

| Result | $\alpha$ | Cost | | | Time(s) |
|---|---|---|---|---|---|
| | | Min | Mean | Med. | |
| This work | 0 | 2035 | 2816 | 2733 | 12 |
| $\delta$=0.05 | 0.25 | 2719 | 3099 | 3272 | 12 |
| | 0.5 | 2036 | 2766 | 2581 | 12 |
| | 0.75 | 2543 | 3101 | 2927 | 12 |
| | 1 | 1996 | 2297 | 2298 | 12 |
| This work | 0 | 2547 | 2834 | 2666 | 6.5 |
| $\delta$=0.1 | 0.25 | 2031 | 2737 | 2577 | 6.5 |
| | 0.5 | 2542 | 2959 | 3072 | 6.5 |
| | 0.75 | 2032 | 2408 | 2297 | 6.5 |
| | 1 | **1907** | **2069** | **2037** | 6.5 |
| TKM + + | | 2261 | 2340 | 3415 | 0.52 |
| KM + + | | 2561 | 4840 | 4145 | 0.2 |
| RAND | | 5443 | 17977 | 13753 | 0.07 |

Table 16: Robust $k$-means++ on Synthetic dataset with $\delta = 0.1$. $n = 1000, d = 2, k = 20$. We delete the farthest 25 as outliers.

| Result | $\alpha$ | Cost | | | Time(s) |
|---|---|---|---|---|---|
| | | Min | Mean | Med. | |
| This work | 0 | 1880 | 1907 | 1912 | 12 |
| $\delta$=0.05 | 0.25 | 1890 | 1908 | 1906 | 12 |
| | 0.5 | 1884 | 1909 | 1904 | 12 |
| | 0.75 | 1892 | 1921 | 1908 | 12 |
| | 1 | 1753 | 1879 | 1893 | 12 |
| This work | 0 | 1882 | 1927 | 1914 | 6.5 |
| $\delta$=0.1 | 0.25 | 1892 | 1908 | 1906 | 6.5 |
| | 0.5 | 1891 | 1902 | 1896 | 6.5 |
| | 0.75 | 1883 | 1905 | 1899 | 6.5 |
| | 1 | **1746** | **1855** | **1887** | 6.5 |
| TKM + + | | 1981 | 2113 | 2211 | 0.48 |
| KM + + | | 1881 | 6113 | 5011 | 0.2 |
| RAND | | 8939 | 28696 | 27168 | 0.07 |

Table 17: Robust $k$-means++ on Synthetic dataset with $\delta = 0.1$. $n = 1000, d = 2, k = 20$. We delete the farthest 50 as outliers.

| Result | $\alpha$ | Cost | | | Time(s) |
|---|---|---|---|---|---|
| | | Min | Mean | Med. | |
| This work | 0 | 3313 | 5135 | 5010 | 12 |
| $\delta$=0.05 | 0.25 | 3283 | 4572 | 4480 | 12 |
| | 0.5 | 4249 | 4808 | 4649 | 12 |
| | 0.75 | 2514 | 4647 | 4449 | 12 |
| | 1 | 2016 | 2780 | 2592 | 12 |
| This work | 0 | 3538 | 4975 | 4673 | 6.5 |
| $\delta$=0.1 | 0.25 | 3306 | 4064 | 4178 | 6.5 |
| | 0.5 | 3284 | 4293 | 4146 | 6.5 |
| | 0.75 | 2737 | 3686 | 3554 | 6.5 |
| | 1 | **1779** | **2656** | **2657** | 6.5 |
| TKM + + | | 2011 | 5272 | 3053 | 0.52 |
| KM + + | | 2188 | 6272 | 6053 | 0.2 |
| RAND | | 7428 | 22902 | 21449 | 0.07 |

Table 18: Robust $k$-means++ on Synthetic dataset with $\delta = 0.1$. $n = 1000, d = 2, k = 20$. We delete the farthest 120 as outliers.