Methods	Datasets							
	cal-housing		abalone		pumadyn		bank-8fh	
L2	1185	(124.59)	7.93	(0.67)	1.24	(0.42)	18.21	(6.57)
L1	1303	(244.85)	7.30	(0.40)	1.29	(0.42)	6.54	(3.09)
Huber	1221	(119.18)	7.73	(0.49)	1.24	(0.42)	7.37	(3.18)
LTS	533	(398.92)	755.1	(126)	0.32	(0.41)	10.96	(6.67)
GemMc	28	(88.45)	2.30	(0.01)	0.12	(0.12)	0.93	(0.80)
[9]	967	(522.40)	8.39	(0.54)	0.81	(0.77)	3.91	(6.18)
AltBndL2	967	(522.40)	8.39	(0.54)	0.81	(0.77)	7.74	(9.40)
AltBndL1	1005	(603.00)	7.30	(0.40)	1.29	(0.42)	1.61	(2.51)
CvxBndL2	9	(0.64)	7.60	(0.86)	0.07	(0.07)	0.20	(0.05)
CvxBndL1	8	(0.28)	2.98	(0.08)	0.08	(0.07)	0.10	(0.07)
Gap(Cvx2)	2e-12	(3e-12)	3e-9	(4e-9)	0.025	(0.052)	0.001	(0.003)
Gap(Cvx1)	0.005	(0.01)	0.001	(0.001)	0.267	(0.269)	0.011	(0.028)

Table 2: RMSE on clean test data for 108 training data points and 1000 test data points, with 10 repeats. Standard deviations shown parentheses. The mean gap values of CvxBndL2 and CvxBndL1, Gap(Cvx2) and Gap(Cvx1) respectively, are given in the last two rows.

alternating can be trapped in poor local minima. The proposal from [9] was not effective in this setting (which differed from the one investigated there).

Next, we conducted an experiment on four real datasets taken from the StatLib repository[9] and DELVE.[10] For each data set, we randomly selected 108 points as the training set, and another random 1000 points as the test set. Here the regularization constant is tuned by 10-fold cross validation. To seed outliers, 5% of the training set are randomly chosen and their X and y values are multiplied by 100 and 10000, respectively. All of these data sets have 8 features, except pumadyn which has 32 features. We also estimated the scale factor on the training set by the mean absolute deviation method, a common method in robust statistics [3]. Again, the ideal parameter $n' = (1 - 5\%)n$ is granted to LTS and 30 random restarts are performed.

The RMSE on test set for all methods are reported in Table 2. It is clear that all methods based on convex losses (L2, L1, Huber) suffer significantly from the added outliers. The method proposed in this paper consistently outperform all other methods with a noticeable margin, except on the abalone data set where GemMc performs slightly better.[11] Again, we observe evidence that the alternating strategy can be trapped in poor local minima, while the method from [9] was less effective. We also measured the relative optimality gaps for the approximate CvxBnd procedures. The gaps were quite small in most cases (the gaps were very close to zero in the synthetic case, and so are not shown), demonstrating the tightness of the proposed approximation scheme.

7 Conclusion

We have developed a new robust regression method that can guarantee a form of robustness (bounded response) while ensuring tractability (polynomial run-time). The estimator has been proved consistent under some restrictive but non-trivial conditions, although we have not established general consistency. Nevertheless, an empirical evaluation reveals that the method meets or surpasses the generalization ability of state-of-the-art robust regression methods in experimental studies. Although the method is more computationally involved than standard approaches, it achieves reasonable scalability in real problems. We are investigating whether the proposed estimator achieves stronger robustness properties, such as high breakdown or bounded influence. It would be interesting to extend the approach to also estimate scale in a robust and tractable manner. Finally, we continue to investigate whether other techniques from the robust statistics and machine learning literatures can be incorporated in the general framework while preserving desired properties.

Acknowledgements

Research supported by AICML and NSERC.

[9] http://lib.stat.cmu.edu/datasets/

[10] http://www.cs.utoronto.ca/ delve/data/summaryTable.html

[11] Note that we obtain different results than [9] arising from a very different outlier process.