

Self Scaled Regularized Robust Regression

Y. Wang, C. Dicle, M. Sznaier, O. Camps

Electrical and Computer Engineering, Northeastern University, Boston, MA 02115.

Given N data points $\mathbf{x}_i \in \mathbb{R}^d$ and corresponding scalars y_i , $i = 1, \dots, N$, linear robust regression seeks to find the parameter \mathbf{r} of a linear mapping from \mathbf{x}_i to y_i such that the number of inliers (i.e. pairs of points where the fitting error of the model is less than a given bound) is maximized, that is:

$$\max_{\mathbf{r}} \text{cardinality}(\mathcal{S}_i(\mathbf{r})) \text{ where} \quad (1)$$

$$\mathcal{S}_i(\mathbf{r}) \doteq \{\mathbf{x}_i: |y_i - \mathbf{x}_i^T \mathbf{r}| \leq \varepsilon\}$$

While this problem is known to be NP hard, several tractable relaxations exist. Perhaps the most commonly used class of algorithms belong to the RANSAC family [1], that relies on random sampling to separate the inliers from the outliers, based on fitting error. While these methods work well in scenarios with relatively low number of outliers, the number of samples needed to guarantee a given probability of success rapidly increases with the percentage of outliers, while performance degrades. Alternative approaches exploit recent advances in compressive sensing, by reformulating robust linear regression as an optimization problem with sparse regularizations [2]. The advantage of these methods is that they admit convex relaxations which can be solved using polynomial-time algorithms. Further, conditions guaranteeing exact recovery of the parameters of the model are available. However, these relaxations may perform poorly in cases where the fitting error for the outliers is large. In addition, these approaches cannot exploit available *a-priori* information, such as co-occurrences.

To circumvent these difficulties, in this paper we present an alternative approach to robust regression. Our main result shows that this approach is equivalent to a "self-scaled" ℓ_1 regularized robust regression problem, where the cost function is automatically scaled, with scalings that depend on the *a-priori* information, motivating the name *Self-Scaled Regularized Robust Regression* (S^2R^3). The starting point is to reformulate (1) as:

$$\mathbf{r}^* = \underset{\mathbf{r}, \mathbf{r}_i}{\text{argmin}} \|\{\mathbf{r} - \mathbf{r}_i\}\|_o \text{ subject to:} \quad (2)$$

$$|y_i - \mathbf{x}_i^T \mathbf{r}_i| \leq \varepsilon, \quad i = 1, \dots, N$$

where $\|\{\mathbf{r} - \mathbf{r}_i\}\|_o$ denotes the number of non-zero elements of the vector sequence $\{\mathbf{r} - \mathbf{r}_i\}_{i=1}^N$. Using the fact that the convex envelope of the cardinality of a vector sequence $\{\mathbf{v}_i\}$ is given by [3] $\|\{\mathbf{v}\}\|_{0,env} = \sum_i \|\mathbf{v}_i\|_\infty$ allows for relaxing (2) to:

$$\min_{\mathbf{r}, \mathbf{r}_i} \sum_{i=1}^N \|\mathbf{r} - \mathbf{r}_i\|_\infty \text{ subject to:} \quad (3)$$

$$|y_i - \mathbf{x}_i^T \mathbf{r}_i| \leq \varepsilon, \quad i = 1, \dots, N$$

As shown below, this problem is equivalent to a suitably scaled traditional ℓ_1 -regularized robust regression:



Figure 1: Face recovery: original and occluded images, best possible recovery with the given basis, proposed, BPRR, cRPCA, LMedS, Mestimator, MLESAC, MSAC, RANSAC, and RR.

Theorem 1. *Problem (3) is equivalent to:*

$$\min_{\mathbf{r}, \eta} \sum_{i=1}^N \frac{|y_i - \mathbf{x}_i^T \mathbf{r} + \eta_i|}{\|\mathbf{x}_i\|_1} \text{ s.t. } |\eta_i| \leq \varepsilon, \quad i = 1, \dots, N \quad (4)$$

From this result, it follows that the S^2R^3 algorithm exhibits noiseless and noisy recovery properties similar to traditional ℓ^1 regularized regression. The main advantage of the proposed formulation, illustrated in Figure 1, is that, due to the automatic scaling, the method is not sensitive to the scale of the outlier errors. In addition, in contrast with previous approaches, the S^2R^3 method can easily handle prior information such as co-occurrence labeling, which in this context simply reduces to using the same variable r_i for all instances of the same label.

- [1] M. A Fischler and R. C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [2] K. Mitra, A. Veeraraghavan, and R. Chellappa. Analysis of sparse regularization based robust regression algorithms. *IEEE Trans. Signal Processing*, 61(5):1249–1257, 2013.
- [3] N.Ozay, M. Sznaier, C. Lagoa, and O. Camps. A sparsification approach to set membership identification of switched affine systems. *Automatic Control, IEEE Transactions on*, 57(3):634–648, 2012.