L1-norm Ranking vs. Angular Embedding on Random Graphs

Jiechao Xiong[1], Xiuyuan Cheng[2] and Yuan Yao[1]

[1] School of Mathematical Sciences, Peking University
[2] Program in Applied and Computational Mathematics, Princeton University

Introduction

Finding a global rating/ranking, as one-dimensional embedding of data, based on pairwise comparisons is a fundamental problem in many fields. Least square (LS) ranking receives rising attention due to its simplicity but subtle structure, such as Hodge decomposition on graphs [?]. Angular Embedding (AE) maps pairwise comparison data onto the circle and find the global ranking score via a primary eigenvector solution [?, ?]. Both work under the presence of small Gaussian noise. Recently Yu (2011) [?] shows that AE is much more robust than least square ranking against sparse outliers.

Although in robust statistics Least Absolute Deviation (L1) has been exploited since 1970s to remove sparse outliers, the L1-formulation in Yu is incorrect for such a purpose and thus inferior to AE in experiments. Therefore, it is unclear what's the robustness property of L1-norm ranking compared with AE in the same setting.

In this report we present a comparative study on L1-ranking and Angular Embedding in the setting of sparse outliers occurred on Erdos-Renyi random graphs. We propose a new L1-norm ranking formulation which solves the Least Absolute Deviation of gradient flows on graphs from pairwise comparison flows which has been explored by Hochbaum [?] and Osher [?] et al. recently. Our results show that the proposed L1-ranking undergoes a phase-transition from exact recovery to failure when the sparsity drops. With some upper and lower bounds such a phase-transition is shown to be optimal up to a logarithmic factor. Experimental studies further show better performance of L1-ranking than AE in exact recovery, against sparse outliers and even with additional small Gaussian noise. These results show that L1-norm statistical ranking is a good candidate for robust ranking.

LS, AE, vs. L1-norm ranking

Given n nodes $V = 1, 2, \ldots, n$ and a set of comparisons $x_{ij} = -x_{ji} : (i,j) \in E$, E is the edge set of comparisons. The statistical ranking problem is to find a potential θ for x_{ij} such that $\theta_i - \theta_j$ is an approximation of x_{ij} satisfied in some sense[?, ?, ?]. LS ranking minimizes the least square

$$(LS): \quad \min_{\theta} \sum_{(i,j)\in E} (\theta_i - \theta_j - x_{ij})^2 \tag{1}$$

L1-norm ranking minimizes least absolute deviation

$$(L1): \quad \min_{\theta} \sum_{(i,j)\in E} |\theta_i - \theta_j - x_{ij}| \tag{2}$$

AE [?, ?] exploits eigenvector decomposition by defining a Hermitian matrix, $H_{ij} = e^{\sqrt{-1}x_{ij}}$ if $(i,j) \in E$ and $H_{ij} = 0$ otherwise. Primary eigenvector of H with maximal absolute eigenvalue, v_1, gives an estimator.

Sparse Perturbation Model on Random Graphs. To compare performances of L1 vs. AE against sparse outliers, consider the model that $(V,E) = G(n,q)$ is a Erdös-Réyni Random Graph, and for $(i,j) \in E$, a sparse perturbation is added such that $x_{ij} = \theta_i - \theta_j + e_{ij}$ and

$$e_{ij} = \begin{cases} 0, & \text{with probability } p \\ U(-L, L), & \text{otherwise} \end{cases}$$

L can be very large, so every edge is an outlier with probability $1-p$.

A Theory of Probabilistic Recovery

Theorem 1 [Upper Bound] For Erdös-Réyni Random Graphs model $G(n,q)$ with uniform perturbation model above, the probability that exact recovery fails for L1-norm ranking is no more than

$$2n \cdot \exp(-c_{p,q}(n-1)qp^2)$$

where $c_{p,q} = 1/[2 + 2p/3 - 4p^2q/3]$.

Remark 1 For any $0 \le p, q \le 1$, $3/8 \le c_{p,q} \le 3/4$, so the probability is at most $2n \cdot \exp(-\frac{3}{8}(n-1)qp^2)$, whence L1-ranking exactly recovers the signal if

$$p \gg \sqrt{\frac{8}{3(n-1)q} \log \frac{2n}{\delta}}$$

Remark 2 For $q = 1$, i.e. G is a complete graph, the probability has a tighter upper bound $2n \cdot \exp(-\frac{1}{2}(n-1)p^2)$.

Theorem 2 [Lower Bound] For discrete model class where $\theta_i \in \Theta(|\Theta| = N \ge 2)$ with sparse perturbation uniform on Θ, the error probability has a lower bound

$$Prob(\hat{\theta} \neq \theta) \succeq 1 - \frac{(n-1)q}{2\log_2 N} I(x_{ij}|\theta_i, \theta_j)$$

where $I(x_{ij}|\theta_i,\theta_j) := \log_2 N - H(x_{ij}|\theta_i,\theta_j) = \frac{1}{2}(N-1)p^2 + O(p^3)$.

Remark 3. A high probability of successful recovery necessarily needs

$$p \ge \sqrt{\frac{4}{(n-1)q} \cdot \frac{\log_2 N}{N-1}}$$

Combining these two theorems, phase-transition of successful recovery for L1-norm ranking is $p_{c,L1} = O(1/\sqrt{nq})$ which is optimal up to a logarithmic factor on n.

Remark 4 (AE) Amit Singer [?] shows by random matrix theory that for the complete graph ($q = 1$), the top eigenvector of AE $\hat{v}_1$ has positive correlation with the vector $z = \frac{1}{\sqrt{n}}(e^{i\theta_1}, \cdots, e^{i\theta_n})^T$ when $p > p_{c,AE,G(n,1)}$, which is the threshold probability$p_{c,AE,G(n,1)} = \frac{1}{\sqrt{n}}$. The correlation $|\hat{v}_1^T z| \to 1$ (but is never exactly 1) when $p \gg p_c$, which means that AE approximately recovers the signals.
This can be extended for general ER random graph $G(n,q)$, with the threshold probability

$$p_{c,AE,G(n,q)} = \frac{1}{\sqrt{nq}}.$$

Thus, both AE and L1-norm ranking have the same phase-transition rates in n up to a logarithmic factor. While, L1 exactly recovers signals when $p > p_{c,L1}$.

Remark 5. In the next section we shall see L1-norm ranking involves a linear programming with $|E|$ dual variables, while AE needs to solve an eigenvalue problem of $|V|$-by-$|V|$ matrix. Therefore in this sense L1-norm ranking is more expensive for the gain of exact recovery.

Experiments

Algorithm. Dual L1-norm ranking is the following maximum flow problem

$$\max \sum_{(i,j)\in E} x_{ij}\xi_{ij}$$
$$s.t. \sum_{j\sim i} \xi_{ij} = 0, \quad |\xi_{ij}| \le 1$$

(1) With only sparse outliers, $|\xi_{ij}| < 1 \Rightarrow \theta_i - \theta_j = x_{ij}$; (2) With both outliers and Gaussian noise, we pursue a least square with those edges such that $|\xi_{ij}| < 1$.

Simulation. Set $n = 80$, $\theta \sim U(-0.2, 0.2)$, $q = 0.5$ and use $cor(\theta, \theta^*)$. $e \sim U(-3,3)$. Notice the magnitude of noise is 25 times of signal. Even with small Gaussian noise, L1-norm ranking still works quite well. Set $\sigma = 0.05$, and $e \sim U(-3,3)$ or $e \sim U(-10,10)$ respectively.

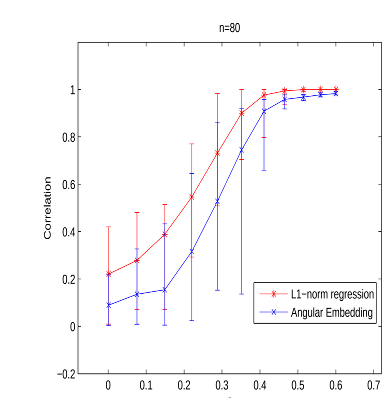

Figure 1: Numerical result without gaussian noise.

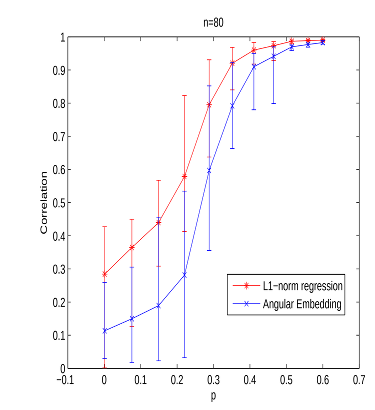

Figure 2: With gaussian noise and $e \sim U(-5,5)$.

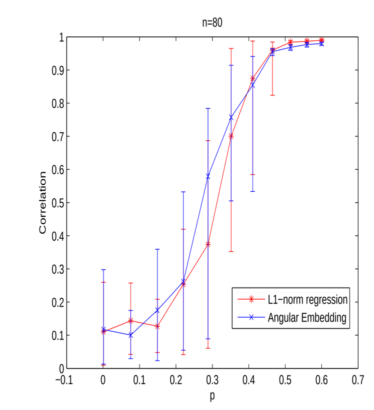

Figure 3: With gaussian noise and $e \sim U(-10,10)$.

The existence of small gaussian noise almost has no effect on L1-norm ranking. AE may warp the perturbations around the circle, whence may have less error when the magnitude of outliers is too big.

Image Reconstruction. Image X as the ground-truth, with an intensity range of 1 over 181×162 pixels. Local comparisons are obtained as intensity differences of X between pixels within a 5×5 neighborhood, added with Gaussian noise of $\sigma = 0.05$. 10% of these measurements are further added with U(-3,3). (i.e. outliers). Both AE and L1 work to reconstruct the image where L1 has better performance (reconstruction error 0.15% for L1 while 1.3% for AE and 6.3% for LS).

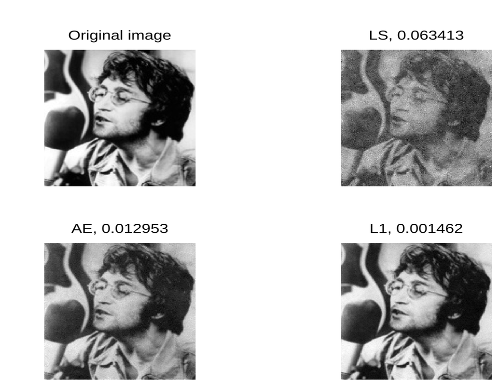

Figure 4: Image reconstruction by AE, LS, L1, from pairwise intensity differences.

References

[1] Hochbaum and Levin, Management Science, 52(9):1394-1404, 2006
[2] Jiang, Lim, Yao, and Ye, Math. Prog., 127(1):203-244, 2011.
[3] Osting, Darbon, and Osher, CAM12-7, 2012
[4] Singer, ACHA, 30(1):20-36, 2011
[5] Yu, IEEE PAMI, 34(1):158-173, 2011