# Distributed Decoding From Heterogeneous 1-Bit Compressive Measurements

Canyi Chen and Liping Zhu
Renmin University of China

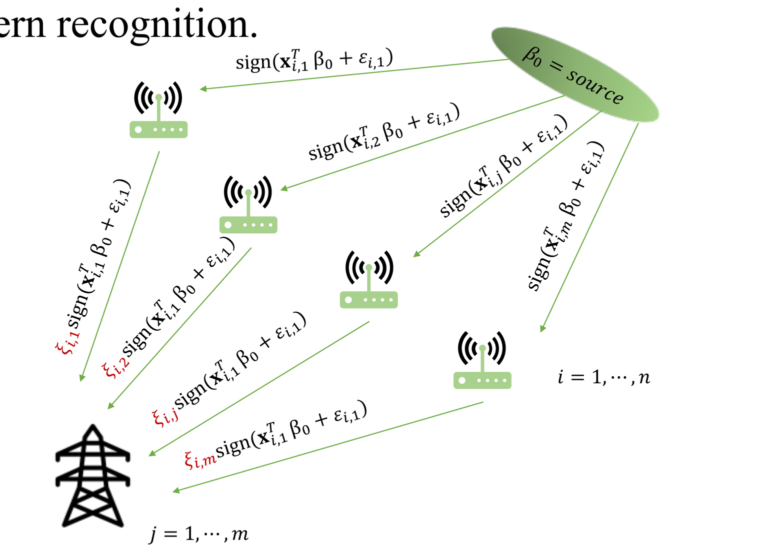
INSTITUTE OF STATISTICS AND BIG DATA

## Abstract

We develop a communication-efficient distributed estimation for the 1-bit compressive sensing where unknown sparse signals are coded into binary measurements with noises and sign flips. We allow for distinctive sign-flipped probabilities and intensities of noises for measurements collected at different nodes, which raises the heterogeneity issue. We suggest a distributed algorithm through penalized least squares to recover the sparse signals. This algorithm is computationally very efficient with only gradient information communicated. The resulting distributed estimate converges after only one iteration even when a lousy initial estimate is provided and achieves a nearly oracle rate after a constant number of iterations. We prove that, up to a proportionality constant, with high probability, the distributed estimate approximates the underlying true sparse signal with precision $\delta$ after a finite number of iterations as long as the total sample size $N$ satisfies $s\log p/(\delta^2 N)=O(1)$, where $p$ is the dimension and $s$ is the number of non-zero elements of the underlying true sparse signals. We also establish statistical guarantee for support recovery. Extensive experiments are provided to illustrate the effectiveness of our proposed distributed algorithm.

## Motivation

Since seminal works of [1, 2], compressive sensing (CS) has become one of the most important approaches to approximate low dimension signals from under-determined and noisy measurements. The 1-bit compressive sensing, which codes each continuous infinite-precision measurement into a single bit, has been found very useful in a variety of applications such as wireless sensor network, cognitive radios and pattern recognition.



**Our contributions.** We provide a communication-efficient algorithm to recover the sparse signal from a heterogeneous distributed system with particular attention paid to the 1-bit compressive sensing problem. The communication cost of our distributed algorithm is $O(mp)$, which is the minimal price a distributed algorithm has to pay. We derive non-asymptotic error bounds for the resultant multi-rounds distributed estimates. After a finite number of iterations, the estimate achieves a nearly oracle rate without severe restrictions on $m$.

## Problem setup and methodology

**Distributed compressive sensing.** The 1-bit compressive sensing assumes that at $j$-th node

$$Y_j=\xi_j \operatorname{sign}(\mathbf{x}_j^T\boldsymbol{\beta}_0+\varepsilon_j), j=1,\cdots,m, \quad \textbf{Model}$$

where $Y_j$ is the 1-bit measurement, $\xi_j$ is a random variable modeling the sign flip of $Y_j$, namely, $\operatorname{pr}(\xi_j=1)=1-\operatorname{pr}(\xi_j=-1)=q_j$, $\mathbf{x}_j$ is a $p$-vector of explanatory variables, $\boldsymbol{\beta}_0$ is the unknown parameter of interest and $\varepsilon_j$ is an independent random error with mean 0 and variance $\sigma^2$.

$$\boldsymbol{\beta}^*=(m\Sigma)^{-1}\sum_{j=1}^m \operatorname{cov}(\boldsymbol{x}_j,Y_j)$$
$$=\operatorname{argmin}_{\boldsymbol{\beta}}\sum_{j=1}^m \mathrm{E}\{(\mathrm{Y_j}-\mathrm{EY_j})-\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta}\}^2$$

is proportional to $\boldsymbol{\beta}_0$. It suffices to estimate $\boldsymbol{\beta}_0$.

## Problem setup and methodology

We assume throughout at each local node, $n$ measurements are collected and denoted as $\{(x_{i,j},Y_{i,j}), i=1,\dots,n, j=1,\dots,m\}$. The subscript $(i,j)$ stands for the $i$-th observation scattered at the $j$-th node. Then we have $N\overset{\text{def}}{=}nm$ measurements evenly scattered at $m$ nodes.
Global loss:

$$\widehat{\boldsymbol{\beta}}_{pen}\overset{\text{def}}{=}\operatorname{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^p}1/(2N)\sum_{j=1}^m\sum_{i=1}^n\left\{(Y_{i,j}-\bar{Y}_j)-(\boldsymbol{x}_{i,j}-\bar{x}_j)^T\boldsymbol{\beta}\right\}^2+\lambda|\boldsymbol{\beta}|_1$$

Surrogate loss:

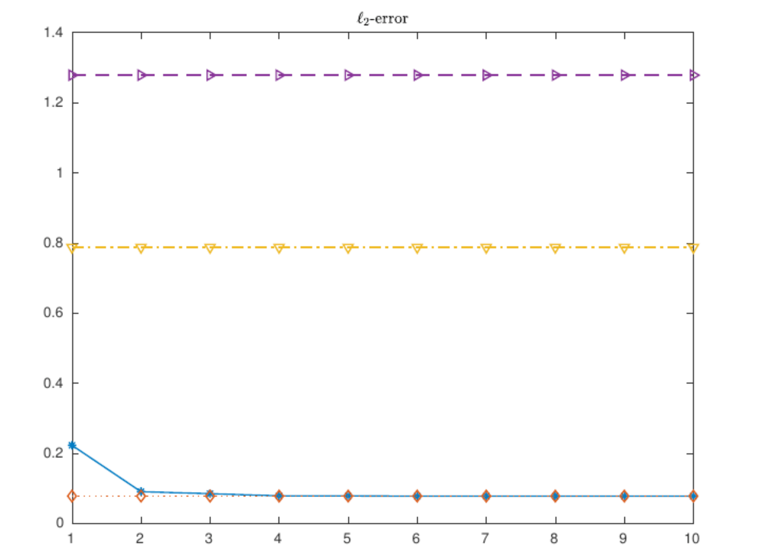
$$\widehat{\boldsymbol{\beta}}^{(1)}\overset{\text{def}}{=}\frac{\boldsymbol{\beta}^T\widehat{\Sigma}_1\boldsymbol{\beta}}{2}-\boldsymbol{\beta}^T\{z_N+(\widehat{\Sigma}_1-\widehat{\Sigma})\,\widehat{\boldsymbol{\beta}}^{(0)}\}+\lambda|\boldsymbol{\beta}|_1 \quad \textbf{Iterations}$$

where $z_N=\sum_{i,j}x_{i,j}Y_{i,j}$.

**$\ell_1$-regularized least squares (Lasso).** Lasso is widely-used for high-dimensional regression models with a sparsity structure.

**Aims for $\boldsymbol{\beta}^*$.** By targeting on $\boldsymbol{\beta}^*$ instead $\boldsymbol{\beta}_0$, we work around the non-smoothness inheriting from the model.

**Surrogate loss function.** Substitute global higher order information with local one that enables a communication-efficient algorithm.



(A): The $\ell_2$-error

FIG 1. *The horizontal axis stands for the number of iterations.*

**Algorithm 1** Distributed Robust Decoding with Lasso

**Input:** Measurements $\{(\mathbf{x}_{i,j},Y_{i,j})_{i=1,\dots,n;j=1,\dots,m}\}$, number of iterations $t$, and regularization parameters $\lambda_0$ and $\lambda_N$.

1: Compute the initial estimator $\widehat{\boldsymbol{\beta}}^{(0)}$ by

$$\widehat{\boldsymbol{\beta}}^{(0)}\overset{\text{def}}{=}\underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\left\{(2n)^{-1}\sum_{i=1}^n(Y_{i,1}-\mathbf{x}_{i,1}^{\mathrm{T}}\boldsymbol{\beta})^2+\lambda_0|\boldsymbol{\beta}|_1\right\}.$$

2: **for** $k=1,\dots,t$ **do**
3: Transmit $\widehat{\boldsymbol{\beta}}^{(k-1)}$ from the first node to the local nodes labeled with $2,\dots,m$.
4: **for** $j=1,\dots,m$ **do**
5: Calculate $\widehat{\Sigma}_j\widehat{\boldsymbol{\beta}}^{(k-1)}$ and $\mathbf{z}_{n,j}$ at the $j$-th node and sends them back to the first node.
6: **end for**
7: Calculate $\widehat{\boldsymbol{\beta}}^{(k)}$ on the first node based on

$$\widehat{\boldsymbol{\beta}}^{(t)}\overset{\text{def}}{=}\underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\,\boldsymbol{\beta}^{\mathrm{T}}\widehat{\Sigma}_1\boldsymbol{\beta}/2-\boldsymbol{\beta}^{\mathrm{T}}\{\mathbf{z}_N+(\widehat{\Sigma}_1-\widehat{\Sigma})\widehat{\boldsymbol{\beta}}^{(t-1)}\}+\lambda_N|\boldsymbol{\beta}|_1.$$

8: **end for**

**Output:** The final estimate $\widehat{\boldsymbol{\beta}}^{(t)}$ obtained from the first node.

**Iteration times of Algorithm 1.** The number of iterations we need to achieve the oracle rate is

$$t\geq\log(Ns/n)/\log\{c_0 n/(s^2\log N)\}, \text{ for some } c_0>0, \quad \textbf{Time}$$

which increase logarithmically with the number of nodes $\boldsymbol{m}$.

**Communication cost of Algorithm 1.** In Algorithm 1, the communication cost mainly resides in the transmission of $z_N$ and $\widehat{\Sigma}\widehat{\beta}^{(t-1)}$. The overall communication cost of Algorithm 1 is of order $O(mp)$.

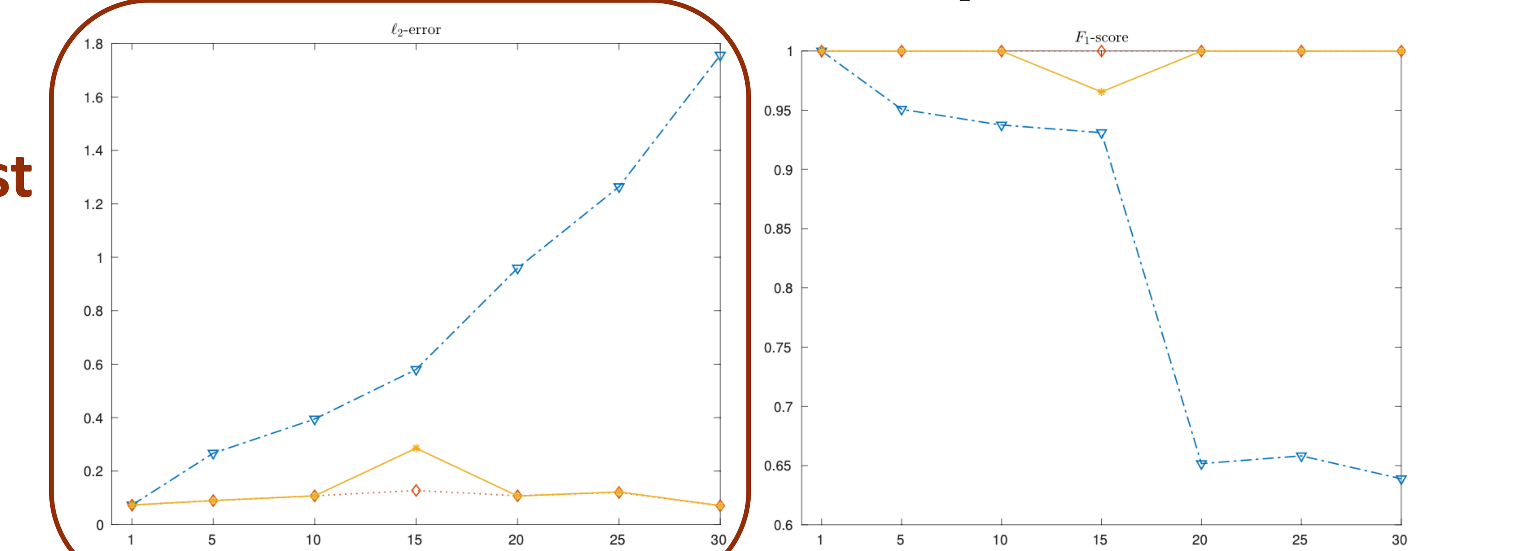
TABLE 1. Both the $\ell_2$-error and the $F_1$-score of the estimates under different sparsity levels $s$. We fix $p=10000$, $n=1000$ and $N=10000$.

| $s$ | divide-and-conquer | | pooled | | distributed | |
|---|---|---|---|---|---|---|
| | $F_1$-score | $\ell_2$-error | $F_1$-score | $\ell_2$-error | $F_1$-score | $\ell_2$-error |
| 5 | 1.0000 | 0.0330 | 1.0000 | 0.0245 | 1.0000 | 0.0245 |
| 10 | 1.0000 | 0.0963 | 1.0000 | 0.0385 | 1.0000 | 0.0385 |
| 20 | 0.8947 | 0.8290 | 1.0000 | 0.0699 | 1.0000 | 0.0702 |
| 30 | 0.5581 | 1.2605 | 1.0000 | 0.1226 | 1.0000 | 0.1232 |
| 50 | 0.4872 | 1.0377 | 0.9583 | 0.3661 | 0.9583 | 0.3627 |
| 100 | 0.3478 | 1.3996 | 0.9950 | 0.3220 | 0.9950 | 0.3231 |

## Effect of local sample size

FIG 2. The horizontal axis stands for the number of nodes m, and the vertical axis stands for the l2-error in (A) and the F1-score in (B) of the divide-and-conquer estimate (dot-dash), distributed estimate (solid) and pooled estimate (dotted), respectively, with the total sample size N = 21600, the sparsity level s = 30, and the dimension p = 5000.

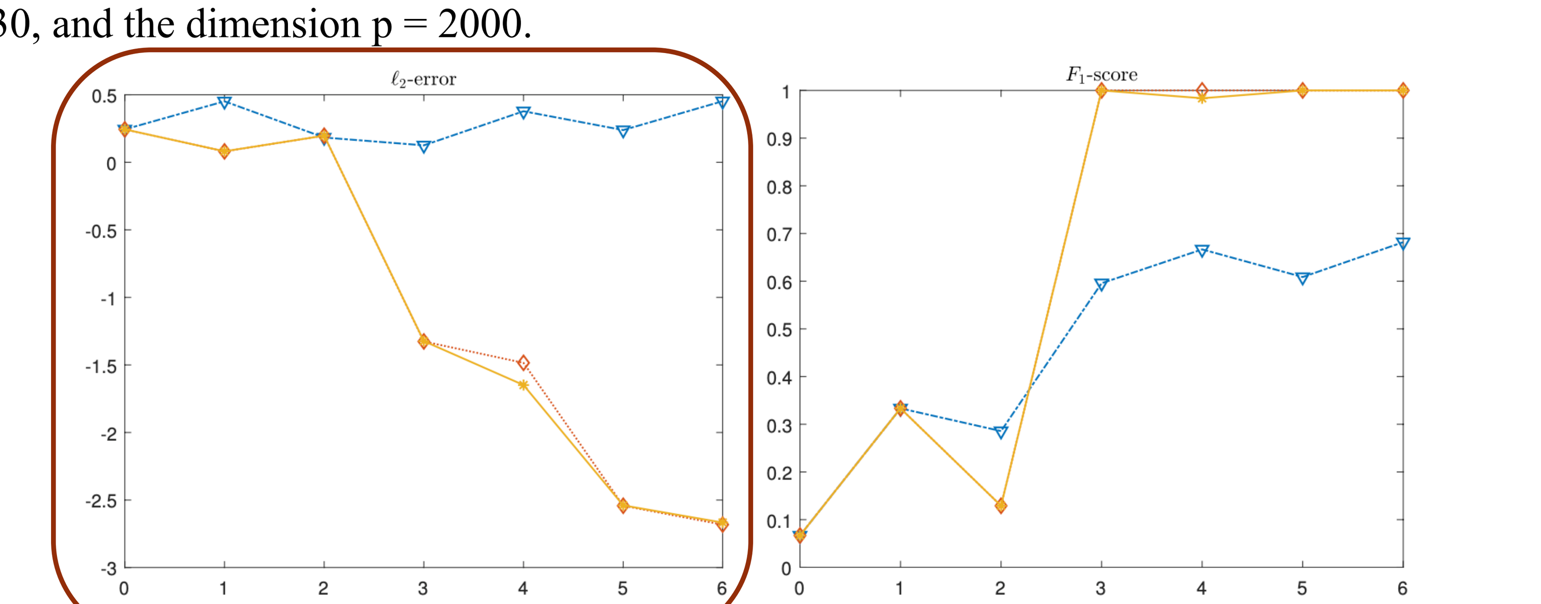
**Robust to $\boldsymbol{m}$**



(A): The $\ell_2$-error (B): The $F_1$-score

## Effect of sample size

FIG 3. The horizontal axis stands for the base-2 logarithm of the number of nodes m in both (A) and (B), and the vertical axis stands for the base-2 logarithm of the l2-error in (A) and the F1-score in (B) of the divide-and- conquer estimate (dot-dash), distributed estimate (solid) and pooled estimate (dotted), respectively. We fix the local sample size n = 800, the sparsity level s = 30, and the dimension p = 2000.



(A): The $\ell_2$-error (B): The $F_1$-score

**Optimality**

## Acknowledgements

Zhu is the corresponding author and his research is supported by Natural Science Foundation of Beijing(Z19J0002) and National Natural Science Foundation of China (11731011, 11931014).

## References

[1] Candes, E. J., Romberg, J., and Tao, T. (2006), "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, 52, 489–509. https://doi.org/10.1109/TIT.2005.862083.

[2] Candes, E. J., Romberg, J. K., and Tao, T. (2006), "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59, 1207–1223.