# On theoretical optimization of the sensing matrix for sparse-dictionary signal recovery

Jianchen zhu, Shengjie zhao, Xu Ma, Gonzalo R. Arce
Key Laboratory of Embedded System and Service Computing, Ministry of Education , Tongji University
Email: zhujianchen@tongji.edu.cn

# outline

01 Compressive sensing model

02 Measurements guarantees

03 Algorithm design

04 Simulation results

# Compressive sensing model

# 01 Compressive sensing model

The basic insight of compressive sensing (CS) is that a small number of linear measurements can be used to reconstruct sparse signals, thus the information we extract from the signal is given by:

$$y = Ax \quad \text{or} \quad y = Ax + e \quad \text{(With noise perturbation)}$$

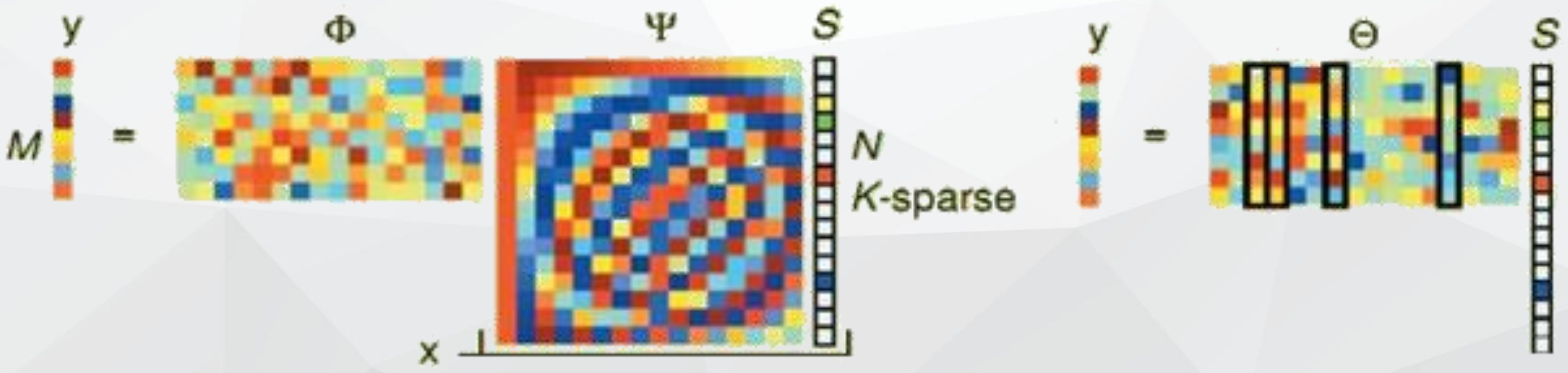



Figure: The compressive sensing process and its domains

# 01 Compressive sensing model

Compressed sensing theory mainly includes three parts:

A. Sparse representation of signals
B. Design measurement matrix
C. Design signal recovery algorithm

Some typical applications in CS theory:

A. Image information security
B. Wireless sensor network (WSN)
C. Magnetic resonance imaging (MRI)
D. Compressive spectral imaging

(a) The original data cube (b) Reconstructions from six shots using Boolean coded aperture

Question: How to recovery the signal from its measurements?

# 01 Compressive sensing model

Given the observation (measurements) of a k-sparse signal from the underdetermined linear system, such signal can be recovered by solving a minimization problem:

$$\min_{\tilde{x}} \|\tilde{x}\|_0 \sim s.t. \|A\tilde{x} - y\|_2 \leq \varepsilon, \sim P_0 \qquad (1)$$

One of the practical and tractable alternatives to this problem can be expressed as:

$$\min_{\tilde{x}} \|\tilde{x}\|_1 \sim s.t. \|A\tilde{x} - y\|_2 \leq \varepsilon, \sim P_1 \qquad (2)$$

Question: How to solve this 1-minimization problem?

1. Bayesian framework
2. Greed pursuit or iteration algorithms
3. Linear programming and so forth

# 01 Compressive sensing model

Studies in [1] [2] have shown that a k-spare signal can be exactly recovered by solving (2) provided that measurement matrix satisfies the restricted isometry property (RIP) conditions, such that

$$(1-\delta_K)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\delta_K)\|x\|_2^2 \quad \text{with} \quad \delta_k \in (0,1) \qquad (3)$$

Many types of random measurement matrices have small RIC with high probability given that the number of measurements is large enough

Question: Dose RIP conditions can apply to the signal which is spare in overcomplete dictionaries?

[1] Candes E J, Tao T. Decoding by linear programming[J]. IEEE Transactions on Information Theory, 2005, 51(12):4203-4215.
[2] Candes E, Romberg J, Tao T. Stable Signal Recovery from Incomplete and Inaccurate Measurements[J]. Communications on Pure & Applied Mathematics, 2010, 59(8):1207-1223.

# Measurements guarantees

# 02 Measurements guarantees

There are numerous signal of interest are sparse in an overcomplete dictionary. More specific, We assume that signal can be sparely using the linear combination of such dictionary, as

$$x = Da \quad \text{with} \quad D \in R^{n \times d}, \sim n \ll d \tag{4}$$

Question: why we use the overcomplete dictionary ?

The D-RIP can be used as a natural extension to the standard RIP such like

$$(1-\delta_K)\|Da\|_2^2 \le \|ADa\|_2^2 \le (1+\delta_K)\|Da\|_2^2 \tag{5}$$

It is well know that random matrices satisfy the D-RIP condition (5) with the number of measurements on the order of

$$k \log(d / k) \tag{6}$$

# 02 Measurements guarantees

First, the random variable has its corresponding expected values; that is

$$E(\|Ax\|_2^2) = \|x\|_2^2 \quad (7)$$

Next, the random variable is strongly concentrated about its expected value, which is give by

$$\Pr(\|Ax\|_2^2 - \|x\|_2^2 \geq \varepsilon \|x\|_2^2) \leq 4e^{-c_0(\varepsilon)m} \quad (8)$$

Thus, our approach can be divided into three steps:

Construct points → Apply (8) to point → Extend to all signals

# 02 Measurements guarantees

**Lemma:** Let A be a random matrix satisfies (8), Then we have:

$$(1-\delta)\|x\|_2 \le \|ADa\|_2 \le (1+\delta)\|Da\|_2 \quad (9)$$

With probability $\ge 1-4e^{-c_0(\delta/2)m}$ (10)

Choose a set of Q points such that

$Q \le (12/\delta)^k$ With probability $\ge 1-4(12/\delta)^k e^{-c_0(\delta/2)m}$ (11)

apply the union bound to (8), extend the result with probability exceeding the right side of (10) such that

$$\Pr(\|Ax\|_2^2 - \|x\|_2^2 \ge (\delta/2)\|x\|_2^2) \le 4(12/\delta)^k e^{-c_0(\delta/2)m}$$ with $\varepsilon = \delta/2$ (12)

$$\binom{n}{k} \le (en/k)^k \qquad \binom{d}{k} \le (ed/k)^k$$

Question: How to extend such result to two cases?

# 02 Measurements guarantees

According to previous steps, the minimal number of measurements only in a single K-dimensional subspace can be expressed as

$$m = O\left(\frac{2k\log(42/\delta)+\log(4/a)}{c_0(\varepsilon)}\right) \quad (13)$$

We then use (9) to go beyond a single k-dimensional subspace in order to acquire the minimal number of measurements when the basis is the orthonormal basis and overcomplete dictionary, respective, such that

$$m = O\left(\frac{2k\log(42en/\delta k)+\log(4/a)}{c_0(\varepsilon)}\right) \qquad m = O\left(\frac{2k\log(42ed/\delta k)+\log(4/a)}{c_0(\varepsilon)}\right) \quad (14)$$

Reason: there are different subspace in two case.

# Algorithm design

# 03 Algorithm design

In general, greedy or iterative related algorithms can break the recover problem into subproblems:

A. Identifying the columns of basis
B. Projecting onto that subspace

An optimal recovery strategy is to solve the problem via  lease-squares:

$$\hat{x} = \arg\min \|y - Az\|_2 \sim s.t. \sim z \in R(\Psi_\Lambda) \qquad (15)$$

More specific, we can compute from (11):

$$\hat{a}_\Lambda = (\hat{A}_\Lambda^H \hat{A}_\Lambda)^{-1} \hat{A}_\Lambda^H y, \sim \hat{a}_{\Lambda^c} = 0 \qquad (16)$$

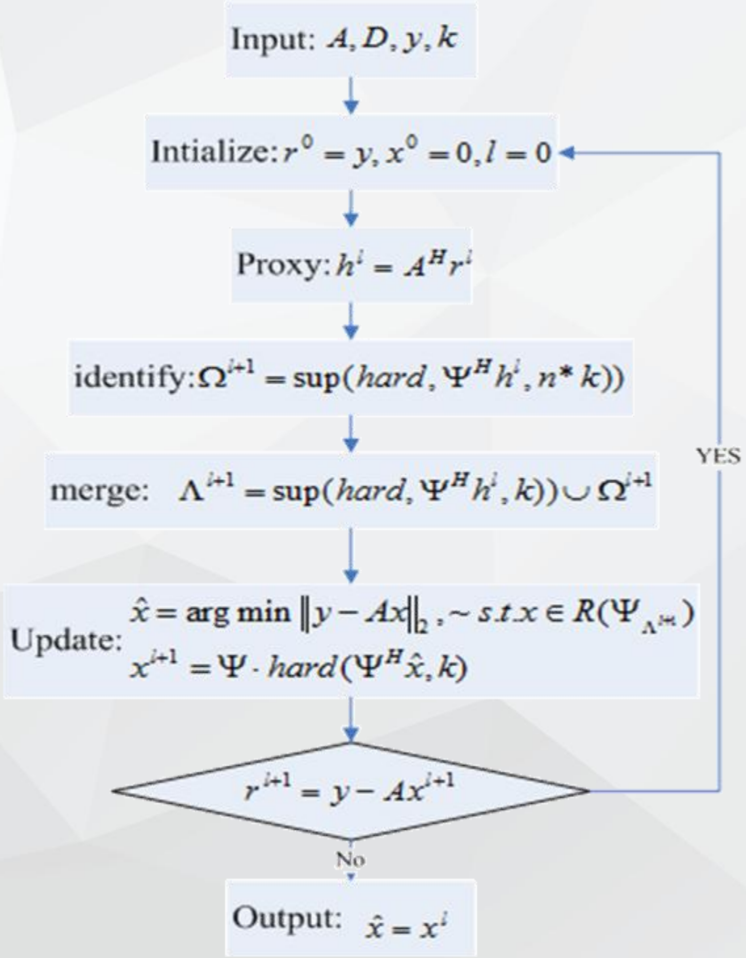
Question: How to implement the GP based algorithm when the basis is not an orthonormal basis?

# 03 Algorithm design

To this end, we envision two natural extension of the canonical GP based algorithms, the flow of the general algorithm can be organized as follow:

Task: This general algorithm is quite flexible and can be invoked in multiple way.

Question: How to find the optimal support in identify step first?



# 03 Algorithm design

The key identification step in our algorithm requires finding the best k-sparse representation of a vector limited by the dictionary constraints:

$$\Omega_{opt} = \arg\min \left\| x - P_{\Lambda} x \right\|_2 \qquad (17)$$

Due to the NP hard problem, we use the near-optimal to approximate the optimal such that

$$\left\| P_{S_D(x,k)} x - x \right\|_2 \le c_1 \left\| P_{\Lambda} x - x \right\|_2 , \left\| P_{S_D(x,k)} x \right\|_2 \ge c_1 \left\| P_{\Lambda} x \right\|_2 \qquad (18)$$

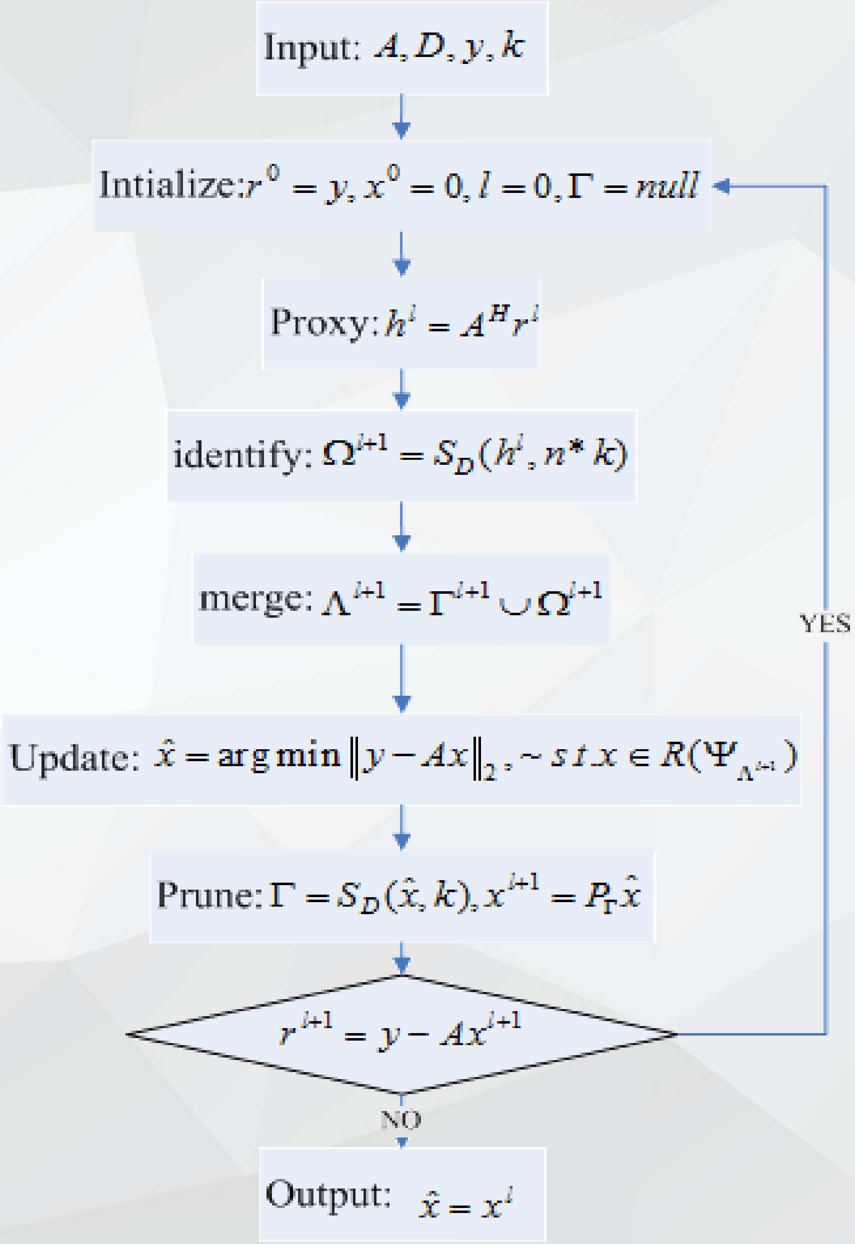
When using some classical CS recovery algorithms including GP and1-norm minimization methods for obtaining the near-optimal projection.

# 03 Algorithm design

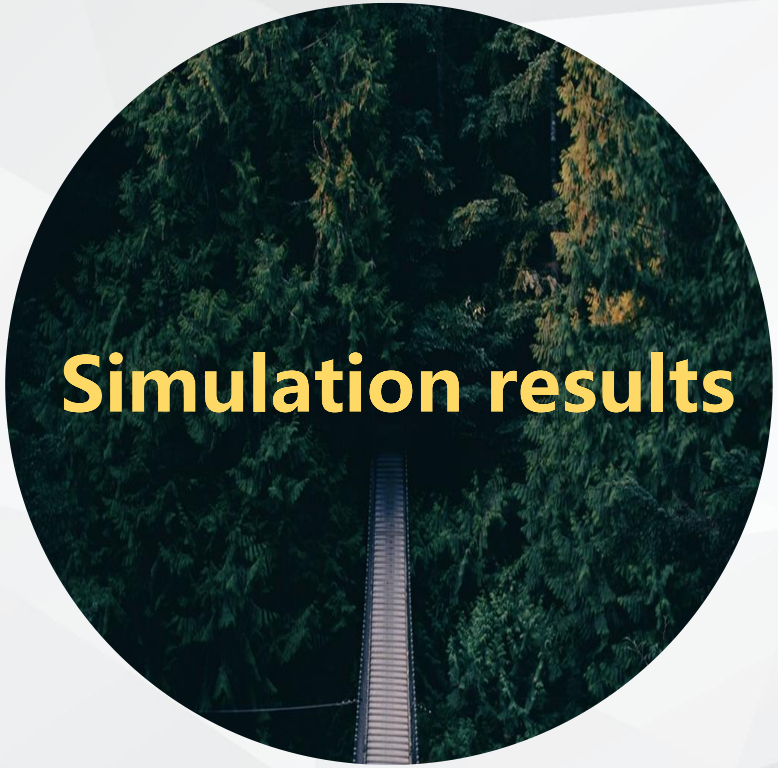
Such projection  is required in the identifying step and another such projection is required in the Prune step, respectively.

Which CS algorithms can be used for the projection and its support acquisition?

1. 1-norm minimization algorithm : linear programming (LP)
2. Greedy pursuit (GP) algorithms: Matching Pursuit and related algorithms



# Simulation results

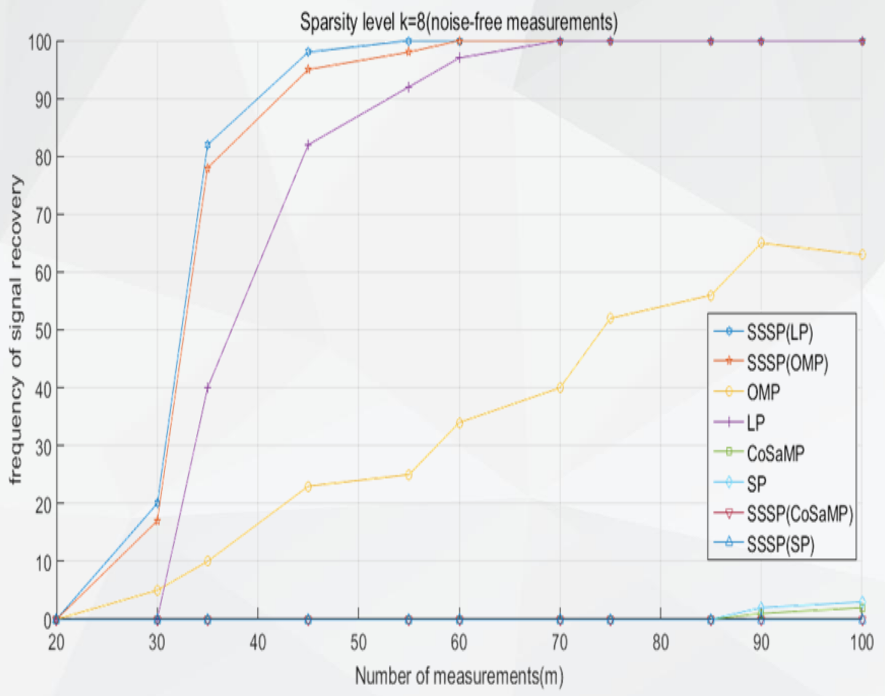


# 04 Simulation results

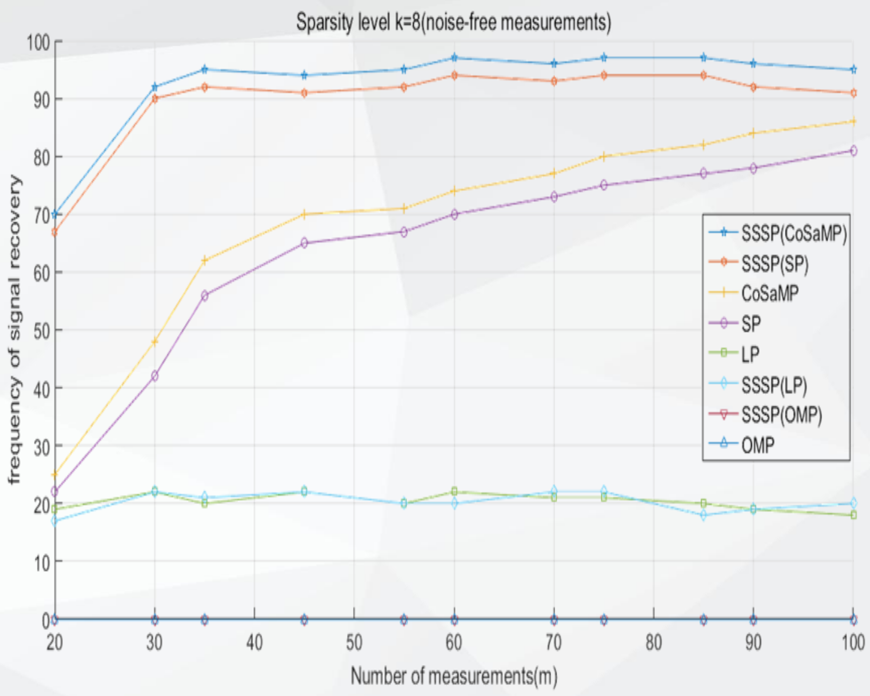
Parameter setting:

1. Length of the signal: $n = 256$
2. The measurement matrix whose dimension is: $64 \times 256$
3. The sparse number is: $k = 8$
4. The overcomplete dictionary is 4 times DFT dictionary: $D \in R^{256 \times 1024}$

# 04 Simulation results




(a)




(b)

Case (a): The non-zero entries of the sparse vector are random positioned and well separated

Case(b): The non-zero entries of the sparse vector are random positioned and cluster together

# Conclusion

1. The number of measurements required guarantees the signal, which is sparse in an over complete dictionary can be recovered from the measurements with high probability.
2. A near-optimal projection strategy is proposed in our algorithm for the near optimal support acquisition such that obtaining the signal estimation.

# THANKS!