PHASE TRANSITION ANALYSIS OF RECENTLY PROPOSED COMPRESSED SENSING RECOVERY ALGORITHMS

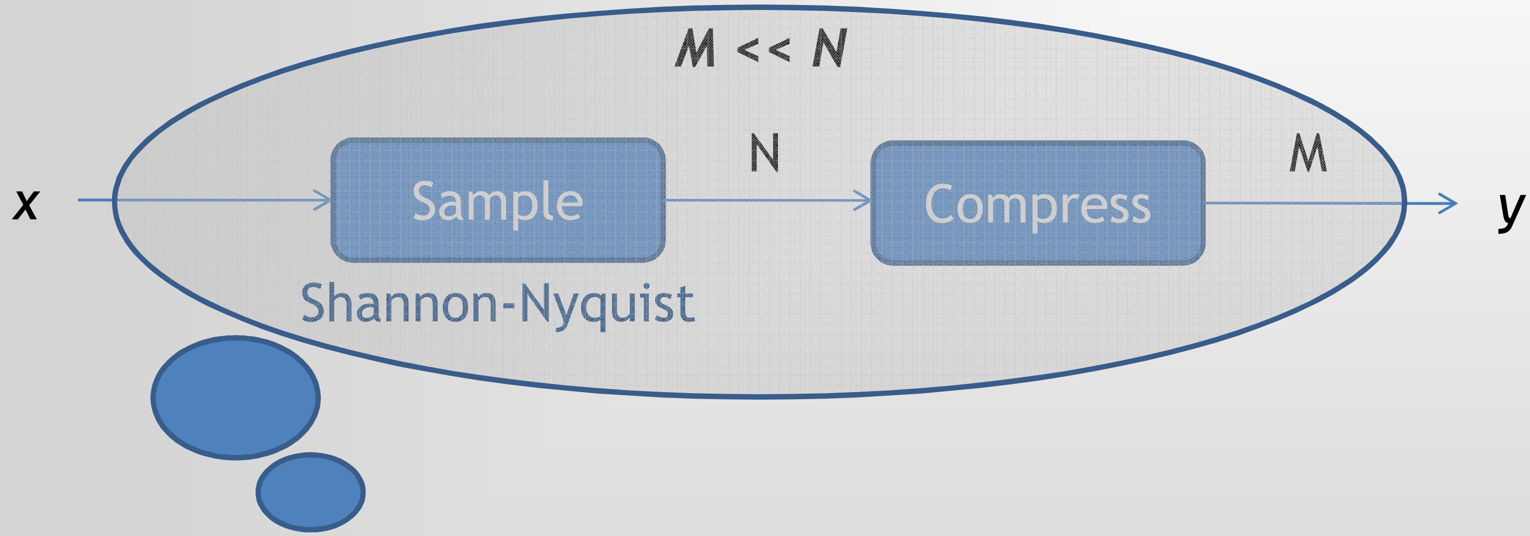
ADRIANA GONZÁLEZ

SIGNAL PROCESSING SEMINAR – FEBRUARY 16TH 2011

SUMMARY

- Introduction on Compressed Sensing and Sparsity
- Optimization Problems in CS
- Some CS Recovery Algorithms
- Analysis and Comparison of the Algorithms
- Conclusions

INTRODUCTION



COMPRESSIVE SENSING

INTRODUCTION

Shannon-Nyquist

sampling frequency
↕
cutoff frequency

Compressive sensing

sampling frequency
↕
information content

- More samples
- More acquisition time
- Higher dimensional data

INTRODUCTION

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} \Phi \end{bmatrix} \begin{bmatrix} x \end{bmatrix} + \begin{bmatrix} n \end{bmatrix}$$

M **x 1**
noisy measurements

M **x** ***N***

N**x1**
signal

M **x 1**
noise

M **<<** ***N***

⇓

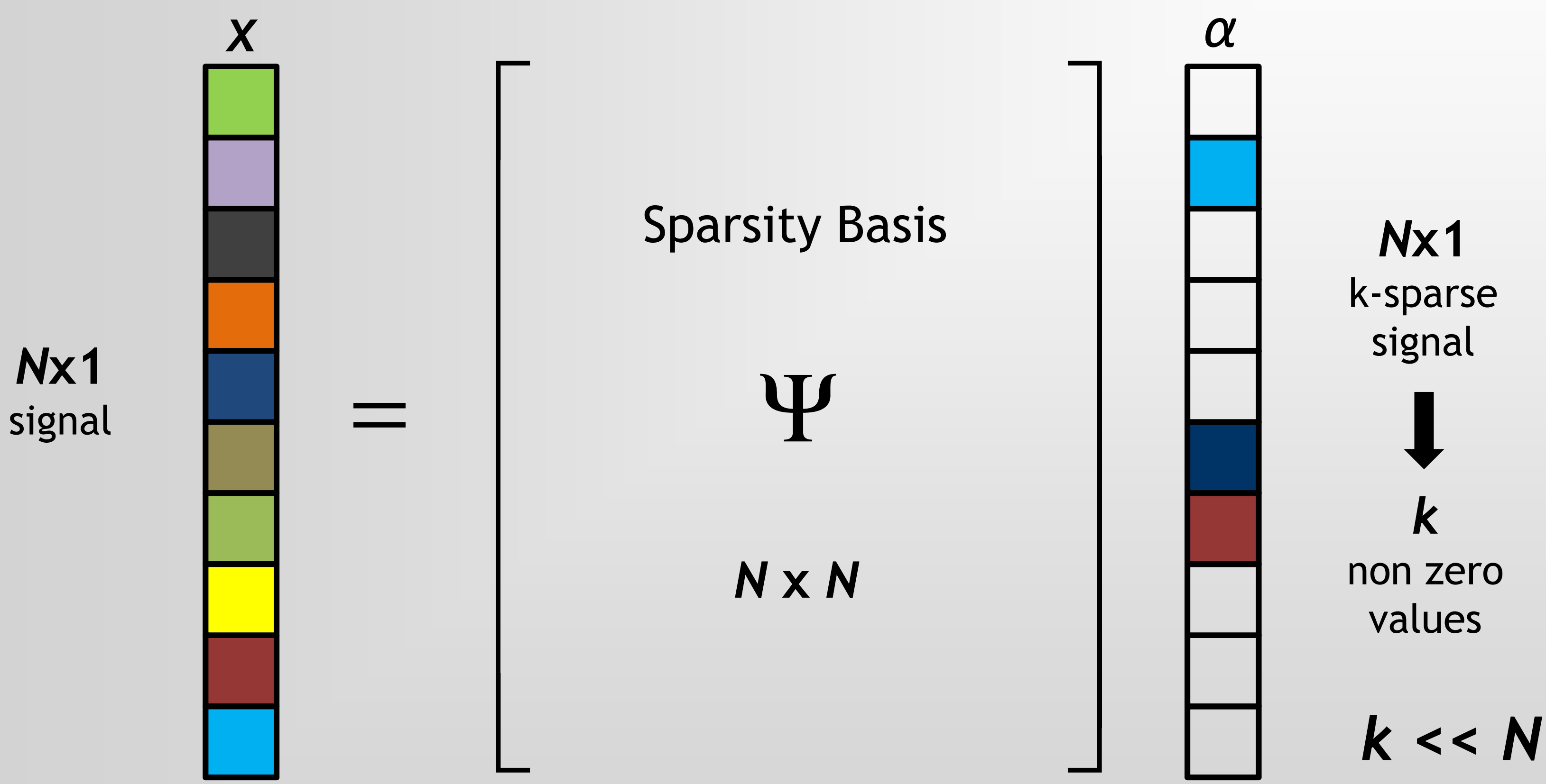
Underdetermined System

⇓

SPARSITY

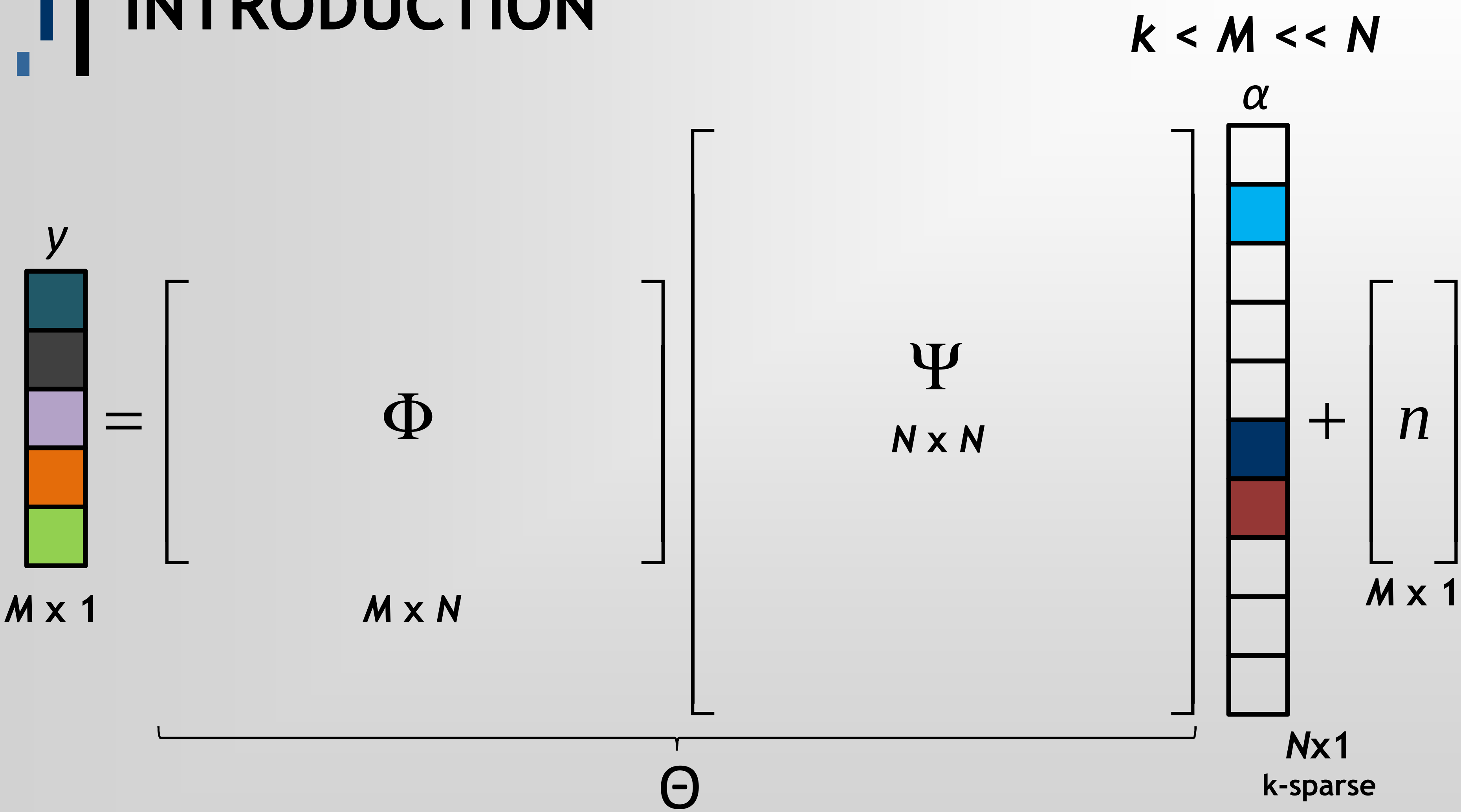
INTRODUCTION

SPARSITY



Sparsity Basis ⟶ Fourier, Wavelets, etc. ⟶ Different basis for different purposes!

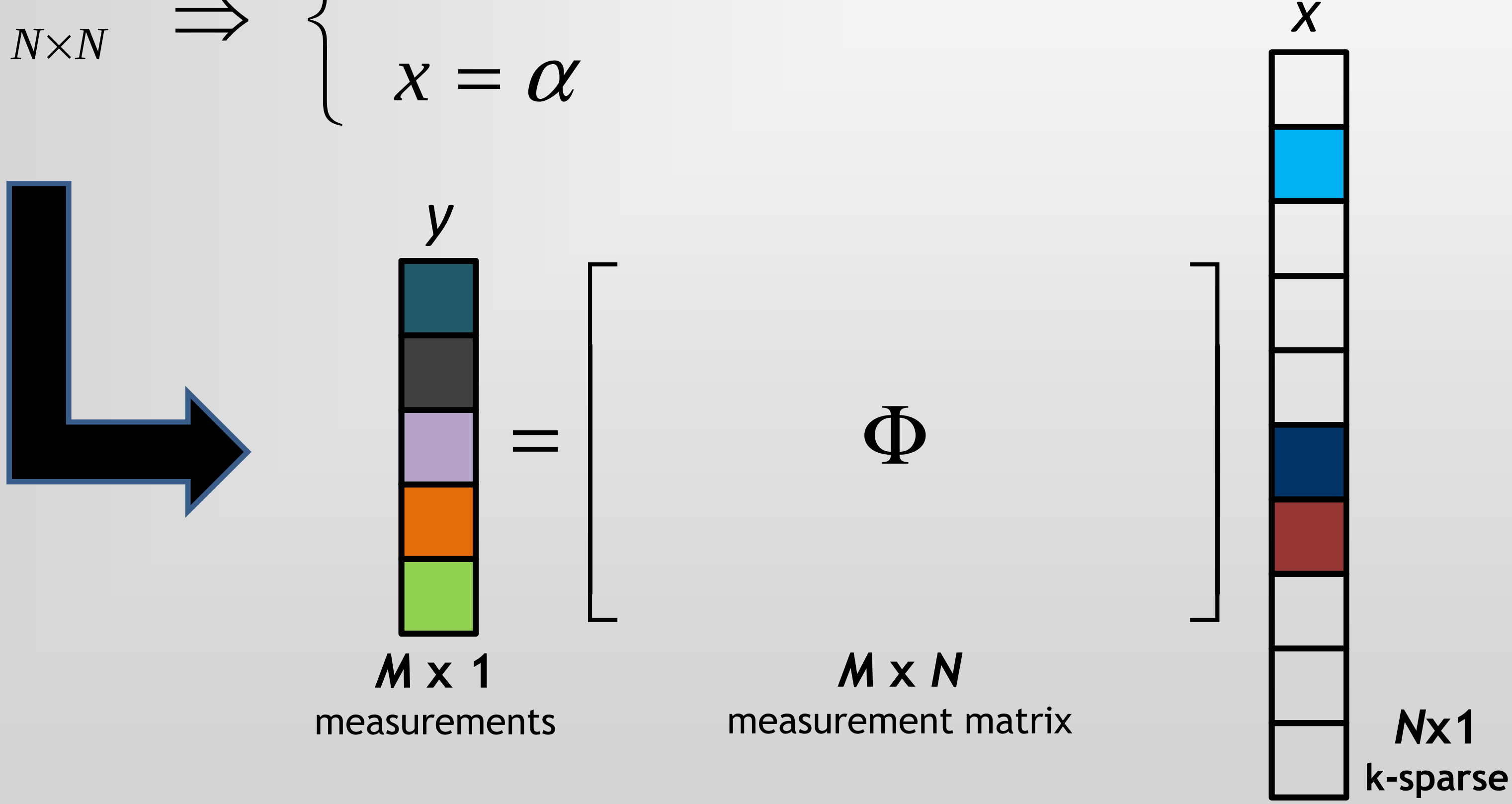
INTRODUCTION



INTRODUCTION

$$k < M << N$$

$$\Psi = I_{N\times N} \Rightarrow \begin{cases} \Theta = \Phi \\ x = \alpha \end{cases}$$



OPTIMIZATION PROBLEMS IN CS

$$\hat{x} = \arg\min_{x \in \Re^N} \underbrace{\| y - \Phi x \|_2^2}_{f(x)\ \text{Fidelity}} \quad s.t. \quad \underbrace{\| x \|_0 = k}_{\text{Sparsity Constraint}}$$

$$\hat{x} = \arg\min_{x \in \Re^N} \underbrace{\| x \|_0}_{\text{Sparsity}} \quad s.t. \quad \underbrace{\| y - \Phi x \|_2^2 \leq \delta^2}_{\text{Fidelity Constraint}}$$

OPTIMIZATION PROBLEMS IN CS

$$\hat{x} = \arg\min_{x \in \Re^N} \| y - \Phi x \|_2^2 \quad s.t. \quad \| x \|_0 = k$$

NP-hard ⇒ **Conditions on matrix Φ to guarantee**

$$\Phi x_1 \neq \Phi x_2 \;\; for \;\; all \;\; k-sparse \;\; x_1 \neq x_2$$

⇓

Depend on the recovery algorithm

CS RECOVERY ALGORITHMS

To find the **optimal** solution to

$$\hat{x} = \arg\min_{x \in \Re^N} f(x) \quad s.t. \quad \| x \|_0 = k$$

in order to reconstruct the original signal ***x***

- IHT
- LIHT
- FLIHT

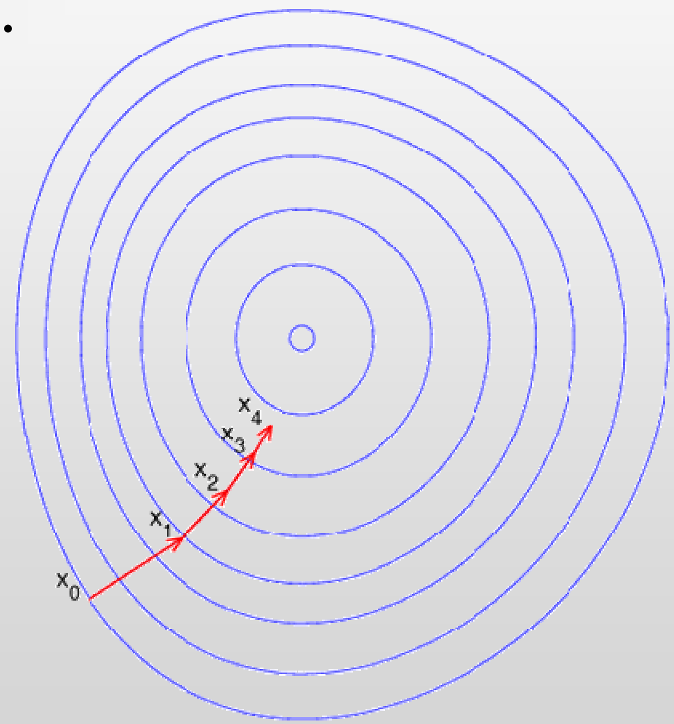
CS RECOVERY ALGORITHMS

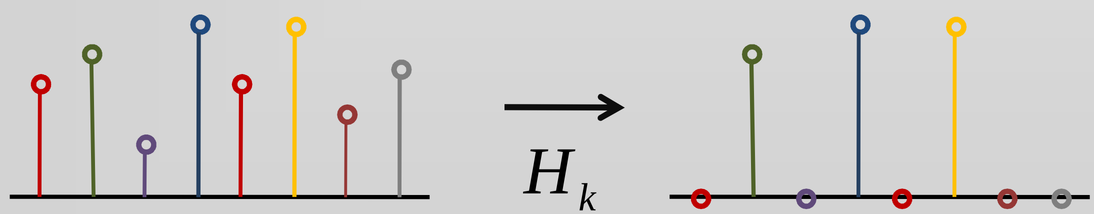
ITERATIVE HARD-THRESHOLDING (IHT)

Based on the gradient descent method.

$$\begin{cases} x_0 = 0 \\ x_{i+1} = H_k\left(x_i - \frac{1}{2}\nabla f(x_i)\right) \end{cases}$$

$$x_{i+1} = H_k\left(x_i + \Phi^*\left(y - \Phi x_i\right)\right)$$





H_k

$$\|x\|_0 = k$$

CS RECOVERY ALGORITHMS

➢ ITERATIVE HARD-THRESHOLDING (IHT)

$$x_{i+1} = H_k\left(x_i - \frac{1}{2}\nabla f(x_i)\right) = H_k\left(x_i + \Phi^*(y - \Phi x_i)\right)$$

- Simple ⇨ One of the most used algorithms
- Condition for matrix Φ ⇨ $\|\Phi\|_2^2 < 1$
- Low convergence rate and not possible to improve.
- For $f(\hat{x}) - f(x) \leq \varepsilon$ ⇨ $\#iter = O\left(\frac{1}{\varepsilon}\right)$

CS RECOVERY ALGORITHMS

V. Cevher, "An ALPS view of sparse recovery". Laboratory for Information and Interference Systems, École Polytechnique Fédérale de Lausanne (EPFL), 2010.

- Lipschitz Iterative Hard-Thresholding (LIHT).
- Fast Lipschitz Iterative Hard-Thresholding (FLIHT).

CS RECOVERY ALGORITHMS

- **Cevher's work - Conditions on matrix Φ.**

Restricted Isometry Property (RIP)

$$(1-c)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+c)\|x\|_2^2$$

Random matrix Φ must be k-RIP ➡ must satisfy RIP with $M = O(k\log(N))$

Incoherent measurements.
Random Gaussian, Random Fourier Ensemble.

CS RECOVERY ALGORITHMS

➢ Cevher’s work - RIP and Lipschitz Continuity

Restricted Isometry Property (RIP)

$$(1-c)\|x\|_2^2 \le \|\Phi x\|_2^2 \le (1+c)\|x\|_2^2$$

Lipschitz Continuous Gradient

$$\|\nabla f(x_1) - \nabla f(x_2)\| \le L\|x_1 - x_2\|$$

$$\|\Phi(x_1 - x_2)\|_2^2 \le \frac{L_{K'}}{2}\|x_1 - x_2\|_2^2$$

$$L_{K'} = 2(1+c)$$

B. Bah, J. Tanner. “Improved bounds on Restricted Isometry Constants for Gaussian matrices”. SIAM J. Matrix Anal. Appl. Vol. 31(5): 2882-2898

$$L_{K'} \le 2\left(1 + \sqrt{\frac{K'}{M}} + \sqrt{\frac{2t}{M}}\right)$$

$$P(success) \ge 1 - e^{-t}$$

CS RECOVERY ALGORITHMS

➢ LIPSCHITZ ITERATIVE HARD-THRESHOLDING (LIHT)

Based on the gradient descent method.

$$\begin{cases} x_0 = 0 \\ x_{i+1} = H_k\left(x_i - \frac{1}{L_{2k}} \nabla f(x_i)\right) \end{cases}$$

⇩

$$x_{i+1} = H_k\left(x_i + \frac{2}{L_{2k}} \Phi^* (y - \Phi x_i)\right)$$

CS RECOVERY ALGORITHMS

➢ LIPSCHITZ ITERATIVE HARD-THRESHOLDING (LIHT)

$$x_{i+1} = H_k\left(x_i - \frac{1}{L_{2k}}\nabla f(x_i)\right) = H_k\left(x_i + \frac{2}{L_{2k}}\Phi^*(y - \Phi x_i)\right)$$

- Simple
- Convergence rate depending on L_{2k}.
- For $f(\hat{x}) - f(x) \le \varepsilon \Rightarrow \#iter = O\left(\frac{1}{\varepsilon}\right)$

CS RECOVERY ALGORITHMS

- **FAST LIPSCHITZ ITERATIVE HARD-THRESHOLDING (FLIHT)**

  Based on the Nesterov's Optimal Gradient Method.

$$\begin{cases} x_{-1} = u_0 = 0 \\ x_i = H_k\left(u_i - \dfrac{1}{L_{3k}}\nabla f\left(x_i\right)\right) = H_k\left(u_i + \dfrac{2}{L_{3k}}\Phi^*\left(y - \Phi x_i\right)\right) \\ a_0 = 1 \\ a_1 = 0.5\left(1 + \sqrt{1 + 4a_i^2}\right) \\ u_{i+1} = x_i + \dfrac{a_i - 1}{a_{i+1}}\left(x_i - x_{i-1}\right) \end{cases}$$

CS RECOVERY ALGORITHMS

FAST LIPSCHITZ ITERATIVE HARD-THRESHOLDING (FLIHT)

$$x_i = H_k\left(u_i - \frac{1}{L_{3k}}\nabla f(x_i)\right) = H_k\left(u_i + \frac{2}{L_{3k}}\Phi^*(y - \Phi x_i)\right)$$

$$u_{i+1} = x_i + \frac{a_i - 1}{a_{i+1}}(x_i - x_{i-1})$$

- Two-level condition.
- Convergence rate depending on L_{3k}.
- For $f(x_{i+1}) - f(x_i) \le \varepsilon \Rightarrow \#iter = O\left(\frac{1}{\sqrt{\varepsilon}}\right)$

ANALYSIS AND COMPARISON

➢ PHASE TRANSITION DIAGRAMS

- CS behavior of the algorithm.

- Amount of measurements M to take in order to recover an N-dimensional signal with k amount of information.

- Analysis in the sparsity-undersampling domain:

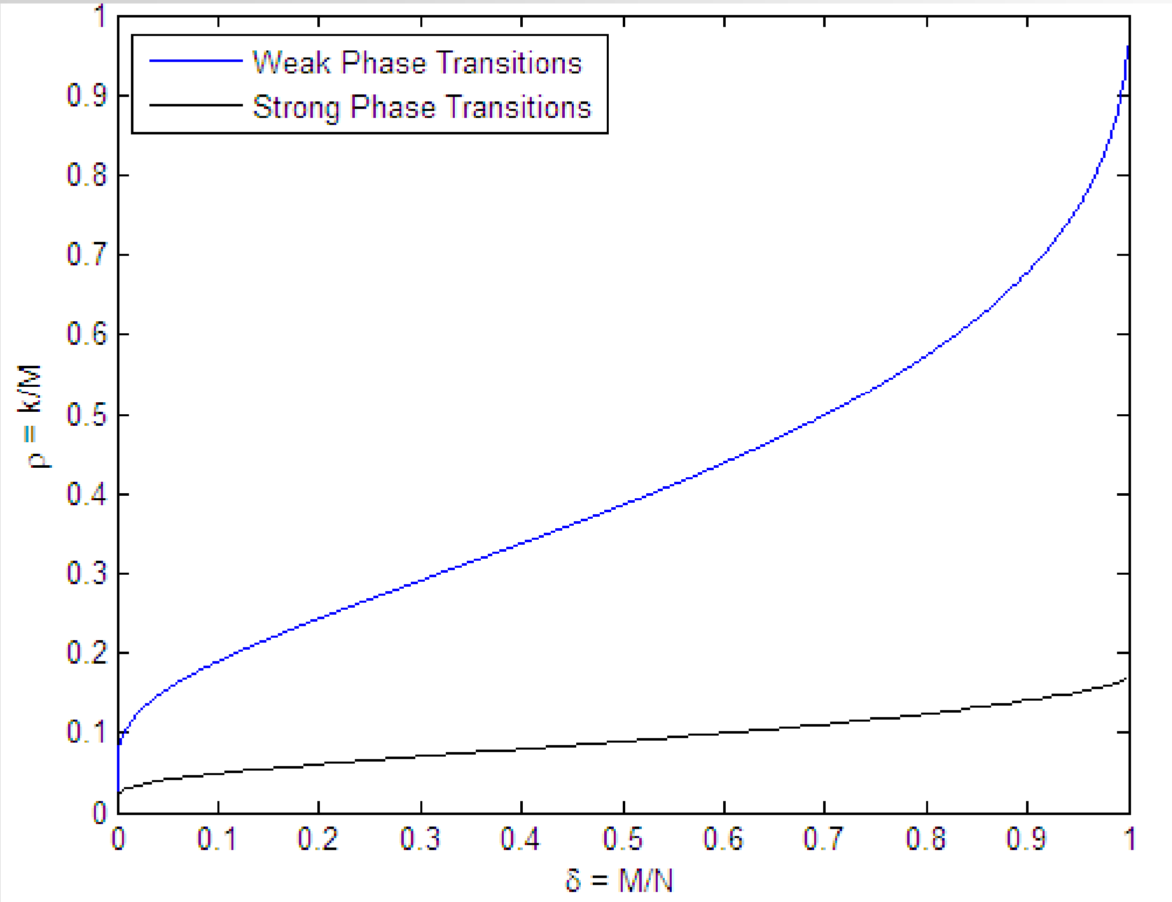
$$\rho = \frac{k}{M}$$

Compression trade-off

$$\delta = \frac{M}{N}$$

Under-sampling ratio

ANALYSIS AND COMPARISON



CROSSPOLYTOPE
Geometric Models of Sparsity

J. Tanner, “Regular Polytopes and Cone,” 2010. http://ecos.maths.ed.ac.uk/polytopes.shtml

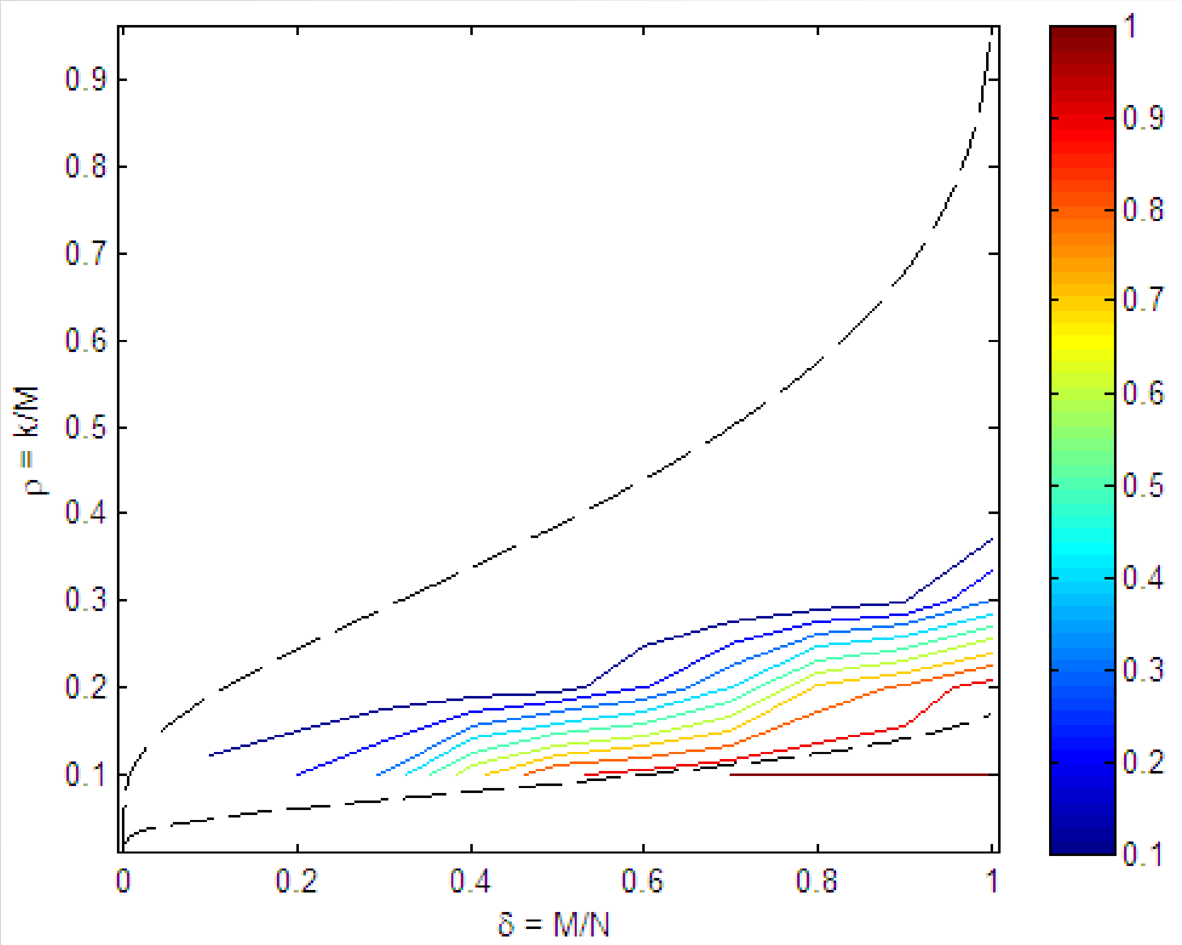
ANALYSIS AND COMPARISON

➢ SIMULATIONS

- N = 1000 and variation of k and M.
- Φ a Random Gaussian Matrix.
- Maximum of 1000 iterations with a stop criterion.
- Averages values for the probability of success over 100 trials.

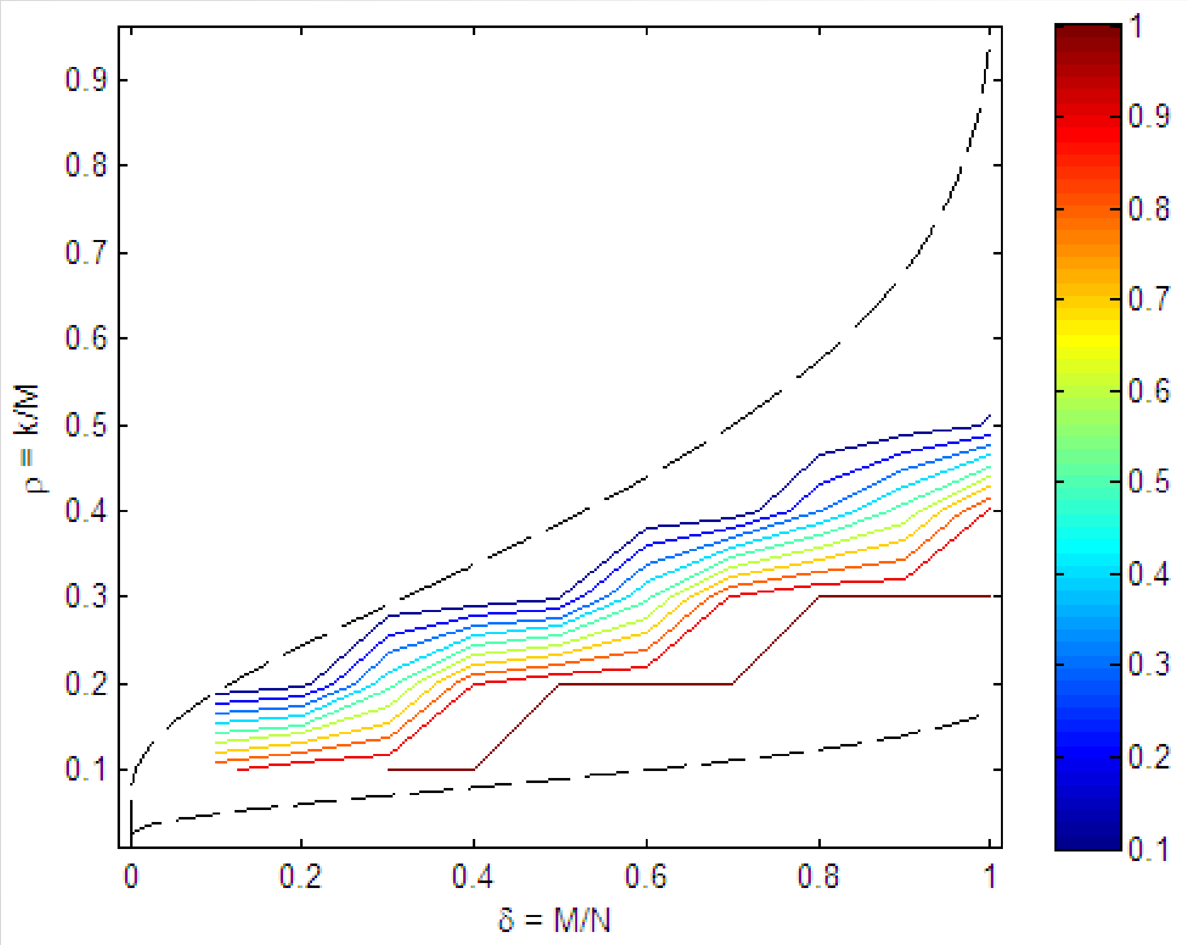
Success ➡ $$\frac{\|\hat{x} - x\|_2^2}{\|x\|_2^2} \leq 10^{-4}$$

ANALYSIS AND COMPARISON



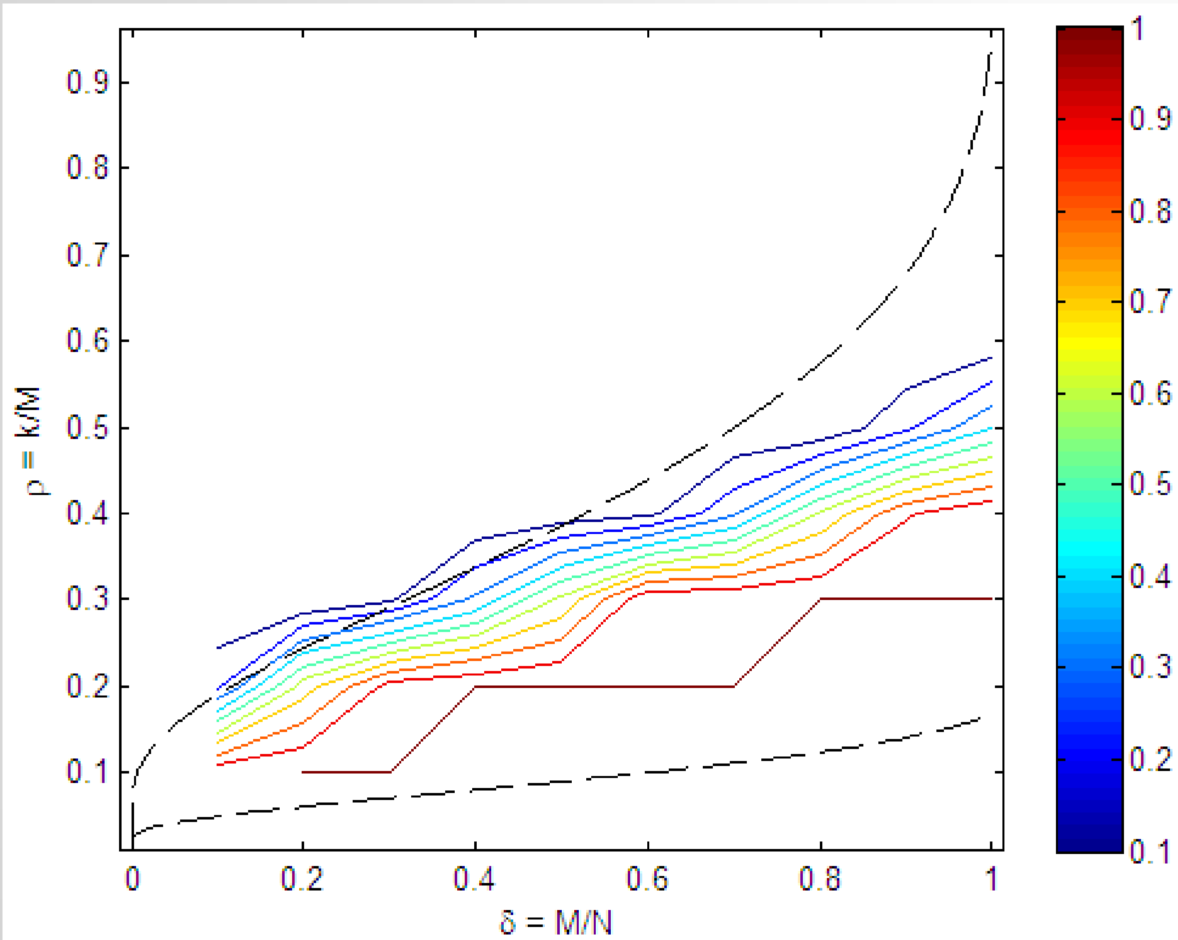
IHT

ANALYSIS AND COMPARISON



L I H T

ANALYSIS AND COMPARISON



F
L
I
H
T

CONCLUSIONS

- **Expected Results:**
  - Phase transitions for FLIHT approximate better to the theoretical curve than the ones for IHT and LIHT, thus providing a higher probability of success for lower values of δ and ρ.
  - Similar when comparing LIHT with IHT.
- **Important advance in CS.**
  - Optimal signal recovery with less measurements and more information.
  - Improve of IHT (basis of a great part of sparse signals reconstruction methods) without a significant increment of the computing and storage complexity of each iteration step.

CONCLUSIONS

- **Results dependency on the properties of the measurement matrix.**
  - If the matrix is not bounded, the algorithms diverge and solving the optimization problem becomes NP-hard.
- **Improvements**
  - Results could be further improved by decreasing the values of Lipschitz gradient constants.
    - ✓ Likely to have better convergence rates and a better approximation to the theoretical curve.
  - Increasing problem dimension.

THANKS FOR YOUR ATTENTION

QUESTIONS?