Distilled Sensing

Active sensing for sparse signal detection and estimation

Rui Castro

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

In collaboration with

Jarvis Haupt and Robert Nowak

CESS Workshop – 2/6/2009

A Sparse Signal Model

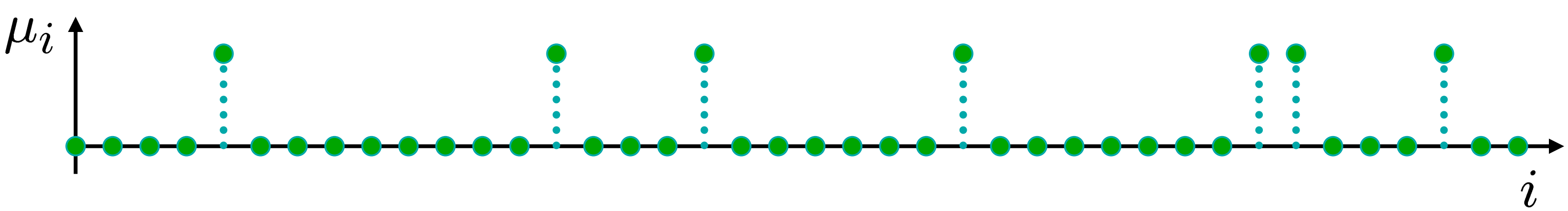
Sparse signal models are extremely useful is a variety of applications (e.g., image reconstruction, compression, etc.)

Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$ be a sparse vector

$$\mu_i = \begin{cases} 0 & i \in I_0 \\ \mu^* & i \in I_S \end{cases} \quad , \text{ where } |I_s| \ll |I_0|$$

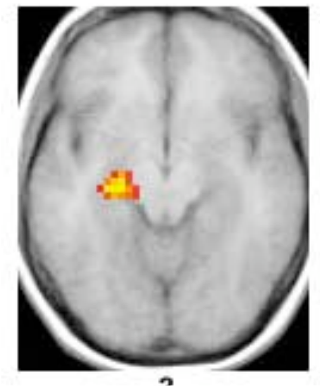
signal **support set**

Example:

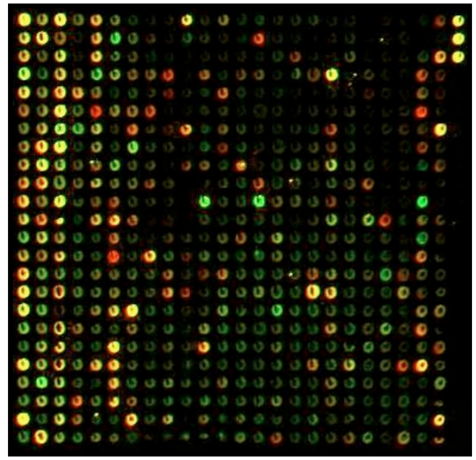


In this talk we will assume $\mu^* > 0$.

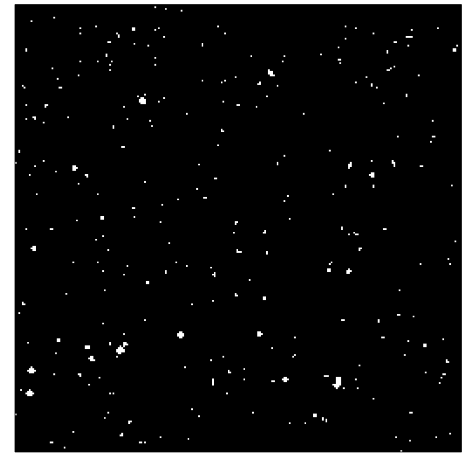
Detection/Estimation of Sparse Signals



fMRI data



Microarray data



Astronomical data

Two fundamental questions:

- Can we efficiently detect sparse signals?
- Can we locate sparse signals efficiently?

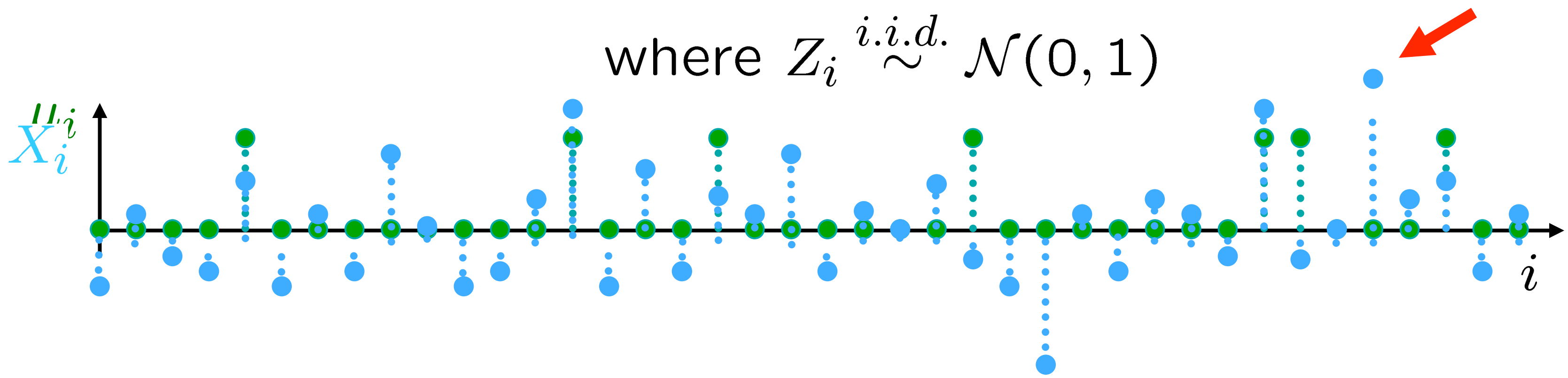
In this talk I will focus on the second question.

A Sparse Signal Model

Observation model:

$$X_i = \mu_i + Z_i, \quad i \in \{0, \ldots, n\},$$

$$\text{where } Z_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0,1)$$

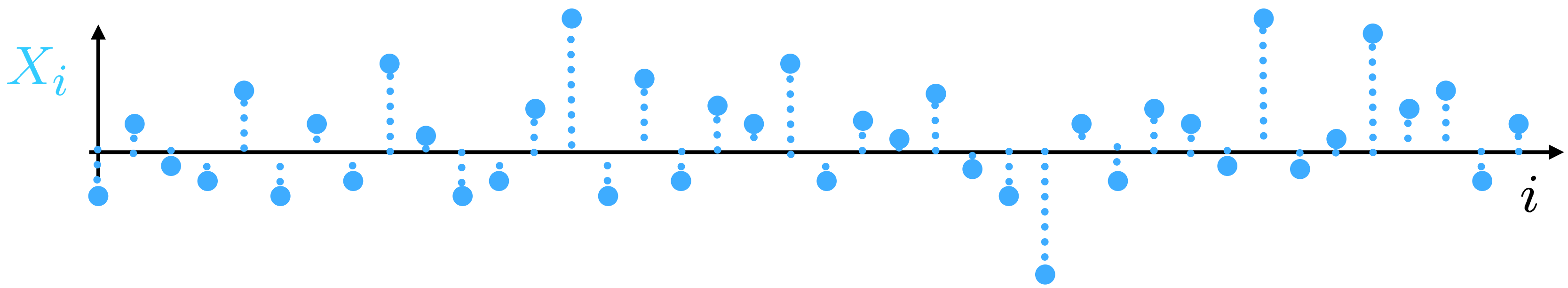


Intuitively signal components correspont to the largest observations...

Because of noise $\max_i X_i \sim \sqrt{2 \log n}$ even if no signal is present

$\boxed{\text{How small can } \mu^* \text{ be so that we can still reliably locate the signal components from the observations?}}$

A Sparse Signal Model



When testing a large number of hypotheses simultaneously we are bound to make errors!!!

Approaches:

- Control the probability of perfect localization of the support (**Bonferroni** correction) **<u>– very conservative</u>**

- Control the relative proportion of errors **(Benjamini & Hochberg '95)**

False Discovery Rate Control

Recall the definition of the **signal support set**

$$I_S = \{i : \mu_i \neq 0\}$$

Goal: Estimate the support as well as possible. Let $\widehat{I_S}(\boldsymbol{X})$ be the outcome of a support estimation procedure.

False **D**iscovery **P**roportion

$$\text{FDP} = \frac{\widehat{I_S}(\boldsymbol{X}) \setminus I_S}{|\widehat{I_S}(\boldsymbol{X})|} = \frac{\text{\# falsely discovered components}}{\text{\# discovered components}}$$

Non **D**iscovery **P**roportion

$$\text{NDP} = \frac{I_S \setminus \widehat{I_S}(\boldsymbol{X})}{|I_S|} = \frac{\text{\# missed components}}{\text{\# true components}}$$

Desirable situation: FDP,NDP ≈ 0

Since *n* is typically very large it makes sense to study **asymptotic** performance, as *n!* 1 .

Known Results (Jin & Donoho '03)

Assume the signal is very sparse:

$$|I_s| = n^{1-\beta}, \text{ where } \beta \in (0,1).$$

Number of signal components

Example: $\beta = 3/4$

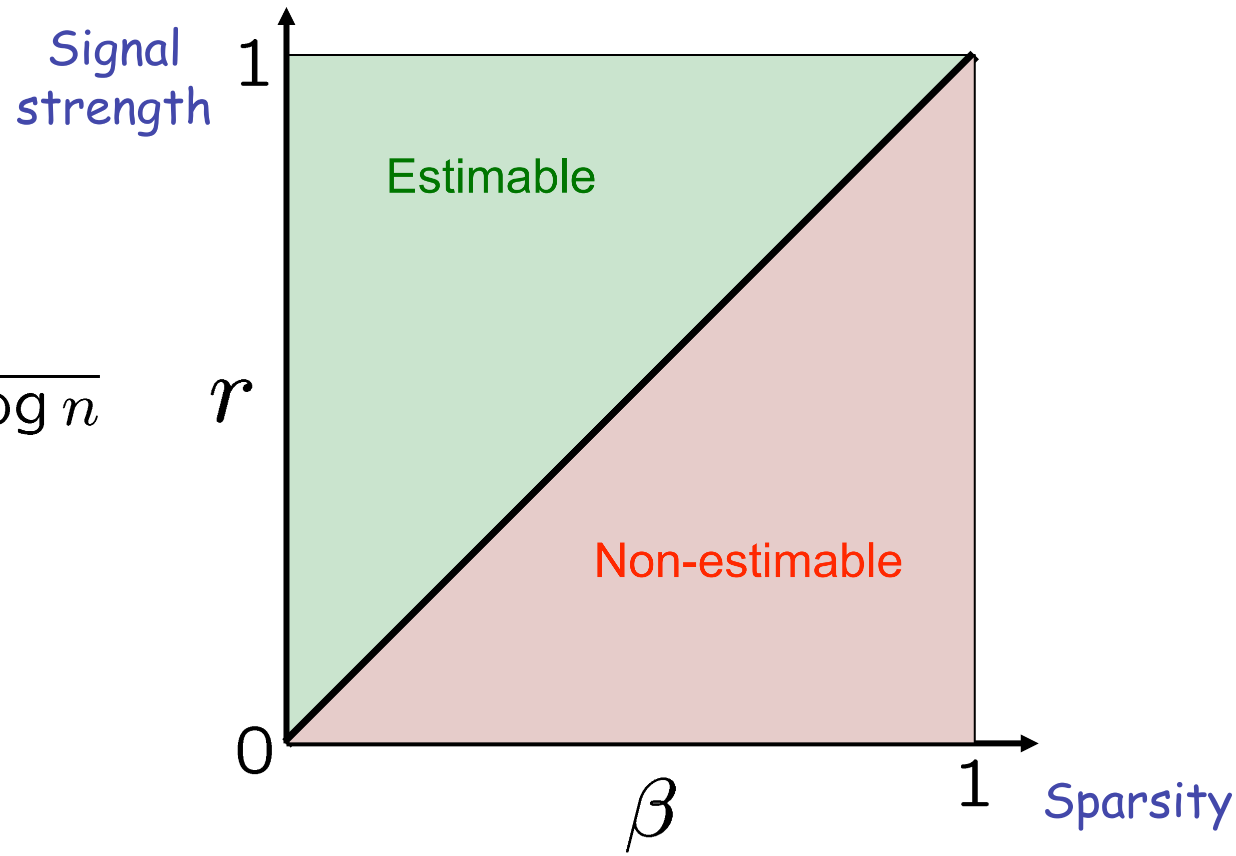
n = 10000) $|I_s|$=10

n = 1000000) $|I_s|$=32

Theorem: If $\mu^* > \sqrt{2\beta \log n}$ then Ben&Hoch thresholding applied to $\boldsymbol{X}$ drives both the FDP and NDP to zero with probability tending to one as $n \to \infty$. Conversely if $\mu^* < \sqrt{2\beta \log n}$ no procedure can control simultaneously the FDP and NDP.

Known Results (Jin & Donoho '03)

$$\mu^* = \sqrt{2r \log n}$$



$\#$ signal components$=n^{1-\beta}$

These **asymptotic** results tell us how strong the signals need to be for reliable signal localization

A Generalization of the Sensing Model

Allow multiple observations...

$$X_i^{(j)} = \phi_i^{(j)} \mu_i + Z_i^{(j)}, \qquad i \in \{1, \ldots, n\}$$

$$\text{where } Z_i^{(j)} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

...subject to a **sampling energy** budget

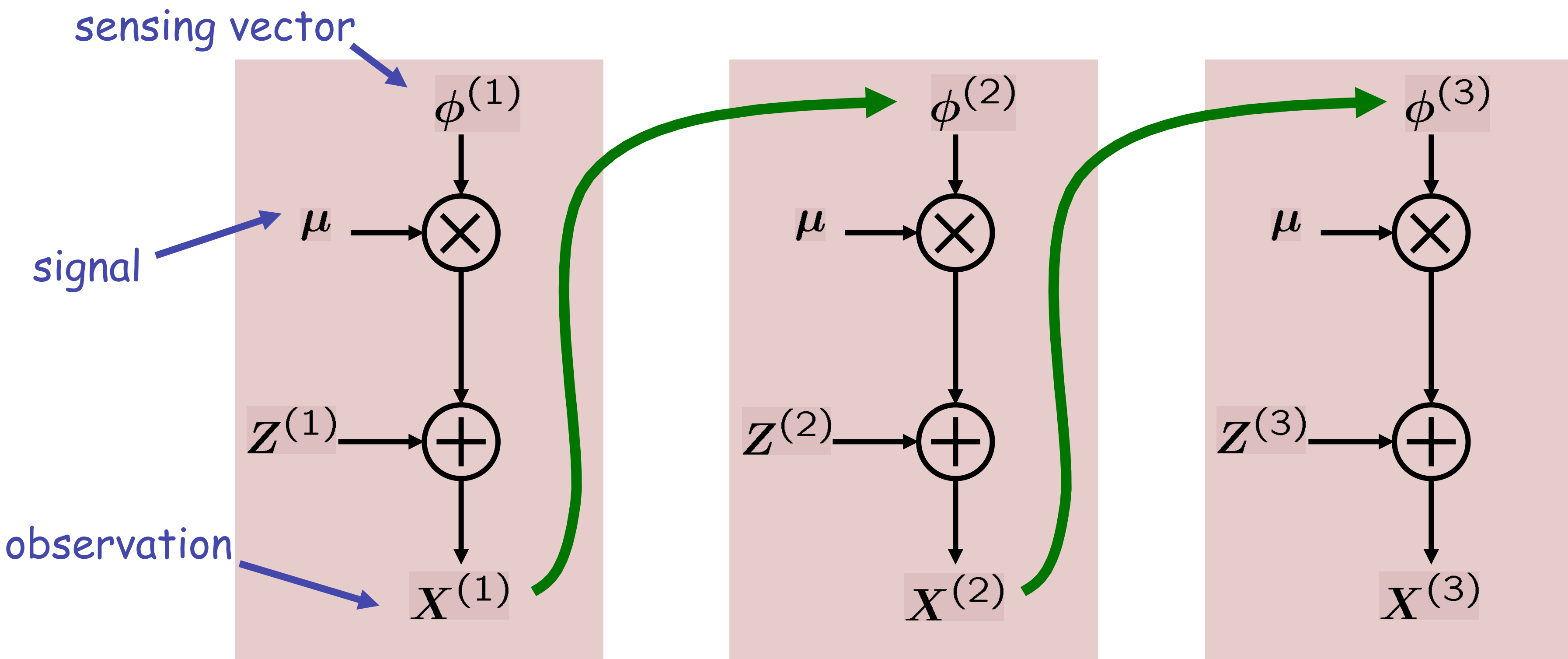
$$\sum_j \sum_{i=1}^n \left(\phi_i^{(j)}\right)^2 \leq n$$

$\boxed{\phi^{(j)} = (\phi_1^{(j)}, \ldots, \phi_n^{(j)}) \text{ are called the } \textbf{sensing vectors}.}$

(**Note:** in the previous work a single observation was considered, where $\phi_i^{(j)} = 1, \; i \in \{1, \ldots, n\}$)

Active Sensing

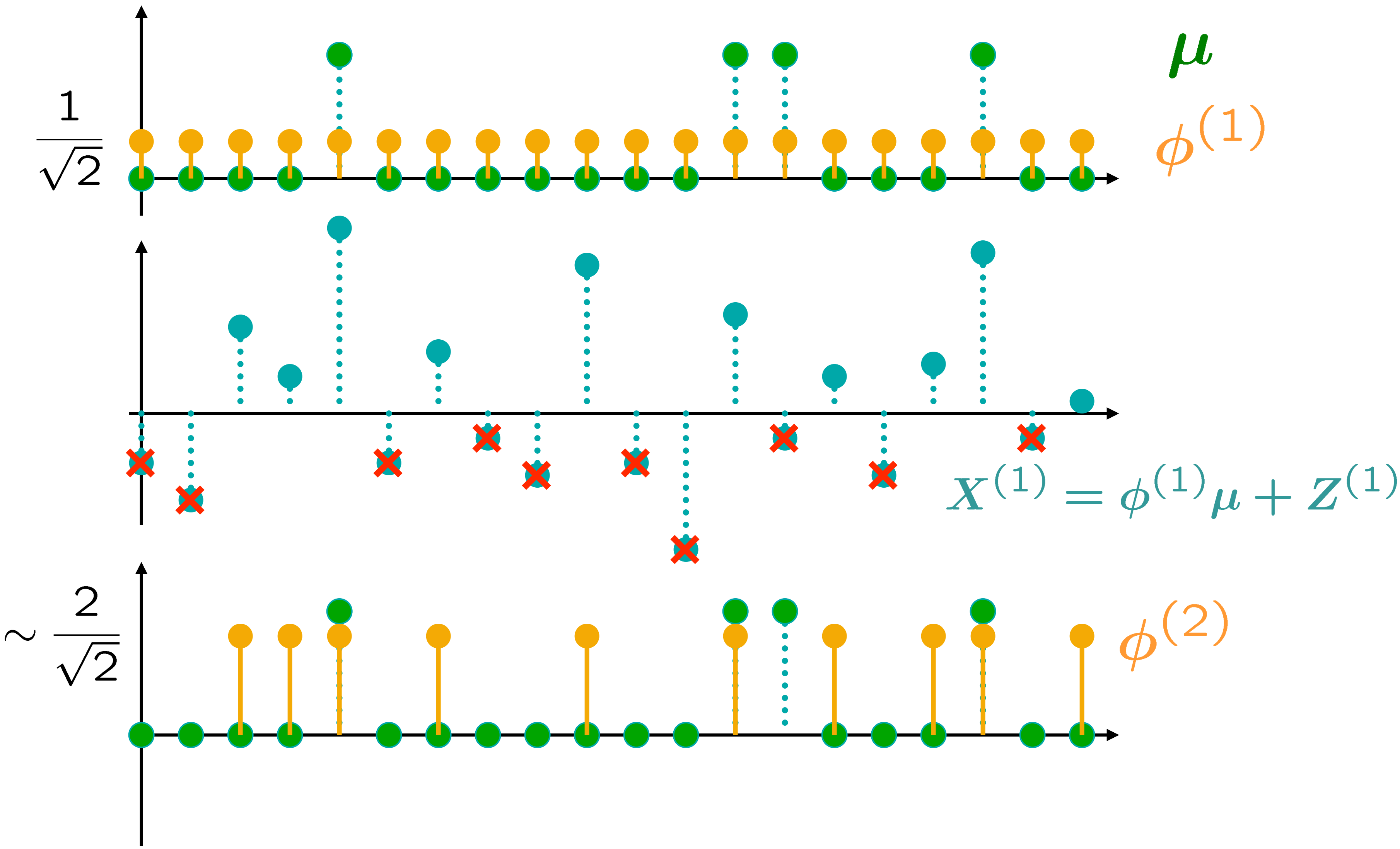
$$\sum_j \|\phi^{(j)}\|^2 = n$$



Key Idea: allow future sensing vectors to depend on past observations:

Dependence on previous observations allows us to focus the sampling energy in promising regions!!!

A Simple Focusing Procedure



Proceeding in this fashion we gradually **focus** on the signal...

Main Result – Part 1

Theorem 1 (J. Haupt, **RC** and R. Nowak)

Consider a $(k+1)$-step approach: at each step retain only the non-negative elements and sense at only those locations in next step (allocate equal fraction of sensing energy to each step). Then if

$$\mu^* > \sqrt{2\beta \frac{k+1}{2^k} \log n}$$

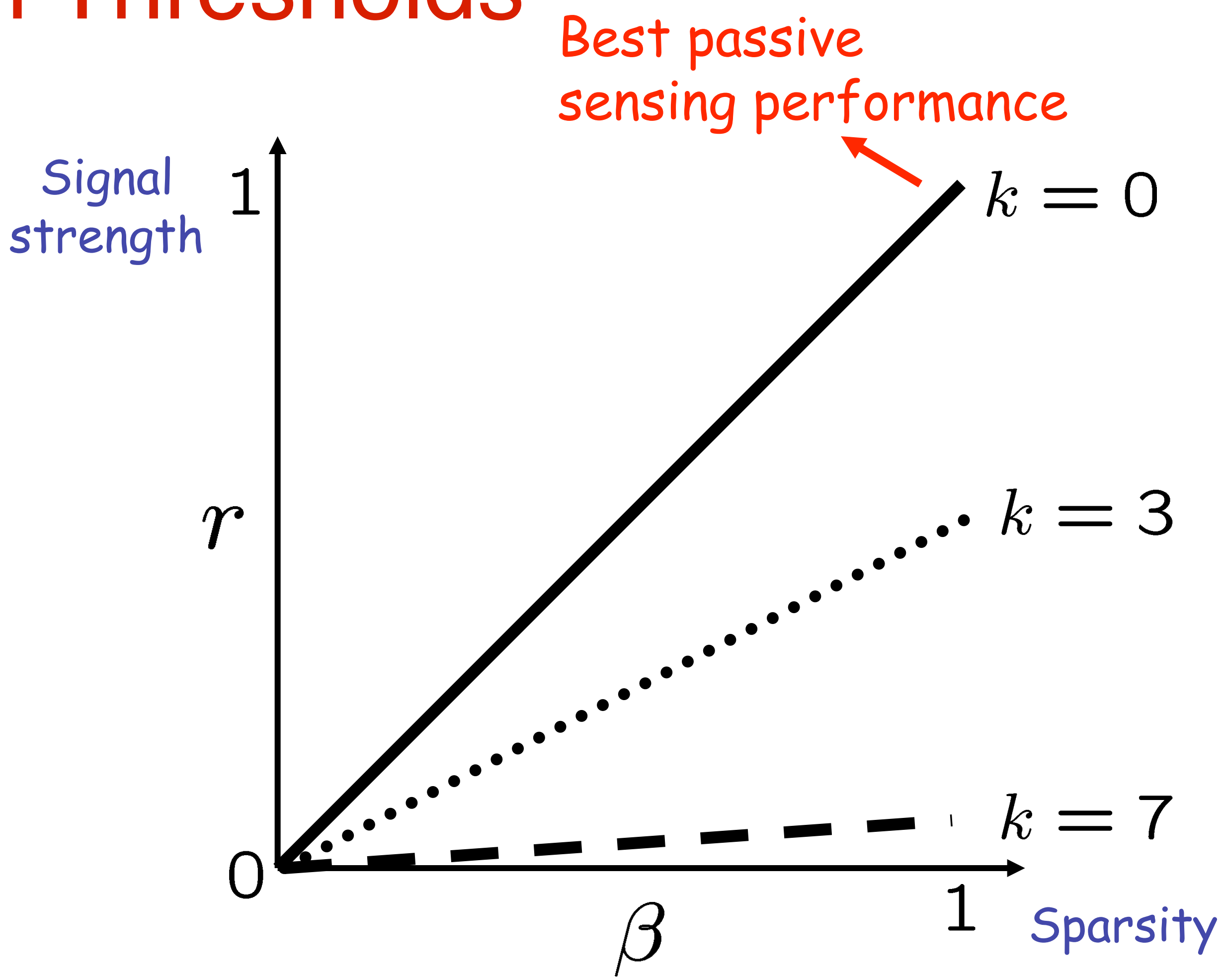
the BH thresholding procedure applied to $\boldsymbol{X}^{(k+1)}$ drives both the FDP and the NDP to zero with probability tending to one as $n \to \infty$.

Furthermore if one does not allow an active sensing scheme then the previous results (equivalent to k=0) cannot be improved.

The improvement is due to the use of feedback, not the fact that we have multiple measurements!!!

Localization Thresholds

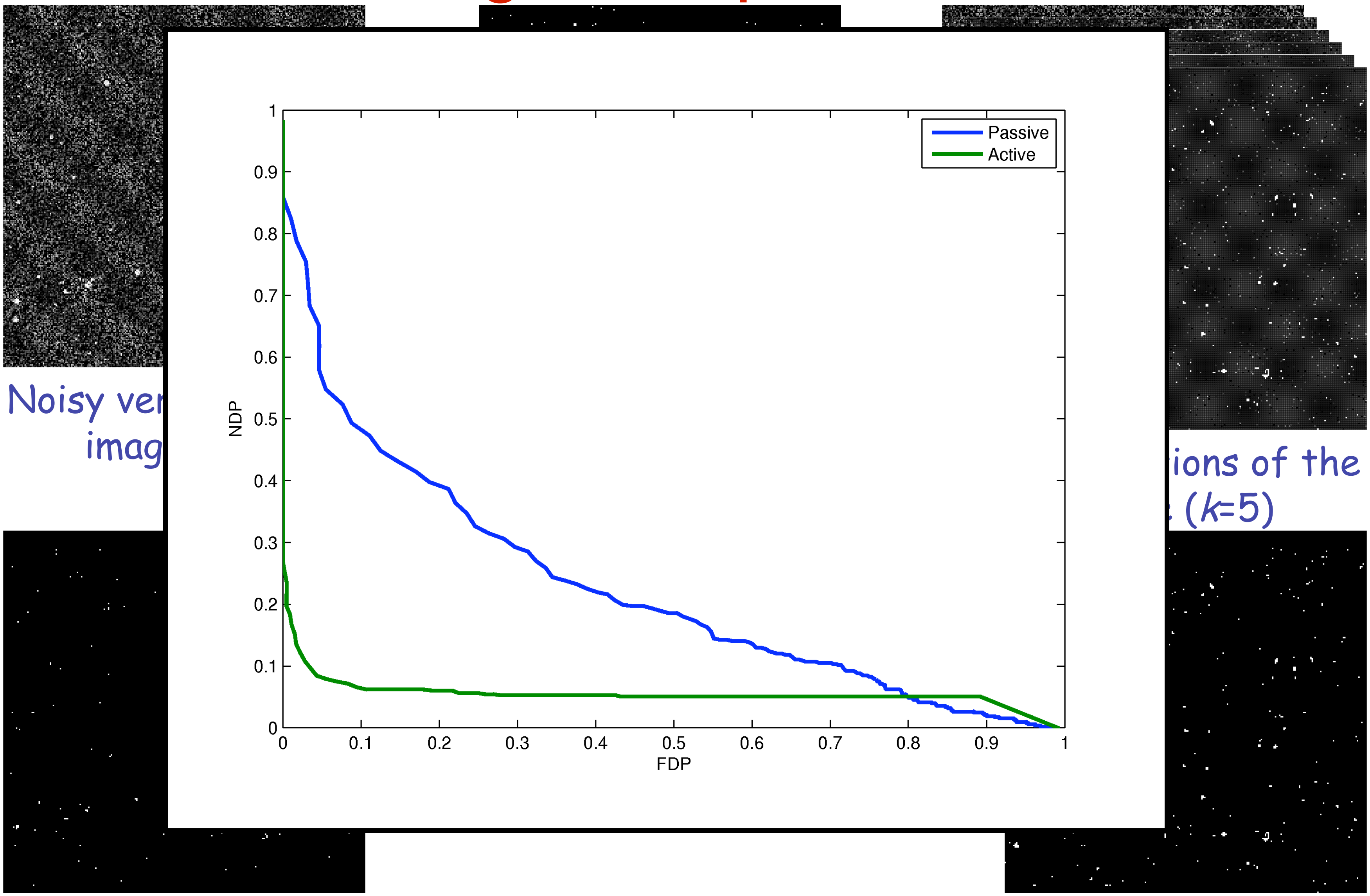
$$\mu^* = \sqrt{2r \log n}$$



\# signal components$=n^{1-\beta}$

These results suggest we might be able to estimate signal with amplitudes growing slower than $\mu^* \sim \sqrt{\log n}$

Distilled Sensing Example



Noisy ver
imag

ions of the
(*k*=5)

Passive sensing recovery
(FDR = 0.01)

Active sensing recovery
(FDR = 0.01)

Main Result – Part 2

Theorem 2 (J. Haupt, **RC** and R. Nowak)

Consider the $(k+1)$-look observation model, where $k(n) = \log_2 \log n$. If

$$\mu^* > \sqrt{2\beta \log_2 \log n}$$

a thresholding procedure applied to $\boldsymbol{X}^{(k(n)+1)}$ drives both the FDP and the NDP to zero with probability tending to one as $n \to \infty$.

The use of sampling feedback greatly decreases the signal strength needed for reliable localization!!!

Proof of Theorem 1 - Sketch

Main Idea: Quantify the effect of distillation j:

$m^{(j)} = |I_S \cap I^{(j)}|$ - Retained true signal locations.
$\ell^{(j)} = |I_0 \cap I^{(j)}|$ - Retained true non-signal locations.

Lemma:

$$\left(1 - \frac{1}{\log n}\right) m^{(j)} \ \leq \ m^{(j+1)} \ \leq \ m^{(j)}$$

$$\left(\frac{1}{2} - \frac{1}{\log n}\right) \ell^{(j)} \ \leq \ \ell^{(j+1)} \ \leq \ \left(\frac{1}{2} + \frac{1}{\log n}\right) \ell^{(j)}$$

with probability tending to one as $n \to \infty$

With high probability each distillation keeps almost and the signal components and rejects about half non-signal components

Proof of Theorem 1 - Sketch

To prove the theorem we need to iterate the application of the lemma. The event

$$\left\{ \begin{array}{l} \left(1 - \frac{1}{\log n}\right)^k n^{1-\beta} \leq m_{k+1} \leq n^{1-\beta} \\ \\ \left(\frac{1}{2} - \frac{1}{\log n}\right)^k n(1 - n^{-\beta}) \leq \ell_{k+1} \\ \\ \ell_{k+1} \leq \left(\frac{1}{2} + \frac{1}{\log n}\right)^k n(1 - n^{-\beta}) \end{array} \right\}$$

holds with probability tending to one as $n \to \infty$.

This implies that the final measurement of distilled sensing is essentially equivalent to the one-observation model where

$n_e \approx \frac{n}{2^k}$ - equivalent signal length

$r_e = r\frac{2^k}{k+1}$ - equivalent signal magnitude

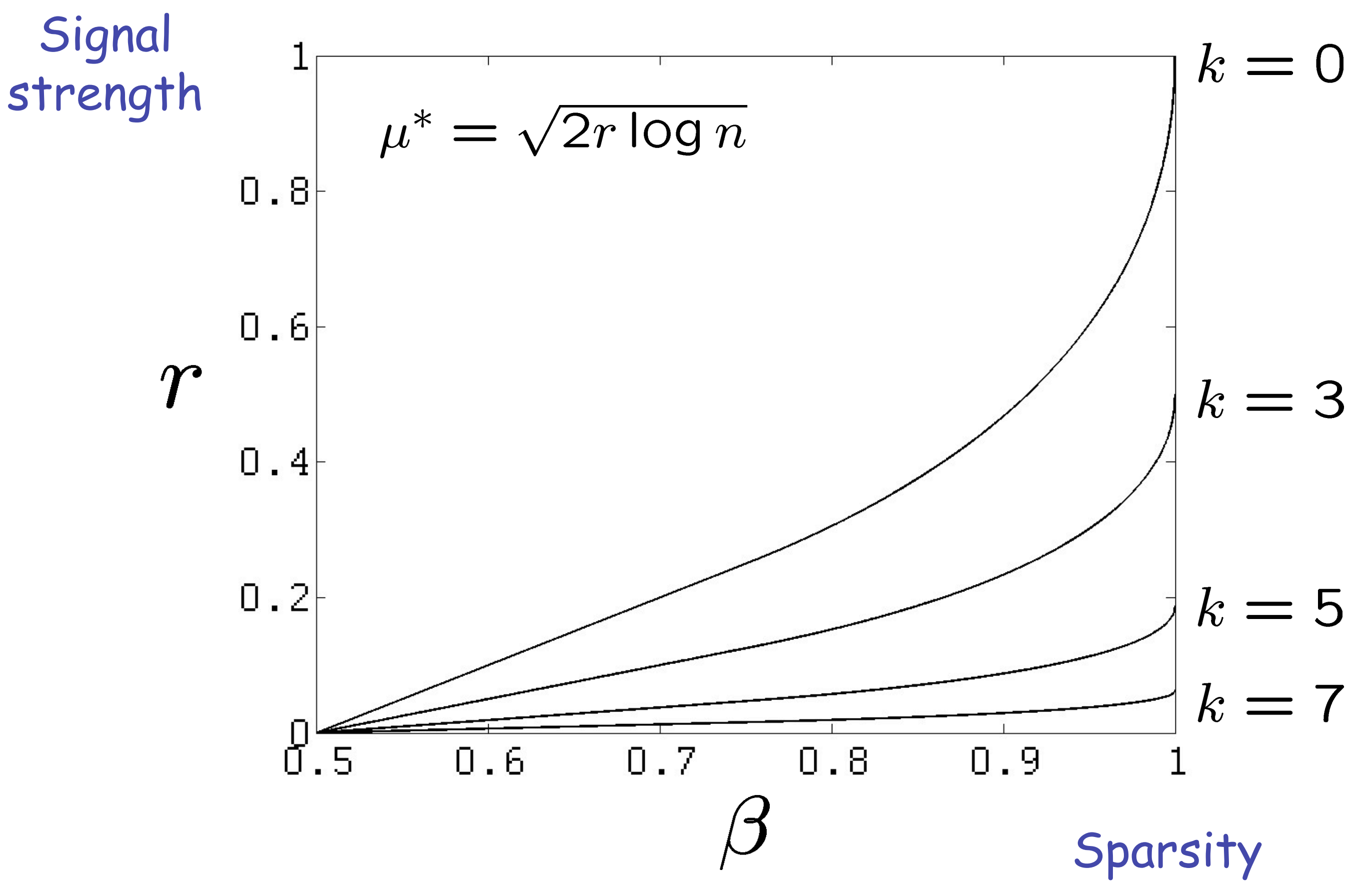
$\beta_e = \beta$ - equivalent signal sparsity

Using BH thresholding on the above signal yields the final result (from Donoho and Jin '03)

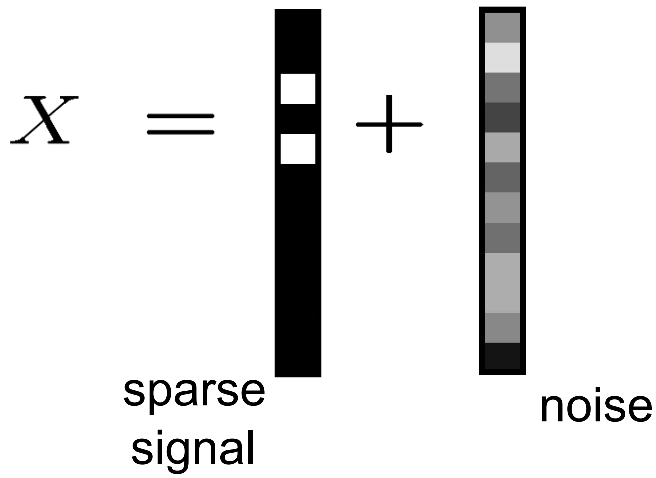
Extensions

The same ideas can be used to signal detection as well:

H_0 - No signal present $\Leftrightarrow \boldsymbol{\mu} = \mathbf{0}$
H_1 - A sparse signal is present



Now you see it, now you don't…

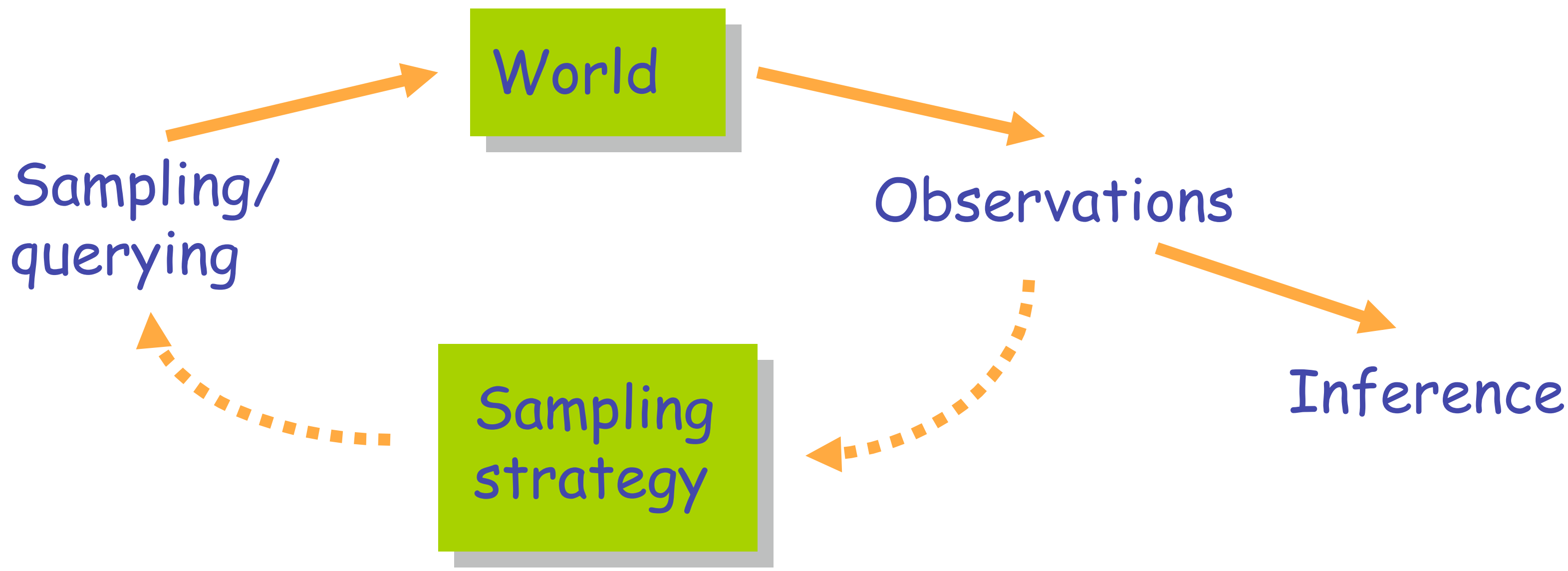


$n \times 1$ vector with $n^{1-\beta}$, $0 < \beta < 1$, non-zero entries of magnitude $\mu > 0$. Can the sparsity pattern be reliably perceived in presense of noise?

Passive sensing: Yes, if $\mu > \sqrt{2\beta \log n}$, otherwise no.
Active sensing: Yes, if $\mu > \sqrt{2\beta \log \log n}$.

Weak signals/patterns are imperceptible without active sensing!!!

Final Remarks



Closing the loop can yield dramatic gains allowing us to even perceive signals that were otherwise imperceptible!!!

rmcastro@ee.columbia.edu http://www.ee.columbia.edu/~rmcastro

Distilled Sensing (DS)

Make $k+1$ observations, where k is the number of *refinement* steps.

Allocate equal fraction of energy to each observation.

Input: Number of refinement steps k

Initialize: Index set $I^{(1)} = \{1, 2, \ldots, n\}$, $j = 1$

while: $j \le k+1$

$$X_i^{(j)} = \phi_i^{(j)} \mu_i + Z_i^{(j)}, \quad \phi_i^{(j)} = \sqrt{\frac{n}{(k+1)|I^{(j)}|}}$$

$$I^{(j+1)} = \{i : X_j(i) \ge 0\}$$

$$j = j+1$$

Output: $\left\{X_i^{(k+1)} : i \in I^{(k+1)}\right\}, \; I^{(k+1)}$

Now we can apply a threshold procedure to the output signal