Overview of Compressed Sensing

with recent results at INFONET Lab

Heung-No Lee, Ph.D.

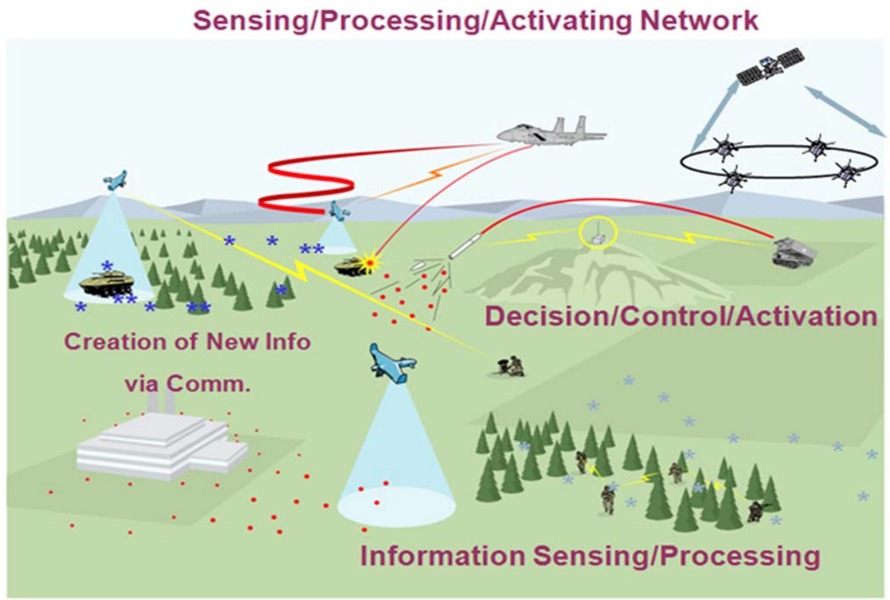
Associate Professor, GIST, KOREA
GIST, Korea
PSIVT 2011

Nov. 23rd, 2011

INFONET Lab.

http://infonet.gist.ac.kr/

- INFOrmation sensing/processing/controlling NETwork
- Wireless Communications, Wireless Networking, Information Theory, Channel Coding Theory, Signal Processing



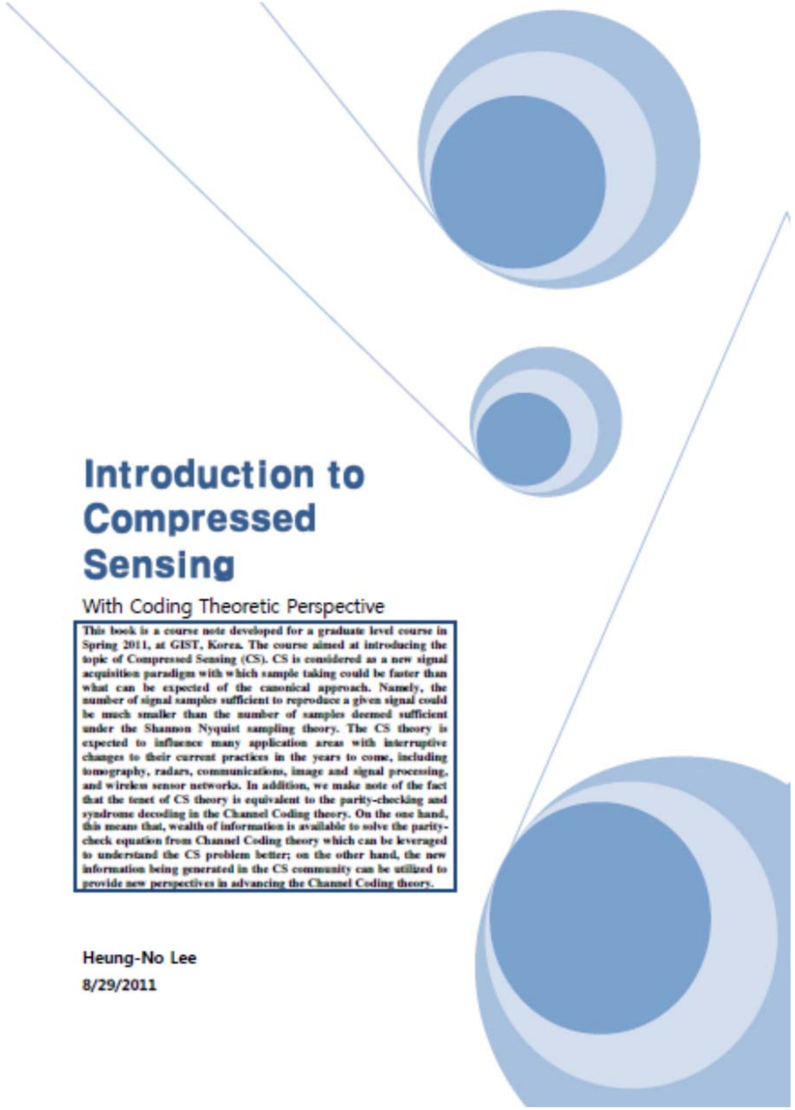
Agenda

- The Shannon-Nyquist Sampling Theorem
- Compressive Sensing
  - New sampling approach?
  - Compressive Sensing narrative
- Recent Results at INFONET
- Applications

- Most materials in this presentation are from the lecture note [Lee11].

Reference

- [Lee11] Heung-No Lee, *Introduction to Compressed Sensing with Coding Theoretic Perspective*, Lecture Note, Spring 2011, GIST, Korea. Available at http://infonet.gist.ac.kr/.


Introduction to Compressed Sensing

With Coding Theoretic Perspective

This book is a course note developed for a graduate level course in Spring 2011, at GIST, Korea. The course aimed at introducing the topic of Compressed Sensing (CS). CS is considered as a new signal acquisition paradigm with which sample taking could be faster than what can be expected of the canonical approach. Namely, the number of signal samples sufficient to reproduce a given signal could be much smaller than the number of samples deemed sufficient under the Shannon Nyquist sampling theory. The CS theory is expected to influence many application areas with interruptive changes to their current practices in the years to come, including tomography, radars, communications, image and signal processing, and wireless sensor networks. In addition, we make note of the fact that the tenet of CS theory is equivalent to the parity-checking and syndrome decoding in the Channel Coding theory. On the one hand, this means that, wealth of information is available to solve the parity-check equation from Channel Coding theory which can be leveraged to understand the CS problem better; on the other hand, the new information being generated in the CS community can be utilized to provide new perspectives in advancing the Channel Coding theory.

Heung-No Lee

8/29/2011

What Donoho said on Compressed Sensing

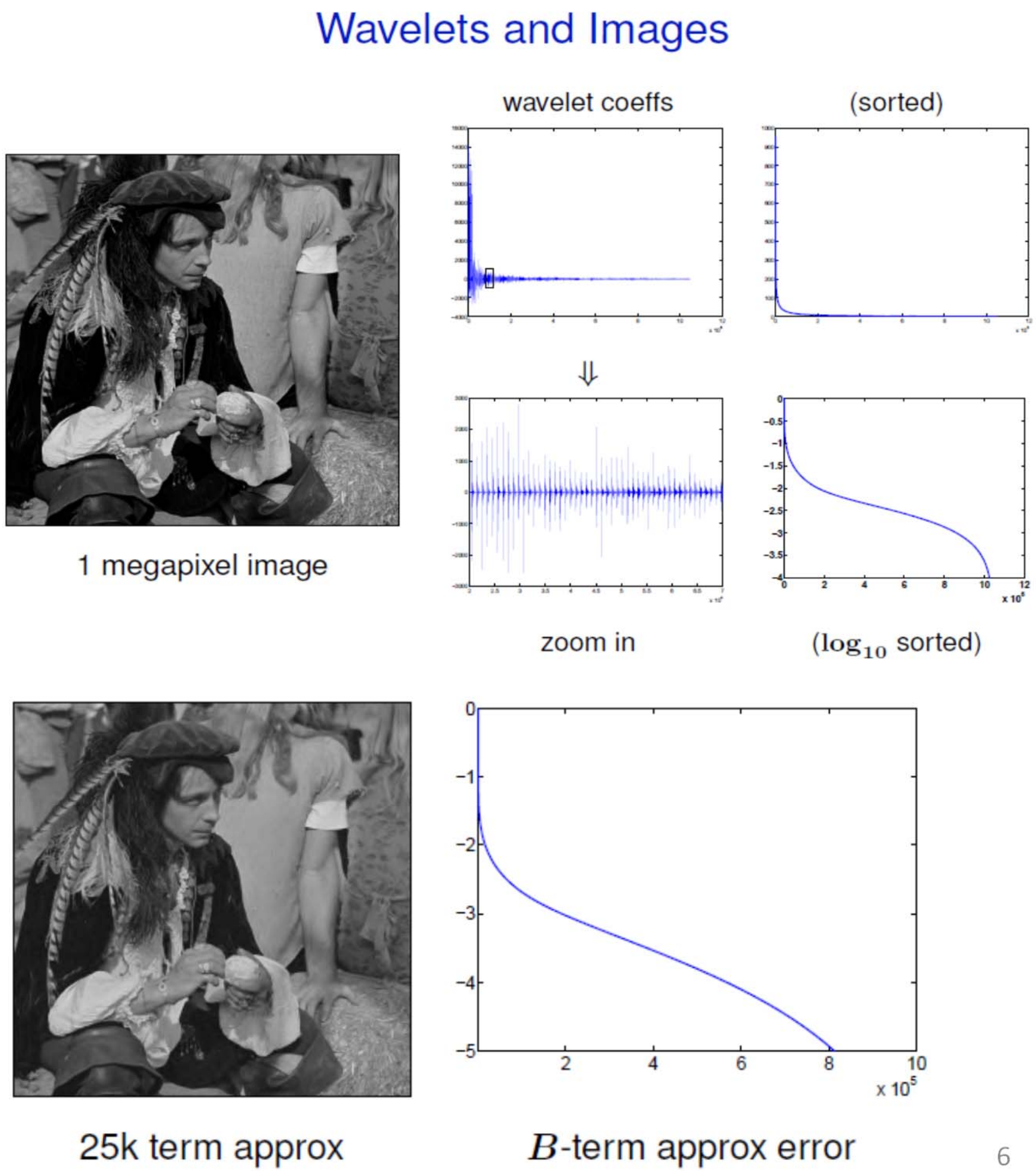
- in his paper Compressed Sensing, "everyone now knows that most of the data we acquire "can be thrown away" with almost no perceptual loss—witness the broad success of lossy compression formats for sounds, images, and specialized technical data. The phenomenon of ubiquitous compressibility raises very natural questions: why go to so much effort to acquire **all** the data when **most** of what we get will be thrown away? Can we not just **directly measure** the part that will not end up being thrown away?" [IEEE TIT 2006]

- in another one of his paper, "The sampling theorem of Shannon-Nyquist-Kotelnikov-Whittaker has been of tremendous importance in engineering theory and practice. Straightforward and precise, it sets forth the number of measurements required to reconstruct any bandlimited signal. However, the sampling theorem is wrong! Not literally wrong, but psychologically wrong. More precisely, it engender[s] the psychological expectation that we need very large numbers of samples in situations where we need very few. We now give three simple examples which the reader can easily check, either on their own or by visiting the website [SparsLab] that duplicates these examples." [Proc. IEEE, 2010]

On Line
CS Tutorials

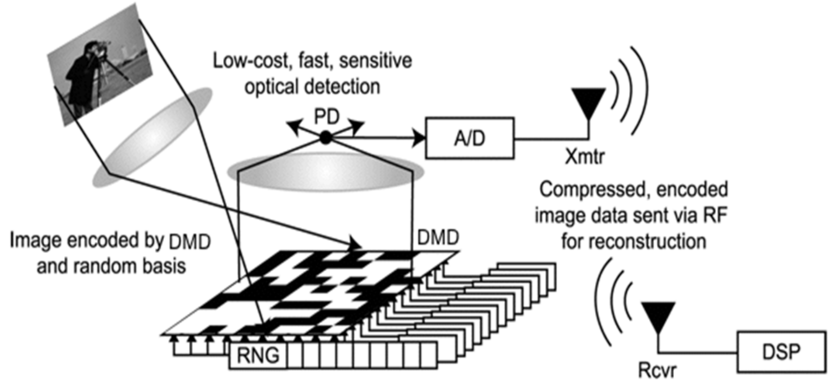
- Many CS Tutorials on line show results verifying what was said in [Donoho06]

- Charts from Romberg-Wakin's CS tutorial, 2007.

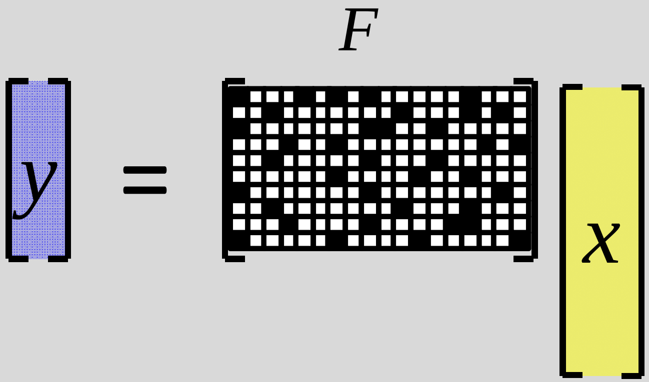


Background on CS

- Compressed sensing (CS)
  - New signal acquisition techniques [Donoho06], cited >4000 times.
  - MIT 2007 Tech Review, “Top 10 Emerging Technologies”
- CS is to find sparse solution from an under-determined linear system.
  - Real, complex field
- Many application areas: Cameras, Medical Scanners, ADCs, Radars, …



$$M < N$$



$$y \in \mathbb{R}^M$$
$$F \in \mathbb{R}^{M \times N}$$
$$x \in \mathbb{R}^N$$

$$\hat{x} = \arg\min_{\mathrm{x}} \|\mathrm{x}\|_{\ell_1} \quad \text{subject to } \mathrm{Ax} = y$$

What Shannon said on Dimension Reduction

Theorem 1. (Shannon's sampling theorem [Shannon48]) *If a function contains no frequencies higher than cps [cycles per second], it is completely determined by giving its ordinates at a series of points spaced seconds apart.*

Signal Dimension ~ $2TW$

Fig. 2—Reduction of dimensionality through equivalence classes.

- But Dimension Reduction from $2TW$ is possible!
- "In the case of sounds, if the ear were completely *insensitive to phase,* then the number of dimensions would be reduced by one-half due to this cause alone. The sine and cosine components and for a given frequency would not need to be specified independently, but only ; that is, the total amplitude for this frequency. The reduction in frequency discrimination of the ear as frequency increases indicates that a further reduction in dimensionality occurs. The vocoder makes use to a considerable extent of these equivalences among speech sounds, in the first place by eliminating, to a large degree, phase information, and in the second place by *lumping groups of frequencies together*, particularly at the higher frequencies." [Shannon48]

Compressed Sensing today

- [Donoho06] has been cited more than 4000 times!!
- RICE University CS repository http://dsp.rice.edu/cs
  - Many tutorials and talks
  - Hundreds of papers
  - Many MATLAB programs downloadable

- Today's focus
  - Review of David Donoho and his colleagues
    - Candes, Romberg, Tao, Baraniuk, …

Shannon’s Sampling Theorem

Shannon Nyquist Sampling Theorem

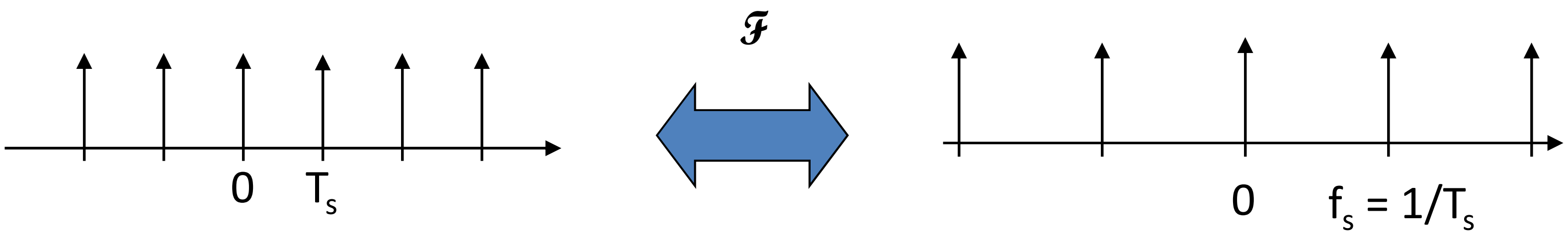
- Consider taking samples of continuous time signal.

- The Sampling Theorem: Any band-limited signals can be represented with the uniform spaced samples taken at a rate greater than twice the max. frequency of the signal.

- Proof: A train of impulses is a train of impulses in frequency

$$\sum_k \delta(t-kT_s) = f_s \sum_n \delta(f - n\, f_s)$$

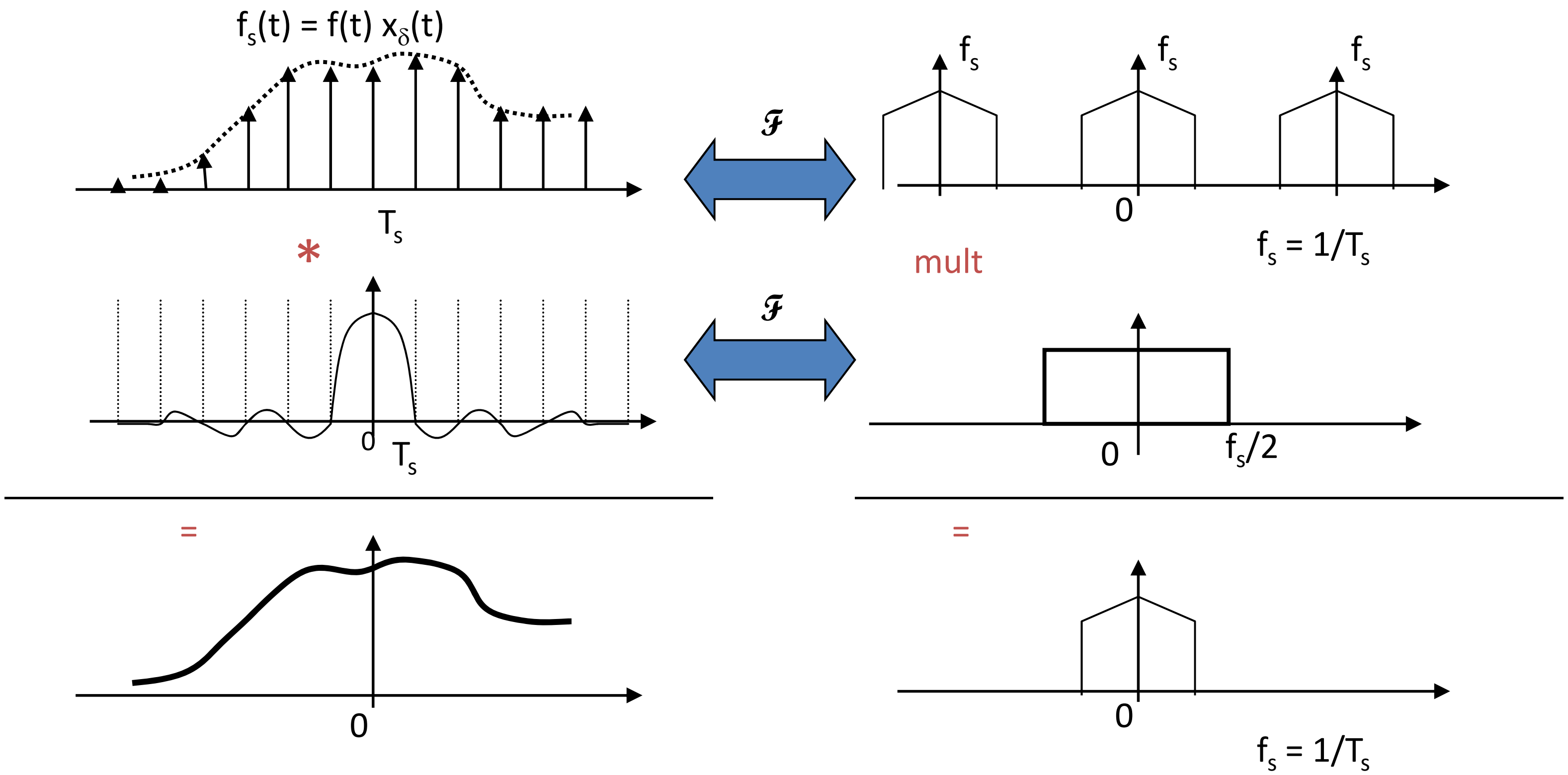
where $f_s = 1/T_s$



Shannon 1948 paper

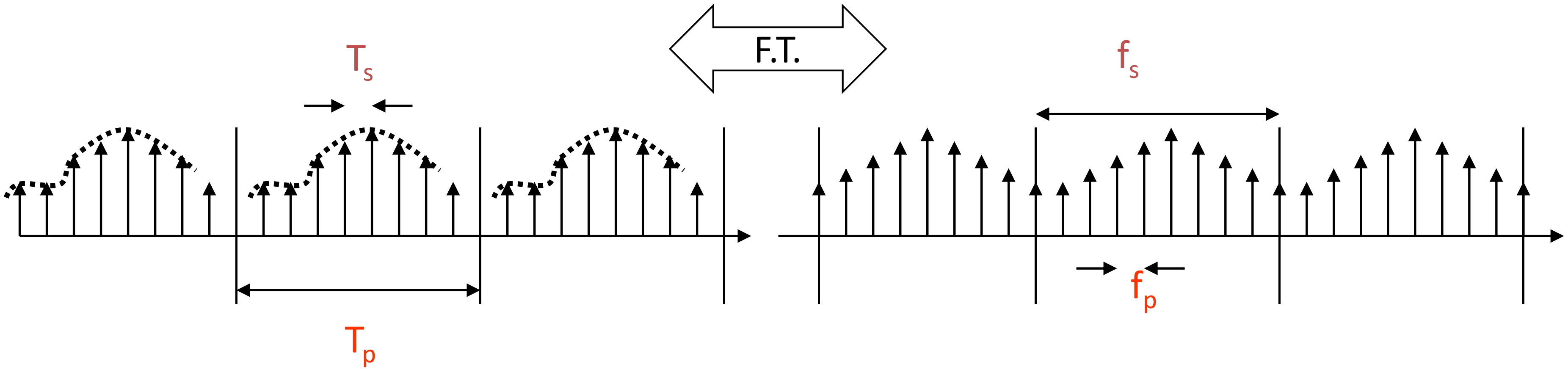
- Theorem 13: Let $f(t)$ contain no frequency over W. Then,

$$f(t) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2W}\right) \frac{\sin\left(2\pi W\left[t - \frac{n}{2W}\right]\right)}{2\pi W\left[t - \frac{n}{2W}\right]}.$$



F.T. vs. Discrete Fourier Transform

- Now, consider taking samples of a frequency spectrum at every f_p in the frequency-domain.
- Thus, in both domains we have periodic and sampled signals.
- Suppose $T_p/T_s = T_p W = N$, an integer.
- Then there are N distinct samples (in each domain).
- The discrete samples of the signal $x(k)$, $k=0, 1, 2, \ldots, N-1$.
- The discrete samples of the Fourier spectrum $X(n)$, $n = 0, 1, 2, \ldots, N-1$.



Discrete Fourier Transform

- DFT

$$X(k) = \sum_{n=0}^{N-1} x(n)\, e^{-j\,(2\pi/N)nk} \rightarrow \; X = Fx$$

- Inverse DFT

$$x(n) = 1/N \sum_{k=0}^{N-1} X(k)\, e^{j(2\pi/N)nk} \rightarrow x = F'X$$

- Using DFT, one can represent the time domain sequence x with the frequency domain sequence X.

- Note that in both domains, we have N signal samples.

What're given so far

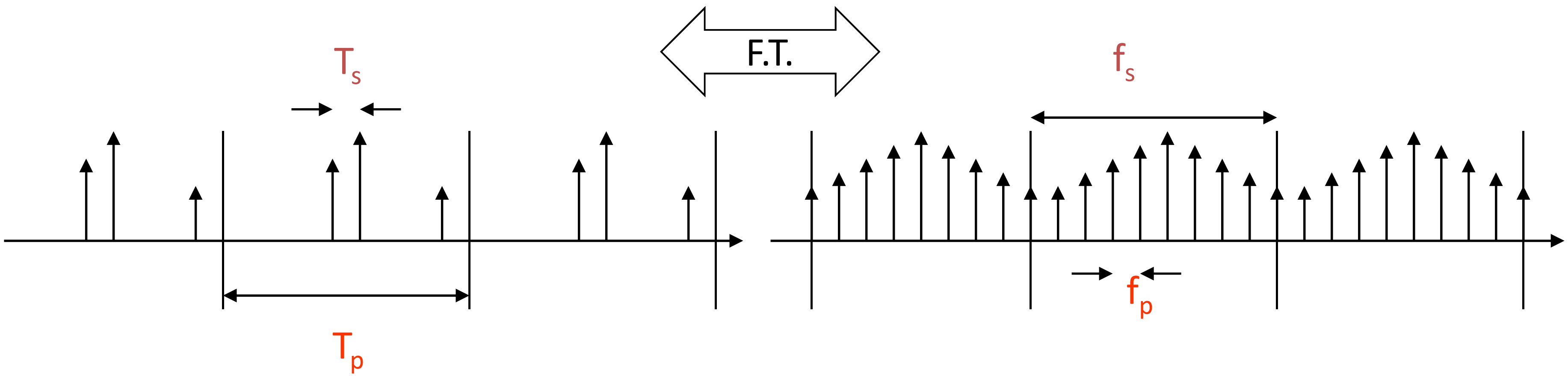
- Are covered in Systems and Signals in Electrical Engineering…

- Let's now move on to the issue of Compressive Sensing

Sparse Signals, Recovery with L1 Minimization

Sparse Signals

- Now suppose that the signal x is *K*-sparse.
  - Only *K* elements of x are non-zero (K << N)



The Big Question: Do we still need all *N* uniform spaced Fourier samples to represent the signal after knowing that it is sparse?

References

- Atomic Decomposition by Basis Pursuit [Chen, Donoho, Saunders 96]
- Uncertainty Principles and Ideal Atomic Decomposition [Donoho01]
- Neighborly Polytopes and Sparse Solution of Underdetermined Linear Equations [Donoho04]
- Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information [Candes, Romberg, Tao 06]
- Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? [Candes, Tao 06]
- Many other papers available at http://dsp.rice.edu/cs.

- These guys say there is a better way to represent the sparse signal!

M<N is good enough!

- DFT again:

  $X(k) = 1/N \ \sum_{n=0}^{N-1} x(n)\, e^{-j(2\pi/N)\, n\, k}$ for all k=0,1,2, …, N-1

  $\mathbf{X} = \mathbf{F}\mathbf{x}$

- We know **x** is sparse. Then,

- Taking only several Fourier l.p. measurements of x(n) is good enough :

  $y(m) = 1/N \sum_{n=0}^{N-1} x(n)\, e^{-j\,(2\pi/N)\, n\, m}$

  $= \langle x(n),\ m\text{-th tone}\rangle$ for m=1, 2, …, M

  where *M* *is the total number of measurements.*

- *We let* $\mathbf{y} = \mathbf{F}_{\mathcal{M}}\mathbf{x}$ be *a subset of* Fourier coefficients of signal **x**,
  where size of the subset $\mathcal{M} \subset \mathbb{Z}_N$ is $M < N$.

- ***M <N* Fourier samples are good enough to represent x!**

Key CS Results

- The number of measurements M required for successful recovery is different under different solution criterion.

- (P0) A K-sparse signal $\mathbf{x}$ can be recovered using the exhaustive search (L_0 min search) [Theorem 1.1, CRT 06].

$$\min \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{F}_{\mathcal{M}}\, \mathbf{x}$$

$$M \geq 2K \Leftrightarrow \text{unique solution}$$ (Prime N)

Key CS Results (2)

- As long as the solution is unique, the L_0 min search finds it exactly.

- Proof: Suppose two K-sparse solutions $\mathbf{x}$, $\mathbf{x}'$. Then, we have

$$\underbrace{\mathbf{F}_{\mathcal{M}}}_{\text{An } M \text{ dimensional map}} \underbrace{(\mathbf{x} - \mathbf{x}')}_{\text{A vector of dimension } <2K} = 0$$

  – This map is injective. Thus, RHS can't be zero unless ($\mathbf{x}$-$\mathbf{x}'$)=0.

Canonical CS Results (3)

- There are ${}_{N}C_{K}$ different ways to choose a set of K columns that accounts for the observation $\mathbf{y}$.
- The complexity of L_0 min search is $\binom{N}{K} \cong 2^{N\,H\left(\frac{K}{N}\right)}$

- (P1) A relaxed approach is L_1 minimization:

$$\min \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{F}_{\mathcal{M}}\,\mathbf{x}$$

$$M \geq cK \log N \Rightarrow \text{unique L1 sol.} = \text{L0 sol.}$$

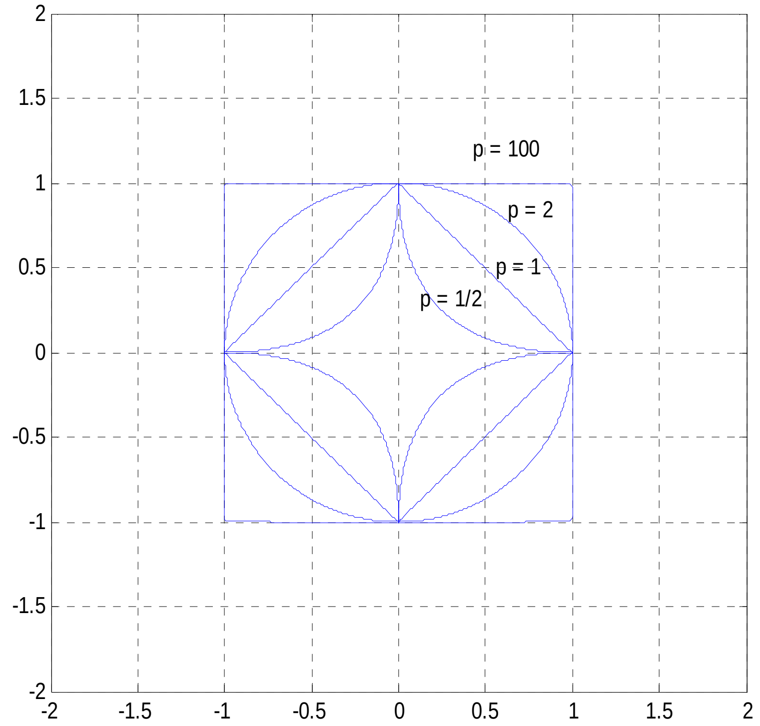
L1 norm?

- The L-p norm of $\mathbf{x}$ is defined for p>0

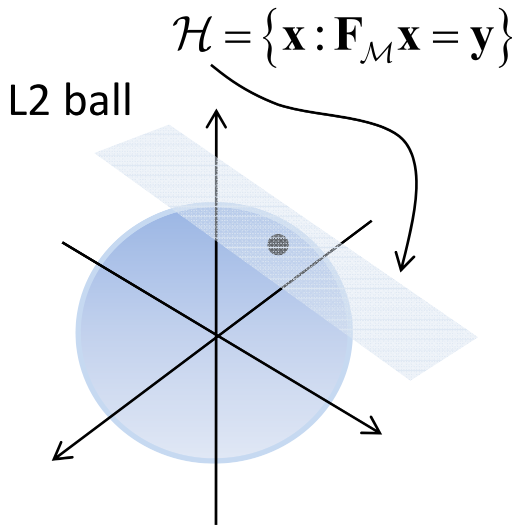
$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^{N} |x_i|^p \right)^{\frac{1}{p}}$$

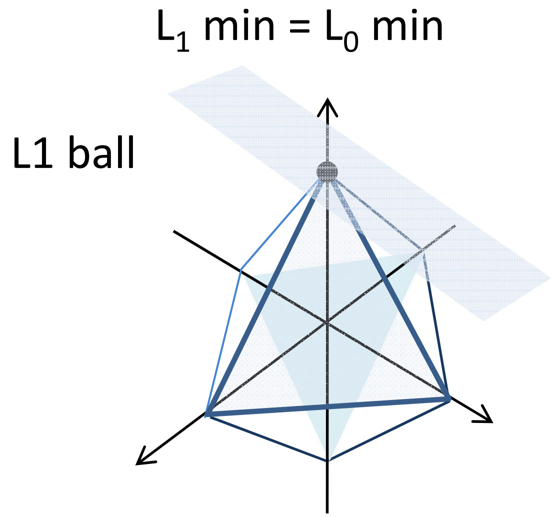
- The L0 norm is not well defined as a norm.
  - Donoho uses it as a "norm" which counts the number of non-zero elements in a vector.

- Let $\mathbf{x} = [1 \quad -1 \quad 0]$ and $\mathbf{y} = [0.5 \quad -0.5 \quad 0.5]$ Which one is bigger?
  - L0 sense
  - L1 sense
  - L2 sense



L_1 vs. L_2 Solution





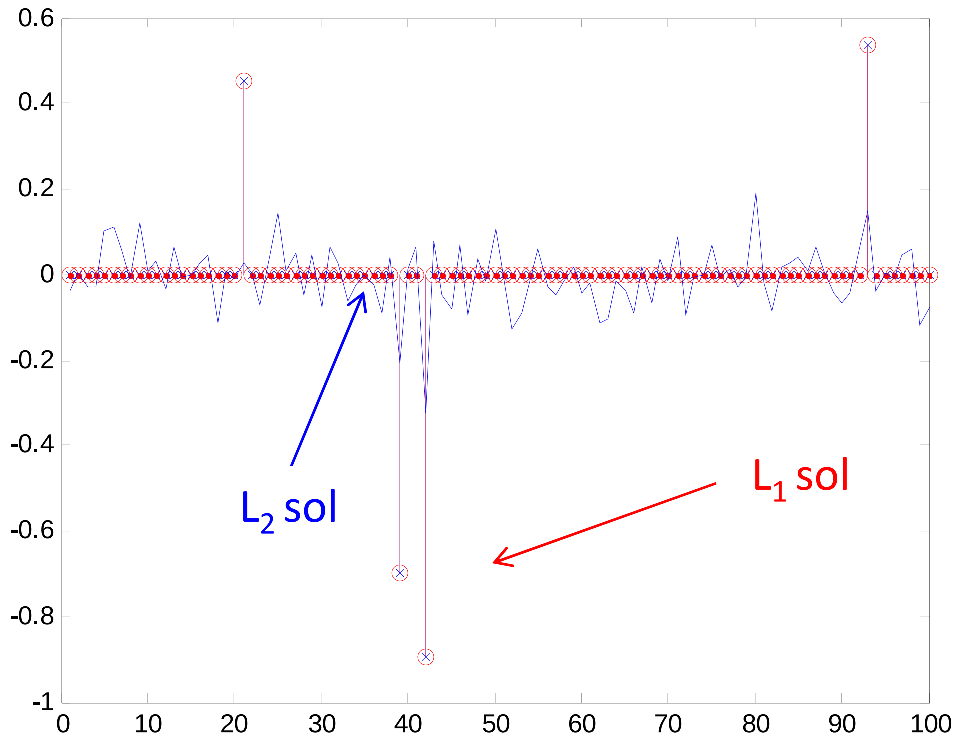
$$x = \arg\min \|x'\|_2 \quad \text{s.t. } y = Fx$$
$$= F^T \left(FF^T\right)^{-1} y$$

$$x = \arg\min \|x'\|_1 \quad \text{s.t. } y = Fx$$

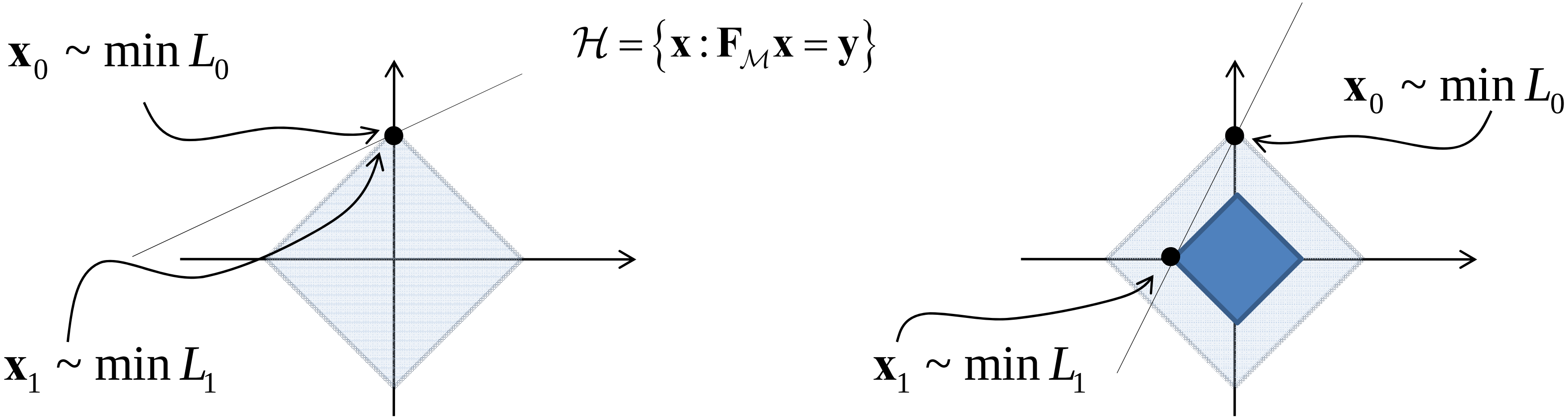
L2 is not suitable but L1 is when the exact solution is sparse.

L_2 vs. L_1 solutions

- L2 solution has energy spread out to everywhere.
- L1 solution attains the sparse signal.



Good vs. Bad



- When the hyperplane cuts through the L_1 ball, L_1 min does not attain the L_0 min.

- We aim to make $\mathbf{F}_{\mathcal{M}}$ so that the bad does not occur (often).

"Uniform Uncertainty Principle"

[Candes, Tao 06]

- If $M \geq cK\lambda$, then for any K-sparse signal x, the following inequality holds with probability close to 1,

$$\frac{1}{2}\frac{M}{N}\|x\|_2^2 \leq \|F_{\mathcal{M}}x\|_2^2 \leq \frac{3}{2}\frac{M}{N}\|x\|_2^2 .$$

- For the Fourier matrix, the bad case won't happen very often if

$$\lambda = (\log N)^6$$

Sparse Representation, Uncertainty Principle, Sparse Signal Recovery with L1 Minimization

Uncertainty Principle for Sparse Representation

[Donoho01]

- Heigenberg's uncertainty principle (UP):
  - momentum and position of a particle cannot be simultaneously determined precisely.

$$\boxed{\Delta p\, \Delta x \geq h}$$

- In sparse representation where the goodness lies in parsimonious representation of a signal of interest, there is an UP as well.
- Suppose a signal x which can be represented by a basis A with sparsity K_A and by a basis B with sparsity K_B. That is,
  - $x = As_A$, sparsity of s_A is K_A
  - $x = Bs_B$, sparsity of s_B is K_B.
- Then,

$$\boxed{K_A K_B \geq \frac{1}{\mu^2}, \quad or \quad \left(K_A + K_B \geq \tfrac{2}{\mu}\right)}$$

where $\mu := \max_{i,j} \left\{ \left| \left\langle a_i, b_j \right\rangle \right| \right\}$ and a_i and b_j are the columns of A and B resp.

- **A signal cannot be sparsely represented in both domains!**

UP and L1 Recovery

- Donoho-Stark 89' then suggest the use of a combined matrix, a dictionary, and of the L1 min routine to find sparse representation of x:

$$(PD) \text{ Find the most sparse representation } s,$$
$$\text{given a signal } x = Ds, \text{ using the dictionary } D = \left[A;\ B\right].$$

- This will be useful when one does not know which basis is more suitable for representing the signal.

- Using the UP, they show that
  - If $\|s\|_0 \le 1/\mu$ and $x = Ds$, then the solution is unique (L0 solution unique).
  - If $\|s\|_0 \le \frac{1}{2}\left(1+1/\mu\right) < 1/\mu$ and $x = Ds$, then the L1 finds the exact solution.

- These classic works done in 80s and 90s provide the foundation for the Compressed Sensing theory.

L1 Minimization Algorithms

- Linear program!
- Basis Pursuit (Chen, Donoho, Saunders 95')

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{s.t.} \quad y = Fx$$

- Recast as an LP

$$\min_{(x,u)} \sum_i u_i = 1^T u + 0^T x = \begin{bmatrix} 0^T & 1^T \end{bmatrix} \begin{pmatrix} x \\ u \end{pmatrix}$$

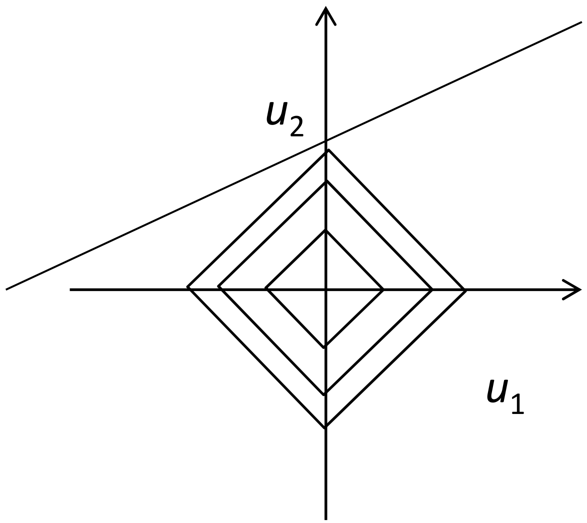
$$\text{s.t.} \quad x - u = \begin{bmatrix} e & -e \end{bmatrix} \begin{pmatrix} x \\ u \end{pmatrix} \preceq 0,$$

$$-x - u = -\begin{bmatrix} e & e \end{bmatrix} \begin{pmatrix} x \\ u \end{pmatrix} \preceq 0,$$

$$\begin{bmatrix} F & 0 \end{bmatrix} \begin{pmatrix} x \\ u \end{pmatrix} - y = 0$$

  - There are many ways to solve this LP problem.
  - L1 magic (Candes-Romberg)
  - CVX (Boyd-Vandenberghe)
  - SparseLab
  - Many others at RICE CS repository

- Primal-Dual Interior Point Method
  1. Write the KKT equation
  2. Linearize it (Newton's method)
  3. Solve for a step direction
  4. Adjust the step size (stay interior : $u > 0, \lambda > 0$)
  5. Iterate until convergence



L1 Minimization Algorithms (2)

- The LP approach is to build the sparse solution from an initial guess which is dense.
  - $O(N^3)$
- If the exact solution is known to be sparse, why don't we start from a null set and build up a sparse solution?
- **Homotopy** [Donoho-Tsaig08']

$$\min_x \ \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda \|x\|_1$$

  - The correct solution is approached when lambda gets smaller.
  - Osborne et al. 2000
  - Tibshirani's LASSO 96'
- K-step property: It finds the solution in K-step if

$$K \le (\mu^{-1}+1)/2.$$

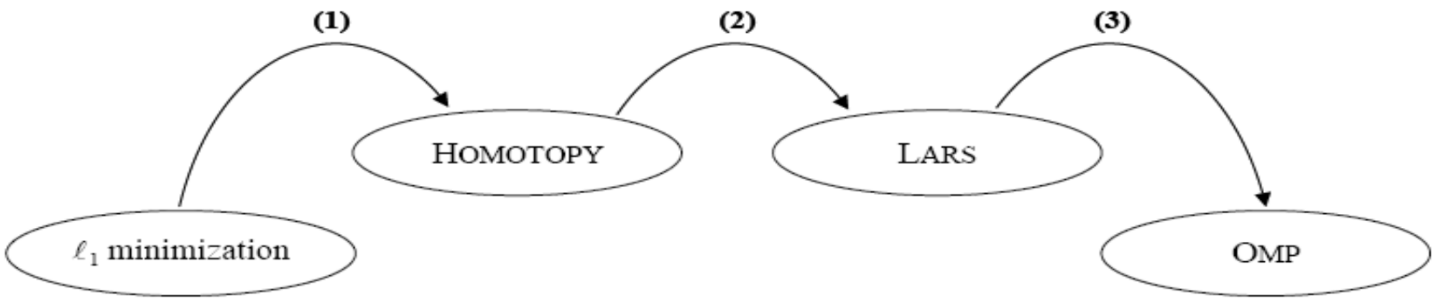
- Homotopy algorithm
  1. Given F and $y = Fx$, set $x_1 = 0$.
  2. Find residual correlation, $c_j = F^T\left(y - Fx_j\right)$
  3. Determine the step direction and size
  4. Update the active set, sol. estimate x_j and the step size.
  5. Stop when the residual correlation is zero; otherwise repeat 2 – 4.

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \end{bmatrix} \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \\ 1 \end{pmatrix}$$

$$F^T y = \begin{pmatrix} 5 \\ 1 \\ -3 \\ 3 \\ -3 \\ 3 \end{pmatrix} \longleftarrow$$

The largest!

Homotopy and others



[Donoho and Tsaig08]

- Homotopy solves L1 minimization problem.
- LARS is obtained from Homotopy by removing the sign constraint check (Only add an element to the active set; no removal of an element)
- OMP and LARS are similar in structure, OMP solves a least squares problem at each iteration, whereas LARS solves a linearly penalized least-squares problem.

Compressed Sensing Narrative

Compressed Sensing Narrative

- Any natural signal x can be sparsely represented in a certain basis:

$$x = Bs \qquad\qquad (s \text{ is } K\text{-sparse})$$

- A sparse signal can be compactly described via a linear transformation:

$$y = Fx = FBs \qquad\qquad (y \text{ is } M \text{ x } 1,\ M < N)$$

- Possible linear transformation matrices for F are many, including
  - Randomly selected rows of the F.T. matrix
  - i.i.d. Gaussian ~ $\mathcal{N}(0, 1/M)$
  - i.i.d. Bernoulli {+1, -1}

- The L1 minimization recovers the signal x perfectly with probability close to 1 as long as the number of measurements are sufficiently large,

$$N > M \geq cK\lambda > 2K$$

  - Where the oversampling factor is

$$\lambda = (\log N)^6 \quad \text{for the FT matrix}$$

$$\lambda = (\log N) \quad \text{for the Gaussian and Bernoulli matrices}$$

Key Ingredients in CS Theory

- *Incoherence* between F and B
  - It is desired to select an F so that it is incoherent to B (imagine the consequence of the opposite case.)
  - Thus, F is usually constructed with the random Gaussian matrix since the statistical property of FB remains the same as that of F when B is unitary (orthogonal).

- *Restricted Isometry Property* (RIP): Candes and Tao define that the K-restricted isometry constant of the sensing matrix is the smallest quantity such that

$$1-\delta_K \le \frac{\|Fx\|_2^2}{\|x\|_2^2} \le 1+\delta_K$$

for any K-sparse vector v sharing the same K nonzero entries as the K-sparse signal x.

  - If a small $\delta_K < 1$ exists for a class of F, then Fx should behave like a unitary transformation (and, y and x are one-to-one)
  - If $\delta_{2K} < 1$, then L0 solution is unique.
  - If $\delta_{2K} < \sqrt{2} - 1$, then L1 solution attains the L0 solution.

Key Ingredients in CS Theory

- RIP is useful for large deviation results as well.
  - Another way to write RIP is : $\left\{ \left| \|Fx\|_2^2 - \|x\|_2^2 \right| \leq \delta \|x\|_2^2 \right\}$
  - Then, one can ask a question that a sensing matrix F selected randomly from an ensemble of M x N sensing matrices (say i.i.d. Gaussian) to have an RIP constant δ .
  - This leads to a large deviation analysis which then leads to the probabilistic statement of the following form:

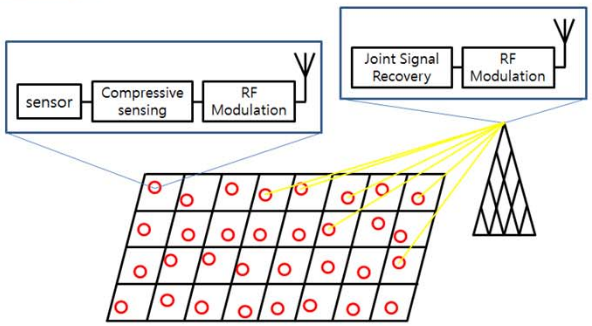
$$\Pr\left\{ \left| \|Fx\|_2^2 - \|x\|_2^2 \right| \leq \delta \|x\|_2^2, \text{ for any } K\text{-sparse } x \right\} \leq \exp\left[-c\left(M - K \log\left(N / K \right) \right) \right]$$

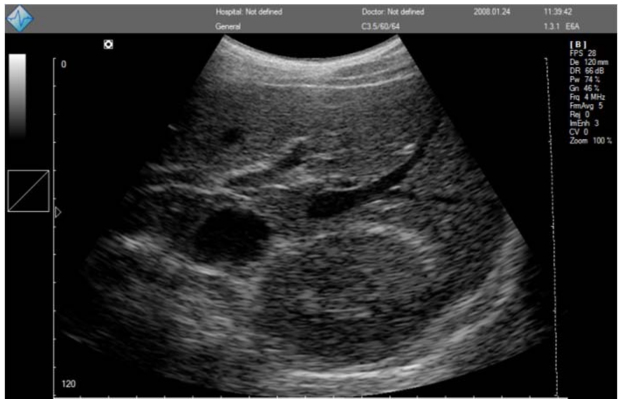
- Stable recovery of L1 minimization.
  - Signals are not exactly sparse (model mismatch).
  - Observations are noisy.
  - L1 recovery provides stable recovery results.
  - The model mismatch and observation noise do not pathologically add in L1 recovery.
  - L1 recovery results are not much worse than the model mismatch and observation errors.

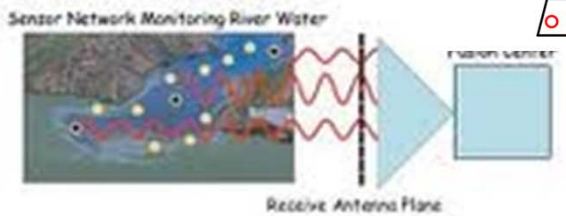
Many Applications

- See http://dsp.rice.edu/cs, a CS repository

- Compressive Imaging
- Medical Imaging
- Analog-to-Information Conversion
- Ultra-wideband radios
- Compressive Spectrum Sensing
- Classification using Sparse Rep
- Super Resolution Imaging







Recent Results of CS at GIST

Compressive Sensing

- CS Basic Equation

  y = F x

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{bmatrix} f_{11} & f_{12} & f_{13} & f_{14} & f_{15} & f_{16} \\ f_{21} & f_{22} & f_{23} & f_{24} & f_{25} & f_{26} \\ f_{31} & f_{31} & f_{33} & f_{34} & f_{35} & f_{36} \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ x_3 \\ 0 \\ 0 \\ x_6 \end{pmatrix}$$

- How to design F?
- How to recover x, fast and robust?

Recast of CS in Channel Coding Context

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{bmatrix} 1\,0\,0\,1\,1\,0 \\ 0\,1\,0\,1\,0\,1 \\ 0\,0\,1\,0\,1\,1 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ x_3 \\ 0 \\ 0 \\ x_6 \end{pmatrix}$$

- Group testing done during the 2nd World War in the US
  - Do not want to call up syphilitic man for service.
  - Do not want to test out all men's blood samples either
  - What to do?
  - Group test
    - Index the blood samples of each man, i=1, 2, …, N.
    - Add blood samples of randomly selected men and test them, M tests.
    - Solve the under-determined set of equations and find all the syphilitic men.

- y is called Syndrome.
- F is a parity-check matrix.
  - A K-error correcting code if SPARK(F)=2K.
  - Any K-error patterns can be found and corrected.

Channel Codes

- Purpose: Add redundancy symbols and offer error-protection
- Message: m
- Codeword: c = Gm
  - Generator matrix G
- Encoding: c = Gm
- Channel: z = Gm + x (x is the channel errors. Errors are sparse!)
- Decoding: find F where FG=0
  - Apply F to z: Fz = FGm + Fx = Fx
  - What's left is y=Fx

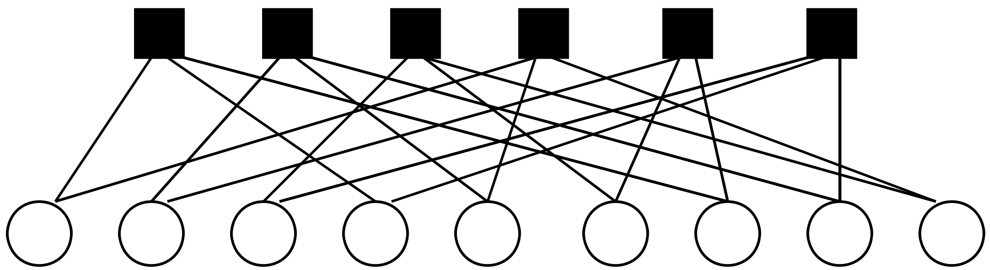
Example of a Channel Code in GF(2)

- Let $G = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$
- Using FG = 0, find F = $\begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$
- Note that SPARK(F) = the size of the smallest subset of columns of F that are l.d. = 3 = d_{min}.
  - The example is a single error correcting code
    = Every single error pattern can be detected
    = All 1-spase signal can be recovered using F.
  - SPARK <= M (The singleton bound)
- Note that UUP is met for F.
  - For all 1-spare signal e, Fe is non-zero.
  - M = 3 > 2K = 2.

LDPC Code/Bipartite Graph

$$M/N = 6/9$$

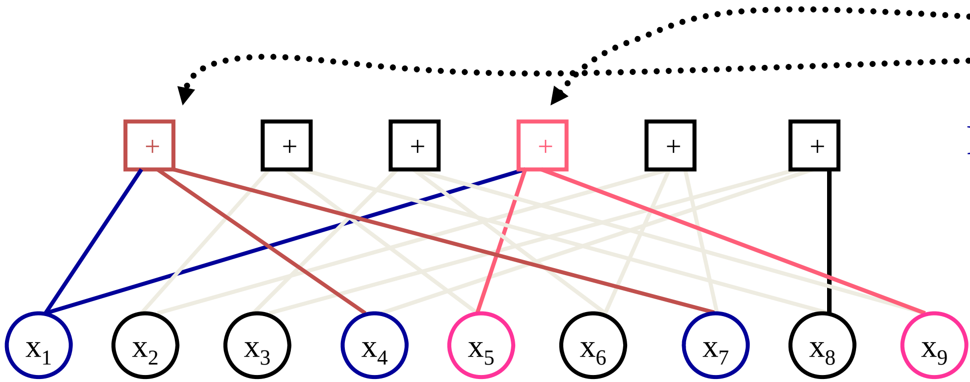
$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_9 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_6 \end{bmatrix}$$



An LDPC code that was shown to achieve the Shannon Limit!

Make the matrix sparse!

Probabilistic Method: GF(2)



$\Pr(S|x_1=0) = \Pr(S_1|x_1=0)\ \Pr(S_4|x_1=0)$

$$P(S \mid x_1 = 1, \mathbf{y}) = \Pr\{\text{odd \# of 1s in } x_4 \text{ and } x_7\} \times \Pr\{\text{odd \# of 1s in } x_5 \text{ and } x_9\}$$
$$= \left\{p_{4,1}(1-p_{7,1}) + (1-p_{4,1})p_{7,1}\right\} \times \left\{p_{5,1}(1-p_{9,1}) + (1-p_{5,1})p_{9,1}\right\}$$

Reed Solomon Codes

- Linear sensing matrix F

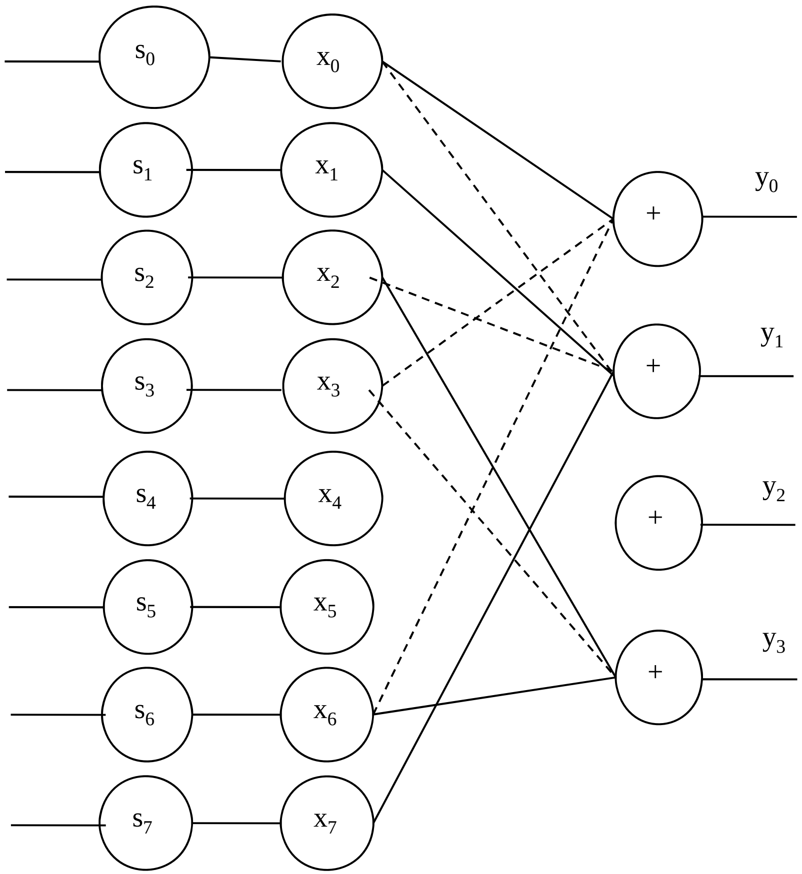
- 2 step recovery
  - Error Locator polynomial
  - OD Matrix inversion

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{bmatrix} f_{11} & f_{12} & f_{13} & f_{14} & f_{15} & f_{16} \\ f_{21} & f_{22} & f_{23} & f_{24} & f_{25} & f_{26} \\ f_{31} & f_{31} & f_{33} & f_{34} & f_{35} & f_{36} \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ e_3 \\ 0 \\ 0 \\ e_6 \end{pmatrix}$$

$$\Big\Downarrow \ \text{Location info}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{bmatrix} f_{13} & f_{16} \\ f_{23} & f_{26} \\ f_{33} & f_{36} \end{bmatrix} \begin{pmatrix} e_3 \\ e_6 \end{pmatrix}$$

Message Passing: State, Value, Matrix, Observation



F. The Message Passing Algorithm

The message passing algorithm is given as the following:

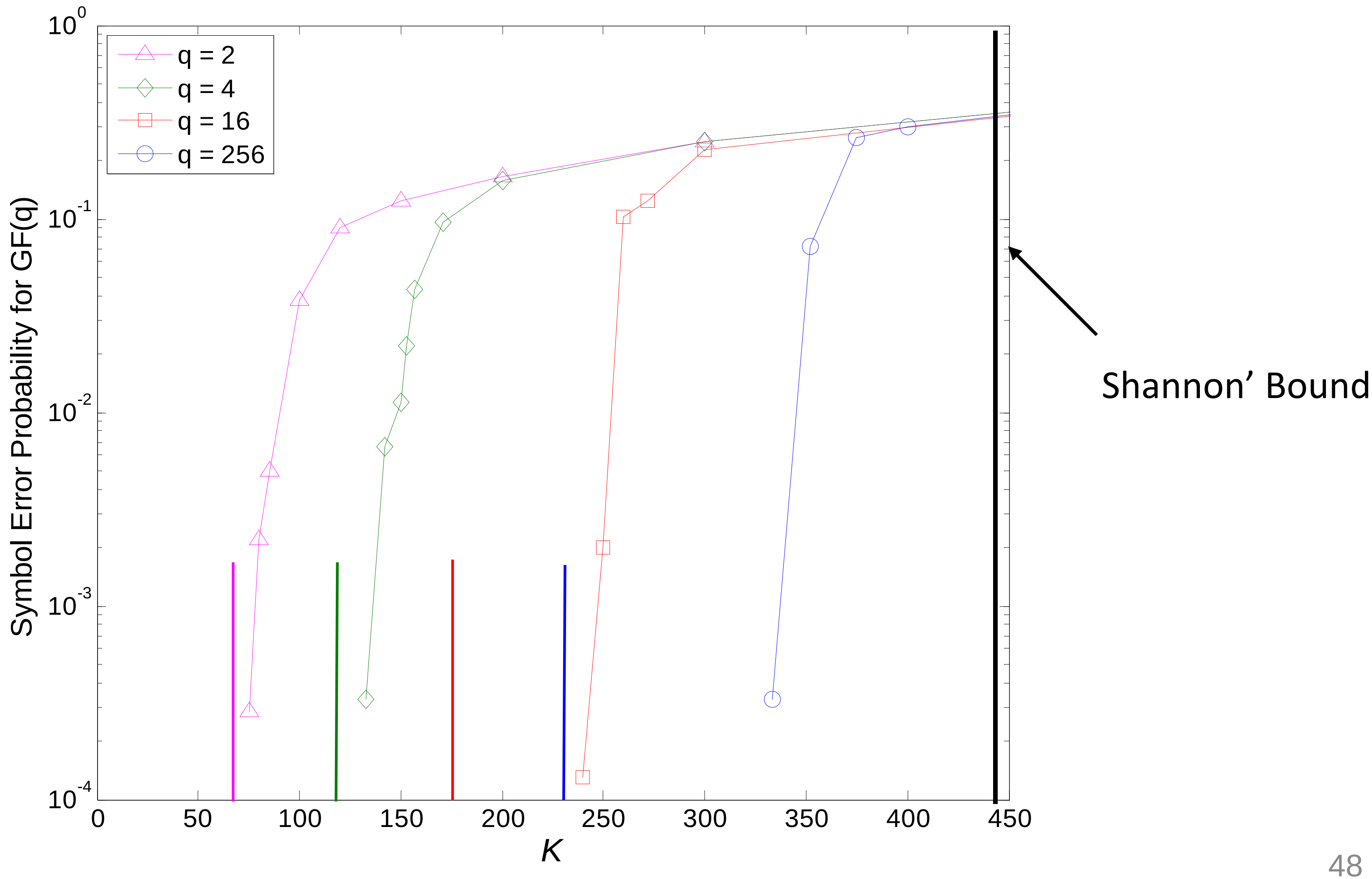
1. Initialization: Set $P(x_t = \tau_t \mid \mathbf{y}) = \frac{K}{N} f_1(x_t = \tau_t) + (1 - \frac{K}{N}) f_0(x_t = \tau_t)$ for all t. Determine a threshold δ for stopping criterion.
2. Run message passing routine: Do the convolution (or the FFT/IFFT) routine for each t, obtaining $\mathrm{P}(x_t = \tau_t \mid \mathbf{y}, C)$ for all t.
3. Run the active set recovery routine. An index t will be decided to be added to the active if the log ratio, $LR(S_t)$, for $t = 0, 1, 2, \cdots, N-1$, is greater than zero, i.e.,

$$I = \{t : LR(S_t) > 0.0\}$$

4. Check if I is $\mathcal{K}$: Run $x_I = \left(A_I^T A_I\right)^{-1} A_I^T y$. When this value is good enough, i.e. $\|r - Ax_I\|_2 \le \delta$ the threshold, the iteration can be put to stop. Otherwise, return to step 2 and repeat.

Simulation Results

- Simulation results of a $C(N = 1200, d_v = 3, d_c = 6)$ code with different field sizes, compared with the GV bounds indicated at 65, 120, 175 and 230.



Support Set Recovery First and then Estimate x

- Good for compressive sensing in the presence of observation noise.
- **"Noisy compressive sensing"** (NCS)

$$z = Ax_0 + n$$

$x_0 \in \mathbb{R}^N$: A deterministic realization of signal x

$z \in \mathbb{R}^M$: The noisy measurements

$n \in \mathbb{R}^M$:The additive noise with $\mathcal{N}(0, \sigma_n^2 \mathbf{I}_M)$

$A \in \mathbb{R}^{M \times N}$: The sensing matrix where $M < N,\ Rank(A) \leq M$

- **Compressive sensing via Bayesian support detection** (CS-BSD)
  - A sparse reconstruction algorithm based on Bayesian approach using the sparse sensing matrix.
  - CS-BSD has *detection-directed estimation* structure which consists of support detection and signal value estimation.
  - CS-BSD is highly robust against noise.
  - CS-BSD achieves the MSE performance of an MMSE estimator which has the support knowledge.

Previous Bayesian Approaches for NCS

- The problem of the sparse reconstruction is modeled based on MAP.

$$\hat{x} = \arg \max_{[\tau_1,..,\tau_N]} p(x = [\tau_1,..,\tau_N] \mid z) \quad s.t. \quad \left\| A\hat{x} - z \right\|_2 \le \varepsilon$$

- 1) **SBL** algorithms for NCS: [08' Ji *et al.*], [10' Babacan *et al.*]:
  - They found the posterior distribution of the signal based on a three-layer hierarchical Bayesian model.
  - The parameters of the posterior is estimated using *Expectation Maximization*.

- 2) Belief propagation (BP) based algorithms for NCS:
  - **CS-BP** [10' Baron et al.]: Updating the signal posterior from two-state Gaussian mixture prior via message-passing algorithm.
  - **SBL-BP** [10' Tan et al.]: Applying BP to the SBL-framework to reduce the cost of *Expectation Maximization*.
  - **SuPrEM** [10' Akcakaya et al.]: Similar to SBL-BP, but use different prior and sensing matrix called low-density frame.

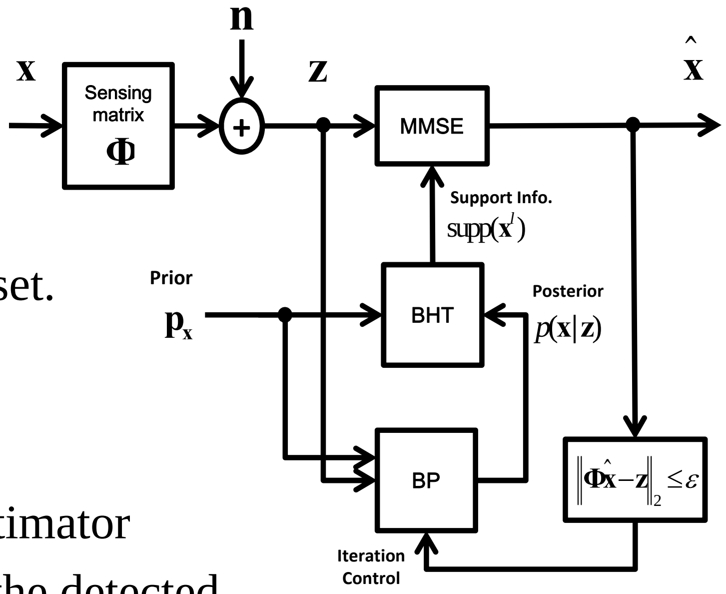
Algorithm of CS-BSD

- CS-BSD has a ***detection-directed estimation*** structure.

$$\hat{x}_i = \begin{cases} 0, & \dfrac{\Pr\{s_i = 0 \mid z\}}{\Pr\{s_i = 1 \mid z\}} \geq 1 \\ \arg\max_{\tau} p(x_i = \tau \mid z, s_i = 1), & \dfrac{\Pr\{s_i = 0 \mid z\}}{\Pr\{s_i = 1 \mid z\}} < 1 \end{cases} \quad \text{for all } i \in \{1, \ldots, N\},$$

where s_i denote state the i-th element

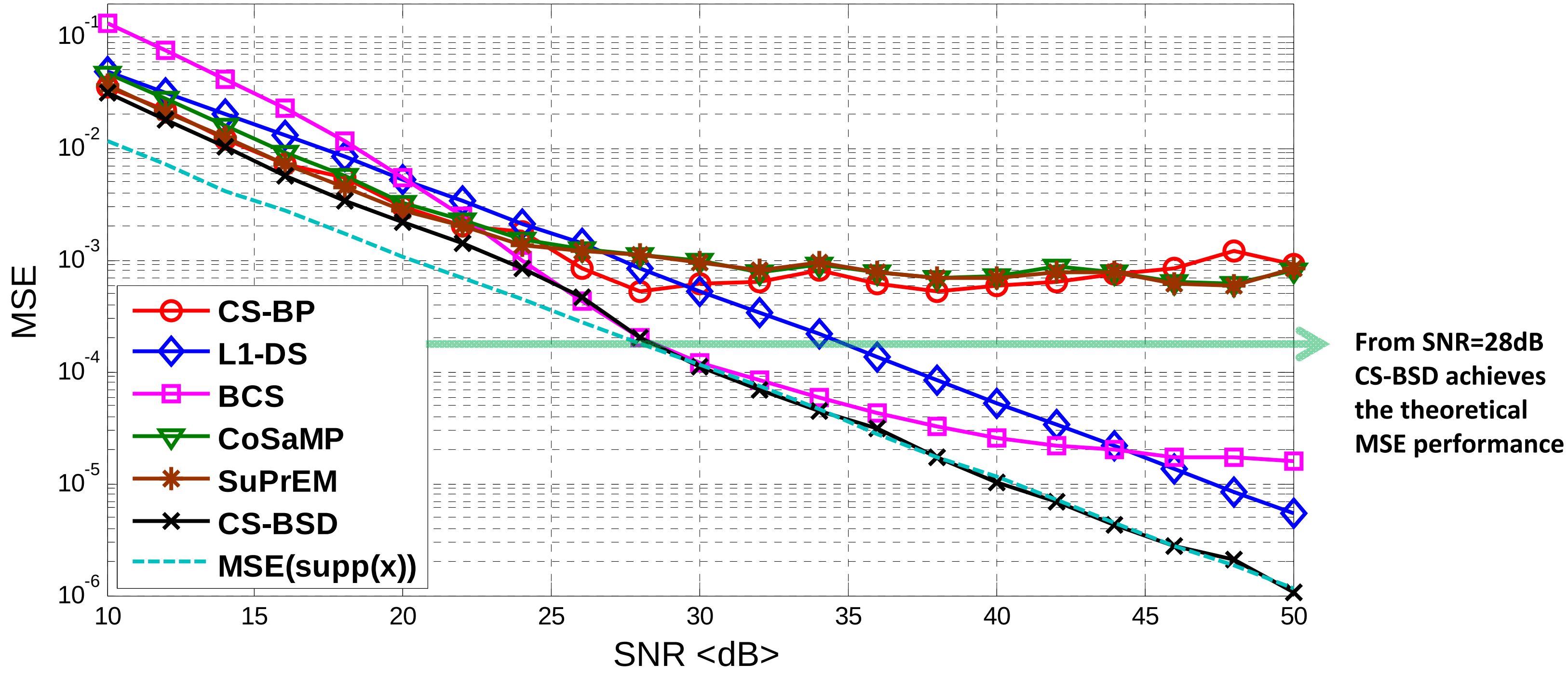
- 1) Detection of support set:
  - *Belief propagation* based iteration provides the posterior of the signal.
  - *Bayesian hypothesis test* detects the support set.



- 2) Estimation of signal values:
  - An *minimum mean square error* (MMSE) estimator provides the value of the signal elements using the detected Support set.

MSE performance comparison

❖ When N=1024, M=512, sparsity rate q=0.05,



❖ The Proposed Method beats all the L1 and its derivates.

❖ The Proposed Method achieves the Cramer Rao bound.

Shannon's CC Theorem vs. CS Theory

Shannon's Channel Coding Theorem

- Rate = 1 – M/N
- "Rate < Capacity" IFF "A matrix F with R and P(e) → 0"
- If Rate < Capacity, there exists a matrix F such that P(e) → 0.
- If Rate < Capacity is not holding, P(e) cannot be 0.

Application to Compressed Sensing

- Channel ~ error rate *K/N*
- Capacity is well known
- 1-M/N < Capacity → M/N> 1 – Capacity.
- For a matrix F with small P(e), M/N > 1 – Capacity

$$P(e) \leq 2^{-N\left[\rho_{comp} - (1-C)\right]}$$

Channel Coding vs. Compressed Sensing

	Compressed Sensing	Channel Coding
y (M x 1)	Observation	Syndrome
F (M x N)	Sensing matrix	Parity check matrix
x (N x 1)	K-sparse signal	Error patterns Error occurs with prob. (K/N)
Theory	If M>K log N, L1 recovers x exactly from y, with high probability.	If R=1-M/N < C, then a good F exists so that P(e) is close to zero. (Converse is true as well)

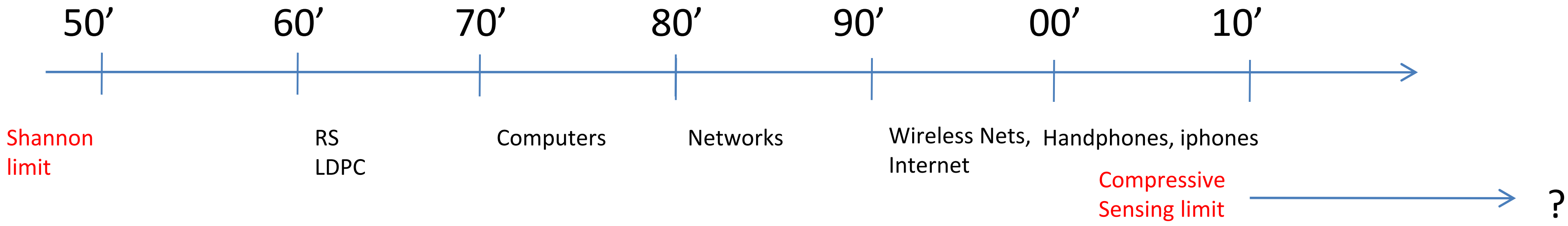
Summary

- We have reviewed CS theory, started from the Sampling Theorem.
  - Sparse representation, UP, L1 routine are building blocks of CS.
  - Uniform spaced samples are not only the option when representing signals.
  - Holistic samples might be more advantageous (e.g. low power sensors, transmitters, imaging devices)

- CS Theory can be understood as the parity check problem in Coding Theory.
  - LDPC codes, Reed-Solomon codes, Reed Muller codes, …
  - We begin to see the emergence of close relation between the CS theory and the Channel Coding Theory.
  - They are not contradictory to each other; instead they are complementing each other and provide new perspectives.

- CS Narrative gave us new perspectives which has been shown useful in many applications including medical imaging, ADCs, spectrum sensing, super resolution areas.

Outlook



- Shannon Theory
  - Shannon limit achieved only in 90s and 00s
  - Served as the lighthouse for technical development in the digital era

- Compressive sensing theory: What would happen in the next 10, 20 years?
  - Dawn of abruptly new generation technology?

Questions & Answers

Home page at http://infonet.gist.ac.kr/
Send comments to Heung-No Lee at heungno@gist.ac.kr.

Other CS Results at INFONET

Improving the Resolution of Spectrometers

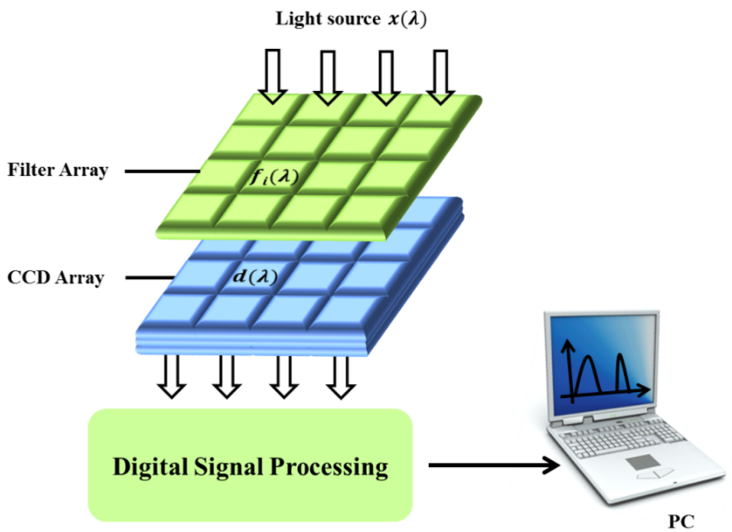
- Problem Statement: We aim to improve the spectral resolution beyond the conventional limit, the number of sensors *M*, using signal processing. Given *M* and the sensing matrix, how much can we improve the spectral resolution?
- Approach: Improving the resolution amounts to solving an underdetermined system of linear equations.

$$\boldsymbol{y} = D\boldsymbol{x} + \boldsymbol{w} = D\Psi\boldsymbol{s} + \boldsymbol{w}$$

- Raw spectrum model :

D – MxN Detector sensitivity matrix
M – # of filters
N – # of spectral components

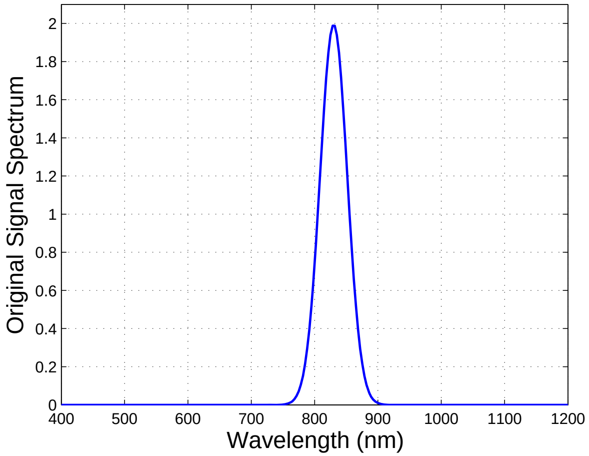
$$M<N$$



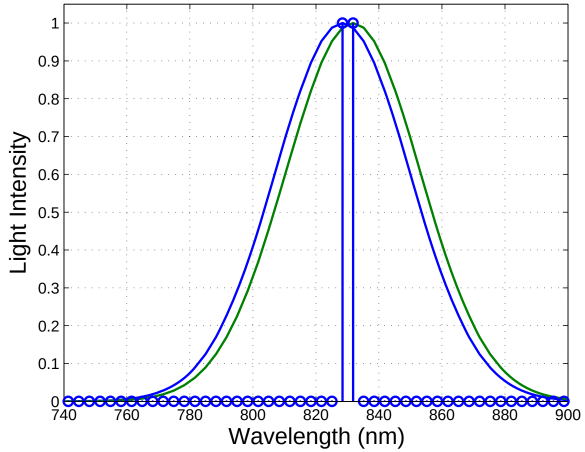
Schematic diagram of a typical miniature spectrometer

Experimental Results

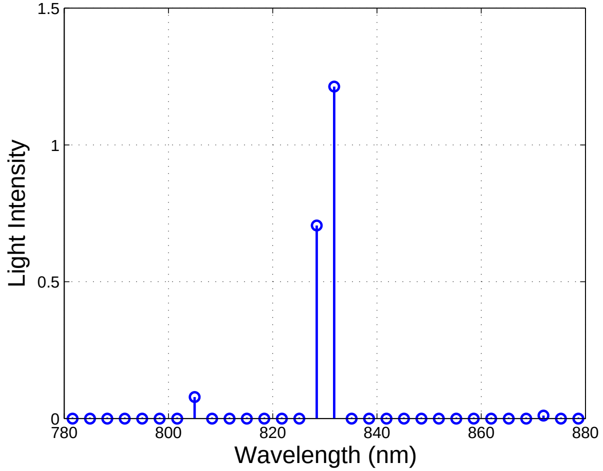
❖ $M=40 \quad N=240$



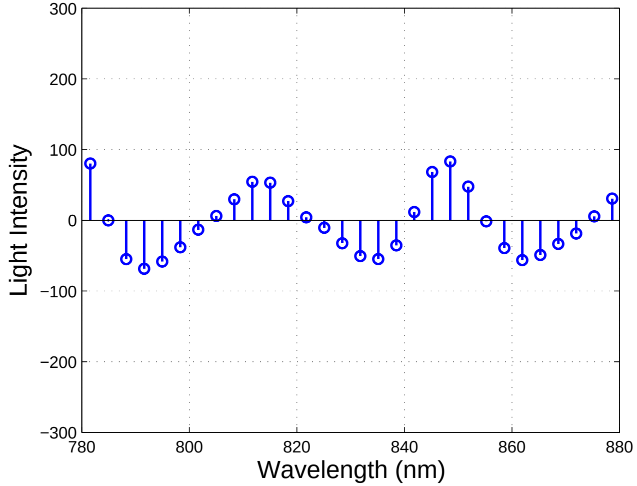
Original signal spectrum



Components of original signal spectrum



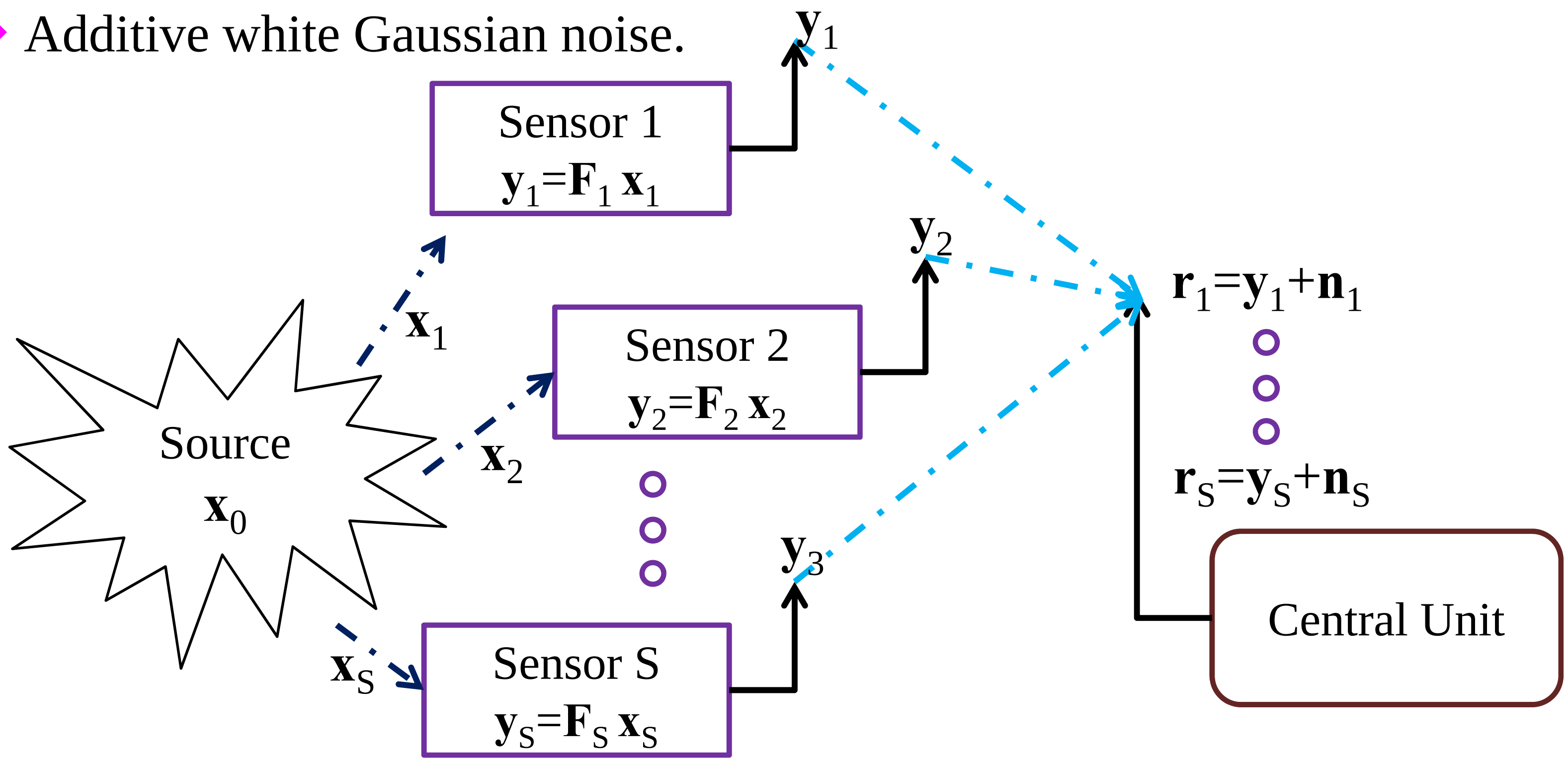
L_1 Signal spectrum estimate



L_2 Signal spectrum estimate

Performance Behavior of the Multiple Sensor System based on Compressive Sensing

- Each M x N $\mathbf{F}_i$ ~ $N(0,1/M)$ elements
- Support set $\mathcal{I}$ is shared, $|\mathcal{I}|=K$.
- Additive white Gaussian noise.



- The SM x 1 vector $\mathbf{r}=[\mathbf{r}^T_1\ \mathbf{r}^T_2 \ldots \mathbf{r}^T_S]^T$ is concatenation of the received vectors.

Previous works & Motivations

- A similar MSS was considered in [Gastpar05]. They concluded that a distortion measure (the mean-squared error) decreases like $1/S$ for the fixed M.

- M converges to K(noise-less case, sensing matrix $\mathbf{F}_i$ is different for each sensor). [Baraniuk05].

- M converges to $2K$ (noisy case, sensing matrices are the same). [Nehorai11].

- Motivations:
  - The analysis under noisy case and different sensing matrices for each sensors is needed.
  - Can we show wheatear the sufficient number of measurements is K or 2K?

Convergence theorem

- Theorem 1: Let rank($\mathbf{F}_{i,\mathrm{J}}$) be K for each i, $M > K$ and $\sigma_{\text{noise}}^2 < \min_s \sum_{i \in \mathcal{I} \setminus \mathcal{J}} x_s(i)^2$. Then, the probability that the support set is not recovered converges to zero as the number of sensors increases, where J⊂[N], |J|=K.

- M converges to K was reported in [Baraniuk05] when there is no noise.

$$
\begin{aligned}
\Pr\{\mathit{failure}\} \le \sum_{\forall \mathcal{J}, |\mathcal{J}|=K} & \exp\left(-\frac{S}{2\sigma_{\min}^2} \times \left((M-K) \times \left(\sigma_{\text{noise}}^2 - \sigma_{\min}^2 \right) + M\delta \right) \right) \times \left(\frac{\sigma_{\text{noise}}^2}{\sigma_{\min}^2} + \frac{M}{M-K} \times \frac{\delta}{\sigma_{\min}^2} \right)^{S \times \left(\frac{M-K}{2} \right)} \\
& + \exp\left(-\frac{SM}{2} \times \frac{\delta}{\sigma_{\text{noise}}^2} \right) \times \left(1 + \frac{M}{M-K} \times \frac{\delta}{\sigma_{\text{noise}}^2} \right)^{S \times \frac{(M-K)}{2}}
\end{aligned}
$$

- Owing to the fact that the support sets are shared, we consider $\sum_{\forall \mathcal{J}, |\mathcal{J}|=K}$ as $\binom{N}{K}$.
- The number of sensors has no relation with the value $\binom{N}{K}$

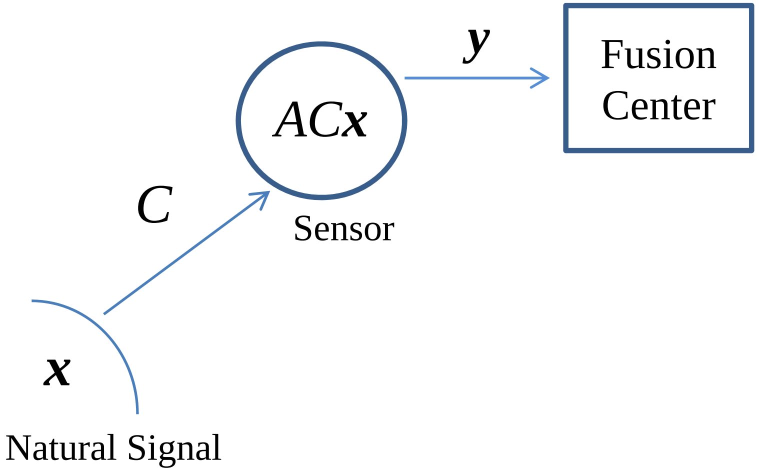
Conclusions & Future work

- Compressive Sensing is a good approach when each sensor independently compresses its signal.
- As the number of sensors increases, the measurement signal size converges to K.

- FW includes
  - solve the problem when the sensing matrices are the same
  - make a connection between our work and Gaspar's work[Gastpar05]
    - Ex) Does distortion measure at each sensor decrease like $1/\exp(S)$ or $1/S$?
  - consider a more realistic problem when the model for inter-sensor correlation is more complex:
    - Ex) Each support set = the shared support set + independent support set.
  - Identify how our work can be connected to Slepian Wolf coding.

A Realistic Compressive Sensing Framework for Multiple Wireless Sensor Networks

- Problem statement: Conventional approaches for signal recovery in a multiple-sensor system assume ideal channel conditions. We deal with effects of a realistic dispersive channel, which encounter when acquiring signals.



Data Model $\boldsymbol{y} = AC\boldsymbol{x} + \boldsymbol{n} = F\boldsymbol{x} + \boldsymbol{n}$

- Challenge: Given the channel matrix C exactly or partially, how to design a sensing matrix A that helps in good signal recovery?
- We first investigate the effect of the channel on the recovery of the sensed signal by analyzing RIP of F in terms of channel parameters

RIP in Realistic Scenarios

- Recall: Goodness of a sensing matrix is measured in terms of its RIP given as:

$$\lambda_{\min}(F^T F) \leq \frac{\|Fs\|^2}{\|s\|^2} \leq \lambda_{\max}(F^T F)$$

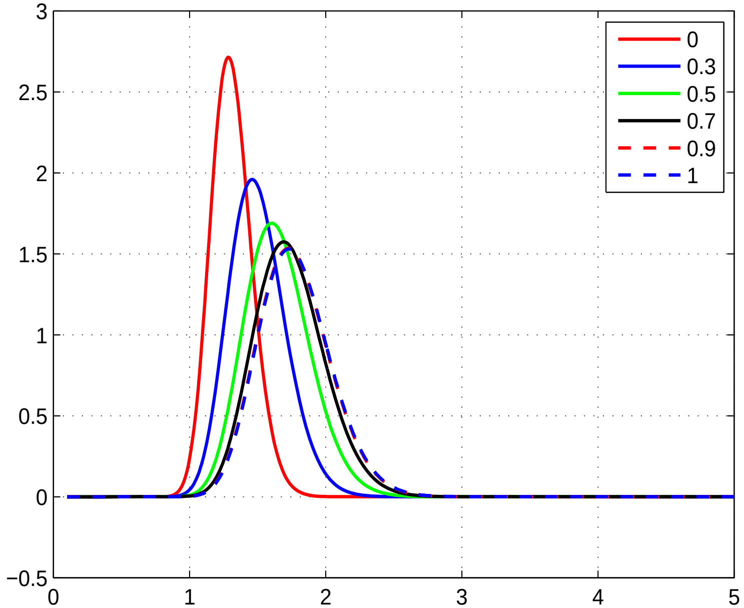
- $F^T F$ is a correlated central Wishart matrix with rank K (the sparsity)

- We note that the RIP depends on the channel Impulse response

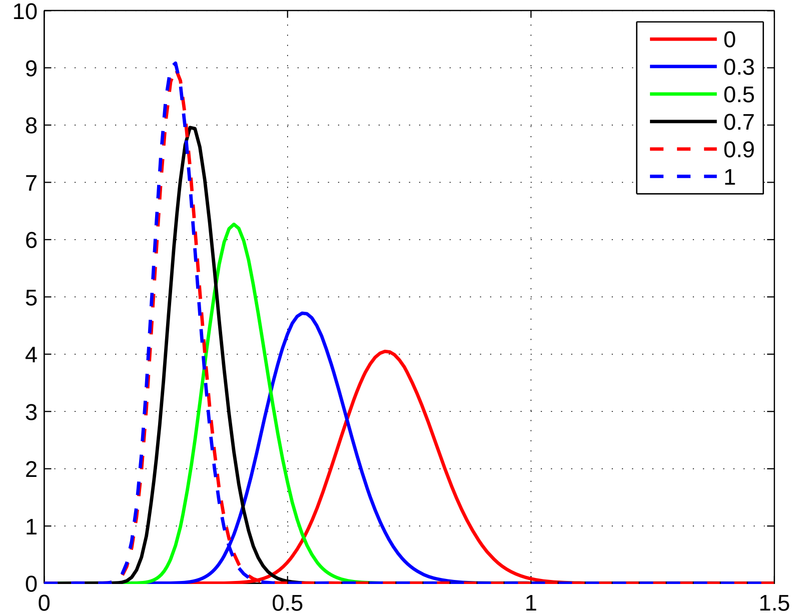
- For a two-path channel , i.e., $\boldsymbol{h} = [1 \;\; h_1]^T$, the RIP depends on the value of h_1

- We obtain the eigenvalue distributions of Wishart matrix as a function of h_1 which helps to obtain a condition for unique signal recovery.

Distributions Of Eigenvalues For Various h_1



Maximum Eigen Value



Minimum Eigen Value

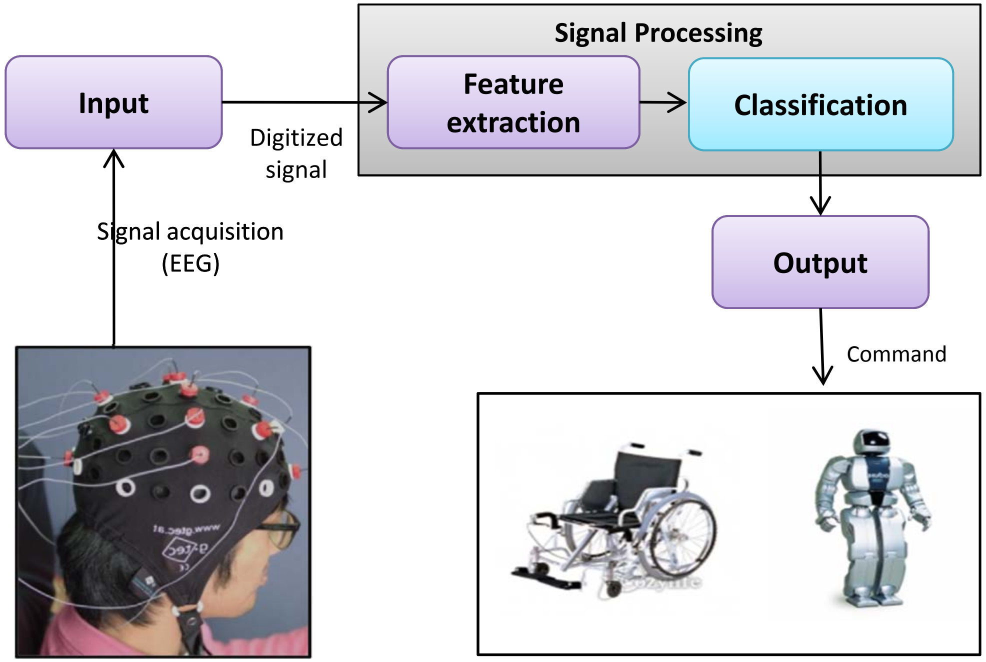
Sparse Representation based Classification (SRC) method for a Brain Computer Interface (BCI) application

Younghak Shin, Seungchan Lee, and Heung-No Lee

Sparse Representation based Classification (SRC) method for a Brain Computer Interface (BCI) application

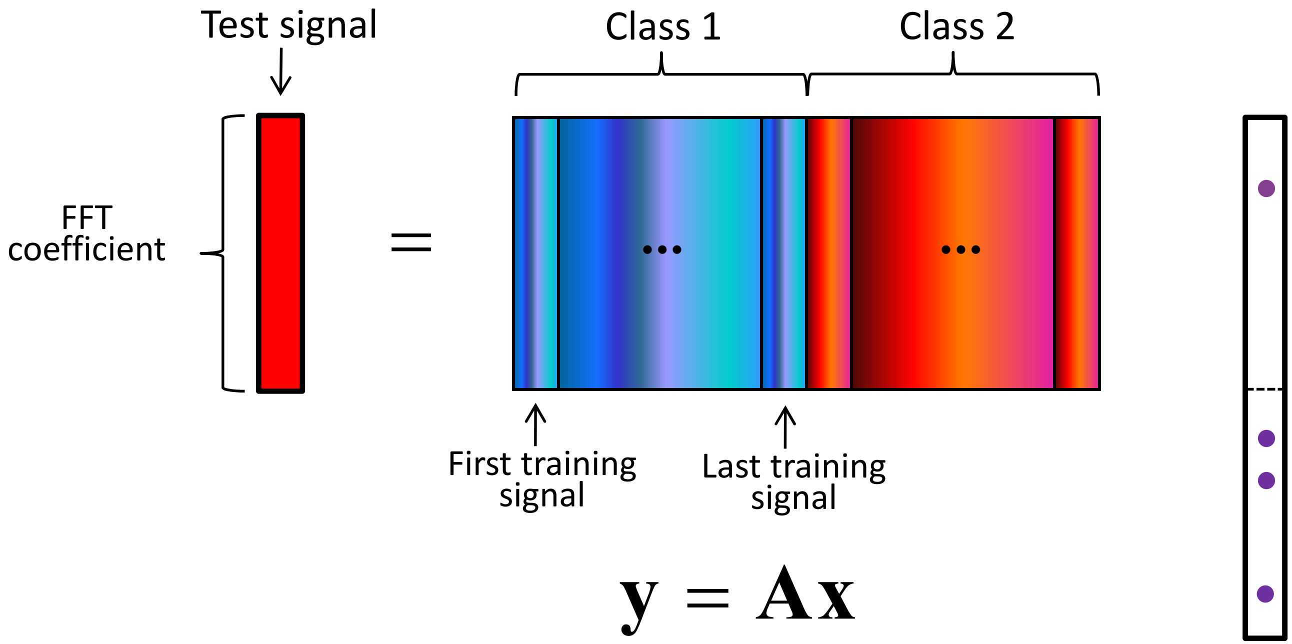
- The Sparse Representation which is used in CS theory can be used for a number of applications including noise reduction, compression, and pattern recognition.
- Recently, Sparse Representation based Classification (SRC) method was studied in Face Recognition [Wright 09] and Speech Recognition area [Gemmeke 11].
- This SRC method have shown superior classification performance
- we apply the SRC method to the Brain Computer Interface application.

Brain Computer Interface System



- BCIs is a new communication and control channel between human brain and an external device.
- In the BCIs, classification is needed to identify a prescribed command from signal features acquired from scalp EEG
- EEG signals are very noisy and non-stationary.
- Thus, powerful signal processing methods are needed
- We propose the sparse representation based classification method.

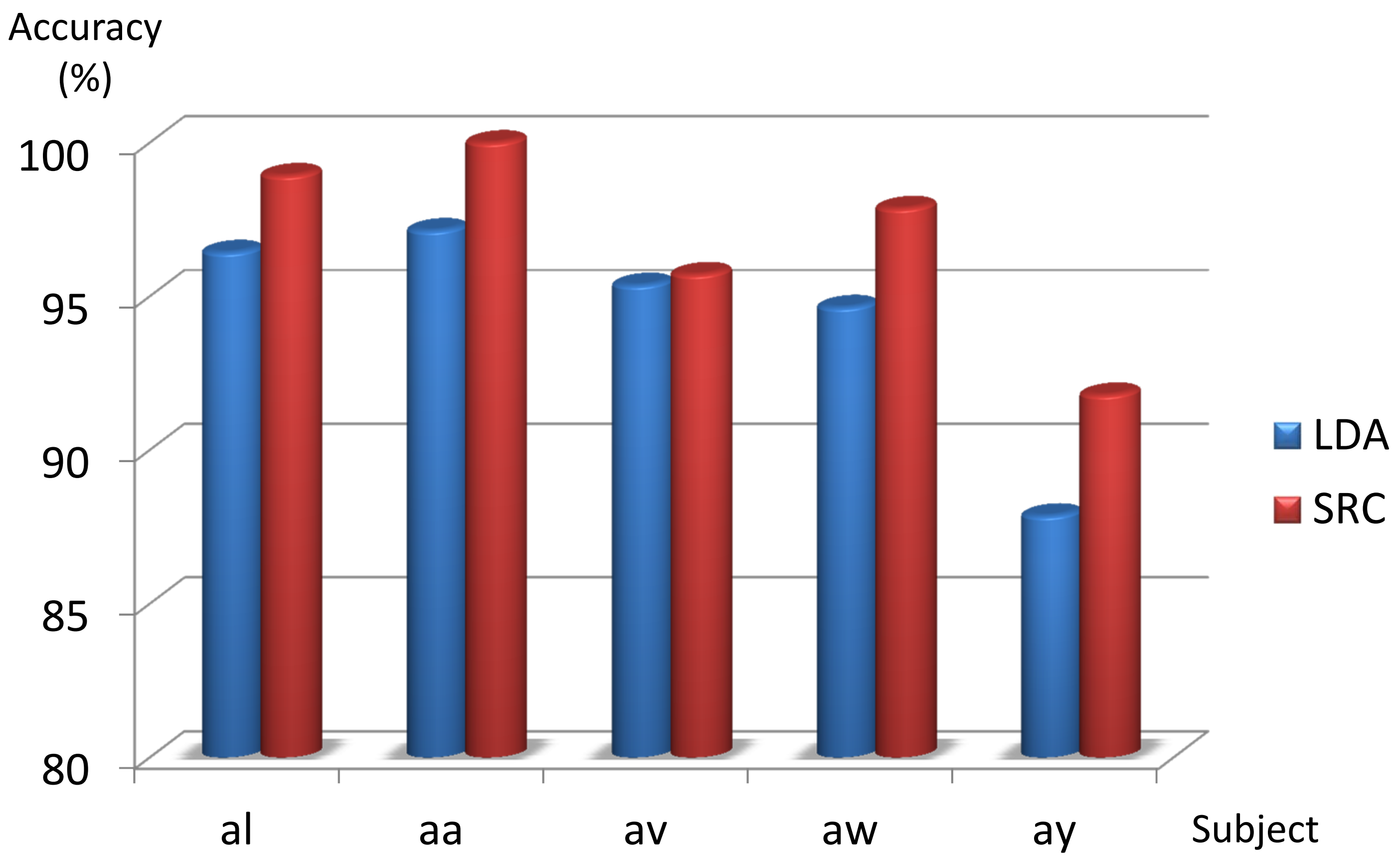
Proposed method



$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

- This sparse representation can be solved by L1 minimization
- For example, a test signal **y** of class 2 can be sparsely represented as the training signals of class 2.

Classification result



- The *t*-test results is $p = 0.0129 < 0.05$.
- Proposed SRC method shows better classification accuracy than conventional LDA classification method.

References

- [Gastpar05]M. Gastpar and M. Vetterli, “Power, spatio-temporal bandwidth, and distortion in large sensor networks”, IEEE J. Sel. Areas Commun., vol. 23, pp. 745 – 754, Apr. 2005.
- [Nehorai10]Gongguo Tang and Arye Nehorai, “Performance Analysis for Sparse Support Recovery”, IEEE Trans. Inform. Theory, vol. 56, no. 3, March. 2010.
- [Baraniuk05]M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin and R. G. Baraniuk, “Distributed Compressed Sensing of Jointly Sparse Signals”, Asilomar Conf. Signals, 2005.

Reference

[Wright09] John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, Yi Ma, “Robust Face Recognition via Sparse Representation” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210~227, February 2009.

[Gemmeke11] J. F. Gemmeke, T. Virtanen and A. Hurmalainen, Exemplar-based sparse representations for noise robust automatic speech recognition. *IEEE Trans. Audio, Speech, and Language Processing,* vol. 19, no. 7, pp. 2067-2080, 2011

[Shin11] Younghak Shin, Seungchan Lee and Heung-No Lee, "A New BCI Classification Method based on EEG Sparse Representation", Signal Processing with Adaptive Sparse Structured Representation, Edinburgh, Scotland, June 27-30, 2011.

References

- [Gastpar05]M. Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks", IEEE J. Sel. Areas Commun., vol. 23, pp. 745 – 754, Apr. 2005.
- [Nehorai10]Gongguo Tang and Arye Nehorai, "Performance Analysis for Sparse Support Recovery", IEEE Trans. Inform. Theory, vol. 56, no. 3, March. 2010.
- [Baraniuk05]M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin and R. G. Baraniuk, "Distributed Compressed Sensing of Jointly Sparse Signals", Asilomar Conf. Signals, 2005.
- [Bruckstein08]A. M. Bruckstein, Michael Elad and Michael Zibulevsky, "***On the Uniqueness of Nonnegative Sparse Solutions to Underdetermined Systems of Equations***", IEEE Transaction on Information Theory, vol. 54, no. 11, pp. 4813 – 4820, Nov, 2008
- [Baraniuk08]Richard Baraniuk, Mark Davenport, Ronald DeVore and Michael Wakin, "***A Simple Proof of the Restricted Isometry Property for Random Matrices***", Constructive Approximation, vol 28, pp. 253-263, 2008.
- [Hormati-Vetteli11]Ali Hormati and Martin Vetteli, "**Annihilating filter-based decoding in the compressed sesning framework**", Proceedings of SPIE, 2007
- [Akcakaya-Vahid Tarokh08]Mehmet Akcakaya and Vahid Tarokh, "**A Frame Construction and a Universal Distrotion Bound for Sparse Represenations**", IEEE Transaction on Signal Proceesing, vol. 56, no. 6, pp. 2443 – 2450, Jun, 2008.
- [Oliver 11] J. Oliver , Woonbi Lee, Sangjun Park, and Heung-No Lee, "Improving Resolution of Miniature Spectrometers by Exploiting Sparse Nature of Signals", Submitted to Optics Express.
- [Chang 08] Chang, C. C., Lee, H. N., "On the estimation of target spectrum for filter-array based spectrometer," Optics Express 16, 1056-1061 (2008).
- [Kurokawa 11] Kurokawa, U., Choi, B. I., and Chang, C. C., "Filter-based miniature spectrometers: spectrum reconstruction using adaptive regularization," IEEE Sensors Journal 11, 1556-1563 (2011).
- [Oliver 11a] J. Oliver and Heung-No Lee, "A Realistic Distributed Compressed Sensing Framework for multiple Wireless Sensor Networks", Signal Processing with Adaptive Sparse Structured Representation, Edinburgh, Scotland, June 27-30, 2011.

References

- Claude E. Shannon, "Communication in the Presence of Noise," *Proceeding of the I.R.E.*, vol. 37, pp. 10-21, January, 1949.
- J. M. Whittaker, "Interpolatory Function Theory," Cambridge Tracts in Mathematics and Mathematical Physics, No. 33, Cambridge University Press, Chapt. IV; 1935.
- David L. Donoho, "Compressed Sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 12891306, Apr. 2006.
- David L. Donoho and Jared Tanner, "Precise Undersampling Theorems," *Proceedings of the IEEE*, vol. 98, pp. 913-924, May, 2010.
- Richard Baraniuk, "Lecture Notes: Compressive Sensing," *IEEE Signal Processing Magazine*, p. 118-121, July, 2007.
- Justin Romberg, "Imaging via compressive sampling," *IEEE Signal Processing Magazine*, 25(2), pp. 14 - 20, March 2008.
- Ashikhmin, A. and Calderbank, A.R., "Grassmannian Packings from Operator Reed-Muller Codes," IEEE Trans. Information Theory, vol. 56, Issue: 11, pp. 5689-5714, 2010.
- Emmanuel Candès and Terence Tao, Near optimal signal recovery from random projections: Universal encoding strategies? (IEEE Trans. on Information Theory, 52(12), pp. 5406 - 5425, December 2006)
- Emmanuel Candès and Terence Tao, "Decoding by linear programming," IEEE Trans. on Information Theory, 51(12), pp. 4203 - 4215, December 2005.
- Emmanuel Candès, Justin Romberg, and Terence Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. (IEEE Trans. on Information Theory, 52(2) pp. 489 - 509, February 2006)
- Dimitri P. Bertsekas, A. Nedic and A.E. Ozdaglar, Convex Analysis and Optimization, Athena Scientific, 2003.
- S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004. D.
- Donoho and X. Huo, "Uncertainty Principles and Ideal Atomic Decomposition," IEEE Trans. on Info. Theory, vol.47, no.7, Nov. 2001.
- M. Elad and A. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," IEEE Trans. Info. Theory, vol. 48, no. 9, Sept. 2002.
- J. Romberg, E. Candes, and T. Tao, "Robust Uncertainty Principles and Optimally Sparse Decompositions," Presentation Downloaded from http://www.ipam.ucla.edu/publications/mgaws1/mgaws1_5184.pdf.
- Leon S. Lasdon, Optimization Theory for Large Systems, Dover publication, 2002.
- Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders, "Atomic Decomposition by Basis Pursuit," Vol. 43, no.1, pp. 129-159, SIAM Review, 2001.
- D.L. Donoho and Y. Tsaig, "Fast Solution of l-1 Norm Minimization Problems When the Solution May Be Sparse," IEEE Transactions on Information Theory, vol. 54, no. 11, pp. 4789-4812, Nov. 2008.

References

- J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” IEEE Trans. Inf. Theory, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- J. A. Tropp, “Just relax: Convex programming methods for subset selection and sparse approximation,” IEEE Trans. Inf. Theory, vol. 51, no. 3, pp. 1030–1051, Mar. 2006.
- M. R. Osborne, B. Presnell, and B. A. Turlach, “A new approach to variable selection in least squares problems,” IMA J. Numer. Anal., vol. 20, pp. 389–403, 2000.
- M. R. Osborne, B. Presnell, and B. A. Turlach, “On the LASSO and its dual,” J. Comput. Graph. Stat., vol. 9, pp. 319–337, 2000.
- R. Tibshirani, “Regression shrinkage and selection via the lasso,” J. Roy. Statist. Soc., vol. 58, no. 1, pp. 267–288, 1996.
- D. L. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” vol.106, no.45, Proceedings of National Academy of Science, Nov. 10, 2009.
- H. Rauhut, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1-92. deGruyter, 2010. http://rauhut.ins.uni-bonn.de/research/pub.php?list=rauhut.
- Y.C. Chen, et. al, “Terahertz Pulsed Spectroscopic Imaging Using Optimized Binary Masks,” Applied Physics Letters 95, 231112, 2009.
- W.L. Chan, M.L. Moravec, R.G. Baraniuk, and D. Mittleman, “Terahertz imaging with compressed sensing and phase retrieval,” Optics Letters, vol. 33, no. 9, May 1, 2008.
- Michael Elad, *Sparse and Redundant Representations*: From theory to applications in signal processing and image processing, Springer Science, New York, 2010.
- D. L. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. submitted to IEEE Trans. Inform. Theory, February 2004.
- E. J. Candès, J. Romberg and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. Comm. Pure Appl. Math., 59 1207-1223.
- Lawson and Hanson, "Solving Least Squares Problems", Prentice-Hall, 1974, Chapter 23, p. 161.
- Donoho and M. Elad, “Optimally Sparse Representation in General Dictionaries via L_1 Minimization, …
- E. J. Candes, “Compressed Sampling,” *Proc. of the Int. Congress of Mathematicians*, Madrid, Spain, 2006.
- S.J. Park and Heung-No Lee, “On the Derivation of RIP for Random Gaussian Matrices and Binary Sparse Signals,” Proc. of International Conference on ICT Convergence 2011 (ICTC 2011), Seoul, Korea, Sept. 28-30, 2011.
- G. Tang and A. Nehorai, “Performance Analysis for Sparse Support Recovery,” IEEE Trans. Info. Theory, vol. 56, no. 3, pp. 1383-1399, March, 2010.