Compressed Sensing and Its Applications:
Seeing Through Computation

Heung-No Lee, Ph.D.

Professor, GIST, KOREA

2015년 12월 28일(월)

Agenda

- 전세계적 연구 동향
- 압축센싱 Narrative
- 사례 소개

강의 요지 및 방향

- 올 해 강의는 2014년에 이어 Seeing through Computation (StC) 부제로 진행하고자 합니다.
- 전 세계적 연구동향을 살펴 보니 Breakthrough에 해당하는 StC 결과가 많이 나왔습니다.
- StC 의 여러 예를 들고 연구자들의 주의를 환기 시키고, 정보, 통신, 신호처리 이론, 전기전자 기초 이론 교육의 중요성을 강조하고자 합니다.

특·집·편·집·기

압축센싱: Seeing through Computation

이흥노 편집위원
(GIST 정보통신공학과)

지난 2011년 1월에는 숭실대학교 신요안 교수님께서 편집자로 수고해 주신 압축센싱에 관한 대한전자공학회 특집호가 출판되었습니다. 거기에서는 압축센싱 기술의 기초이론과 통신 및 초음파 시스템 응용에 대한 연구 동향을 살펴 보는 기회를 가졌습니다. 지난 특집호 이후 삼년 여가 지난 지금 어떤 새로운 연구가 진행되었는지 되돌아 보며, 압축센싱이라는 주제를 다시 한 번 다루어 보고자 합니다. 이렇게 압축센싱이라는 연구분야가 계속해서 여러 연구자들의 관심을 끄는 이유는 압축센싱 분야에서 오랫동안 사용되어 온 전통적 방식의 신호 획득 방법 및 그에 의한 시스템 디자인 방식을 무너뜨리는 전혀 새로운, 와해적인(interruptive) 시스템과, 혁신적(beakthrough)인 결과들이 속속 발표되고 있기 때문입니다.

그 예로서, 렌즈가 필요 없어서 어디든지 부착하고 사용할 수 있는 초소형 Lens Free 카메라, 위치 측위 정확도의 전통적 한계를 무너뜨리는 레이더, 전기 코일 통속에 앉아서 기다리는 시간을 10배 이상 획기적으로 줄인 고해상도 fMRI, Shannon-Nyquist 표본화와 같은 전자공학 상식을 뛰어넘는 Analog-to-Digital Converter, 기존의 방식으로는 상상할 수 없을 만큼의 넓은 주파수 대역을 모니터링 할 수 있는 인지무선 라디오의 스펙트럼 센싱 방법, 광 대역 통신 시스템에서 간섭을 제거하는 방법, 정확도를 크게 향상시킨 무선 채널 추정 방법 등을 들어볼 수 있습니다.

이번호에서는 이런 여러 가지 연구결과들을 Seeing through Computation이라는 부제를 붙이고 바라보고자 합니다. 압축센싱 이론을 응용측면에서 간략히 말하면, 신호측정은 간결히, 신호복원은 신중하고 정확하게 라고 할 수 있습니다. 후 신호 처리 Computation을 통해 복원을 잘 할 수 있으므로, 신호 측정을 하는 센서부분은 상대적으로 훨씬 간결하게 만들 수 있다는 의미입니다. 즉, 센서 숫자를 줄이거나 측정시간을 줄이거나 하는데 쓰일 수 있습니다. 혹은 센서 수나 신호 측정량을 고정시킬 경우, 센서 수 대비 정보 해상도 또는 정확도를 향상 시킬 수 있다는 것입니다. 현미경, 영상장치, ADC, 스펙트럼 센싱, 무선 채널 추정, 간섭신호 추적 등은 사물을 보는 것, 즉 Seeing 을 의미합니다. 후 처리 연산을 통해 신호를 잘 복원할 준비가 되어 있으면, 즉 Computing 자원을 확보하고 있으면, 센서를 통해 사물을 보는 시스템을 간결하고 쉽게 만들 수 있다는 것입니다.

이번 특집호에서는 저의 연구실과 Seeing through Compuation이라는 주제에 잘 맞는 연구를 진행해 오신 다섯 분의 연구자를 선별하여 초청하고 특집호를 구성 하였습니다. 먼저 최진호 교수님께서는 "Compressed Sensing Radar 연구 동향"을 통해 압축센싱의 레이더 시스템 적용 및 레이더의 분해능 증대 방안 연구 동향을 살펴보았습니다. 전병우 교수님께서는 "압축센싱과 영상처리 응용"의 제목으로, 단일 화소 카메라, Lens Free 카메라 등 압축센싱의 영상처리 응용 최신 동향을 살펴봐 주셨습니다. 황도식 교수님께서는 "다중변수 압축 MRI 기술 연구 동향"의 제목으로 다중변수 압축 MRI의 기본 원리와 관련 분야 연구 동향을 살펴 주셨습니다. 신요안 교수님께서는 "무선통신에서의 압축센싱 응용" 이란 제목의 논문에서, 압축센싱이 무선 통신에서 Analog-to-Digital Conversion, 무선 채널 추정, 인지 라이오 스펙트럼 센싱, 무선랜 다중 표적 무선 측위 등에서 정확도를 높일 수 있는 데 쓰일 수 있음을 보여주었습니다. 예종철 교수님께서는 "신호처리를 이용한 초고해상도 형광 현미경"이라는 제목의 논문에서 압축센싱이 형광 현미경의 해상도 증진에 쓰일수 있음을 보여주었습니다. 마지막으로 저의 연구실에서 광 대역 통신 시스템의 협 대역 간섭 제거를 위한 압축센싱 기술 연구 동향을 살펴 보았습니다.

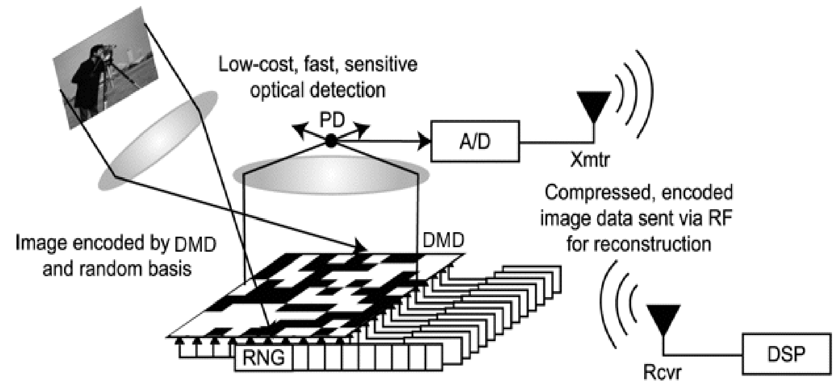
이 자리를 빌어 연구에 전념하느라 바쁘신 중에도 저의 원고 초대에 흔쾌히 응해 주시고, 촉박한 일정에도 불구하고, 기한 내에 원고를 마감해주신 다섯 분의 교수님들과, 심의와 조언을 아끼지 않아 주신 강의성 학회지 편집 위원장님과 편집위원님들께 감사의 말씀을 드립니다.

17

전자공학회지 2014. 6_497

Background

- 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, …



$$M < N$$

$$y = F x$$

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

$$\hat{x} = \arg\min_x \|x\|_{\ell_1} \quad \text{subject to } F x = y$$

전세계적 연구 동향

- (압축센싱이란) 2006년 정보이론과 신호처리 분야에 소개된 압축센싱이론은 응용 측면에서 **한 마디로 요약하면 '영상 및 음성 등의 자연신호를 압축적으로 센싱(Encoding) 할 수 있으며**, 그렇게 했을 때 **적은 량의 측정샘플 만으로도 신호를 복원(Decoding) 할 수 있다'** 정도로 표현 할 수 있음.
  - 2008 Rice대학 Single-Pixel Camera, 단 한 개의 포토센서로 영상을 압축하고 복원할 수 있음 시연하고 이론의 실제성 입증.

- 압축적 샘플링(Encoding)과 신호 복원(Decoding) 문제는 하나의 간단한 연립방정식 $y = Ax$ 으로 표현 가능함.
  - $A \sim M \times N$ 측정행렬이라 칭함.
  - $M<N$ 일 때 압축이 일어남.

- **신호복원**은 측정행렬 A를 알고 있는 상황에서, y 로부터 x를 찾아내는, 즉 **역문제를 푸는 문제임**.

- 이 역문제가 잘 풀리는 경우는?

혁신 기술 출현

- (혁신기술 출현) 압축센싱은 또한 "**적은 수의 센서로도**" 혹은 "**짧은 시간 동안의 측정으로도**" 신호를 **높은 해상도로 복원**하는데에도 응용 가능함이 밝혀졌음.

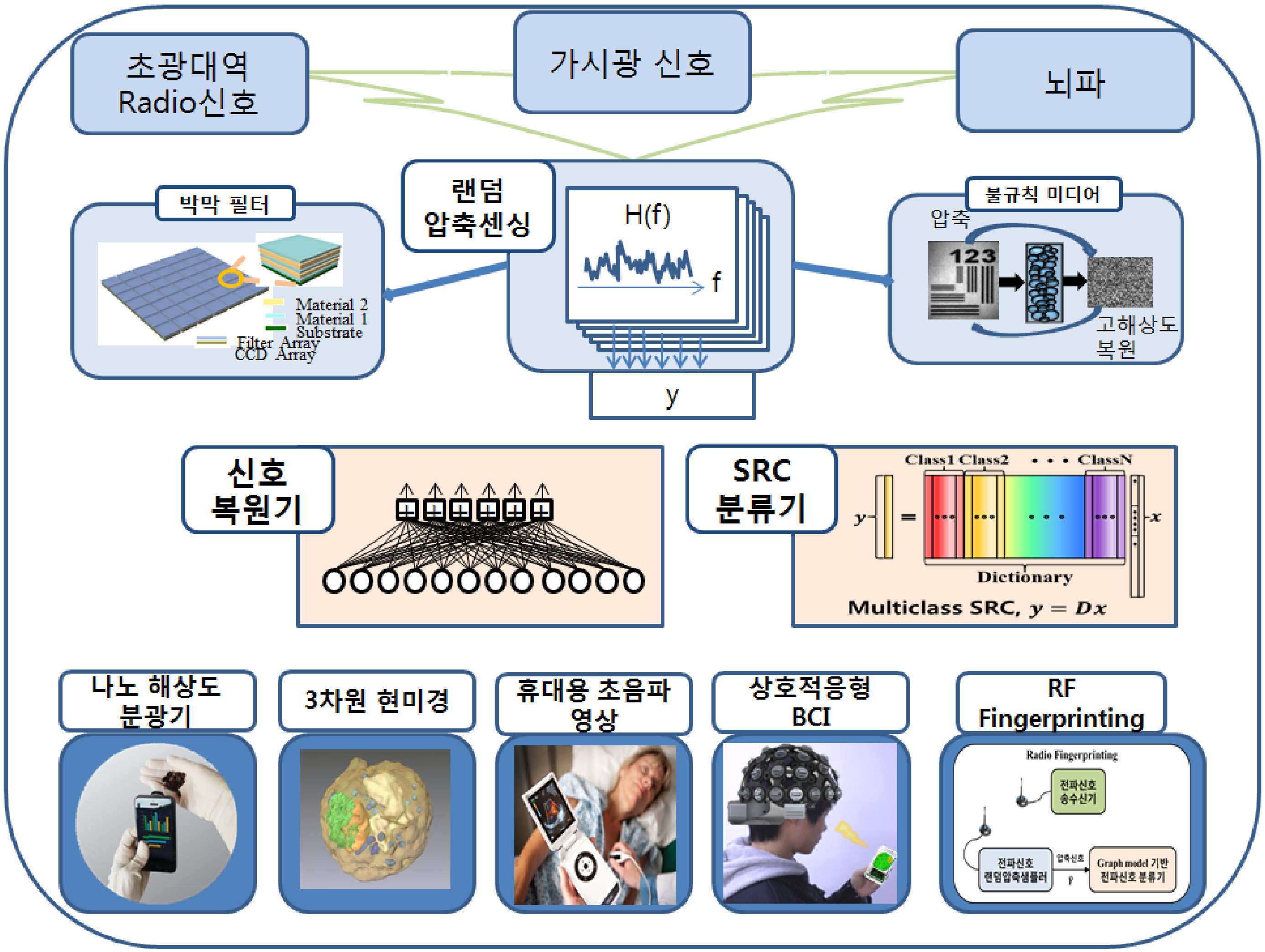
- 레이더, fMRI, 현미경등 영상장치들은 **신호가 센서에 도달하기까지 어떻게 변화하는지를 나타내는 광물리학적 전달 함수가** 잘 알려져 있는데, 이 점을 사용하여 **Decoding에 필요한 측정행렬 *A*를 만들어내고**, Diffraction Limit등을 돌파하는 **초 고해상도 신호복원이 가능함**이 알려짐.

- 이런 응용연구 결과의 발표는, 2010년 전후로부터 현재까지 지속적으로 이루어지고 있으며, 초 고해상도 레이더, 초 광대역 신호 측정, 촬영시간이 크게 단축된 fMRI 머신, 초 고해상도 현미경 등의 **혁신 기술의 출현을 가능케** 하고 있음.

Sparse Representation

- 압축센싱분야에서 개발된 Decoder 알고리즘은 또한 **신호를 희소하게 표현 하는 방법** (Sparse Representation) **으로도 응용되어 왔음**.
- 즉, 신호의 고유정보를 잃지 않고 신호의 **Dimensionality를 크게 Reduction 하는 방법**으로 쓰이고 있음.
- 나아가, 신호를 알려진 몇 가지의 **클래스로 구분하는데 뛰어난 성능**을 보임이 보고됨.
- CCTV 카메라에 찍힌 사람의 이상 행동 양식 구분, 사람의 얼굴 검색 등 Big Data 처리 및 분석 등에도 Sparse Representation Classification이라는 이름으로 응용되고 있음.
- 한 예로, 사용자의 뇌파 신호를 구분하여 컴퓨터 명령을 내릴 수 있는 Brain-Computer Interface 용 분류알고리즘을 들 수 있음.

Focus Areas in this talk



Encoding 연구의 혁신성

- 압축센싱의 **Encoding 문제는 아날로그 신호 *x*(*t*)를 아날로그 도메인 상에서 압축적으로 샘플링하여 디지털 샘플 벡터 *y*를 얻는 전기/전자/재료 공학적 장치를 만들어 내야하는 문제임**.

- 다시 말해, 압축센싱이론 논문에서 흔히 다루는, 이미 디지털 샘플링이 완료된 신호 벡터 *x*를 측정행렬 *A*와 곱하여 *y*를 얻는 것은 Encoding 문제가 아님.

- 아날로그 도메인에서 연속적인 신호 *x*(*t*)를 랜덤한 반응 패턴을 가진 함수와 곱하고, 그 결과를 더해서 샘플을 취하는 장치를 만드는 연구 및 실제 구현이 혁신적 성과를 내기 위해 꼭 필요함.

Decoding 연구 지속의 중요성

- 많은 Decoder가 현재 개발 되었음
  - Gaussian, FFT, Bernoulli 센싱등 문제가 수학적으로 잘 알려진 형태 일 때

- 특정 문제에 특화된 Decoder 개발 필요 함
  - Photonics, Spectrometers, Electronic Eyes, …
  - Non negative signal sensing
  - No general sensing matrices

FUNDAMENTAL QUESTIONS ON DECODING PROBLEM

Shannon’s Sampling Theorem

Shannon Nyquist Sampling Theorem

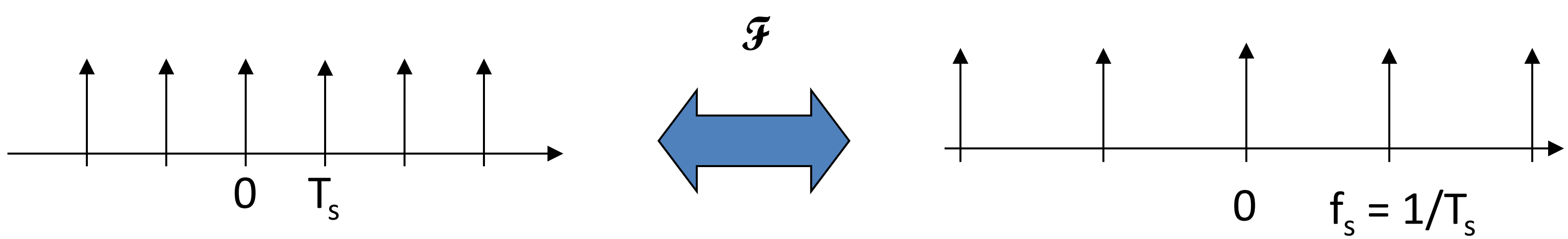
- Consider taking samples of continuous time signal.

- 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

$$\sum_{k=-1}^{1} \delta(t-kT_s) \text{ , } f_s \sum_{n=-1}^{1} \delta(f - n f_s)$$

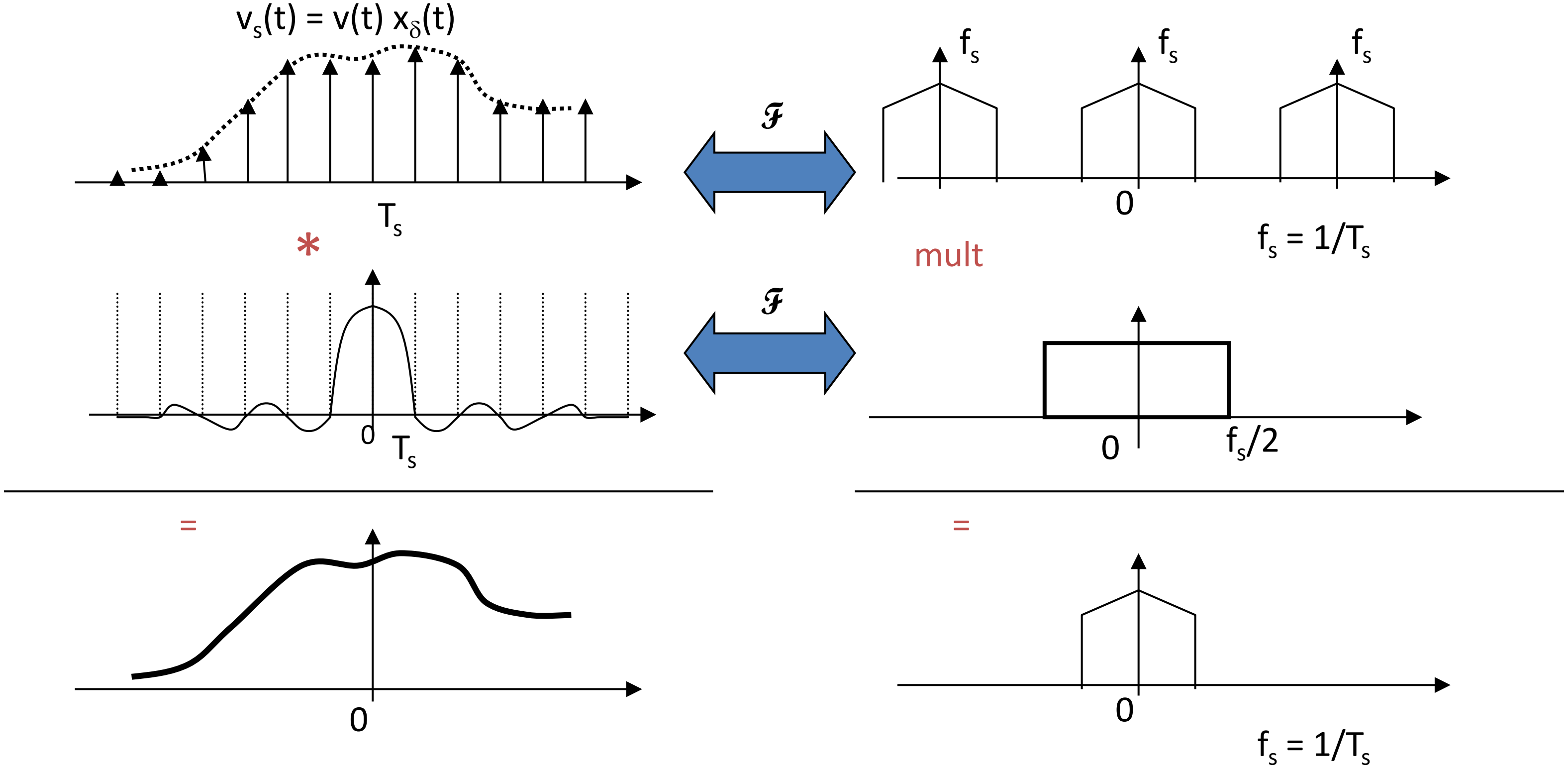
where $f_s = 1/T_s$



Shannon 1948 paper

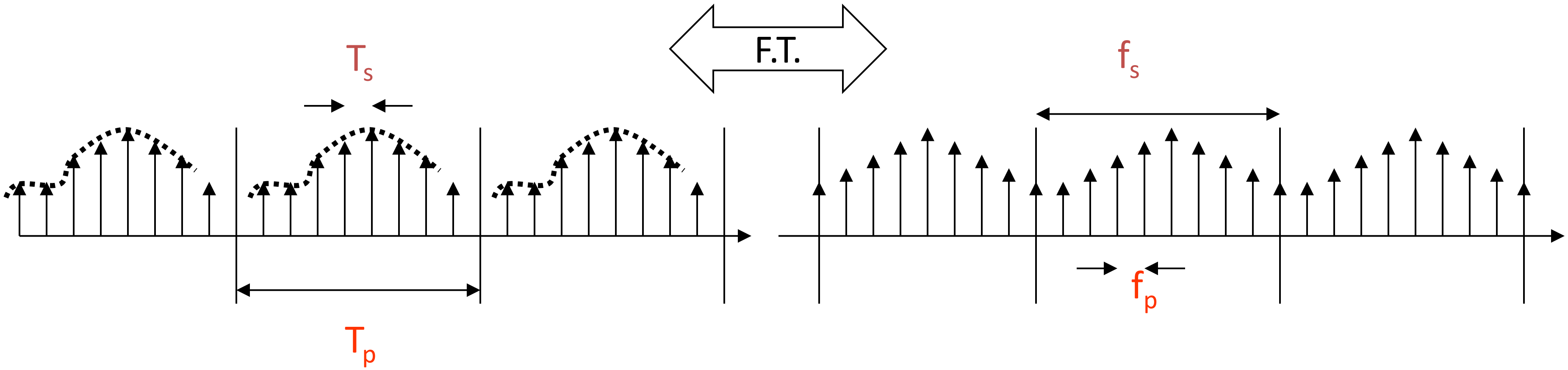
- Theorem 13: Let $f(t)$ contain no frequency over W. Then,

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



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 choosing $T_p/T_s = T_p W = N$, an integer.
- Then there are N distinct samples (in each domain).
- The discrete samples of the signal x(k), k=0, 1, 2, …, N-1.
- The discrete samples of the Fourier spectrum X(n), n = 0, 1, 2, …, N-1.



Discrete Fourier Transform

- DFT

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

- Inverse DFT

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

- 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 given so far

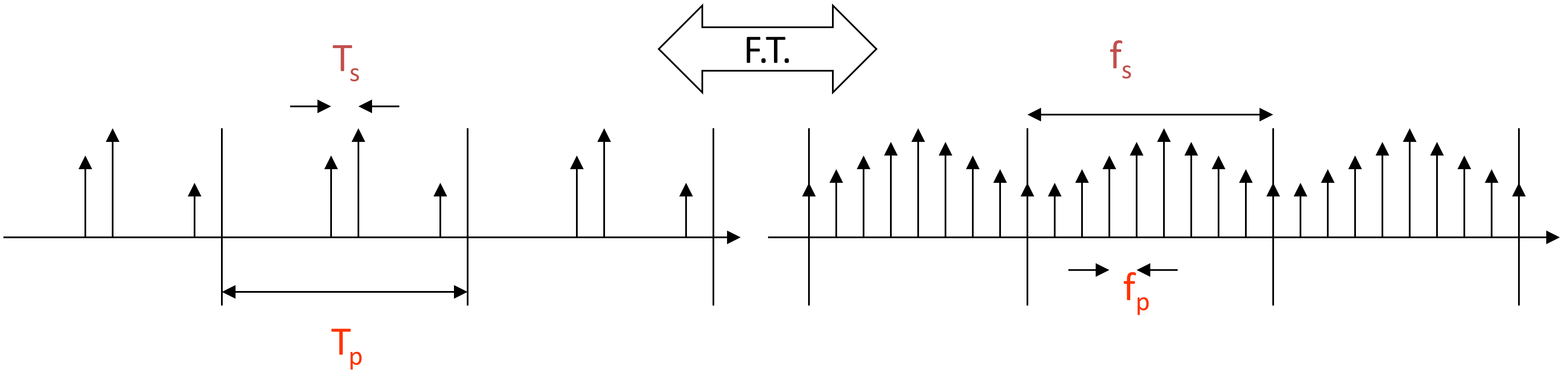
- Are covered in Systems and Signals in Electrical Engineering…

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

Sparse Signals, Recovery with L1 Norm 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 $k = 0,1,2, \ldots, N-1$

  $\mathbf{X} = \mathbf{Fx}$

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

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

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

  $= \langle x, m\text{-th tone} \rangle$ for $m=1, 2, \ldots, M$

  where M is the total number of measurements.

  $\mathcal{M} \subset \mathbb{Z}_N$

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

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

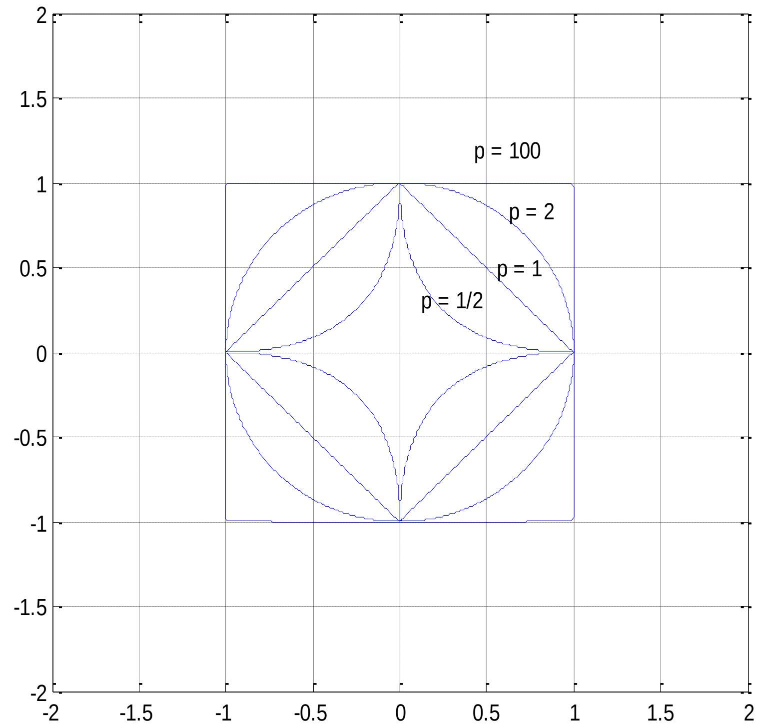
L1 norm?

- The L-*p* norm of **x** is defined for p>0

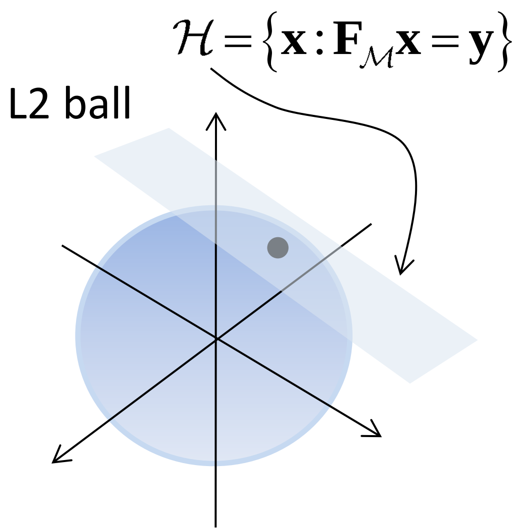
$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^{N} |x_i|^p \right)^{\frac{1}{p}}$$

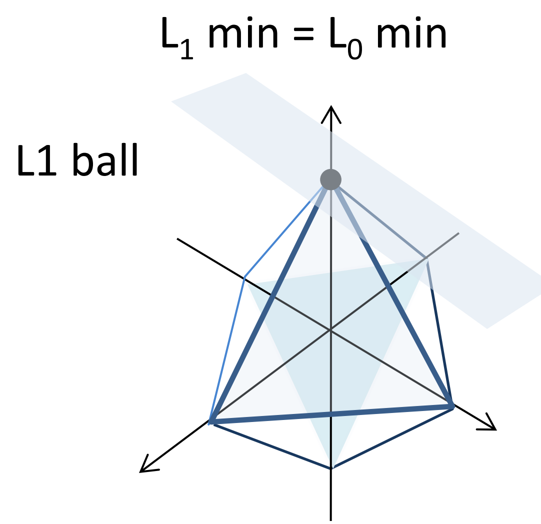
- The L0 norm is not well defined as a norm.
  - Donoho uses it as a "norm" whi ch counts the number of non-z ero elements in a vector.

- 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





$$x = \arg\min \|x'\|_2 \quad \text{s.t. } y = Fx$$

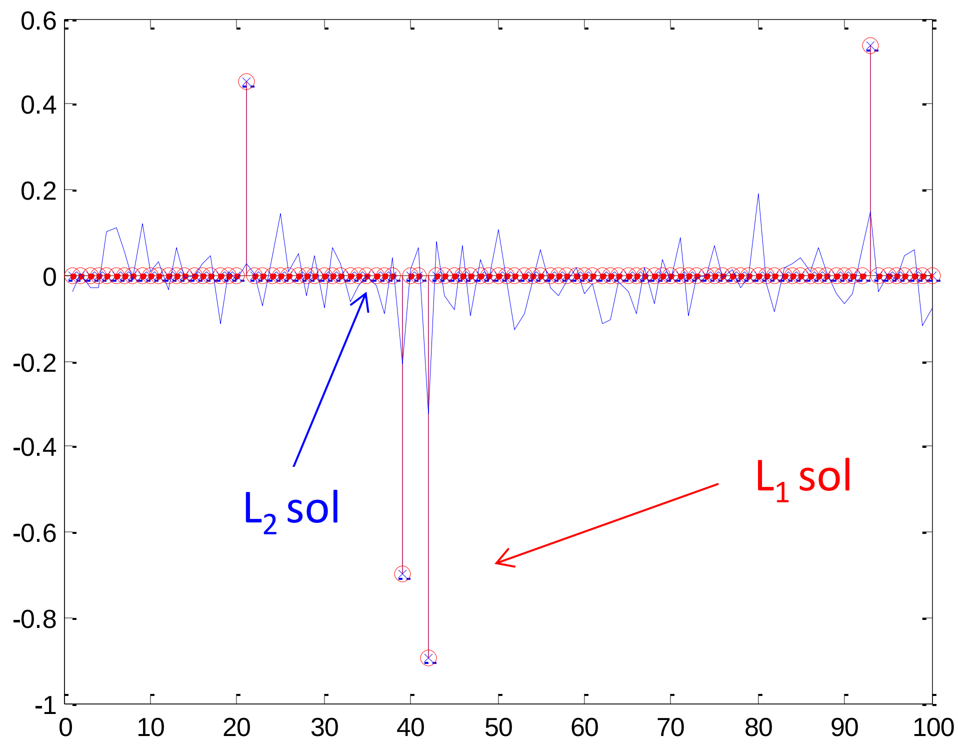
$$= F^T \left(FF^T\right)^{-1} y$$

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

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

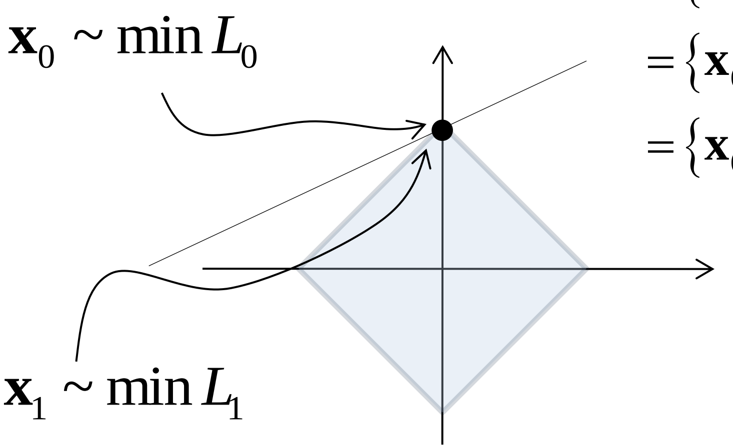
L_2 vs. L_1 solutions

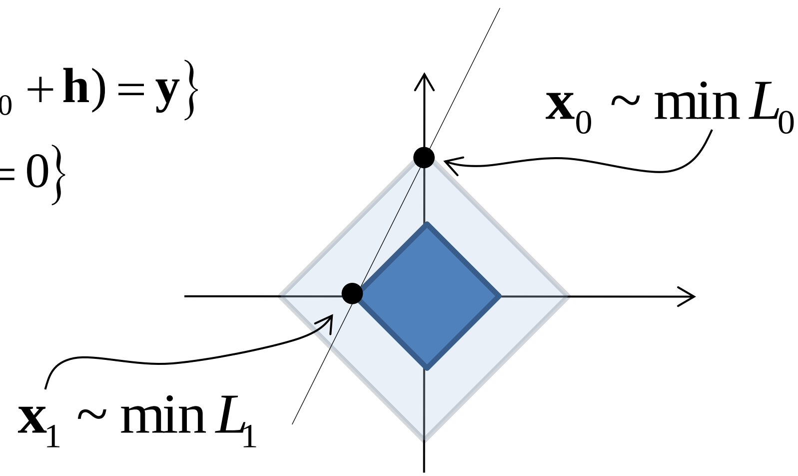
- L2 solution has energy spread out to everywhere.
- L1 solution attains the sparse signal.



Good vs. Bad

$$\mathcal{H} = \{\mathbf{x} : \mathbf{F}_{\mathcal{M}}\mathbf{x} = \mathbf{y}\}$$
$$= \{\mathbf{x}_0 + \mathbf{h} : \mathbf{F}_{\mathcal{M}}(\mathbf{x}_0 + \mathbf{h}) = \mathbf{y}\}$$
$$= \{\mathbf{x}_0 + \mathbf{h} : \mathbf{F}_{\mathcal{M}}\mathbf{h} = 0\}$$





- 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 frequently if $\lambda = O(\log N)^6$

Basic Decoding Problem

- How many number of measurements *M* is required for successful recovery?

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

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

$$M \geq 2K \Leftrightarrow \text{unique solution}$$

Key CS Results (2)

- As long as the solution is unique, a search algorithm can find it exactly.

- Proof: Suppose two *K*-sparse solutions **x**, **x**'. Then, we have

$$\mathbf{F}_{\mathcal{M}}(\mathbf{x}-\mathbf{x}')=0$$

An *M* dimensional map

A vector of dimension <=2*K*

- This map is injective. Thus, RHS can't be zero unless (**x**-**x**')=0.

Canonical CS Results (3)

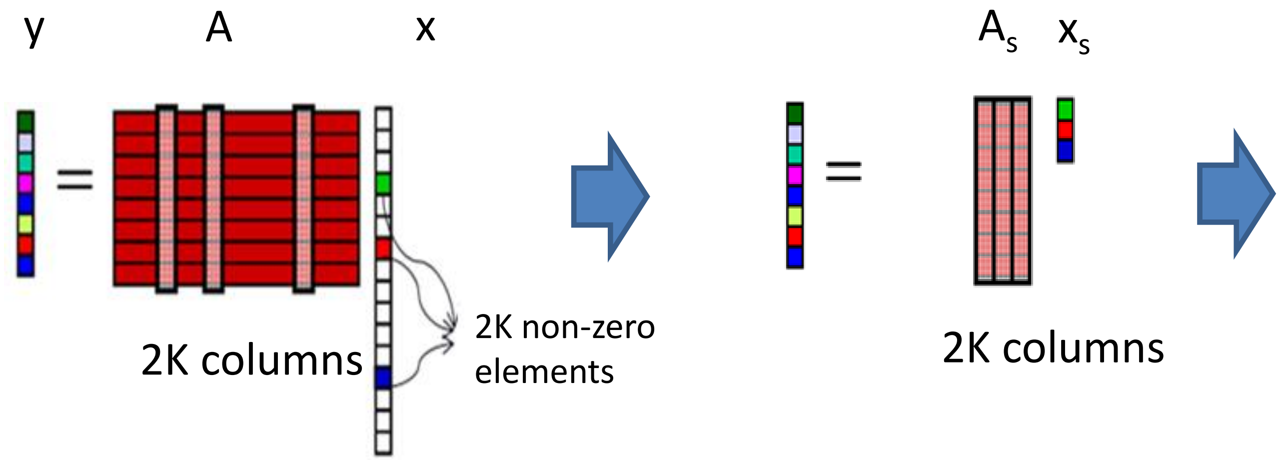
- There are ${}_N C_K$ different ways to choose a set of K columns that accounts for the observation **y**.
- The complexity of L_0 min search is $\binom{N}{K} \cong 2^{N H\left(\frac{K}{N}\right)}$

- (P1) A relaxed approach is L_1 minimization:

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

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

When the support set is known



$$x_S = \left(A_S^T A_S\right)^{-1} A_S^T y$$

- Decoding is easy!!
  - as long as 2K columns ~ linearly independent
- If any 2K selections of columns of A ~ lin. Ind., the solution is unique.
- Spark of A = the size of the smallest subset of columns of A that are l.d.
  - 예: Spark of A = 5.

Spark of Sensing Matrix

- It determines how good the sensing matrix is in a fundamental way.
- Suppose an A with Spark = 5.
- Then, any 2-sparse signal can be uniquely transformed into y.
- This means an exhaustive search (L0 decoding) will give perfect recovery.
- Note that M>=Spark-1 (Singleton bound)

Loose Thoughts on the Number of Measurements

- Suppose random A
- Total number of support sets ~ 2^M
- M ~ K log N/K
- A measurement is either zero or non-zero.
- A zero measurement ~ useless
- A non-zero measurement ~ useful measurement

$$\binom{N}{K} \cong 2^{N H\left(\frac{K}{N}\right)}$$

$$\approx 2^{K \log N/K}$$

$$H\left(\tfrac{K}{N}\right) \sim \tfrac{K}{N} \log \tfrac{N}{K}$$

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

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

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

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

$$\begin{bmatrix} A & 0 \end{bmatrix} \begin{pmatrix} x \\ u \end{pmatrix} - b = 0$$

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

- Basic approach
  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)

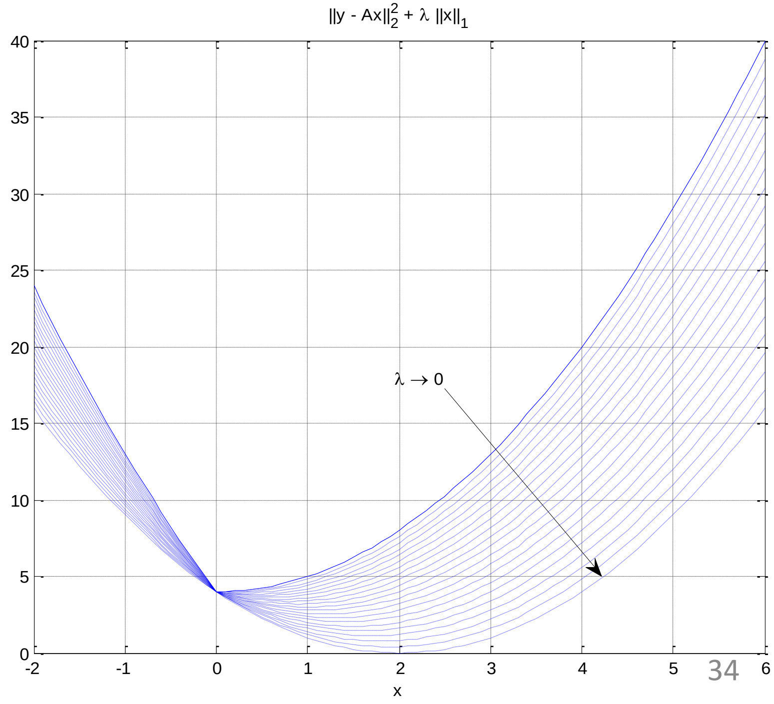
- The LP approach is to build the sparse solution from an initial guess which is dense.
  - $O(N^3)$
- If the exact solution is known to be sparse, why don't we start from a null set and build up a sparse solution?
- **Homotopy** [Donoho-Tsaig08']

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

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

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

- 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 xj and the step size.
  5. Stop when the residual correlation is zero; otherwise repeat 2 – 4.



Compressed Sensing Narrative

Compressed Sensing Narrative

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

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

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

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

- 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 t o 1 as long as the number of measurements are sufficiently large,

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

  - Where the oversampling factor is

$$\lambda = \left(\log N\right)^6 \quad \text{for the FT matrix}$$
$$\lambda = \left(\log N\right) \quad \text{for the Gaussian and Bernoulli matrices}$$

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 con sequence of the opposite case.)
  - Thus, *F* is usually constructed with the random Gaussian matrix since t he statistical property of *FB* remains the same as that of *F* when *B* is un itary (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, then *Fx* should behave like a unitary transformat ion (i.e., *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 for the probability that a sensing matrix *F* selected randomly from an ensemble of *M* x *N* matrices (say i.i.d. Gaussian) to have a given RIP constant δ.
  - This gives a large deviation analysis which then leads to the probabilistic statement of the following form: *for any K−sparse x*

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

- 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 i n L1 recovery.
  - L1 recovery results are not much worse than the model mismatch and observation errors.

Recasting CS to Channel Coding Theory context

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

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{bmatrix} 100110 \\ 010101 \\ 001011 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ x_3 \\ 0 \\ 0 \\ x_6 \end{pmatrix}$$

- 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 + 1.
  - 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 output: $z = c + x$ (x is the channel errors)
- 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)

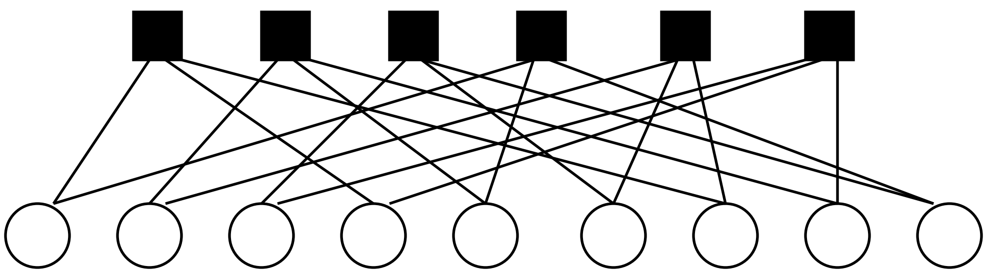
- Let $G = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$
- Using FG = 0, find F = $\begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$
- Note 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+1* (The singleton bound)
- Note that UUP is met for F.
  - For all 1-spare signal x, Fx is non-zero.
  - *M* = 3 > 2K = 2.

When SPARK is defined for real valued matrix, and d_{min} is for binary field. Then, SPARK >= d_{min} See [Lui13]

LDPC Code/Bipartite Graph

$$M/N = 6/9$$

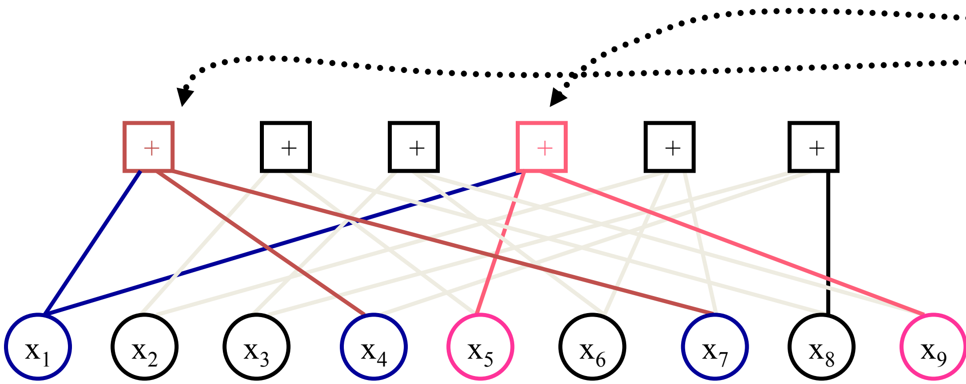
$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_9 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_6 \end{bmatrix}$$



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)

$p_{1,1}$ $p_{2,1}$ $p_{3,1}$ $p_{4,1}$ $p_{5,1}$ $p_{6,1}$ $p_{7,1}$ $p_{8,1}$ $p_{9,1}$

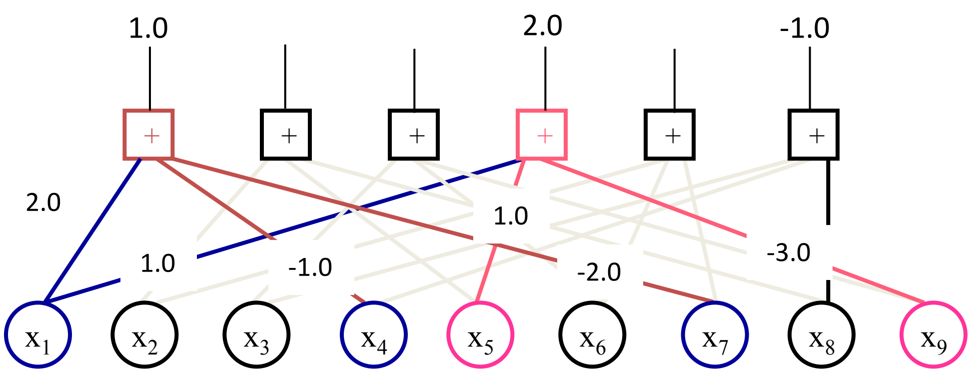
$p_{i,1} \coloneqq \Pr\{xi = 1\}$

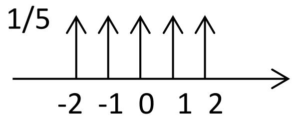
$$P(S \mid x_1 = 1, \mathbf{y}) = \Pr\{\text{odd \# of 1s in } x_4 \text{ and } x_7\} \times \Pr\{\text{odd \# of 1s in } x_5 \text{ and } x_9\}$$

$$= \left\{ p_{4,1}(1 - p_{7,1}) + (1 - p_{4,1}) p_{7,1} \right\} \times \left\{ p_{5,1}(1 - p_{9,1}) + (1 - p_{5,1}) p_{9,1} \right\}$$

Let’s do this problem

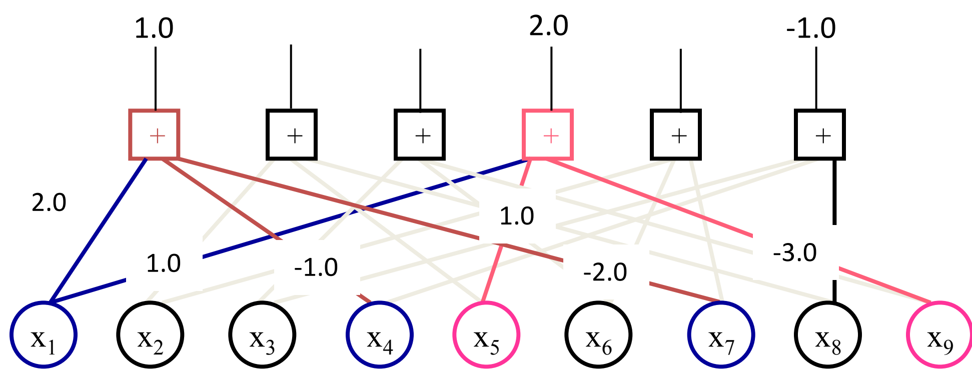
- Determine the pdf at x_1 given the pdfs at x_4, x_7, x_5, x_9 and y.
- Find the pdf at x_1





Let's do this problem (2)

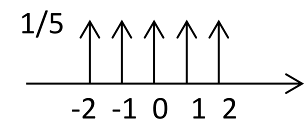
- Determine the pdf at x_1 given the pdfs at x_4, x_7, x_5, x_9 and y.
- How to find pdf at x_1
- An example at x_1 at -2



- Via the first check ($1.0 = 2x_1 - x_4 - 2x_7$)

  $P(S|x_1 = -2, y_1=1.0) = P(1.0 = -4 - x_4 - 2x_7) = P(5= - x_4 - 2x_7) =$ sum_x_4 $P(5= - x_4 - 2x_7) = 1/5\{P(5 = 2 - 2x_7) + P(5 = 1 - 2x_7) + P(5 = 0 - 2x_7) + P(5 = -1 - 2x_7) + P(5 = -2 - 2x_7)\}$

  $= 1/5\{P(x_7 = -3/2) + P(x_7 = -4/2) + P(x_7=-5/2) + P(x_7=-6/2) + P(x_7=-7/2) = 1/5(0+1/5+0+0+0) = 1/25$



- Via the 4^{th} check ($2.0 = x_1 + x_5 - 3x_9$)

  $P(S|x_1 = -2, y_4=2.0) = P(2 = -2 + x_5 - 2x_9)$

  $= P(4=x_5 - 2x_9)$

  $=$ sum_x_5 $P[x_9 = (x_5-4)/2]$

  $= 1/5\{P[x_9 = -6/2]+P[x_9 = -5/2]+$

  $P[x_9 = -4/2]+P[x_9 = -3/2]+P[x_9 = -2/2]\}$

  $= 1/5\{0+0+1/5+0+1/5\} = 2/25$

- Combine the two $P(S|x_1 = -2, y) = (1/25)*(2/25)$

- Further examples at other points in x_1 is straight forward ~ the product of the convolutions of two pdfs

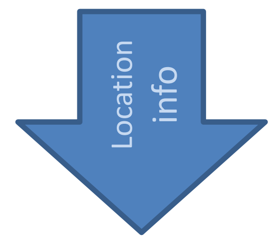
Leads to BHT and AMP
(The 4^{th} Discussion in this presentation)

- Pdfs of other variables can be similarly obtained

Reed Solomon Codes

- Design of (*N*, *NR*, *K*) RS code
  - Selecting the 2*K* consecutive n-th root of unity as the roots of the generator polynomial g(x).
  - The resulting syndrome equation is Vandermonde system
    - Spark = 2K +1
    - Achieves the Singleton bound (max. spark for given *N* and code rate *R*.

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



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

- Decoding is done in 2 step
  - Error locator polynomial
  - Over determined matrix inversion

Generator/Parity Check of BCH codes

- $g(x)$ = LCM of minimal polynomials for the D consecutive powers of α.
  - LCM is the sufficient condition that the roots of $g(x)$ are the D consecutive powers of α.

- A code polynomial $c(x)$ is a multiple of generator polynomial. ($\Leftrightarrow$ $c(\alpha^b) = c(\alpha^{b+1}) = \ldots = c(\alpha^{b+D-1}) = 0$ for some $b = 1, 2, \ldots$)
  - *Nomenclature: Narrow sense* ($b = 1$) and *primitive* ($n = q^m - 1$)

- Consider $\mathbf{s} = \mathbf{H}\mathbf{c}^T$:

$$\begin{pmatrix} 1 & \alpha^b & \alpha^{2b} & \alpha^{3b} & \ldots & \alpha^{(n-1)b} \\ 1 & \alpha^{b+1} & \alpha^{2(b+1)} & \alpha^{3(b+1)} & \ldots & \alpha^{(n-1)(b+1)} \\ 1 & \alpha^{b+2} & \alpha^{2(b+2)} & \alpha^{3(b+2)} & \ldots & \alpha^{(n-1)(b+2)} \\ 1 & \alpha^{b+D-1} & \alpha^{2(b+D-1)} & \alpha^{3(b+D-1)} & \ldots & \alpha^{(n-1)(b+D-1)} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ \ldots \\ c_{(n-1)} \end{pmatrix} = \mathbf{0}$$

Are you now being convinced to witness the relation between the compressed sensing and channel coding theory?

Let us discuss some specific examples.

Several Examples

1. Prony method [Prony1795]
2. Eigenvalue distribution of Gram matrix and RIP
3. Super-Resolution
4. Support set detection using hypothesis testing on belief propagation results
5. Number of measurements needed in multiple correlated measurement cases
6. Sparse vs. dense matrices for compressed sensing over GF
7. Brain Computer Interface with EEG and SR Classification
8. Turbid lens imaging, Communications Problems, Radars, Comp-Eyes, X-ampling

FIRST ONE

Gaspard Clair François Marie Riche de Prony
(July 22, 1755 - July 29, 1839)

Prony's Method [1795] in CS framework

- Classical Prony's method can be cast into CS framework [Vetterli07]

- The signal model in Prony's method consists of linear combination of K ex ponentials with unknown amplitudes $\{c_i\}$

$$y_m = \sum_{i=0}^{N-1} c_i u_i^m = \sum_{p=0}^{K-1} c_{i_p} u_{i_p}^m \quad c_i \in R \quad u_i \in C \quad m = 1, 2, \cdots, M$$

- Given $\{y_m\}$, our aim is to find the unknowns non-zero coefficients$\{c_{i_p}\}$ and its locations$\{i_p\}$.

- The above equation is similar to CS system with K non-zeros values $\{c_i\}$, which can be written in matrix-vector form as $y = Vc$.

- The Vandermonde matrix V, which acts a sensing matrix, is known at the recovery point.

Prony's Method Contd.

- Prony gave a solution to find the unknowns in 1795 called the annihilation filter method.
  - Annihilation filter is a sequence which when convolved with a given sequence results in zero always. It can be constructed for the measurement $\{y_m\}$ as well.

- Call $\{h_m\}$ the filter with Z-transform $H(z) = \sum_{i=0}^{K} h_i z^{-i} = \prod_{i=1}^{K} \left(1 - u_i z^{-1}\right)$
  - $\{u_i\}$ are called (zeros) roots of the filter, that is, $H(u_i) = 0$.

- If the roots of the filter are the same as K exponentials that constitute $\{y_m\}$, then $y_m * h_m = 0$, i.e.,

$$y_m * h_m = \sum_{l=0}^{K-1} h_l y_{m-l}$$

$$= \sum_{l=0}^{K-1} h_l \left(\sum_{p=0}^{K-1} c_{i_p} u_{i_p}^{m-l} \right) = \sum_{p=0}^{K-1} c_{i_p} u_{i_p}^{m} \underbrace{\sum_{l=0}^{K-1} h_l u_{i_p}^{-l}}_{H\left(u_{i_p}\right)=0} = 0$$

Prony's Method Contd.

- Given $\{y_m\}$ how to find an annihilation filter $\{h_m\}$?
- What we need is $y_m * h_m = 0$, which can be written as

Let m = K here.

$$\underbrace{\begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ y_m & y_{m-1} & \cdots & y_0 \\ y_{m+1} & y_m & \cdots & y_1 \\ \vdots & \vdots & \ddots & \vdots \\ y_{2m-1} & y_{2m-2} & \cdots & y_{m-1} \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix}}_{Y} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{K-1} \end{bmatrix} = 0$$

- If at least $2K + 1$ values of y are available, then the above system admits a solution when rank(Y)=K
- Taking h_0= *1,* the above system can be solved for $\{h_1, \ldots, h_{K-1}\}$.

Prony's Method Contd.

- Given the coefficients $\{1, h_1, \ldots, h_{K-1}\}$, we can get the unknown locations $\{i_p\}$ by polynomial root finding [Vetterli02].

- That is, by finding the zeros of the filter $H(z)$ as $\{h_0, h_1, \ldots, h_{K-1}\}$ a nd $\{u_{i_p}\}$ are related by

$$H(z) = \sum_{i=0}^{K} h_i z^{-i} = \prod_{p=1}^{K} \left(1 - u_{i_p} z^{-1}\right)$$

- Polynomial rooting can be done in $O(K^2)$ operations.

Prony's Method Contd.

- K non-zeros values $\{c_i\}$ can then be obtained by solving linear system of equati ons (Classic Vandermonde system)

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ u_{i_0} & u_{i_1} & \cdots & u_{i_{K-1}} \\ \vdots & \vdots & \ddots & \vdots \\ u_{i_0}^{K-1} & u_{i_1}^{K-1} & \cdots & u_{i_{K-1}}^{K-1} \end{bmatrix} \begin{bmatrix} c_{i_0} \\ c_{i_1} \\ \vdots \\ c_{i_{K-1}} \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{K-1} \end{bmatrix}$$

- The above system has the unique solution when $u_p \neq u_q$, for $p \neq q$.

- In summary, by using Prony's method, only 2K +1 measurements are needed t o decode a K-sparse signal.

Summary

- Sensing a real-valued sparse signal with the Vandermonde system is very good.
- This method gives the best performance with the least number of measurements, $M > 2K$.
- This works for real valued unknown sparse signal.
- The recovery process of sparse signal is similar to RS decoder.

Eigenvalue distribution

SECOND ONE

Eigenvalue in Compressive Sensing

- In CS, restricted isometry constant (RIC) of a sensing matrix has an intriguing connection to the eigenvalues

- RIC of a sensing matrix measures the goodness of the matrix for sensing and recovery of sparse signals

$$\left(1-\delta_K\right) \leq \frac{\left\|F_{\mathcal{K}} \boldsymbol{x}_{\mathcal{K}}\right\|^2}{\left\|\boldsymbol{x}_{\mathcal{K}}\right\|^2} \leq \left(1+\delta_K\right) \qquad \lambda_{\min}\left(F_{\mathcal{K}}^T F_{\mathcal{K}}\right) \leq \frac{\left\|F_{\mathcal{K}} \boldsymbol{x}_{\mathcal{K}}\right\|^2}{\left\|\boldsymbol{x}_{\mathcal{K}}\right\|^2} \leq \lambda_{\max}\left(F_{\mathcal{K}}^T F_{\mathcal{K}}\right) .$$

- $\mathcal{K}$ denotes the support set

- We can say probabilistic statements about the RIC δ_K if we know the eigenvalues of the $K \times K$ matrix $F_{\mathcal{K}}^T F_{\mathcal{K}}$

- Aim: To derive novel, tractable eigenvalue distributions

Wishart Matrices and RIP

- When F is an $M \times K$ Gaussian matrix, $F^T F$ is a Wishart matrix (popular in multivariate statistics and MIMO communications)

- We derive the extreme eigenvalue distributions $f(\lambda_{\max})$ and $f(\lambda_{\min})$ of the Wishart matrix from the joint distribution $f(\lambda)$ of the eigenvalues:

$$f(\lambda) = \frac{\pi^{K^2/2} \rho^{-KM/2}}{2^{KM/2} \Gamma_K\left(\frac{M}{2}\right) \Gamma_K\left(\frac{K}{2}\right)} \left|V(\lambda)\right| \prod_{i=1}^{K} \lambda_i^{(M-K-1)/2} \, e^{\frac{-\lambda_i}{2\rho}}$$

ρ - Variance of Gaussian random variable

Multivariate Gamma function

Vandermonde Matrix of eigenvalues $\lambda_{\max}, \lambda_2, \lambda_3, \cdots, \lambda_{\min}$

Maximum and Minimum Eigenvalues

- Maximum eigenvalue is obtained by

$$f(\lambda_{\max}) = \int_0^{\lambda_{\max}} \int_0^{\lambda_2} \cdots \int_0^{\lambda_{K-1}} f(\boldsymbol{\lambda})\, d\lambda_{\min} \cdots d\lambda_3 d\lambda_2$$

- Minimum eigenvalue is obtained by

$$f(\lambda_{\min}) = \int_{\lambda_{\min}}^{\infty} \cdots \int_{\lambda_3}^{\infty} \int_{\lambda_2}^{\infty} f(\boldsymbol{\lambda})\, d\lambda_{\max} d\lambda_2 \cdots d\lambda_{K-1}$$

- After substituting the joint distribution in the above expressions, we follow two key steps:
  1. Expansion of the Vandermonde determinant along the desired eigenvalue.
  2. Multiple integration of sub-determinants using the theory of sk ew-symmetric matrices.

Maximum and Minimum Eigenvalues

- Maximum eigenvalue distribution

$$f\left(\lambda_1\right) = c\sum_{n=1}^{K}(-1)^{n+1}\ \lambda_1^{K-n+(M-K-1)/2}e^{-\frac{\lambda_1}{2\rho}}\ \mathrm{PF}\left(\mathrm{B}_n\right)$$

  - PF is a Pfaffian of skew-symmetric matrix ($A = -A^T$)
  - $\mathrm{PF}\left(A\right) = \sqrt{\det\ A}$
  - The (i,j)th entry of B_n for odd K is

$$b_{i,j} = \int_0^{\lambda_1}\int_0^{\lambda_1}\theta_i\left(\lambda_i\right)\theta_j\left(\lambda_j\right)\mathrm{sgn}\left(\lambda_j - \lambda_i\right)d\lambda_i d\lambda_j\ ,$$

$$\theta_i\left(\lambda_i\right) = \lambda_i^{r_{n,i}}\lambda_i^{(M-K-1)/2}e^{-\frac{\lambda_i}{2\rho}}$$ $r_{n,i}$ is an non-negative integer

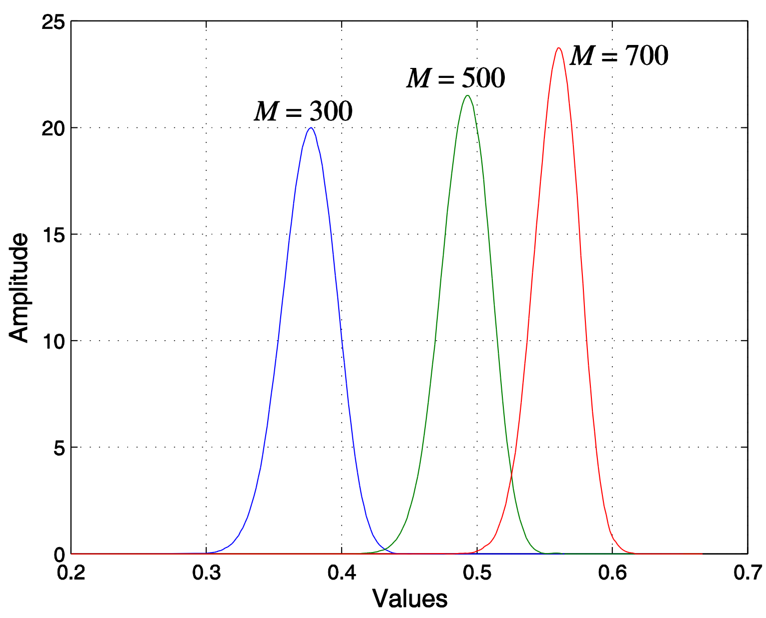
- Minimum eigenvalue distribution

$$f\left(\lambda_K\right) = c\sum_{n=1}^{K}(-1)^{n+K}\ \lambda_K^{K-n+(M-K-1)/2}e^{-\frac{\lambda_K}{2\rho}}\ \mathrm{PF}\left(\mathrm{D}_n\right)$$

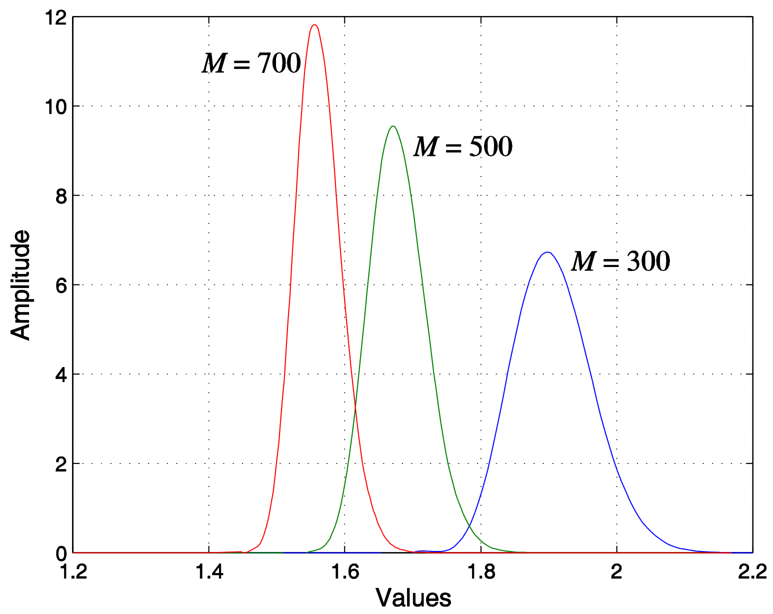
Plots of Eigenvalues

❖ Simulation set-up

$$K = 51\ , \quad M = 300, 500 \text{ and } 700$$



Minimum eigenvalue of Wishart matrix



Maximum eigenvalue of Wishart matrix

Role of RIC in Compressive Sensing

- A sensing matrix with a good RIC is deemed good for sensing and reconstruction of sparse signals

- If a sensing matrix F satisfies the RIP

$$(1-\delta_K) \le \frac{\left\| F_{\mathcal{K}} \boldsymbol{x}_{\mathcal{K}} \right\|^2}{\left\| \boldsymbol{x}_{\mathcal{K}} \right\|^2} \le (1+\delta_K)$$

with $\delta_K < 1$, then F is said to satisfy *RIP of order K.*

- We aim to state RIP of order K using eigenvalues

Eigenvalue in Undersampling Analysis

- Since

$$\lambda_{\min}\left(F_{\mathcal{K}}^{T} F_{\mathcal{K}}\right) \leq \frac{\left\|F_{\mathcal{K}} \boldsymbol{x}_{\mathcal{K}}\right\|^{2}}{\left\|\boldsymbol{x}_{\mathcal{K}}\right\|^{2}} \leq \lambda_{\max}\left(F_{\mathcal{K}}^{T} F_{\mathcal{K}}\right) .$$

  the RIC can be related to minimum eigenvalue as

$$1-\delta_{K} = \min_{\mathcal{K}} \lambda_{\min}\left(F_{\mathcal{K}}^{T} F_{\mathcal{K}}\right)$$

- We say that a matrix F satisfies the RIP of order K if

$$\Pr\left\{\min_{\mathcal{K}} \lambda_{\min}\left(F_{\mathcal{K}}^{T} F_{\mathcal{K}}\right) > a\right\} > 1-\eta$$

  and we call such matrix a well-conditioned matrix

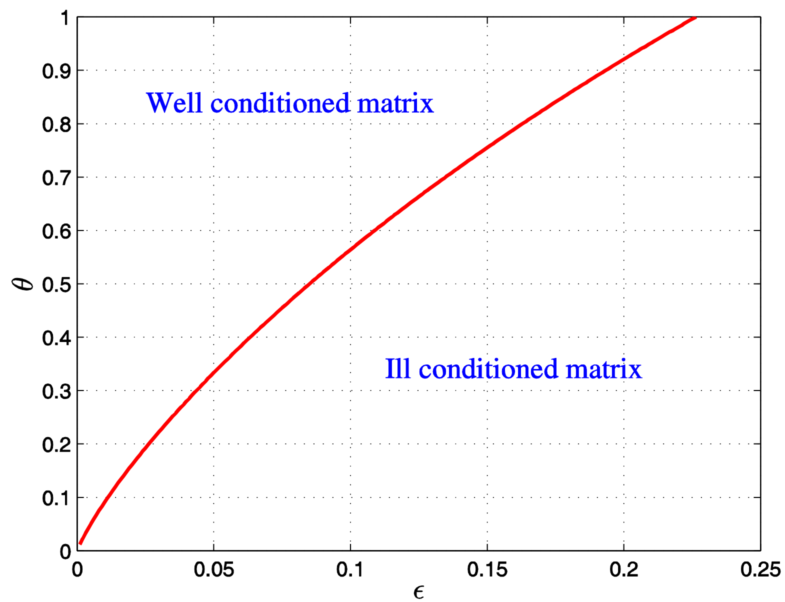
Probability of Well-Conditioned Matrix

$$\Pr\{\text{Well-conditioned matrix}\} = \Pr\left\{\min_{\mathcal{K}} \lambda_{\min}\left(F_{\mathcal{K}}^T F_{\mathcal{K}}\right) > a\right\}$$

$$= \Pr\left\{1-\delta_K > a\right\}$$

$$\geq 1 - e^{-N\, E_K}$$

- The exponent is a function sparsity ratio $\varepsilon := \frac{K}{N}$ and undersampling ratio $\theta := \frac{M}{N}$



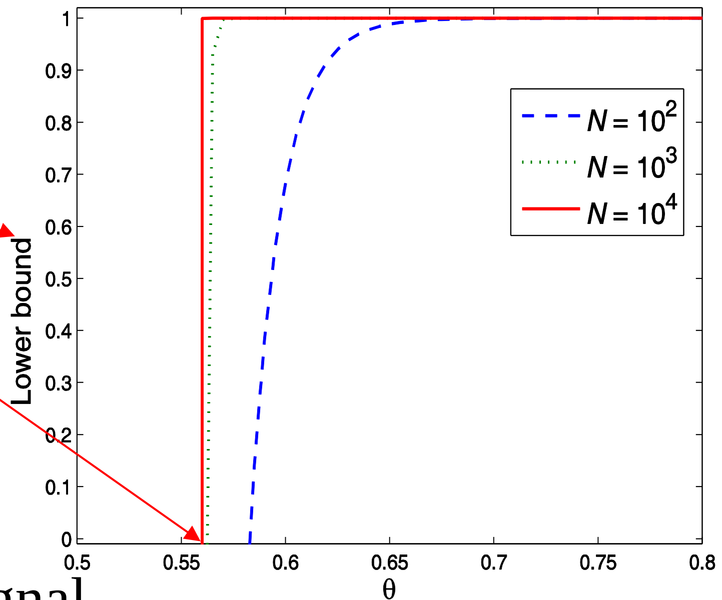
Under-sampling Analysis

- Undersampling analysis : Aims to find the minimum number of measurements needed by using a matrix with a specific RIC: $\delta_K < \delta$.

$$\Pr\left\{\delta_K < \delta\right\} \geq 1 - e^{-N\,E_K}$$

- For OMP, [Davenport10] advised that a matrix with $\delta_{K+1} < \frac{1}{3\sqrt{K}}$ is good for sparse signal recovery.

$$\Pr\left\{\delta_{K+1} < \frac{1}{3\sqrt{K}}\right\} \geq 1 - e^{-N\,E_{K+1}}$$



- As $\lim\limits_{N\to\infty} \Pr\left\{\delta_{K+1} < \frac{1}{3\sqrt{K}}\right\} \to 1 \quad \theta \to \theta_{\text{th}}^{\infty}$

- Thus, $M > \theta_{\text{th}}^{\infty} N$ are sufficient for a Gaussian matrix to recovery a sparse signal

Summary

- We have derived new eigenvalue distributions of Wishart matri ces

- Our distributions are exact, compact and are useful for the eig envalue analysis of small and large systems

- We have related the RIC of a sensing matrix to its eigenvalues for the purpose of undersampling analysis

- We have shown that for every RIC condition there exists a thre shold above which finding a Gaussian matrix is easy.

References

- [Murihead 1982] R. J. Muirhead, Aspects of Multivariate Statistical Theory, John Wiley & Sons Inc., 1982.
- [Verdu 2004] Antonia Maria Tulino and Sergio Verdu, Random Matrix Theory and Wireless Communications. Now Publishers Inc, 2004.
- [Alan 1964] Alan T. James, “Distributions of matrix variates and latent roots derived from normal sample,” Annals of Mathematical Statistics, vol. 35, no. 2, pp. 475-501, 1964.
- [Khatri 1969] C. G. Khatri, “Non-central distribution of i-th largest characteristic roots of three matrices concerning complex multivariate multivariate normal populations,” Journal of Institute of Ann. Statistical Math., vol. 21, pp. 23-32, 1969.
- [Krishnaiah 1971] P. R. Krishnaiah and T. C. Chang , “On the exact distribution of the smallest root of the Wishart matrix using zonal polynomials,” Ann. Math. Stat., vol. 23, pp. 293-295, 1971.
- [Alouini 2004] Ming Kang and Mohamed-Slim Alouini, “A comparative study on the performance of MIMO MRC systems with and without cochannel interference,” IEEE Trans. on Comm., vol. 52, no. 8, pp 1417-1425, 2004.
- [Sugiyama 66] Sugiyama, T. and Fukutomi, K. “On the distribution of the extreme characteristic roots of matrices in multivariate analysis,” Repts. Stat. Appl. Res., Union of Japanese Scientists and Engineers, vol. 13, 1966.
- [Edleman 1989] Alan Edleman, “Eigenvalues and conditional number of random matrices,” Ph.D Thesis, MIT, 1989.
- [Chiani 2003]Marco Chiani, Moe Z. Win, and Alberto Zanella, “On the capacity of spatially correlated MIMO Rayleigh-fading channel,” IEEE Trans. on Info. Theory, vol. 49, no. 10, pp. 2363-2371, 2003.
- [Chinai 2009]Alberto Zanella, Marco Chiani, and Moe Z. Win, “On the marginal distribution of the eigenvalues of Wishart matrices,” IEEE Trans. on Comm., vol. 57, no. 4, pp. 1050-1060, 2009.

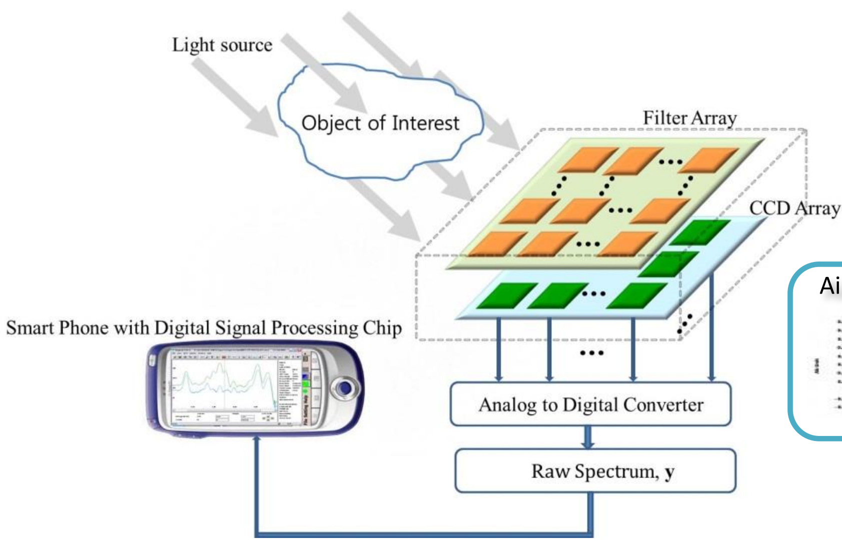
THIRD ONE

Compressive Spectrometers for Super Resolution

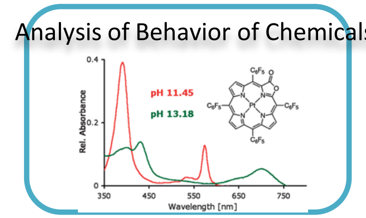
- Spectrometer: Used to find the spectrum of an optical signal

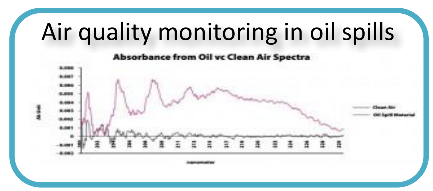
- It takes in the light, breaks it into its spectral components, and displays them in a portable device such as smart phones.

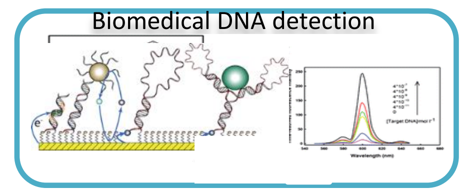
State-of-the art portable spectrometers



Applications







- The ability of the spectrometer in revealing fine information is determined by its "Resolution"
- Problem: Resolution is limited by the number of filters

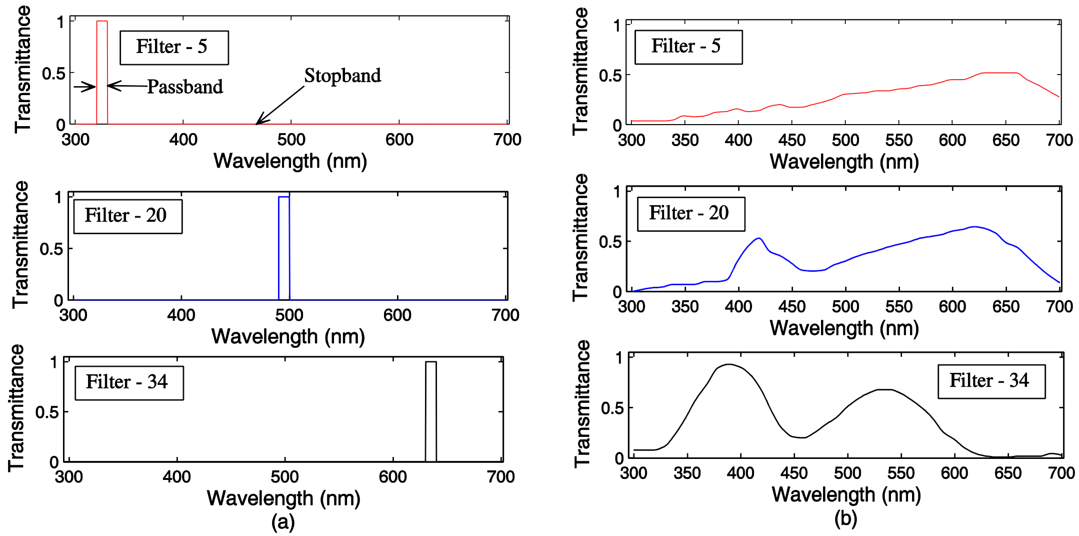
Compressive spectrometers Contd.

- How to improve the resolution for a fixed set of filters in a spectrometer?
- Solution: Compressive Spectrometers!!!
- Innovations
  - Spectrum acquisition using random filters design (using thin-films) [Lee13s]
    - Analog domain acquisition (Our design is first of a kind)
    - A set of M filers with good auto and cross covariance
  - Spectrum reconstruction using a new L_1 algorithm [Lee12s]
- We model the spectrometer output as underdetermined linear system $\mathbf{y}=D\mathbf{x}$
- The matrix D is transmittance matrix is determined from the filter manufact uring process
  - MEMS, non-ideal filters (conventional)
  - Thin-film, random filters (our method)

Various Transmittance Functions

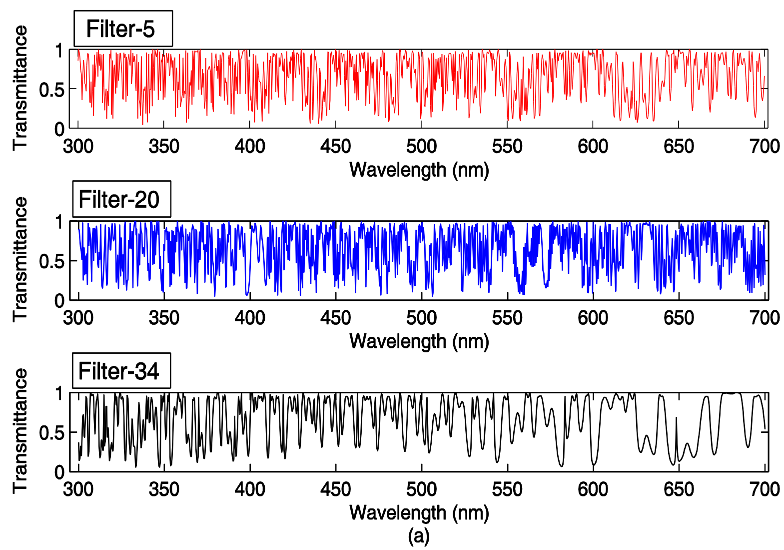
Conventional Design

1. Stringent filter design
2. Local sampling



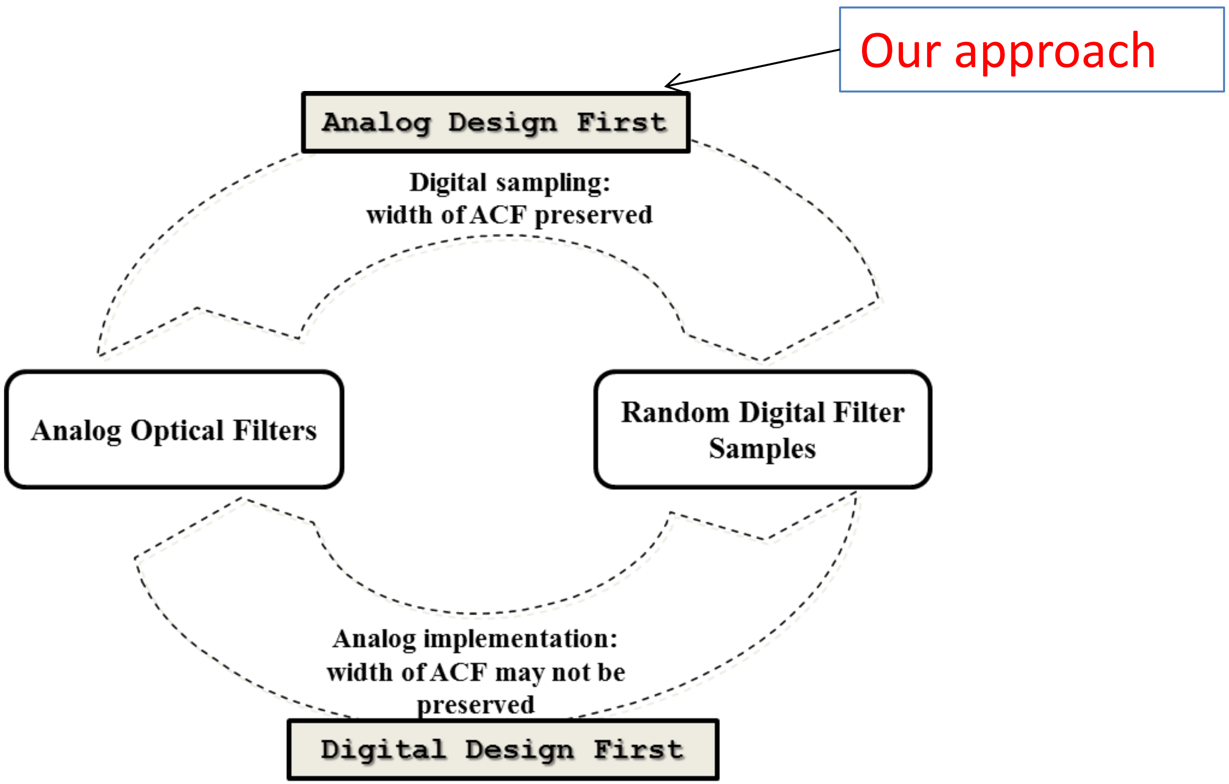
Our approach

1. Ease of filter design
2. Holistic sampling



Digital Vs. Analog-design-first Approach

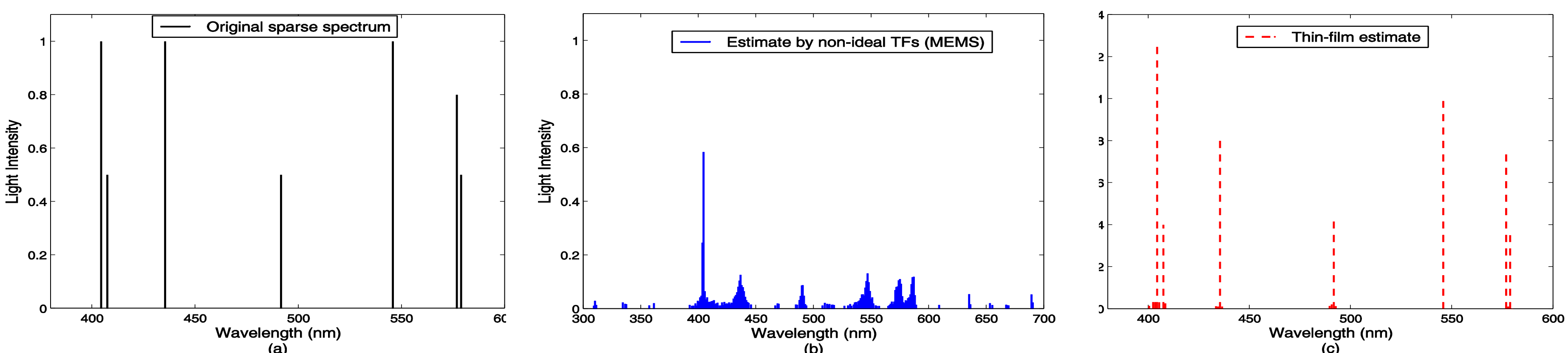
- How to design random transmittances, analog or digital? Analog is the answer!



- Digital filter design first approach
  - May not preserve the auto-covariance function (ACF)
  - May result in a random structure that cannot be implemented via analog designs

Mercury lamp spectral lines estimation

- The least separation among the 7 mercury spectral lines (Fig. (a)) is 2.106 nm (which is between the wavelengths 576.959 nm and 579.065 nm).
- Thin-film filter based spectrometer resolves the least separated spectral lines (Fig. c).
- Where as the MEMS based non-ideal filters cannot resolve even the dominant spectral lines (Fig. (b)).
- Resolution limit = 10nm, whereas compressive spectrometer achieves 0.99 nm, 10 times better!!!

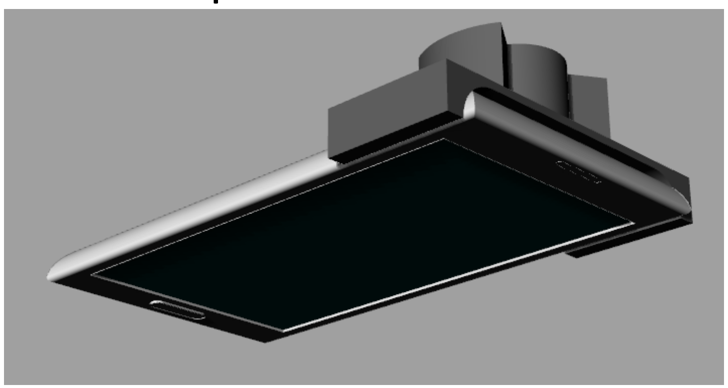


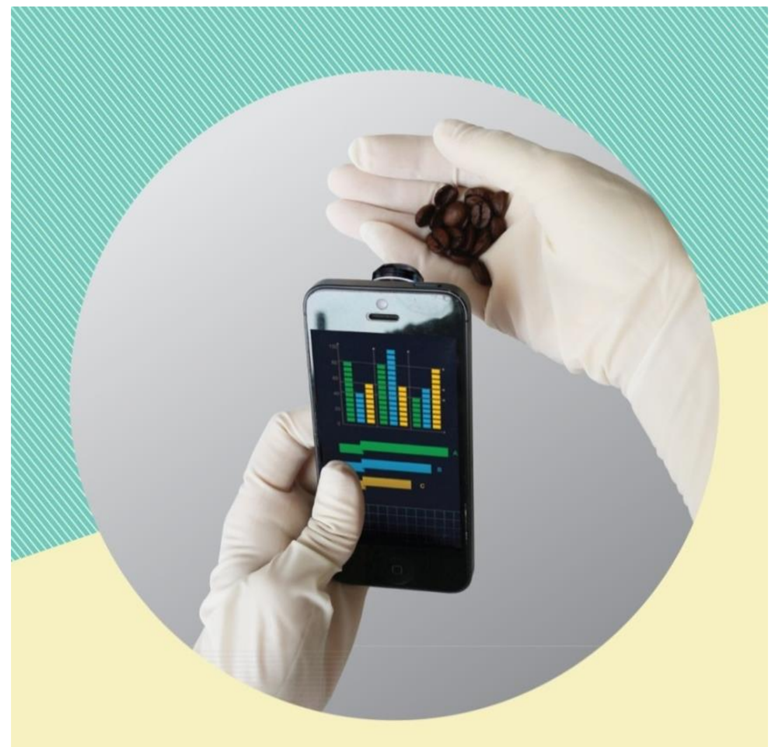
Summary

- 신호처리로 hardware의 단점을 보완하는 시스템 개발 Needs가 증가 함.
- 좋은 신호 처리 알고리즘을 사용하여 센서 및 센서 시스템을 간단하면서도 정확하게 작동할 수 있게 할 수 있음.
- Lensfree camera, single pixel camera등의 application 등 이 있음.

Future works

- Implementing random filters with thin-film technology varying thickness and reflective indices
- Ultimate Goal: Smartphone attachable high resolution spectrometers and microscopes

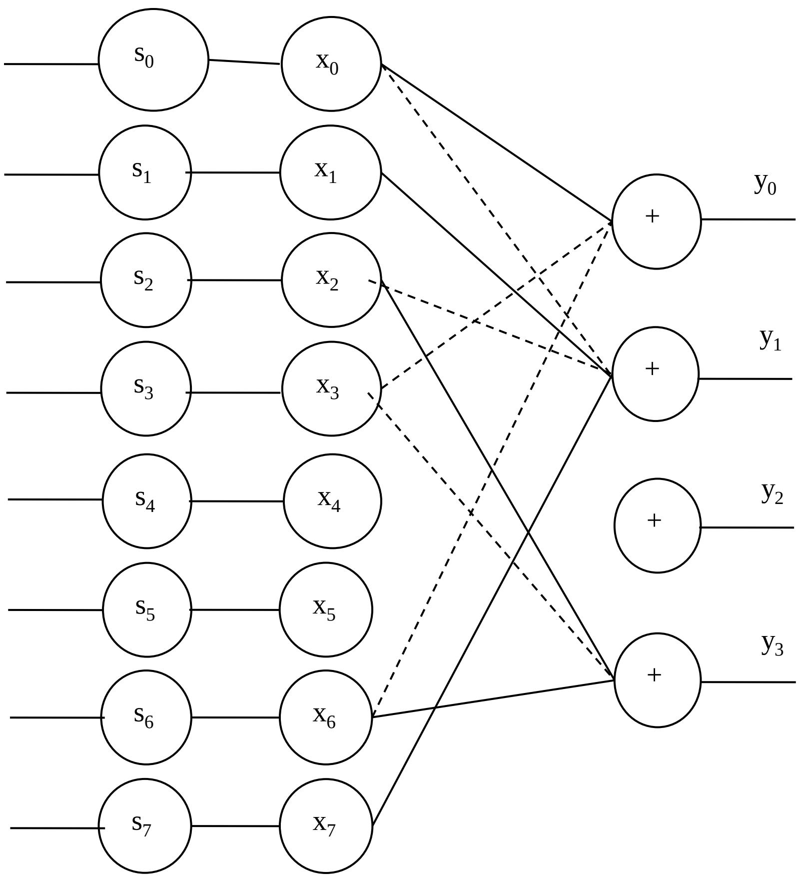




FOURTH ONE

J. KANG, HEUNG-NO LEE, K. KIM, “PHASE TRANSITION ANALYSIS OF SPARSE SUPPORT DETECTION FROM NOISY MEASUREMENTS, HTTP://ARXIV.ORG/ABS/1303.6388/.

Message Passing: State, Value, Matrix, Observation



F. The Message Passing Algorithm

The message passing algorithm is given as the following:

1. Initialization: Set $P(x_t = \tau_t \mid \mathbf{y}) = \frac{K}{N} f_1(x_t = \tau_t) + \left(1 - \frac{K}{N}\right) f_0(x_t = \tau_t)$ 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.

Signal Detection Algorithms

- The sparse signal can be reconstructed from the following criterion

$$\tilde{x}_t := \arg\max_{\tau_t \in GF(q)} P\left(x_t = \tau \middle| y, C\right) \qquad \text{s.t.} \quad y = Fx$$

Theorem 2 : The *aposteriori* probability (AP) that the first value, $x_0 = \tau_0 \in GF(q)$

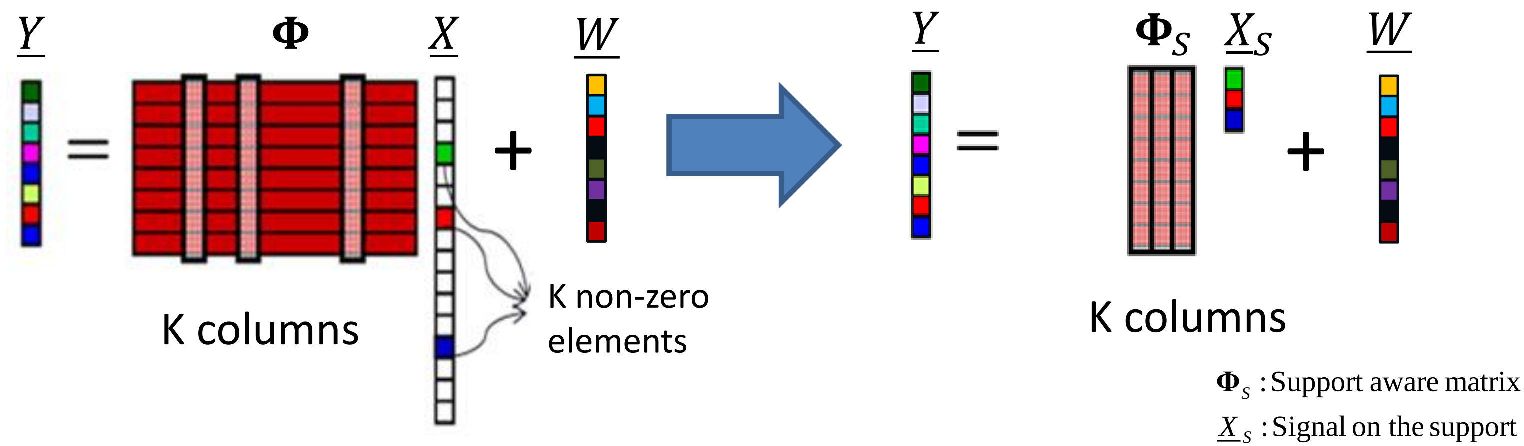
Given the observation y and enforcing the checks (checks should be satisfied),

is given by

$$P\left(x_t = \tau \middle| y, C\right) = \frac{P\left(x_t = \tau \middle| y\right)}{P\left(C \middle| y\right)} \prod_{p=1}^{d_c} \left[\sum_{x_{t,p}} P\left(C_{ip} \middle| x_t = \tau_0, x_{t,p}, y\right) P\left(x_{0,0} \middle| y\right) \right]$$

Sparse Support Detection first!

- Once the sparse support is known, the uncertainty to the recovery of $\underline{X} \in \mathbb{R}^N$ is confined to the additive noise $\underline{W} \in \mathbb{R}^M$.



- Nevertheless, most recovery algorithms, such as LASSO [Tibshirani'96], OMP [Tropp'06] and CS-BP [Baron'10], to date for the problem have been developed under auspices of signal estimation rather than support detection.

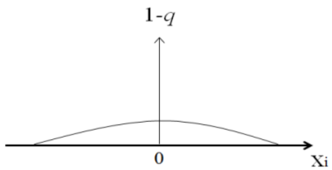
Difficulties and Breakthroughs

- In LDPC decoding, only binary messages are usually decoded.
- In compressed sensing, non-binary cases as well as real-valued and complex valued codes are used.
- Our approaches
  - Non-binary cases, i.e., GF(q): can do up to block lengths of thousands.
  - Real- or complex-valued:
    - Use quantization
    - Do the belief propagation and obtain the posteriors
    - Find the support set from the posterior
    - Form the over determined set of equations and find the signal values
- Breakthrough: Approximate Message Passing algorithm by Donoho

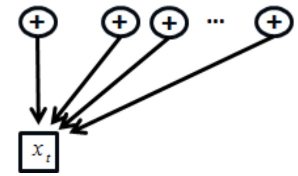
Marginal posteriors, spike-and-slab again!

- From Th. 1

Spike-and-slab prior is assumed
$P(x_t = \tau | y)$
$= qN(\tau; 0, \sigma_X^2) + (1-q)\delta(\tau))$



Each message from check converges to Gaussian PDF $N(\tau; x_{0,t}, \sigma_W^2)$ **by CLT**



$$P\left(x_t = \tau \middle| y, C\right) \propto P\left(x_t = \tau \middle| y\right) \prod_{p=1}^{d_c} \left[\sum_{x_{t,p}} P\left(C_{ip} \middle| x_t = \tau_0, x_{t,p}, y\right) P\left(x_{0,0} \middle| y\right) \right]$$

$$\propto qc_1 \mathcal{N}\left(\tau; \frac{Lx_{0,t}\sigma_X^2}{L\sigma_X^2 + \sigma_W^2}, \frac{\sigma_X^2 \sigma_W^2}{L\sigma_X^2 + \sigma_W^2}\right) + (1-q)c_2\delta(x)$$

The product of Gaussian PDFs is a scaled Gaussian PDF

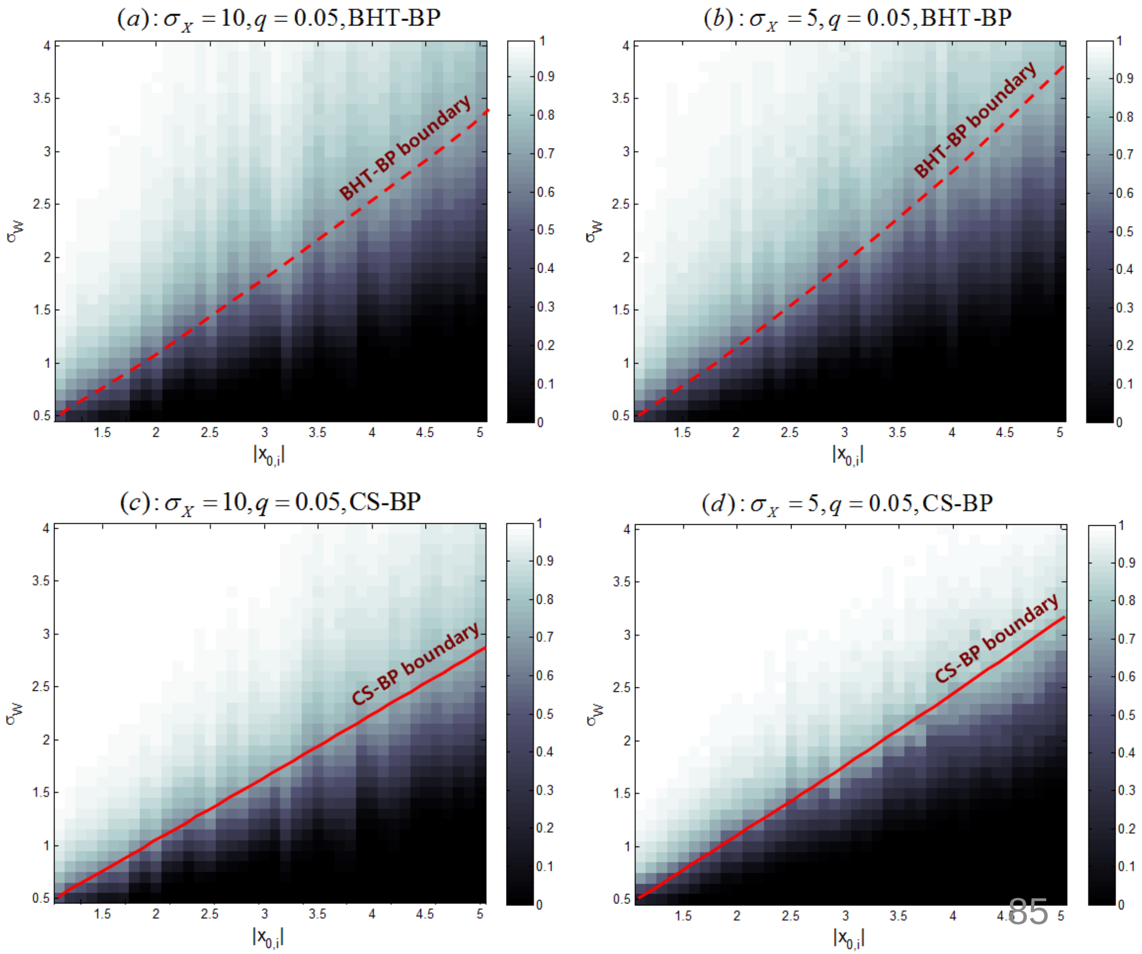
A slab density **A spike**

→ Marginal posterior has a shape of spike-and-slab PDF !

Sharp Comparison with PT Analysis

New observation: State detection failures occur only at non-zero states ($S_i = 1$)

- It led to a sharp transition analysis between estimation based approach and the detection based approach.
- Namely, it shows how much benefit there is with the detection based approach (BHT-BP), compared to the estimation based approach (CS-BP).
- The PT-diagram provides an exact border line between success and failure of the algorithm on the plane of the noise level and the signal magnitude



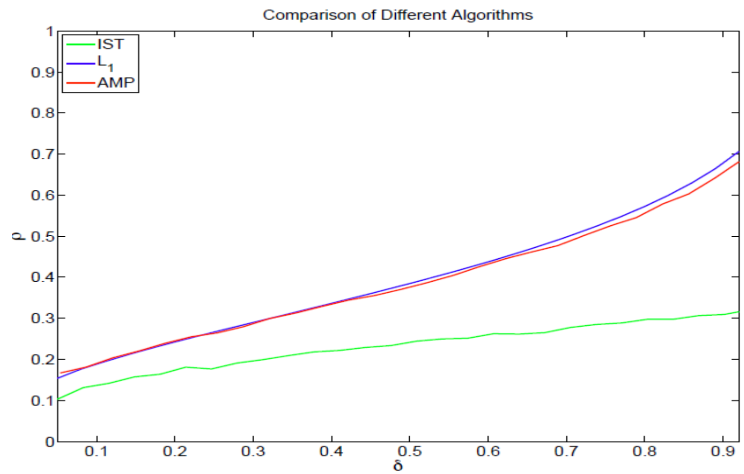
Approximate-Message Passing (AMP)

- Donoho developed a remarkable low-computational solver to compressed sensing recovery, called AMP [Donoho'09] [Donoho'10-1] [Donoho'10-2].

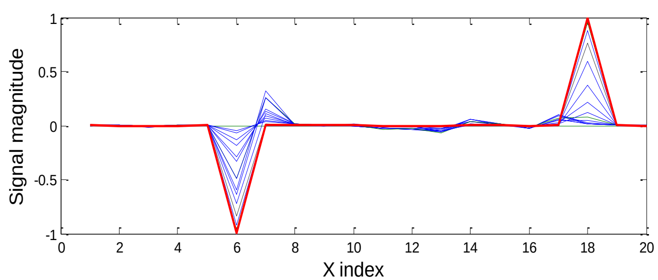
AMP algorithm
Init.: set $\underline{x}^{t=0} = \underline{0}, \underline{z}^{t=0} = \underline{y}, \hat{\tau}^{t=0}$
Threshold update: $\hat{\tau}^t = \dfrac{\hat{\tau}^{t-1}}{N\delta}\sum_{i=1}^{N}\eta'\left((\mathbf{A}^T\underline{z}^{t-1})_i;\hat{\tau}^{t-1}\right)$
Signal update : $\underline{x}^t = \eta\left(\mathbf{A}^T\underline{z}^{t-1} + \underline{x}^{t-1};\hat{\tau}^{t-1}\right)$
Residual update : $\underline{z}^t = \underline{y} - \mathbf{A}\underline{x}^t + \dfrac{1}{\delta}\underline{z}^{t-1}\left\langle \eta'\left((\mathbf{A}^T\underline{z}^{t-1}) + \underline{x}^{t-1};\hat{\tau}^t\right)\right\rangle$

- AMP has the following properties

1) Working with dense measurement matrices, e.g., standard Gaussian matrices
2) Achieving the equivalent phase transition performance to Lasso [Tibshirani'96] as $N \to \infty$ under noise/noiseless cases
3) $O(M+N)$ computation per iteration for the recovery (cf) Lasso has $O(MN^2)$ computation)



Experimental phase transition diagram comparing *Iterative soft thresholding (IST), L1-solver (L1), AMP* [Donoho'10-2]



An example of AMP recovery when N=20, M=10, K=2 20 iterations.

Construction of AMP [Donoho'10-1]

- Step I: Construct a joint distribution over the signal $\mathbf{s} = [s_1, \ldots, s_N]$:

$$p(\mathbf{s}, \mathbf{y}) = \frac{1}{Z} \prod_{i=1}^{N} \exp(-\beta \,|\, s_i \,|) \times \prod_{a=1}^{M} \exp\left(-\frac{\beta}{2} (y_a - (\mathbf{A}\mathbf{s})_a)^2 \right)$$

A joint distribution

$$v_{i\to a}^{t+1}(s_i) \cong e^{-\beta|s_i|} \prod_{b\neq a} c_{b\to i}^{t}(s_i)$$

$$c_{a\to i}^{t}(s_i) \cong \int \exp\left(-\frac{\beta}{2} (y_a - (\mathbf{A}\mathbf{s})_a)^2 \right) \cdot \prod_{j\neq i} v_{i\to a}^{t}(s_i)(ds_j)$$

Classical BP update rule

- Step II: For large system limit (N, M→∞), the classical BP message $c_{a\to i}^{t}(s_i)$ can be approximated to a Gaussian PDF with mean $z_{a\to i}^{t}$ and the variance $\hat{\tau}_{a\to i}^{t}$.

$$z_{a\to i}^{t} \triangleq y_a - \sum_{j\neq i} A_{aj} x_{j\to a}^{t}, \quad \hat{\tau}_{a\to i}^{t} \triangleq \sum_{j\neq i} A_{aj}^2 \tau_{j\to a}^{t}$$

Mean and variance of $c_{a\to i}^{t}(s_i)$

  – Then, the message $v_{\mathrm{i}\to a}^{t+1}(s_i)$ is approximated by the product of a Gaussian and a Laplace PDF, given as

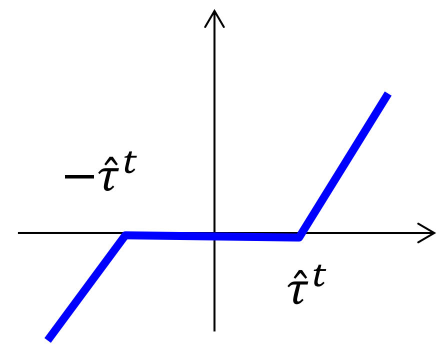
$$v_{i\to a}^{t+1}(s_i) \cong \frac{1}{c_\beta} \exp\left(-\beta \,|\, s_i \,| - \frac{\beta}{2\hat{\tau}^{t}} (s - \sum_{b\neq a} A_{bi} z_{b\to i}^{t})^2 \right)$$

Set $\mathbf{E}_{v_{i\to a}^{t+1}}\left[s_i^{t+1} \right] = x_{i\to a}^{t+1}, \ \mathbf{Var}_{v_{i\to a}^{t+1}}\left[s_i^{t+1} \right] = \tau_{i\to a}^{t+1}$ **Mean and variance of $v_{i\to a}^{t+1}(s_i)$**

Construction of AMP [Donoho'10-1]

- Step III: Take the limit $\beta \to \infty$ and get a thresholding function $\eta(\,)$ for calculation of mean $x_{i\to a}^{t+1} = \mathbf{E}_{v_{i\to a}^{t+1}}[s_i]$.

$$\mathbf{E}_{v_{i\to a}^{t+1}}(s_i^{t+1}) \overset{as\ \beta\to\infty}{\approx} \eta\Big(\sum_{b\neq a} A_{bi} z_{b\to i}^t ; \hat{\tau}^t\Big) = \begin{cases} \sum_{b\neq a} A_{bi} z_{b\to i}^t - \hat{\tau}^t & \text{if } \sum_{b\neq a} A_{bi} z_{b\to i}^t \geq \hat{\tau}^t \\ \sum_{b\neq a} A_{bi} z_{b\to i}^t + \hat{\tau}^t, & \text{if } \sum_{b\neq a} A_{bi} z_{b\to i}^t \leq -\hat{\tau}^t \\ 0, & \text{otherwise} \end{cases}$$



Soft thresholding

- Step IV: With large N (M/N fixed) and the1st order-Taylor series approx of $\eta(\,)$, AMP is obtained.

The mean-passing rule

$$x_{i\to a}^t = \eta\left(\sum_{b\neq a} A_{bi} z_{b\to i}^{t-1} ; \hat{\tau}^{t-1}\right)$$

$$z_{a\to i}^t = y_a - \sum_{j\neq i} A_{aj} x_{j\to a}^t$$

$$\hat{\tau}^t = \frac{\hat{\tau}^{t-1}}{N\delta} \sum_{i=1}^{N} \eta'\left((\mathbf{A}^T \mathbf{z}^{t-1})_i ; \hat{\tau}^{t-1}\right)$$

The AMP rule

$$\underline{x}^t = \eta\left(\mathbf{A}^T \underline{z}^{t-1} + \underline{x}^{t-1} ; \hat{\tau}^{t-1}\right)$$

$$\underline{z}^t = \underline{y} - \mathbf{A}\underline{x}^t + \frac{1}{\delta}\underline{z}^{t-1}\left\langle \eta'\left((\mathbf{A}^T \underline{z}^{t-1}) + \underline{x}^{t-1} ; \hat{\tau}^t\right)\right\rangle$$

$$\hat{\tau}^t = \frac{\hat{\tau}^{t-1}}{N\delta} \sum_{i=1}^{N} \eta'\left((\mathbf{A}^T \underline{z}^{t-1})_i ; \hat{\tau}^{t-1}\right)$$

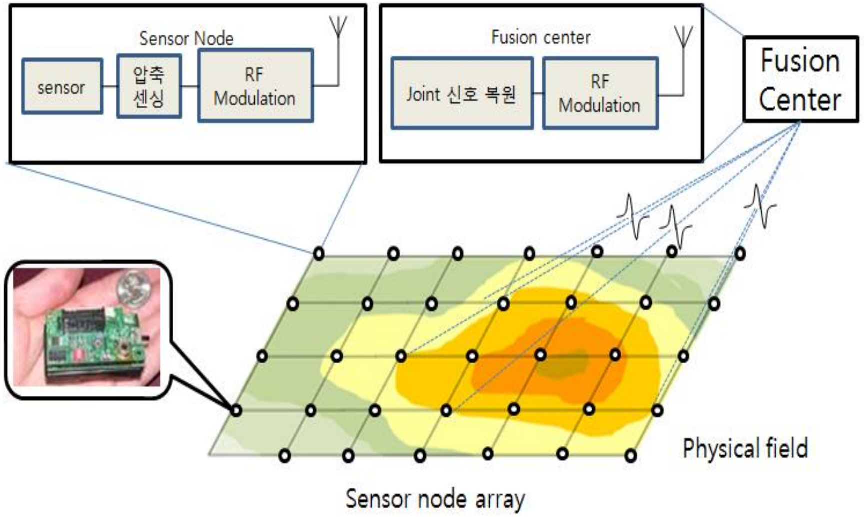
Summary of the fourth one

- Sparse support detection is most crucial for signal recovery in CS.
- We introduce a new detection-oriented belief propagation algorithm for CS.
  - Show how much the proposed method is superior to conventional estimation based approach.
- AMP is considered a breakthrough, taking only O(M+N) computation, and working for real- and complex-valued CS parity check problems.
  - Possible for application in ADCs, network codes, distributed compression, radars, tomography, medical imaging, microscopes, …

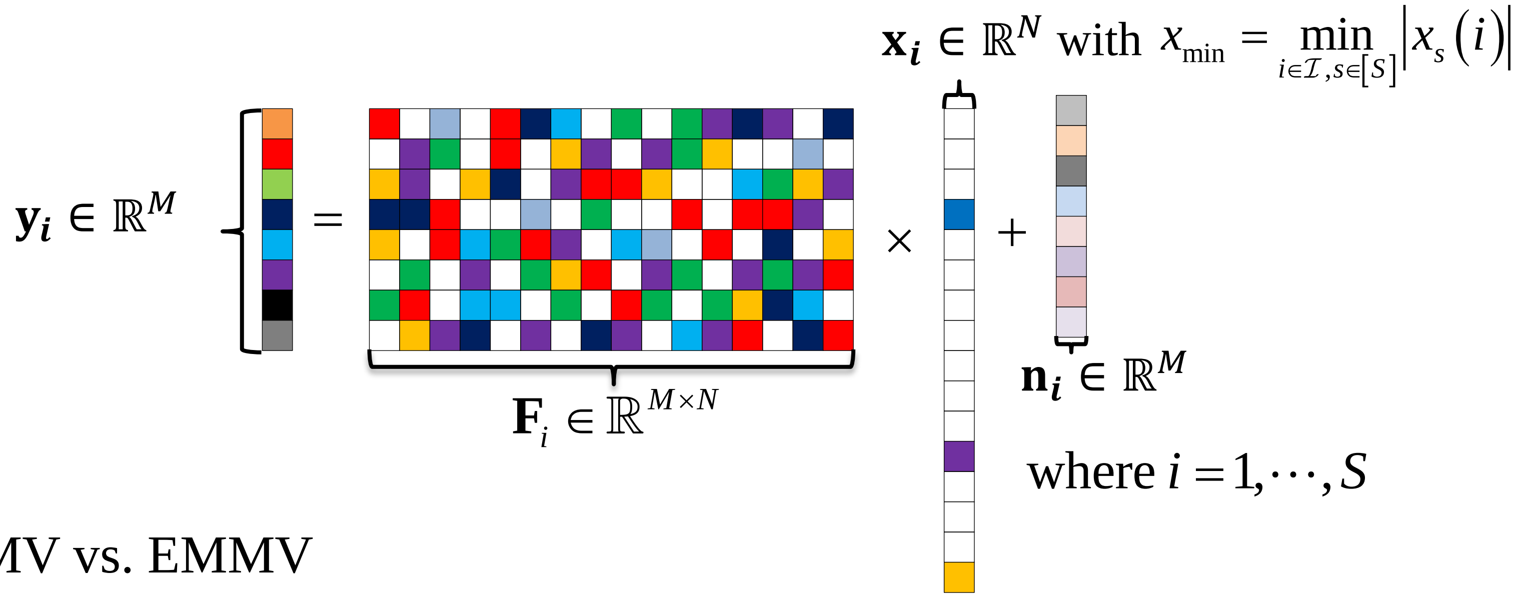
FIFTH ONE

Multiple Sensor Problems

- One more application on multiple sensor systems
  - Correlated measurement: $H(X_1, X_2) < H(X_1)+H(X_2)$.
  - Slepian Wolf, Wyner-Ziv coding ~ *distributed* source coding and joint decoding.
  - We aim at using sparse representation and achieve *distributed* source coding.



Multiple Measurement Vectors



- MMV vs. EMMV
  - The same support set for all sensors -- A new way to represent correlation.
  - When all measurement matrices are the same, the EMMV model is the MMV model.

- Our goal
  - Is to jointly reconstruct the support set of each sparse vector, with the knowledge of the sensing matrices
  - Determine how many measurements M are needed as S scales up?

A Joint Typical (JT) decoder[Lee12] [Lee13] - 1

- A decoder used in here is called joint typical (JT) decoder which explores all possible subsets to find the correct support set.

- It yields a set as decoded output, when all measurement vectors and all measurement matrices are known. Namely,

$$\text{JT decoder}: \left\{\forall_s : \left(\mathbf{y}_s, \mathbf{F}_s\right)\right\} \mapsto \mathcal{J},$$
$$\text{where } \mathcal{J} \subset \{1, \cdots, N\} \text{ and } \left|\mathcal{J}\right| = K.$$

A Joint Typical (JT) decoder[Lee12] [Lee13] - 2

- If a following condition is satisfied for a set $\mathcal{J}$, then this set becomes the output of the JT decoder

$$\left| \left(\sum_{s=1}^{S} \frac{\left\| \mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right)\mathbf{y}_s \right\|^2}{SM} \right) - \frac{\left(M-K\right)\sigma_n^2}{M} \right| < \delta,$$

where $\mathbf{Q}\left(\mathbf{F}\right) := \mathbf{I} - \mathbf{F}\mathbf{F}^{\dagger}$, σ_n^2 is the noise variance, $\delta > 0$ and $\mathbf{F}_{s,\mathcal{J}}$ is constructed by collecting column vectors of $\mathbf{F}_s$ corresponding to indices of $\mathcal{J}$.

The probability of the failure event – 1

- A failure event $\mathrm{E}(D_{failure})$ of the JT decoder is

$$\mathrm{E}\left(D_{failure}\right) = \mathrm{E}\left(\mathbf{Y},\mathcal{I},\delta\right)^{c} \bigcup_{\forall \mathcal{J}\neq\mathcal{I},|\mathcal{J}|=K} \mathrm{E}\left(\mathbf{Y},\mathcal{J},\delta\right),$$

where $\mathrm{E}\left(\mathbf{Y},\mathcal{I},\delta\right)^{c}$ is an event where the JT decoder makes failure such that the correct support set $\mathcal{I}$ is not $\delta-$ jointly typical with $\mathbf{Y}$, $\mathrm{E}\left(\mathbf{Y},\mathcal{J}\neq\mathcal{I},\delta\right)$ is an event where the JT decoder declares that an incorrect support set $\mathcal{J}$ is δ $-$jointly typical with $\mathbf{Y}$.

The probability of the failure event – 2

- Owing to the union bound, we have

$$\mathbb{P}\left\{\mathrm{E}\left(D_{failure}\right)\right\} \le \mathbb{P}\left\{\mathrm{E}\left(\mathbf{Y},\mathcal{I},\delta\right)^{c}\right\} + \sum_{\forall \mathcal{J}\neq\mathcal{I},|\mathcal{J}|=K} \mathbb{P}\left\{\mathrm{E}\left(\mathbf{Y},\mathcal{J},\delta\right)\right\}.$$

- Obtaining the exact probabilities at the right hand side in the equation is non-trivial.
- Thus, we deiced to obtain their upper bounds by using the Chernoff bound.

The probability of the failure event – 3

- Let us define a random variable by $Z_{\mathcal{I}} := \sum_{s=1}^{S} \left\| \mathbf{Q}\left(\mathbf{F}_{s,\mathcal{I}}\right)\mathbf{y}_s \right\|^2 \big/ \sigma_n^2.$

- Then, it is readily seen that this random variable is a quadratic random variable.
- Also, its mean, variance and moment generating functions are

$$\mathbb{E}\left[Z_{\mathcal{I}}\right] = S\left(M-K\right),\ \mathbb{V}\left[Z_{\mathcal{I}}\right] = 2S\left(M-K\right),\ \mathbb{E}\left[\exp\left(tZ_{\mathcal{I}}\right)\right] = \left(1-2t\right)^{-S(M-K)/2}.$$

The probability of the failure event – 4

- Then, the following probability is bounded by

$$\mathbb{P}\left\{\mathrm{E}\left(\mathbf{Y},\mathcal{I},\delta\right)^{c}\right\} = \mathbb{P}\left\{\mathrm{Z}_{\mathcal{I}} \le W_1\right\} + \mathbb{P}\left\{\mathrm{Z}_{\mathcal{I}} \ge W_2\right\}$$

$$\overset{(a)}{\le} \sum_{i=1}^{2} \underbrace{\mathbb{E}\left[\exp\left(t_i Z_{\mathcal{I}}\right)\right]\exp\left(-t_i W_i\right)}_{:= f\left(t_i;W_i\right)} \qquad (a) : \text{the Chernoff bound.}$$

$$= \sum_{i=1}^{2} f\left(t_i;W_i\right)$$

where $W_i = S\left(\left(M-K\right)+\left(-1\right)^i M\delta/\sigma^2\right), i = 1,2.$

- The function $f(t_i;W_i)$ is convex with respect to t_i.
- Thus, the optimal value t_i is obtained by investigating the first derivative of this function.

The probability of the failure event – 5

- Let the optimal values be

$$t_i^* = \left(1 - S\left(M-K\right)/W_i\right)/2, i = 1, 2.$$

These optimal values are obtained from $t_i^* = \arg_{t_i} f^{(1)}\left(t_i; W_i\right) = 0.$

- Then, we finally have

$$\mathbb{P}\left\{\mathrm{E}\left(\mathbf{Y}, \mathcal{I}, \delta\right)^c\right\} \leq \sum_{i=1}^{2} f\left(t_i^*; W_i\right)$$

$$\overset{(a)}{\leq} 2 f\left(t_2^*; W_2\right)$$

$$= 2\left(\frac{(M-K)\sigma_n^2 + M\delta}{(M-K)\sigma_n^2}\right)^{\frac{S(M-K)}{2}} \exp\left(-\frac{SM\delta}{2\sigma_n^2}\right).$$

(a) : due to $f\left(t_1^*; W_1\right) \leq f\left(t_2^*; W_2\right)$.

- Similarly, an upper bound on the probability $\mathbb{P}\left\{\mathrm{E}\left(\mathbf{Y}, \mathcal{J}, \delta\right)\right\}$ for a given set $\mathcal{J}$ is obtained.

The probability of the failure event – 6

- Then, the failure probability is given by

$$
\begin{aligned}
\mathbb{P}\left\{\mathrm{E}\left(D_{failure}\right)\right\} &\le \mathbb{P}\left\{\mathrm{E}\left(\mathbf{Y},\mathcal{I},\delta\right)^c\right\} + \sum_{\mathcal{J}\in\mathcal{S}\setminus\{\mathcal{I}\}} \mathbb{P}\left\{\mathrm{E}\left(\mathbf{Y},\mathcal{J},\delta\right)\right\} \\
&\le 2\left(\frac{(M-K)\sigma_n^2+M\delta}{(M-K)\sigma_n^2}\right)^{\frac{S(M-K)}{2}} \exp\left(-\frac{SM\delta}{2\sigma_n^2}\right) \\
&\quad + \binom{N}{K}\left(\frac{(M-K)\sigma_n^2+M\delta}{(M-K)\left(x_{\min}^2+\sigma_n^2\right)}\right)^{\frac{S(M-K)}{2}} \exp\left(-\frac{S\left(M\delta-(M-K)x_{\min}^2\right)}{2\left(x_{\min}^2+\sigma_n^2\right)}\right) \\
&:= p\left(S,N,M,K,\sigma_n^2,x_{\min}^2,\delta\right)
\end{aligned}
$$

- Detailed explanations are given in [Lee12][Lee13].

A sufficient condition for infinite *S* [Lee13]

- Theorem 1: $M = \Omega\left(\frac{K}{S}\log\left(N/K\right)+K\right)$ is sufficient to jointly reconstruct the support set as N goes to infinity if $x_{\min}^2 > M\delta/\left(M-K\right)$.

- The above theorem suggests an inversion relation between M and S.
- In [Hyder09], the authors empirically show that M is decreased as S is increased.
- Thus, the theorem verifies this empirical result.
- Note that similar relations are given in [Nehorai09][Rao13]. But, these relations are made under more general conditions.

A sufficient condition for a finite *S* [Lee13]

- Theorem 2: $M > K$ is sufficient to jointly reconstruct the support set as S goes to infinity if $x_{\min}^2 > M\delta/(M-K)$.

- The above theorem suggests that K + 1 measurements suffices to jointly reconstruct the support set when S goes to infinity.
- In [Duarte13], the authors show that K + 1 measurements are sufficient to jointly reconstruct the support set as well. But, these authors do not consider the presence of noises.
- It suggests that taking more measurements vectors reduces effects of noises.
- A similar conclusion is given in [Rao13].

Summary of the fifth one

- We have verified the empirical simulation results as reported in [Hyder09]
  - M is decreased as S is increased
  - $K + 1$ measurements suffices to jointly reconstruct the support set when S is suf ficiently large.

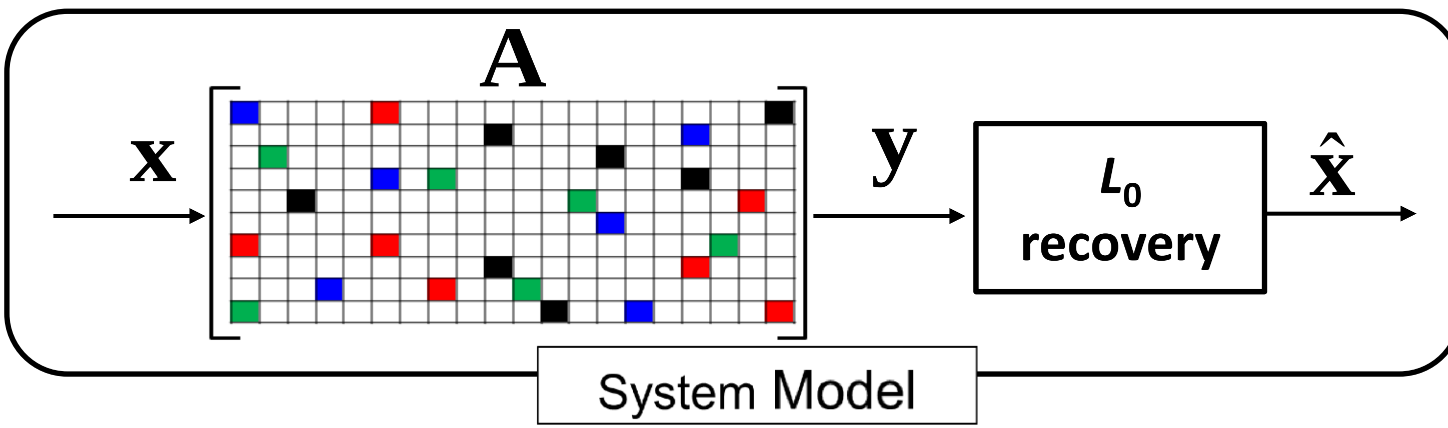
- We have shown that taking more measurements vectors reduces effects of n oises.
  - The sufficient condition given in Theorem 2 is accordance with that in [Duarte13].
  - A similar result is given in [Rao13]

SIXTH ONE

Recovery Bounds on Sparse Signals over Finite Fields

- **Research Goal**
  - Theoretical performance limits of compressive sensing problems over finite fields

- **Approaches**
  - Upper and lower bounds on recovery performance using L_0 minimization
  - Impact of sparseness of sensing matrices on recovery performance



  - where $\mathbf{X}$ is sparse, $\mathbf{A}$ is a sensing matrix, $\mathbf{y}$ is a measurement signal, $\hat{\mathbf{x}}$ is an estimated signal.

Compressed Sensing over Finite Fields

- The measured signal $\mathbf{y}$ is given by

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

- The elements of the sensing matrix are i.i.d.,

$$\Pr\{A_{ij} = \alpha\} = \begin{cases} 1-\gamma & \alpha = 0, \\ \gamma/(q-1) & \alpha \neq 0, \end{cases}$$

- The sparse signals $\mathbf{x}$ is randomly and uniformly selected form the set $\mathcal{L}$
  - $\mathcal{L} := \bigcup_{k_1=1}^{K} \mathcal{L}_{k_1}$, where $\mathcal{L}_{k_1}$ denotes the set of signals $\mathbf{x}$ of length N with sparsity k_1.

Probability of Error for L_0 Minimization

- L0 Minimization decision,

$$(L_0) \quad \hat{\mathbf{x}} = \min \|\bar{\mathbf{x}}\|_0 \quad \text{subject to} \quad \mathbf{A}\bar{\mathbf{x}} = \mathbf{y},$$

- The error event is for $\mathbf{x} \neq \hat{\mathbf{x}}$

$$\mathcal{E}_0(\mathbf{x}, \hat{\mathbf{x}}) := \{\mathbf{A} : \ \mathbf{A}\mathbf{x} = \mathbf{A}\hat{\mathbf{x}}\}$$

- This error event is upper bounded as follows

$$\mathcal{E}_0(\mathbf{x}, \hat{\mathbf{x}}) \subseteq \bigcup_{\bar{\mathbf{x}} \in \mathcal{L}:\ \mathbf{A}\bar{\mathbf{x}} = \mathbf{y}} \mathcal{E}_0(\mathbf{x}, \bar{\mathbf{x}})$$

- Probability of error averaged on all sparse signals **x,**

$$\mathcal{E}_0 := \bigcup_{\mathbf{x} \in \mathcal{L}} \mathcal{E}_0(\mathbf{x}, \hat{\mathbf{x}}) \qquad \mathcal{E} := \bigcup_{\mathbf{x} \in \mathcal{L}} \bigcup_{\bar{\mathbf{x}} \in \mathcal{L}:\ \mathbf{A}\bar{\mathbf{x}} = \mathbf{y}} \mathcal{E}_0(\mathbf{x}, \bar{\mathbf{x}})$$

- Thus, due to $\mathcal{E}_0 \subseteq \mathcal{E}$, $\Pr\{\mathcal{E}_0\} \leq \Pr\{\mathcal{E}\}$

Upper Bounds on Probability of Error

$$\Pr\{\mathcal{E}_0\} \le \Pr\left\{\bigcup_{\mathbf{x}\in\mathcal{L}}\ \bigcup_{\bar{\mathbf{x}}\in\mathcal{L}:\ \mathbf{A}\bar{\mathbf{x}}=\mathbf{y}} \mathcal{E}_0(\mathbf{x},\bar{\mathbf{x}})\right\}$$

$$\overset{(a)}{\le} \frac{1}{|\mathcal{L}|}\sum_{\mathbf{x}\in\mathcal{L}}\sum_{\substack{\bar{\mathbf{x}}\in\mathcal{L}\\ \mathbf{x}\ne\bar{\mathbf{x}}}} \Pr\left\{\mathbf{A}:\mathbf{A}\mathbf{x}=\mathbf{A}\bar{\mathbf{x}}\middle|\mathbf{x}\right\}$$

$$\overset{(b)}{=} \frac{1}{|\mathcal{L}|}\sum_{\mathbf{x}\in\mathcal{L}}\sum_{h=1}^{2K}\sum_{\bar{\mathbf{x}}\in\bar{\mathcal{L}}_h} \Pr\left\{\mathbf{A}:\mathbf{A}\mathbf{x}=\mathbf{A}\bar{\mathbf{x}}\middle|\mathbf{x}\right\}$$

$$\overset{(c)}{=} \frac{1}{|\mathcal{L}|}\sum_{\mathbf{x}\in\mathcal{L}}\sum_{h=1}^{2K}\left(\left|\bar{\mathcal{L}}_h\right|\Pr\left\{\mathbf{A}\mathbf{d}_h=0\right\}\right)$$

$$= \frac{1}{|\mathcal{L}|}\sum_{k_1=1}^{K}\sum_{\mathbf{x}\in\mathcal{L}_{k_1}}\sum_{h=1}^{2K}\left(\left|\bar{\mathcal{L}}_h\right|\Pr\left\{\mathbf{A}\mathbf{d}_h=0\right\}\right)$$

$$\boxed{\overset{(d)}{=} \frac{1}{|\mathcal{L}|}\sum_{h=1}^{2K} N_h \Pr\left\{\mathbf{A}\mathbf{d}_h=0\right\},}$$

- (a): Union bound
- (b): Partition of the set
- (c): The probability is identically the same with each other,

$$\Pr\left\{\mathbf{A}:\mathbf{A}\mathbf{x}=\mathbf{A}\bar{\mathbf{x}}\middle|\mathbf{x}\right\} = \Pr\left\{\mathbf{A}\mathbf{d}_h=0\right\}$$

- (d): Collection of N_h difference vectors with the same Hamming weight

Upper Bounds on Probability of Error

Theorem 1 (Upper Bound) : For any sensing matrix with i.i.d. elements, an upper bound of probability for L_0 recovery is given by

$$\Pr\{\mathcal{E}_0\} \le \frac{1}{|\mathcal{L}|} \sum_{h=1}^{2K} \sum_{k_1=1}^{K} \binom{N}{k_1} (q-1)^{k_1} \sum_{k_2=1}^{k_1} \sum_{t=0}^{k_2} N_{h,k_1,k_2,t} P_h^M .$$

- where $N_{h,k_1,k_2,t} = \binom{N-k_1}{t} (q-1)^t \binom{k_1}{k_2-t} \binom{k_2-t}{h-2t-k_1+k_2} (q-2)^{h-2t-k_1+k_2}$,

$$P_h = q^{-1} + \left(1-q^{-1}\right)\left(1-\frac{\gamma}{1-q^{-1}}\right)^h .$$

Upper Bounds on Probability of Error

- When the sensing matrix is uniformly random, the upper bound is

$$\begin{aligned}
\Pr\{\mathcal{E}\} &\le \frac{1}{|\mathcal{L}|}\sum_{h=1}^{2K} N_h q^{-M} \\
&\overset{(a)}{=} \left(|\mathcal{L}|-1\right) q^{-M} \\
&< Kq^{-M}\binom{N}{K}(q-1)^K \\
&\le 2^{\log_2 K - M\log_2 q + NH_b(K/N) + K\log_2(q-1)},
\end{aligned}$$

Corollary 2 (Sufficient condition on *M*) : If the following holds,

$$M \ge \frac{\log_2 K + NH_b\left(K/N\right) + K\log_2\left(q-1\right)}{\log_2 q}$$

then $\Pr\{\mathcal{E}_0\} \to 0$ as $N \to \infty$

Lower Bounds on Probability of Error

- Using the Fano's inequality, the probability of error is lower bounded

$$\Pr\{\mathcal{E}_0\} \geq \frac{H(\mathbf{x}|\mathbf{y},\mathbf{A})-1}{\log_q|\mathcal{L}|} = \frac{H(\mathbf{x})-H(\mathbf{y}|\mathbf{A})-1}{\log_q|\mathcal{L}|},$$

  – using $H(\mathbf{y}|\mathbf{A}) \leq H(\mathbf{y}) \leq MH(y_1) \leq M\log_q q = M$

- The lower bounds is

$$\Pr\{\mathcal{E}_0\} \geq 1 - \frac{M+1}{\log_q|\mathcal{L}|}.$$

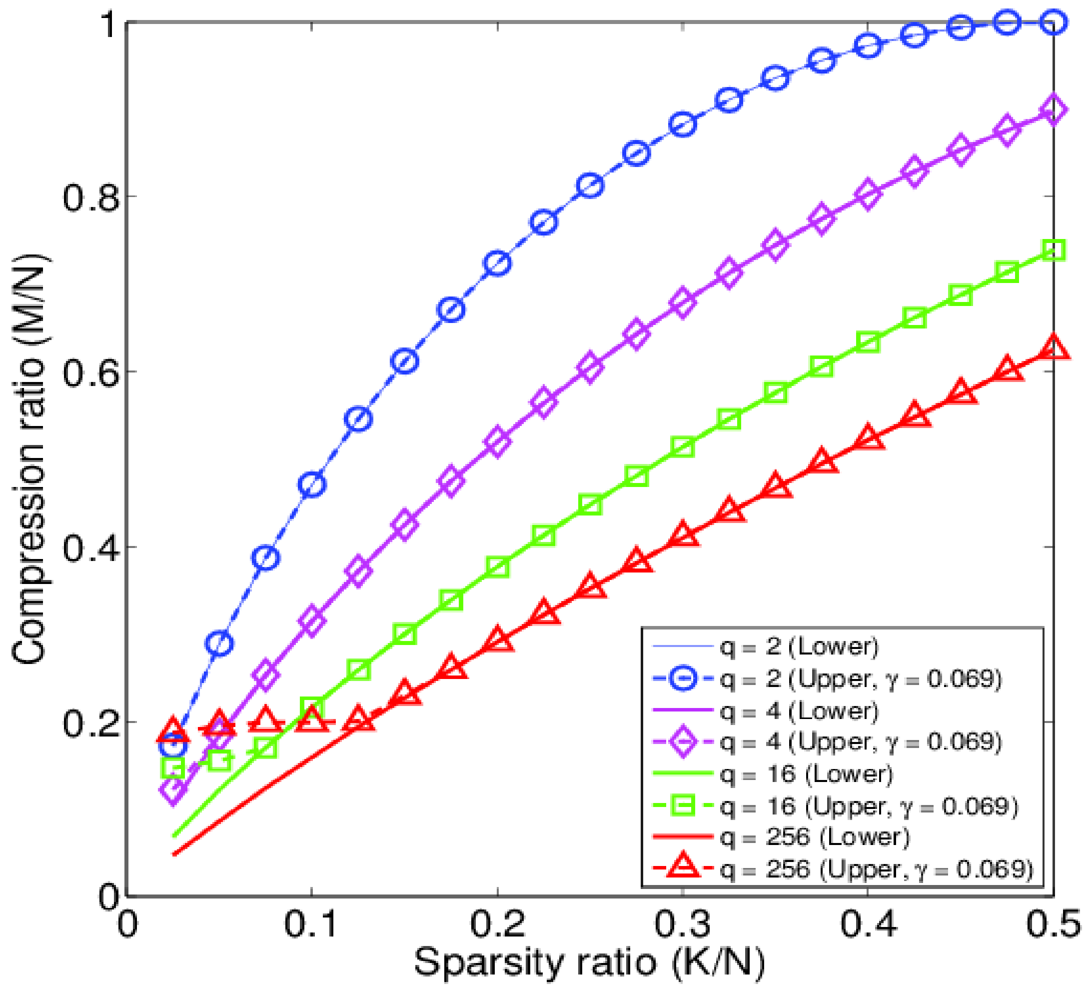
Theorem 3 (Necessary condition on *M*). For a probability of error arbitrarily small, the following

$$M > \frac{NH_b(K/N) + K\log_2(q-1) - \log_2(N+1)}{\log_2 q},$$

is a necessary condition.

Upper and Lower bounds

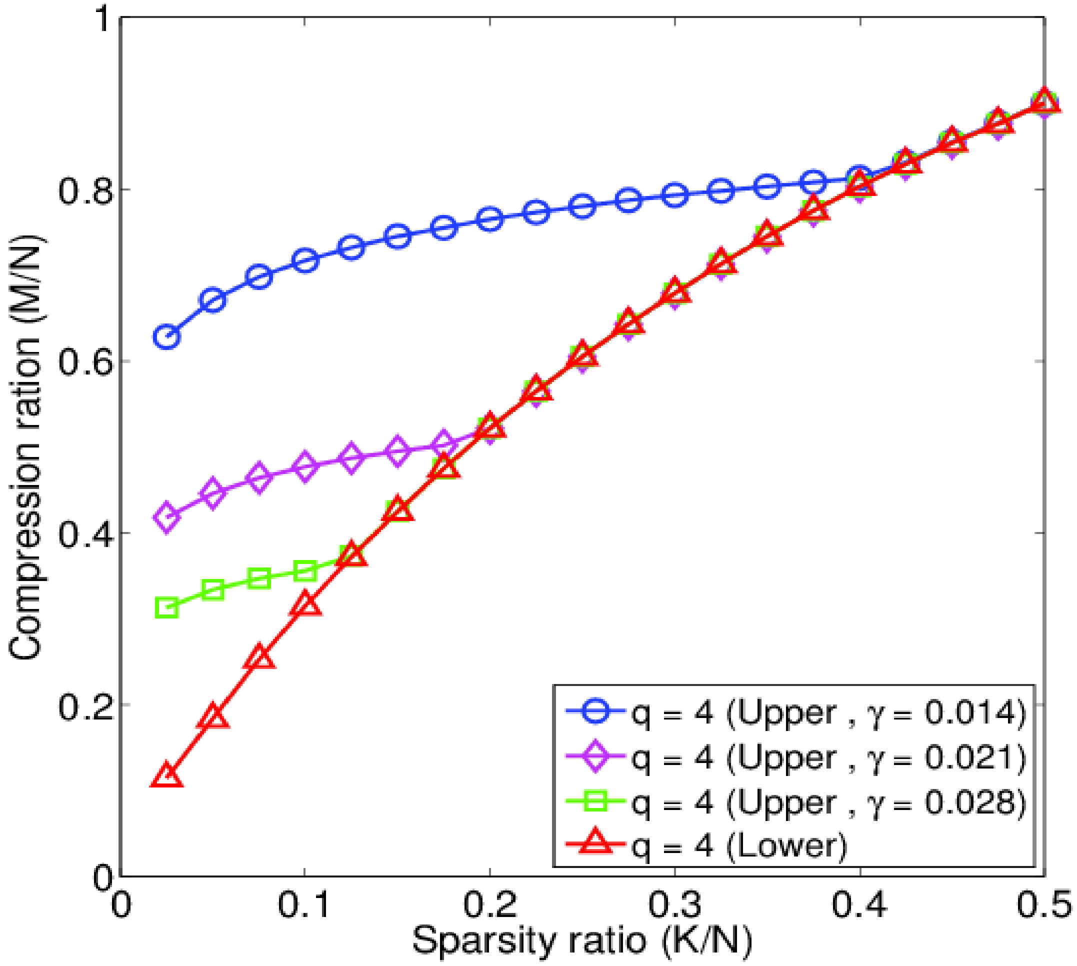
- **Results**



Lower (solid) bounds, and upper (dashed) bounds for N = 1000 with sparse factors γ =0.069

Upper and Lower bounds

❖ **Results**



Lower bounds and upper bounds for $N = 1000$ with different sparse factors.

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

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

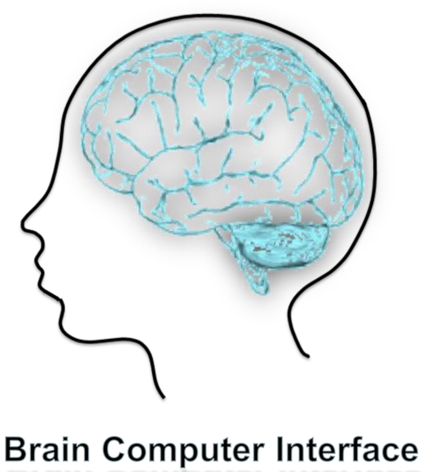
Summary

- Novel information theoretic results
  - Entropy of sparse signals
  - Mutual information between signal and the measurement
  - Led to Fano's inequality
- Novel results on the number of measurements needed over GF
  - Simplified decoders
  - Upper bounds
  - Combinatorial analysis technique
- New results on the density of sensing matrix
  - To measure sparse signals, a sufficiently dense matrix is needed!

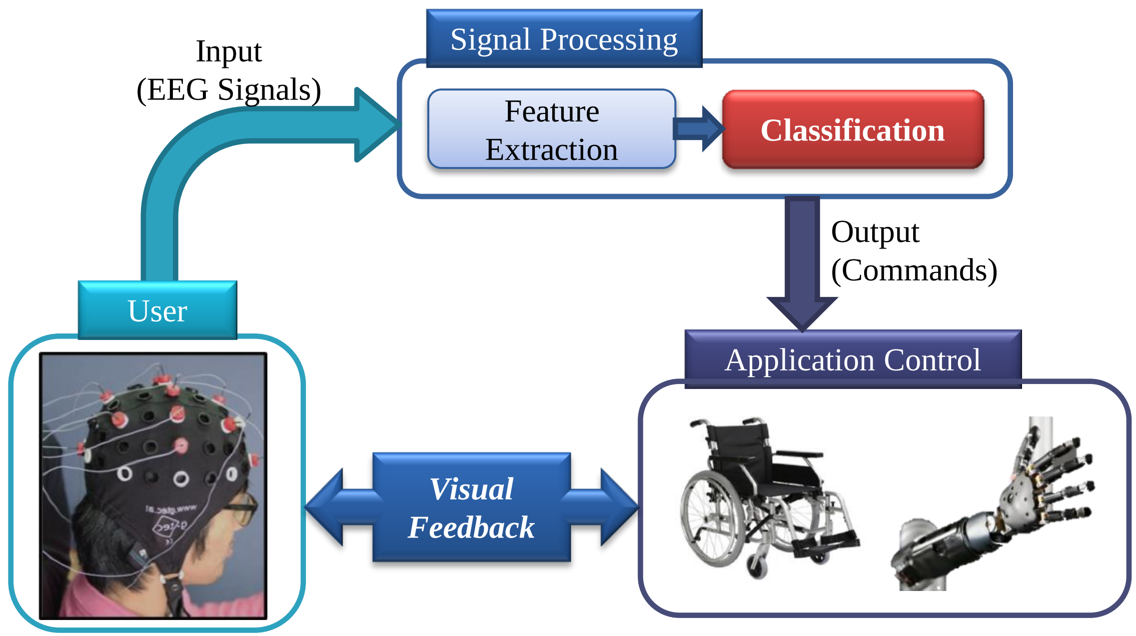
Sparse Representation based Classification Method for Motor Imagery based BCI Systems

Agenda

- EEG based BCI system
- Sparse Representation based Classification [Shin 2012 JNE]
  - Introduction
  - Motivation and purpose
  - Methods
  - Results
  - summary
- Evaluation of SRC method [current work]
  - Motivation and purpose
  - Methods
  - Results
  - Discussions
  - Summary
- Future work



EEG based BCIs

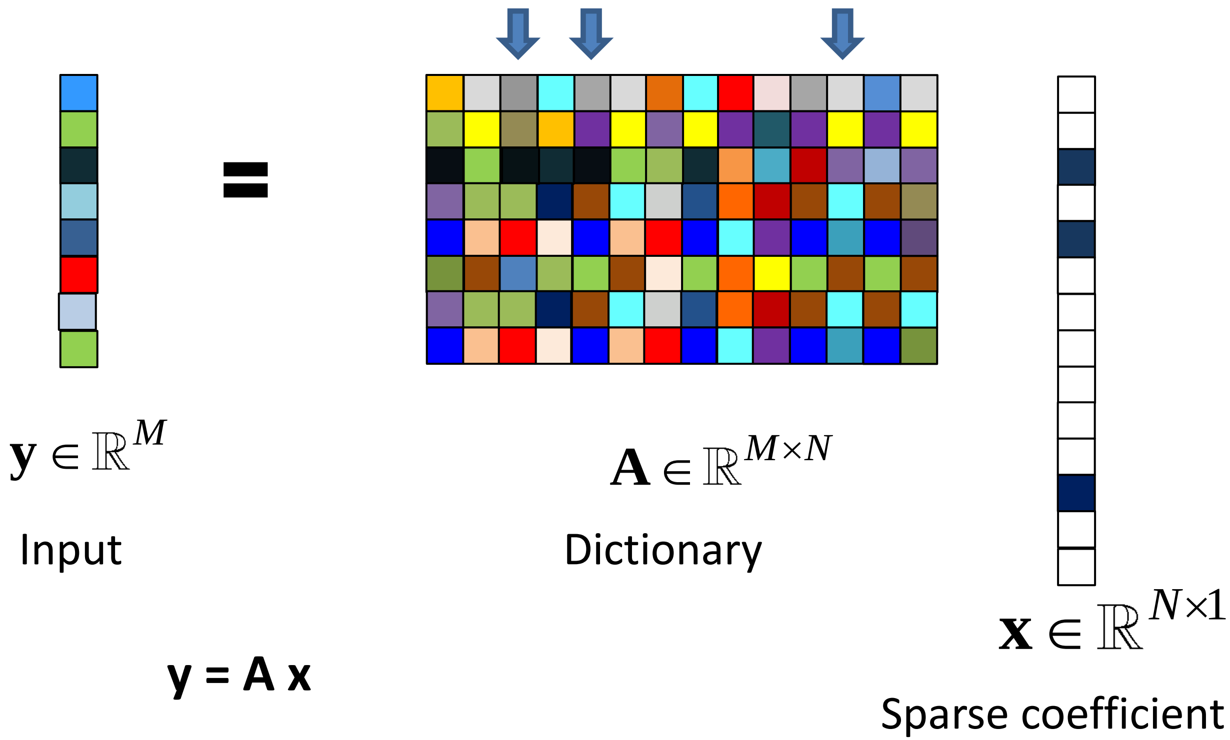


- BCI is a novel communication and control channel between person and external world.
- BCIs allow user-to-computer communication only using user's intention or imagination instead brain's normal output pathways of peripheral nerves and muscle.
- In the BCIs, classification is needed to transform the extracted feature of a user's intention into a computer command to control the external device.
- However, EEG signals are very noisy and have non-stationary characteristics. Therefore, powerful signal processing methods are needed.
- In this study we focus on BCI classification method.

Sparse Representation based Classification

Sparse Representation (SR)

- Recently, Sparse Representation has received a lot of attention in signal processing and machine learning field.
- The problem of SR is to find the most compact representation of a signal in terms of linear combination of atoms in an over-complete dictionary [Huang 2006].

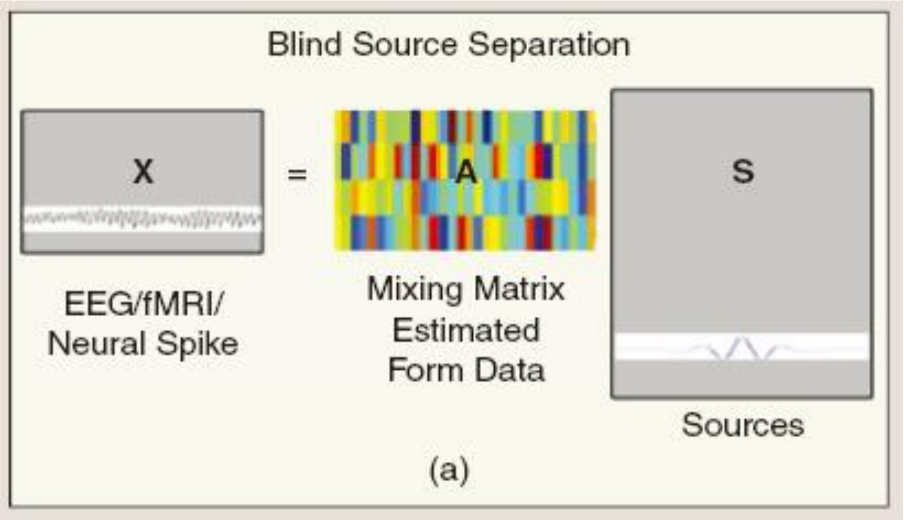


y = A x

Sparse representation for brain signal processing

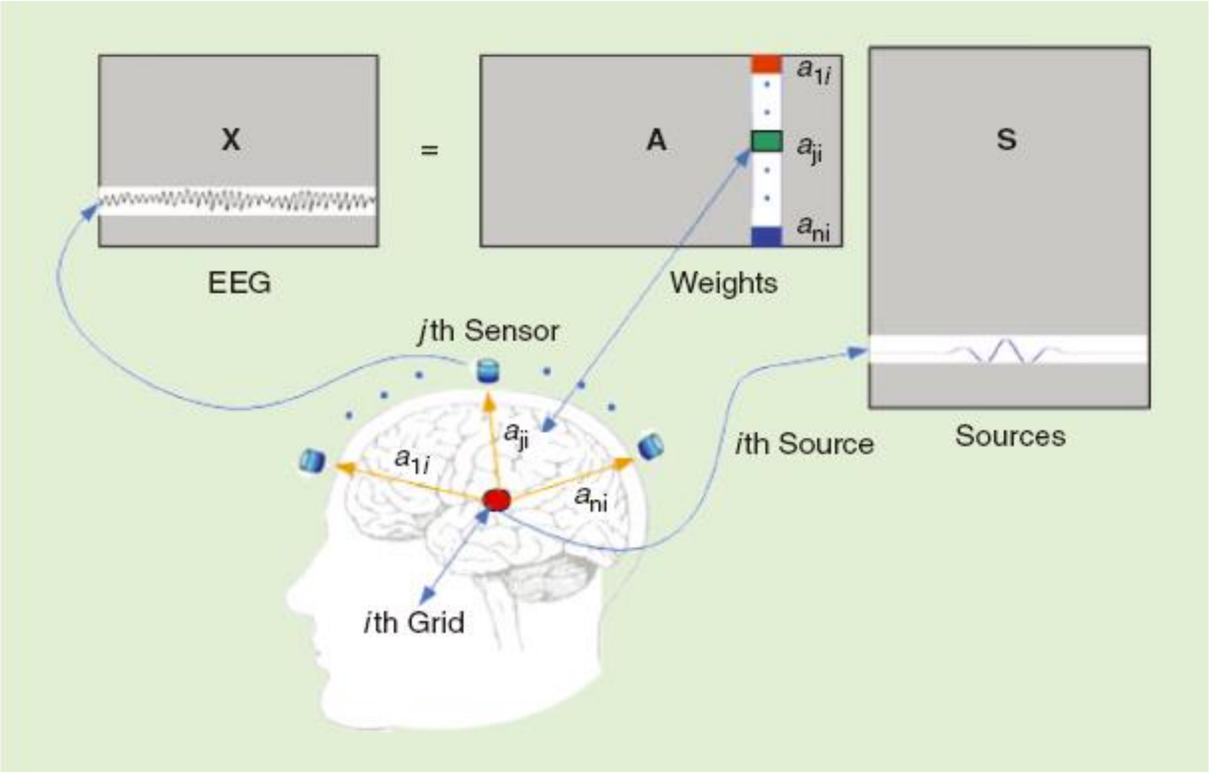
[Yuanqing 2014]

- Blind source separation
  - EEG signals can be considered as linear mixtures of unknown sources with an unknown mixing matrix.
  - The brain sources can be assumed to be sparse in a certain domain such as the time or the time-frequency domain
  - The true sources can be obtained through sparse representation-based BSS
  - The mixing matrix is estimated using, e.g., a clustering algorithm.



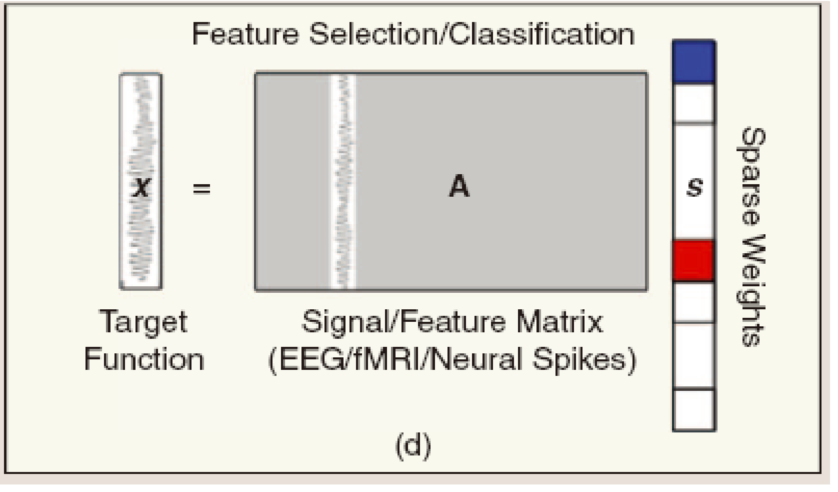
Sparse representation for brain signal processing
[Yuanqing 2014]

- EEG inverse imaging
  - The brain sources can be obtained and localized by sparse representation-based EEG inverse imaging where the mixing matrix A is first estimated based on a head model, and the brain sources are then separated and localized



Sparse representation for brain signal processing

- Feature selection and classification
  - Sparse representation-based classification (SRC) can be conducted as shown below [see Figure 1(d)].
  - The target function is a test sample/feature vector and each column of the data matrix is a training sample/feature vector of a certain class
  - These problems in brain signal processing can be solved under the framework of sparse representation.

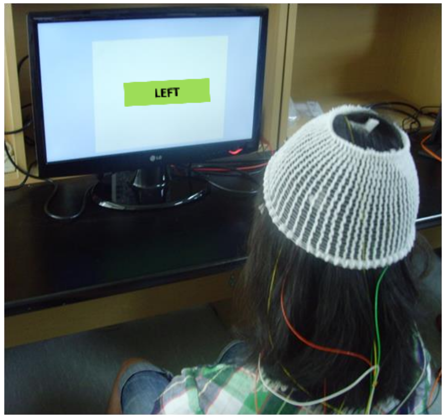


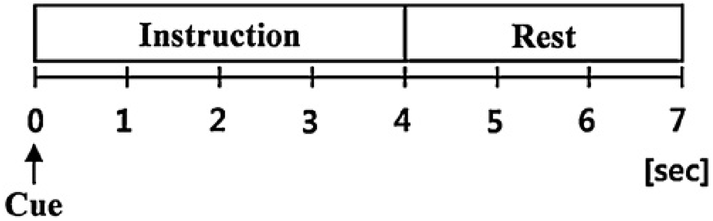
Motivation and Purpose

- Sparse representation can be used for a number of applications including noise reduction, source localization, and pattern recognition.
- Recently, classification based on Sparse Representation has received a lot of attention in face recognition and image processing [Wright 2009].
- This SR based classification shows satisfactory classification performance in many applications.
- In this study, we firstly apply SR to the motor imagery based BCI classification.
- Using Mu and Beta rhythms as a feature of MI BCI, we aim to develop a new Sparse Representation based Classification (SRC) method.

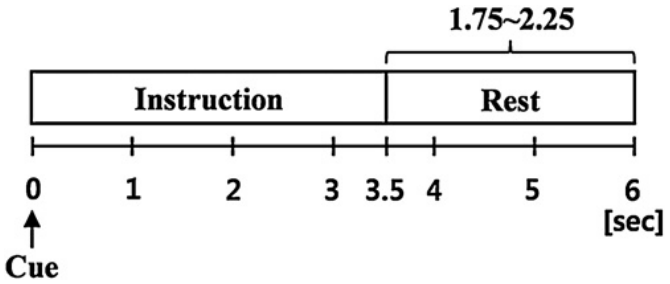
Data acquisition

- We use two different datasets
  - INFONET dataset
    - Five healthy subjects(average age = 22±6.85)
    - Right hand and left hand imaginations
    - 16 EEG channels
    - 80 trials per class

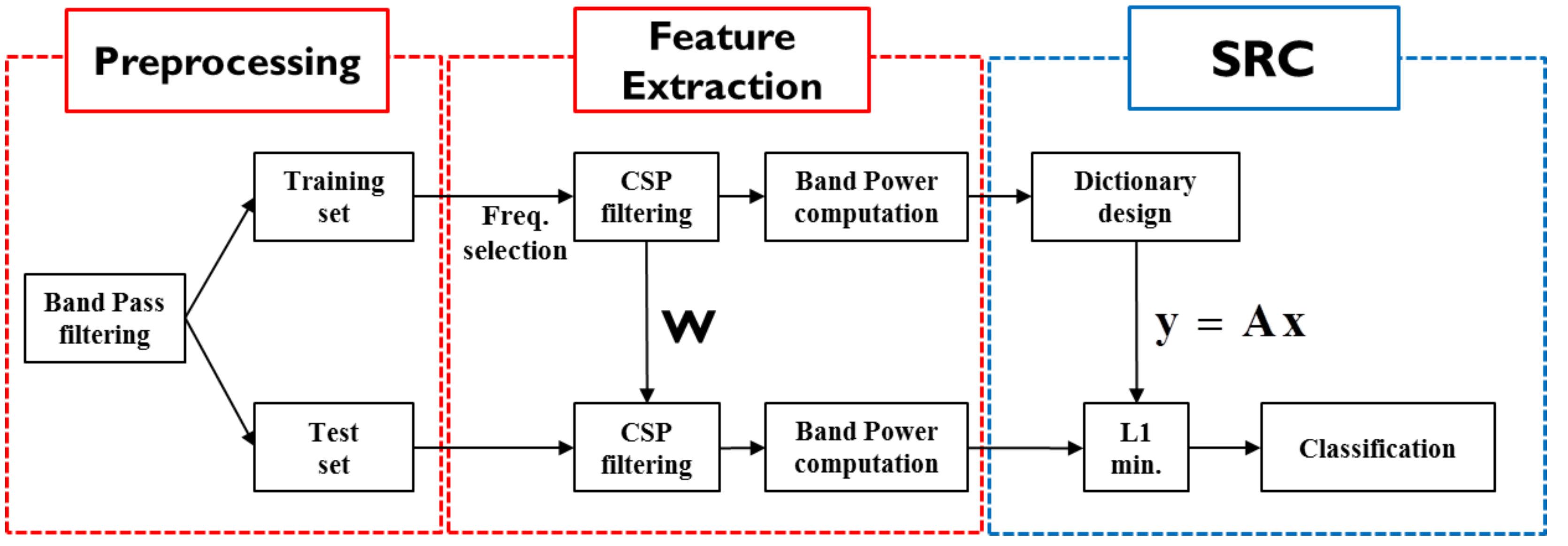




  - Berlin dataset
    - BCI competition dataset (Data set IVa)
    - Five healthy subjects
    - Right hand and right foot imaginations
    - 118 EEG channels
    - 140 trials per class

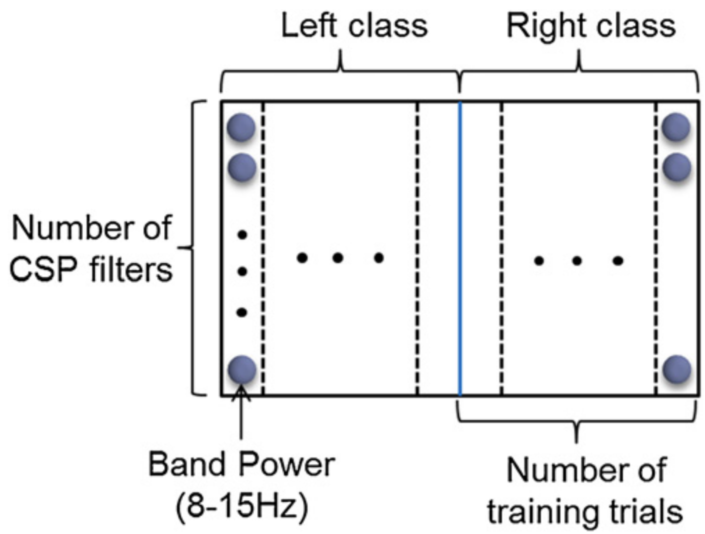


Proposed SRC scheme



- We use a band pass filtering as a preprocessing method.
- We designed dictionary **A** using CSP filtering.
- To use a mu rhythm as a BCI feature, we compute the power of mu band.
- To find coefficient vector **x**, we use the L1 minimization tool for test signal **y**.

Incoherent Dictionary



$$\mathbf{A} := [\mathbf{A}_L ; \mathbf{A}_R]$$

$$\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \ldots, \mathbf{a}_{i,N_t}]$$

where, i is class, N_t is total trials

$$M(\mathbf{A}_L, \mathbf{A}_R) \triangleq \max\left\{ \left| \left\langle \mathbf{a}_{L,j}, \mathbf{a}_{R,k} \right\rangle \right| : j,k = 1,2,\ldots,N_t \right\}$$

- *M* is the measure of mutual coherence of two component dictionaries; when *M* is small, we say that the dictionary is incoherent .
- The incoherent dictionary promotes the sparse representation of the test signal under the L1 minimization [Donoho 2003].
- We use the CSP filtering to design an *incoherent* dictionary.
- When a dictionary is incoherent, a test signal from one particular class can be predominantly represented by the columns of the same class.

Uncertainty Principle for Sparse Representation

- In quantum mechanics, Heigenberg's uncertainty principle (UP) state s that the momentum and the position of a particle, say of an electr on, cannot be simultaneously determined precisely.

$$\Delta p \, \Delta x \geq h$$

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

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

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 represented sparsely in both domains!**

UP and L1 Recovery

- Donoho-Stark in 89' then suggested the use of a combined matrix, a dictionary, and of the L1 min routine to represent the signal x:

$$(PD)\ \ \text{Find the most sparse representation } s,$$

$$\text{given a signal } x = Ds,\ \text{using the dictionary } D = \left[A;\ B\right].$$

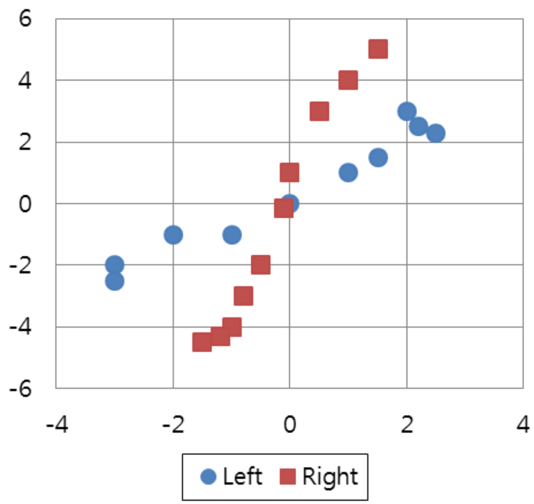
- This will be useful when one does not know which basis is more sui table for representing the signal.

- Using the UP, they show that
  - If $\|s\|_0 \le \frac{1}{\mu}$, then the equation has the unique solution (L0 solution unique) .
  - If $\|s\|_0 \le \frac{1}{2}\left(1+\frac{1}{\mu}\right) < \frac{1}{\mu}$, then the L1 solution attains the exact solution.

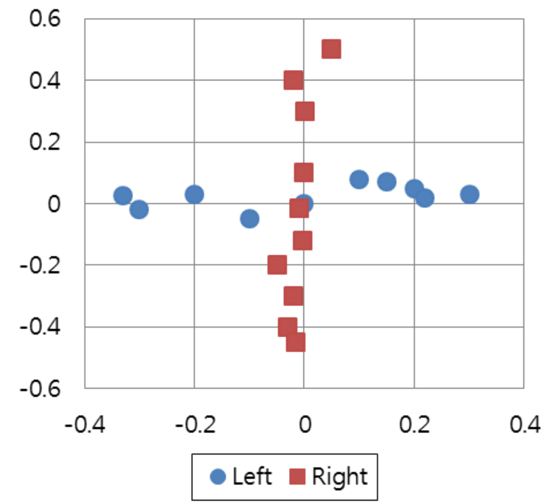
- These classic works done in 80s and 90s provides the foundation for the Compressed Sensing theory.

CSP(Common Spatial Pattern) filtering

- CSP filtering is a powerful signal processing technique suitable for EEG-based BCIs [Blankertz 2008].
- CSP filters maximize the variance of the spatially filtered signal for one class while minimizing it for the other class.
- The CSP filtering was used to produce high *incoherence* between the two group of columns in the dictionary.
- Using the CSP filter, we form maximally uncorrelated feature vectors between the two classes.

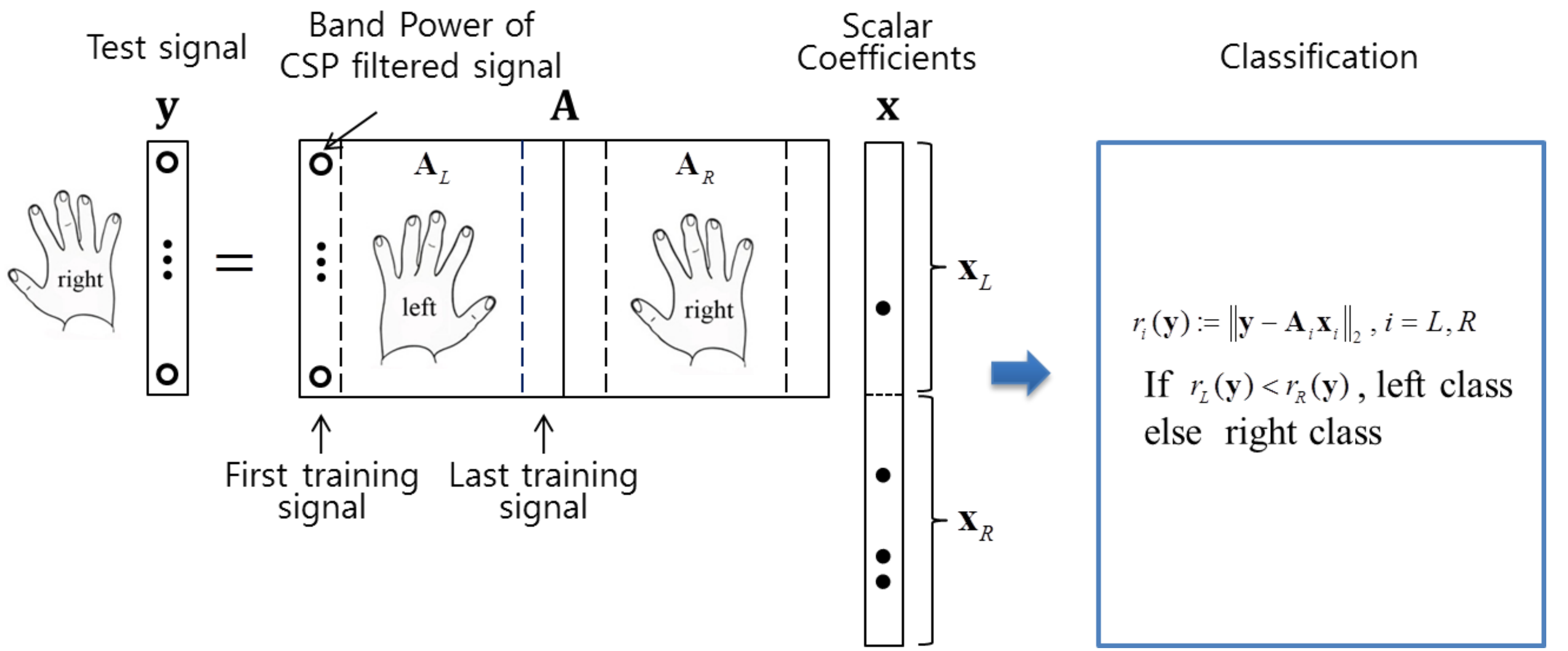


[before CSP filtering]



[after CSP filtering]

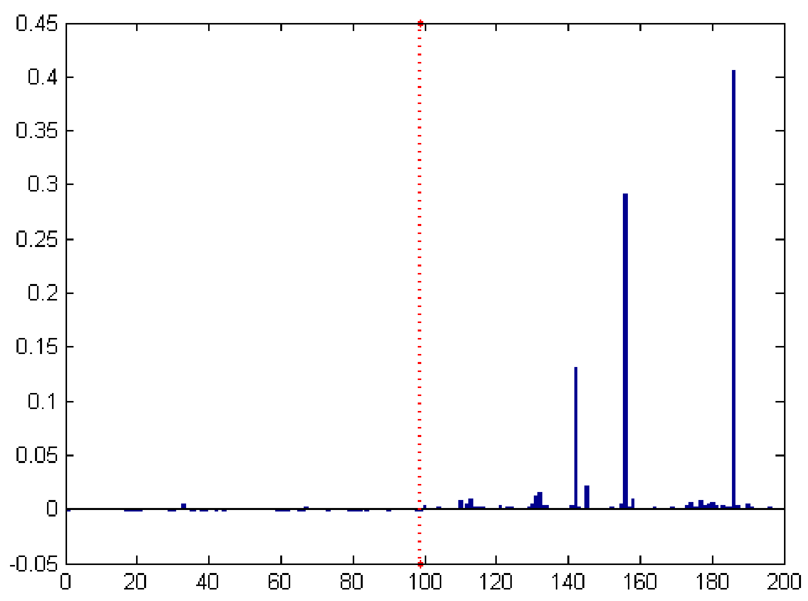
Sparse Representation and Classification



- The sparse representation can be solved by L1 minimization [Candès 2006].
- For example, a test signal **y** of the right class can be sparsely represented as the training signals of the right class.
- However, EEG signals are very noisy, nonzero coefficients may appear in the indices corresponding to the left class.
- We use a minimum residual classification rule.

Sparse representation results

- EEG Sparse representation
  - Sparse representation of real EEG signals for one subject.
  - X-axis represents the number of total training trials (the number of columns of dictionary A).
  - Y-axis represents the recovered coefficients $\mathbf{x}$ in $\mathbf{y} = \mathbf{A}\mathbf{x}$.
  - The class of the test trial was the right hand imagery.
  - The test signal of the right class is sparsely represented with several training signals of the right class.



Classification accuracy of INFONET dataset

Subject	SRC Accuracy [%]	LDA Accuracy [%]
A	95.63	93.13
B	63.75	61.87
C	68.14	67.50
D	80	76.25
E	71.25	68.12
Mean (SD)	75.75 (12.60)	73.37 (12.18)

- We use 2 CSP filters out of 16.
- For all subjects, the accuracy of the proposed SRC is better than conventional LDA method.

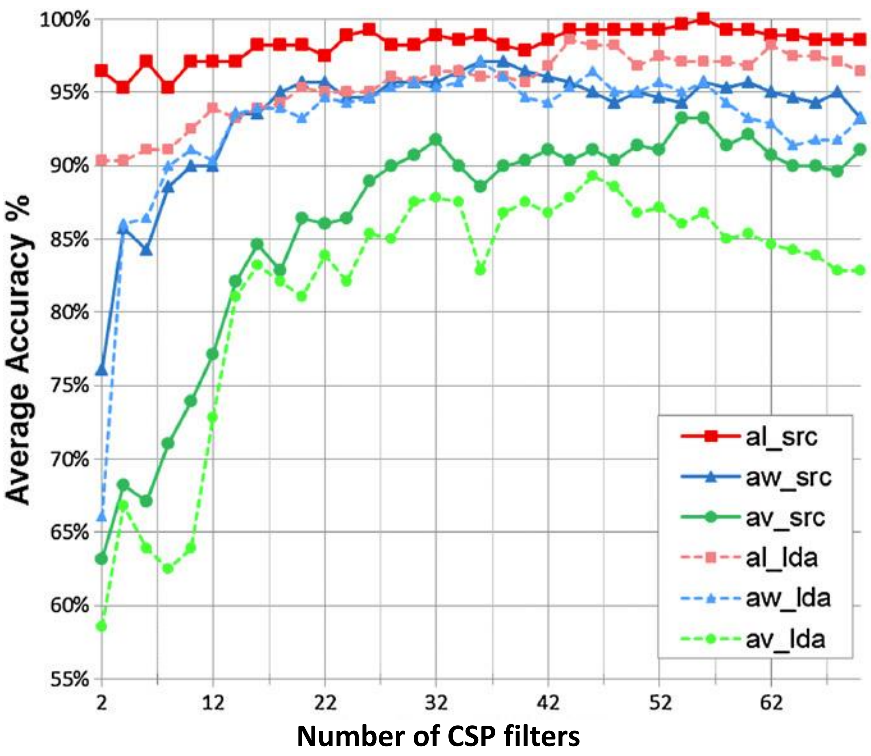
Classification accuracy of Berlin dataset

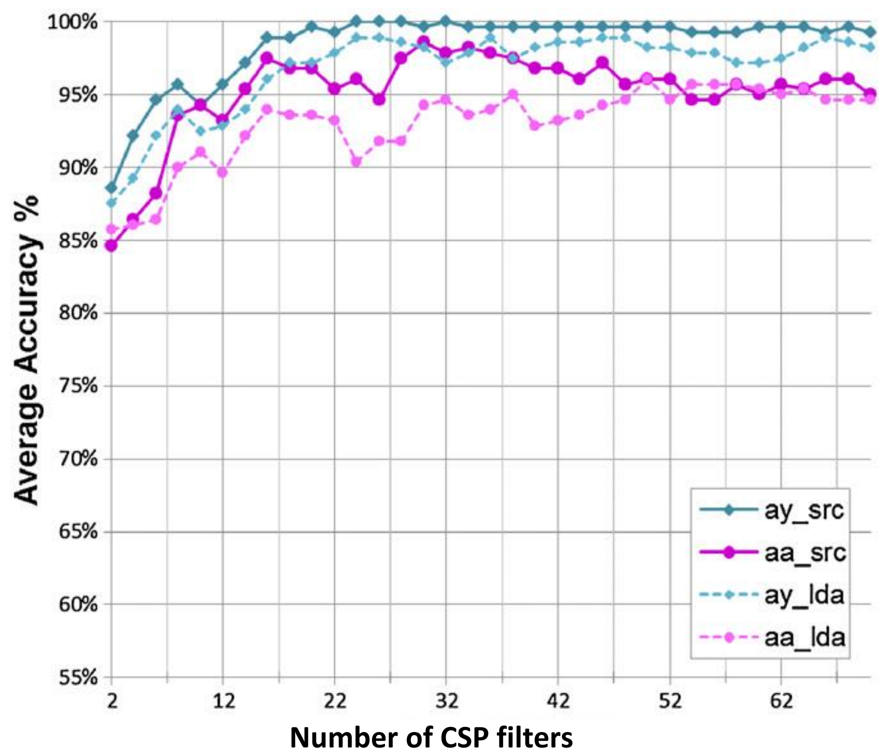
Subject	SRC Accuracy [%]	LDA Accuracy [%]
al	98.93	96.43
ay	100	97.14
aw	95.71	95.36
aa	97.86	94.64
av	91.79	87.86
Mean (SD)	96.85 (3.25)	94.29 (3.72)

- We use 32 CSP filters out of 118.
- For all subjects, the accuracy of the proposed SRC is better than conventional LDA method.

Classification results

- Berlin dataset
  - We examine classification accuracies of SRC and LDA as a function of the number of CSP filters (feature dimensions) for each subject.





Summary

- We propose a sparse representation based classification (SRC) method for the motor imagery based BCI system.
- The SRC method needs a well-designed dictionary matrix made of a given set of training data.
- We use the CSP filtering to make the dictionary uncorrelated for two different classes.
- The SRC method is shown to provide better classification accuracy than the LDA method.

Reference

[Wolpaw1991] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-based brain-computer interface for cursor control," Electroencephalogr. Clin. Neurophysiol., vol. 78, pp. 252-259, 1991.

[Shin 2012] Younghak, S., Seungchan, L., Junho, L., Heung-No, L., 2012. Sparse representation-based classification scheme for motor imagery-based brain-computer interface systems J. Neural Eng. 9, 056002.

[Huang 2006] K. Huang and S. Aviyente, "Sparse Representation for Signal Classification," Neural Information Processing Systems, 2006.

[Donoho2006] D.Donoho, "Compressed sensing," IEEE Trans. Information Theory, vol. 52, pp. 1289-1306, 2006

[Wright2009] 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

[Gemmeke 2011] Gemmeke, J.F., Virtanen, T., Hurmalainen, A., 2011. Exemplar-based sparse representations for noise robust automatic speech recognition. IEEE Trans. Audio, Speech, Lang. Proc. 19 (7), 2067–2080.

[Blankertz2008] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, Klaus-Robert Müller. "Optimizing Spatial Filters for Robust EEG Single-Trial Analysis," IEEE Signal Proc. Magazine, 25(1):41-56, January 2008.

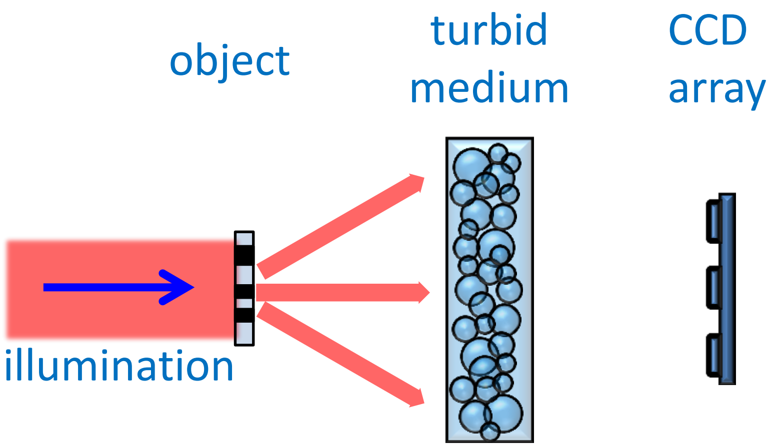
[Candès 2006] E. Candès, J. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," Comm. Pure and Applied Math., vol. 59, no. 8, pp. 1207-1223, 2006

[Kübler 2001] Kübler, A., Neumann, N., Kaiser, J., Kotchoubey, B., Hinterberger, T., and Birbaumer, N. P. (2001). Brain-computer communication: self-regulation of slow cortical potentials for verbal communication. Arch. Phys. Med. Rehabil. 82, 1533–1539.

Compressive Sensing for Imaging through turbid media

Compressive Sensing for Imaging through turbid media

- See object hidden under turbid media [Mosk2012]
  - Turbid media: biological tissues, white paint
  - It may become possible to have an early disease diagnosis with optical imaging

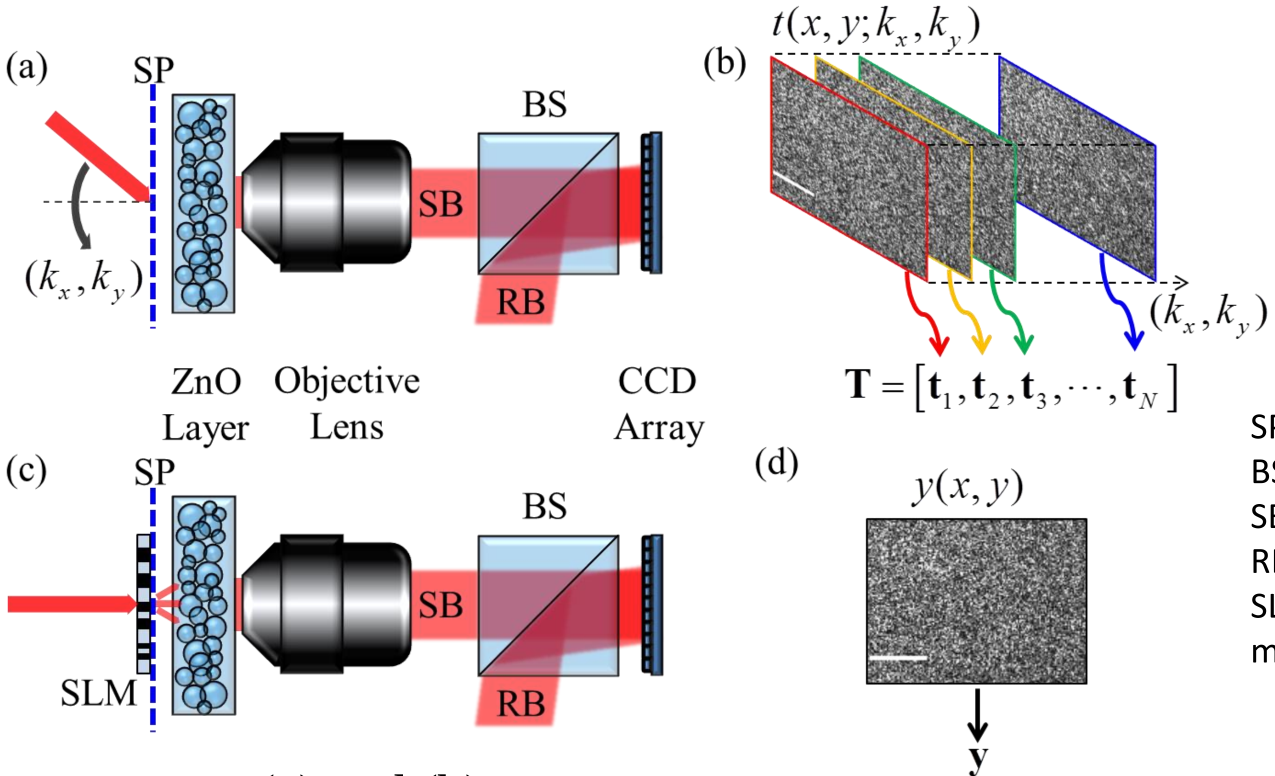


- Due to the multiple scattering, the outgoing object waves are spatially scrambled and become a speckle field (SF) at an observation plane
- For image recovery, the multiple scattering should be suppressed; the object image should be recovered

Compressive Sensing for Imaging through turbid media

- The wave propagation is a time reversible (TR) process [Mosk2012], [Yaqoob2008]
- The multiple scattering in turbid media can be reversed by a TR operator
- Phase conjugation (PC) is the monochromatic version of the TR operator
  - A *de facto* standard method to date for imaging through turbid media
  - PC compensates the phase variations due to multiple scattering in turbid media by recording the SFs and back-propagating the complex conjugates of them through the media so that the phase variations are cancelled; a photorefractive crystal is used as a phase conjugate mirror.
- Computational PC
  - PC can be done virtually through computational estimation
    - This requires the so called transmission matrix (TM) of the medium [Popoff2010]
    - TM-based image recovery
  - SFs are recorded at the CCD array and the recovery is made in digital signal processing
  - A number of advantages over the optical PC for it has an image data format which is reproducible [Cui2010]

Compressive Sensing for Imaging through turbid media



SP: sample plane,
BS: beam splitter,
SB: sample beam,
RB: reference beam,
SLM: spatial light modulator

- TM acquisition: (a) and (b)
  - A collection of plane waves each with different incident angle is used as a basis
  - The SF for each plane wave is obtained and stored as a column in TM
- Object speckle field acquisition: (c) and (d)
  - The object SF (OSF), which is the output SF of turbid medium with the object wave, instead of the plane wave, is then obtained

Compressive Sensing for Imaging through turbid media

- System model

$$\mathbf{y} = \mathbf{T}\mathbf{a} + \mathbf{n}$$

where $\mathbf{y} \in \mathbb{C}^M$, $\mathbf{a} \in \mathbb{C}^N$, $\mathbf{n} \in \mathbb{C}^M$ are the vector representations of $y(x, y)$, $a(k_x, k_y)$, $n(x, y)$, and each column of $\mathbf{T} \in \mathbb{C}^{M \times N}$ is the vector representation of the for a given (k_x, k_y). Each element of **T** is a CSCG random variable.

Compressive Sensing for Imaging through turbid media

- The estimate by PC:

$$\hat{\mathbf{a}}_{\mathrm{PC}} = \mathbf{T}^{*}\mathbf{y}$$
$$= \mathbf{T}^{*}\mathbf{T}\mathbf{a} + \mathbf{T}^{*}\mathbf{n}$$

- PC is not good
  - For correlated cases, each element of the estimated angular spectrum is contributed not only from the angular spectrum element with the considered angle but also from those with the other angles whose SFs are correlated to that with the considered angle.
  - Thus, erroneous estimation is made even in noiseless cases.
  - Note that turbid media do not provide orthogonal TMs for they have memory effects among the SFs of the input waves whose incident angles are not separated enough [Freund 1988]
- It appears to have insufficient speckle suppression in the image recovered by PC [Popoff2010]
  - This requires an additional procedure such as temporal ensemble averaging over multiple exposures
  - In time-critical cases or in the case of imaging a moving object, its applicability can be limited

Compressive Sensing for Imaging through turbid media

- CS framework is suitable for imaging through turbid media
  - Compressibility
    - Most natural object images are well approximated by only several terms in the Fourier domain [Bruckstein2009].
    - We see that the basis signals in TLI are plane waves with different angles and the image is an angular spectrum in the Fourier domain
    - Thus angular spectrum is expected to be well approximated by small number of elements
  - Isometry
    - Checking the isometry of a matrix is a NP hard problem.
    - But, the Gaussian distributed matrices are proven to have an optimal isometry [Bruckstein2009], [Candès2011]
      - Through the random walk analysis, it was found that the SF in the transmission geometry is complex-valued Gaussian distributed provided that the number of elementary contributions is large [Goodman1976]

Compressive Sensing for Imaging through turbid media

- The SSE, an oracle-like estimation, can be made by solving the following L1 norm minimization problem [Bruckstein2009], [Candès2011]

$$\hat{\mathbf{a}}_{\mathrm{SSE}} = \arg\min_{\mathbf{a}} \|\mathbf{a}\|_1 \ \text{ subject to } \mathbf{y} = \mathbf{T}\mathbf{a}.$$

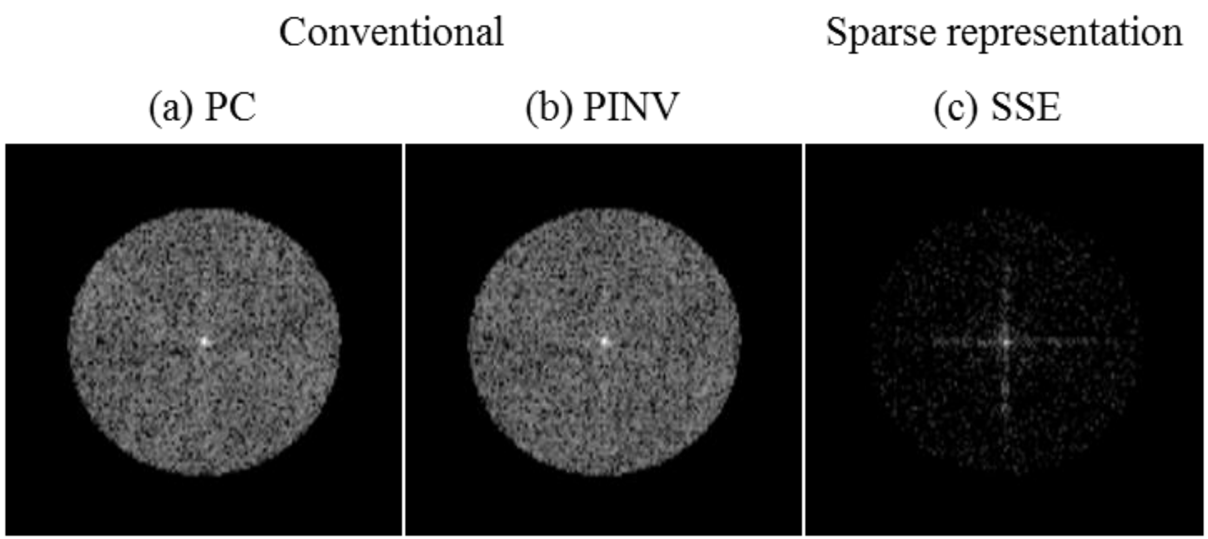
  – The SSE aims to find the solution which has the smallest number of nonzero elements, $\|\mathbf{a}\|_0$, (with a compact representation)
    - This is NP hard problem

  – But, the L1 norm minimization promotes the estimate to be close to a compressible signal which has a compact representation.

$$\|\mathbf{a}\|_1 := \sum_i |a_i|$$

Compressive Sensing for Imaging through turbid media

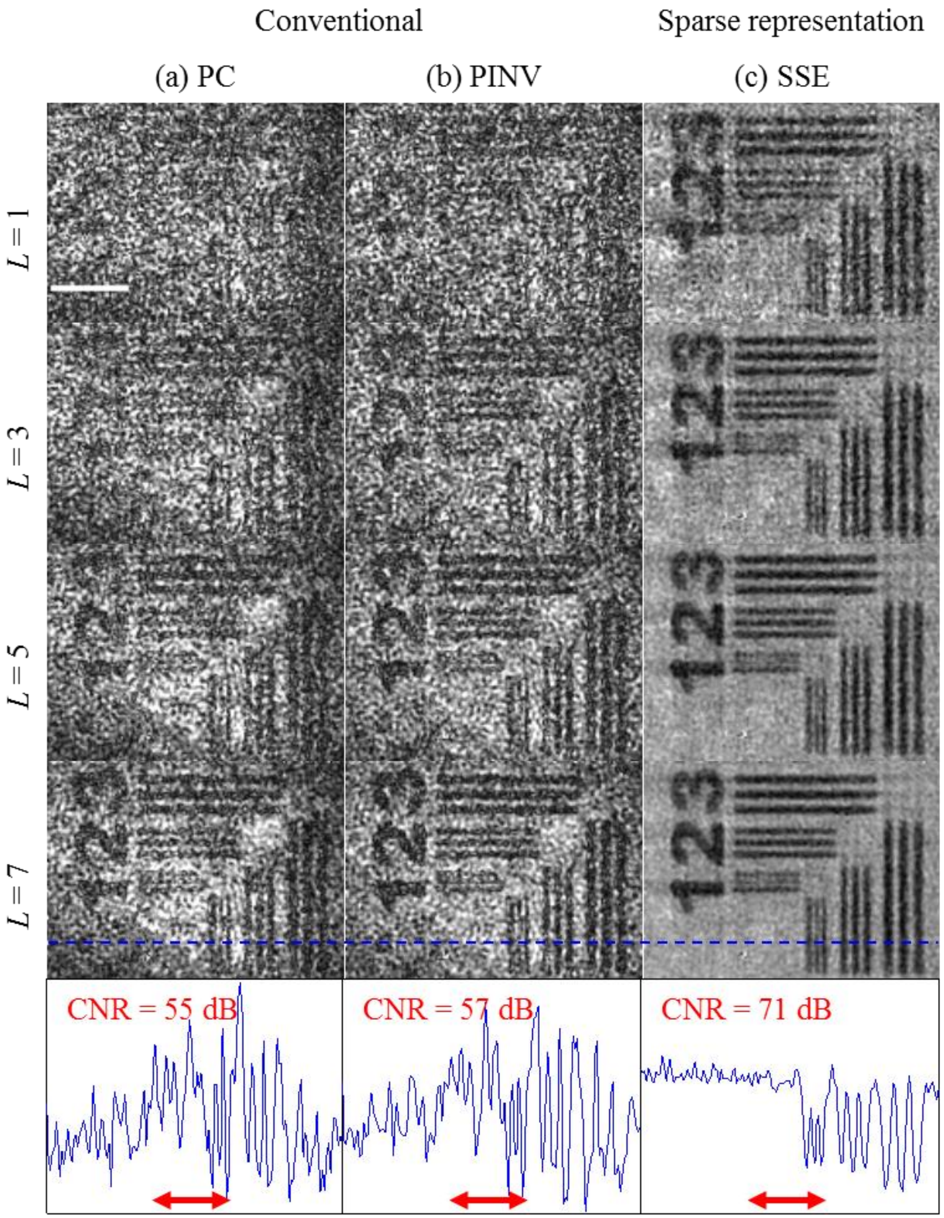
❖ Angular spectrum estimation



Estimated angular spectra using (a) PC, (b) PINV, and (c) SSE, respectively. Here, M = 4389, N = 20000. All angular spectra are represented in log scale for better visibility.

❖ Most error terms in the estimated angular spectrum by SSE are reduced considerably

- Image recovery
  - Recovered amplitude images averaged over one, three, five, and seven samples
  - Cross sections of them averaged over seven samples
  - Constrast-to-noise ratios (CNRs) are calculated in the subsets (red arrow lines) of the cross sections.
    - Here, M = 4389, N = 20000 and K is less than 147.
    - Scale bar: 10 μm.



Cell-Imaging Thru Biological Tissue

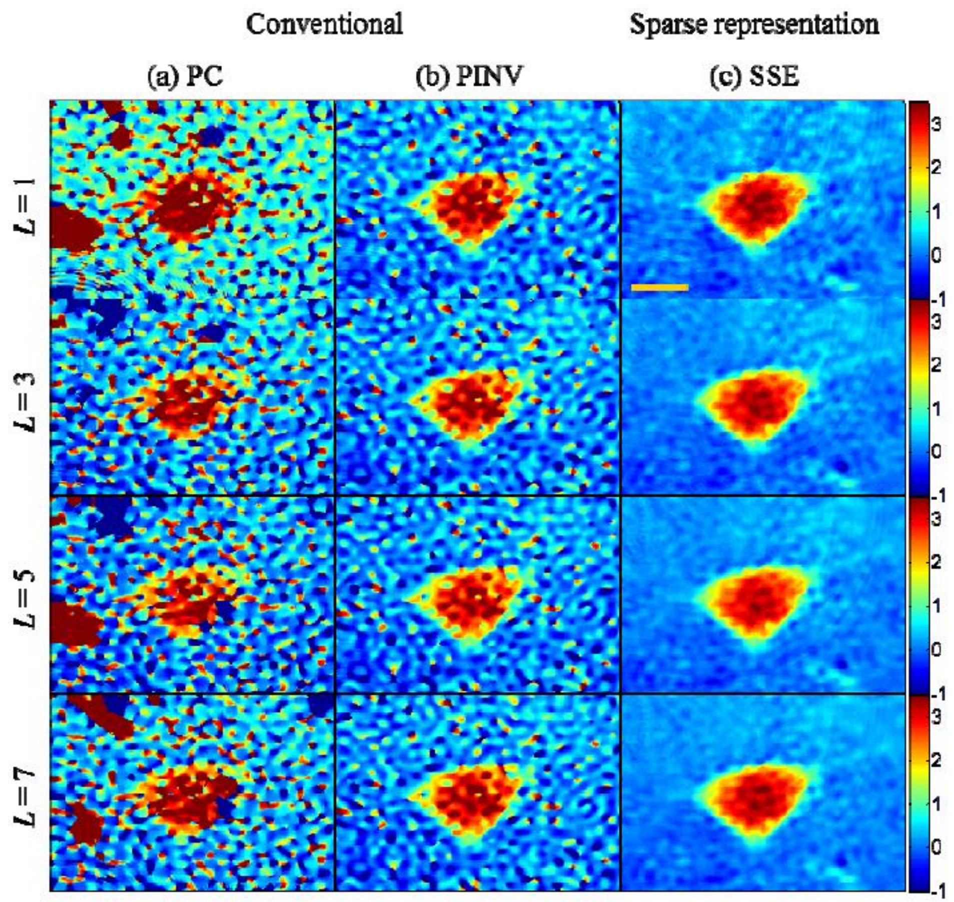


Fig. 5. Recovered quantitative phase images of a live cell averaged over one, three, five, and seven samples using (a) PC, (b) PINV, and (c) SSE, respectively. Here, $M = 696$, $N = 5000$. Color bar: phase in radians. Scale bar: 10 μm.

Summary of turbid lens

- MIMO 통신에서는, random scattering 채널이 있습니다. Scattering이 많은 지상 채널이 깨끗한 채널보다 통신 용량이 클 수 있고, degree of freedom이 사용하는 안테나 수 만큼 커져서, rate-diversity를 trade-off 할 수 있는 여지도 제공하고, 여러 가지 장점이 있음이 알려졌고, 실제 시스템 개발에 적용되어 왔습니다.

- 광학에서는, 최근에, turbid 매터리얼을 렌즈로 사용하면, 광학시스템의 resolution 및 field of view등 을 크게 높일 수 있음이 실험적으로 증명되었습니다. 그러나, 이론적 접근 측면에서 많은 연구가 필요한 부분입니다.

- 본 연구에서는 CS를 turbid렌즈 이메징에 적용하고 input신호의 estimation 결과를 높일 수 있음을 실험적으로 보였습니다.

- 정보 및 신호처리이론의 Information-Estimation Theory를 활용하여 estimation의 검출 능력 한계 분석 및 새로운 검출 방법 등을 제안 할 수 있는 새로운 연구분야 입니다.

References

- [Jang2014-2] H. Jang, C. Yoon, E. Chung, W. Choi, and H.-N. Lee, "Speckle suppression via sparse representation for wide-field imaging through turbid media," Accepted for publication in Optics Express .
- [Mosk2012] A. P. Mosk, A. Lagendijk, G. Lerosey, and M. Fink, "Controlling waves in space and time for imaging and focusing in complex media," Nature Photonics. 6, 283–292, 2012.
- [Yaqoob2008] Z. Yaqoob, D. Psaltis, M. S. Feld, and C. Yang, "Optical phase conjugation for turbidity suppression in biological samples," Nature Photonics. 2, 110–115, 2008.
- [Popoff2010] S. M. Popoff, G. Lerosey, R. Carminati, M. Fink, A. C. Boccara, and S. Gigan, "Measuring the transmission matrix in optics: an approach to the study and control of light propagatin in disordered media," Phys. Rev. Lett. 104, 100601, 2010.
- [Cui2010] M. Cui and C. Yang, "Implementation of a digital optical phase conjugation system and its application to study the robustness of turbidity suppression by phase conjugation," Opt. Express. 18, 3444–3455, 2010.
- [Choi2011] Y. Choi, M. Kim, C. Yoon, T. D. Yang, K. J. Lee, and W. Choi, "Overcoming the diffraction limit using multiple light scattering in a highly disordered medium," Phys. Rev. Lett. 107, 023902, 2011.
- [Popoff2010] S. M. Popoff, G. Lerosey, M. Fink, A. C. Boccara, and S. Gigan, "Image transmission through an opaque material," Nature Communications. 1, 1–5, 2010.
- [Bruckstein2009] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," SIAM Review. 51, 34–81, 2009.
- [Goodman1976] J. W. Goodman, "Some fundamental properties of speckle," J. Opt. Soc. Am. 66, 1145–1150, 1976.
- [Freund 1988] I. Freund and M. Rosenbluh, "Memory effects in propagation of optical waves through disordered media," Phys. Rev. Lett. 61, 2328 (1988).
- [Candès2011] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," Applied and Computational Harmonic Analysis. 31, 59–73, 2011.

Compressed Sensing Radar

Motivations of compressed sensing radar

- **레이더는 원거리 공간에 있는 표적을 전자기파로 탐색하는 센서 시스템**
  - 레이더의 측정 샘플 수는 한정되어 있고, 탐색해야 하는 공간은 넓음
  - 레이더 문제는 ill-conditioned 선형 역 문제를 푸는 문제
  - 2000년대 중반, 이러한 ill-conditioned 역 문제를 Sparse Representation 과 Compressed Sensing을 통하여 푸는 혁신적인 방안이 신호처리 및 정보이론 분야에서 발표

- **압축 센싱 레이더는 이 방법을 레이더 문제에 적용한 것**
  1. 원거리 공간에 있는 표적은 성긴(sparse) 신호로 표현 (Sparse Representation)
  2. 이러한 성긴 신호는 한정된 압축 샘플로도 완벽하게 복원 가능
  3. 분해능 향상 및 여러 표적 동시 추적 능력 증대

Surprising results of compressed sensing radar

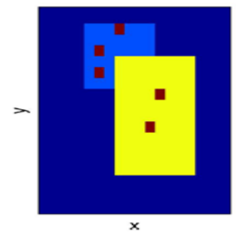
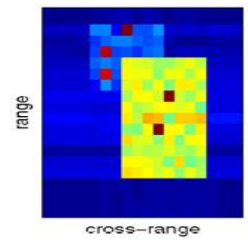
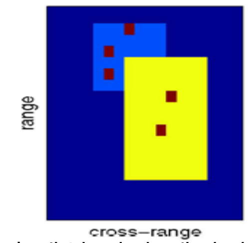
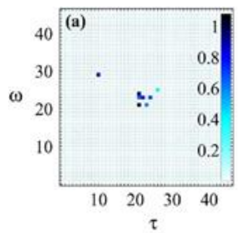
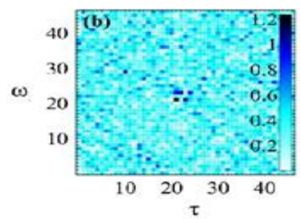
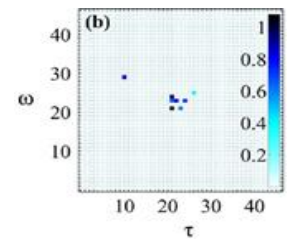
- 기존의 정합필터(matched filter) 기반 레이더는 **신호의 대역폭(B)**에 의해 **거리 해상도(range resolution,$\Delta r \geq \frac{c}{2B}$, $c \approx 3 \times 10^8$)가 결정됨**

- 압축센싱 기반 레이더는 **상기의 거리 해상도를 뛰어 넘을 수 있다는 결과**가 보고됨 [Strohmer09][Pi11][Ender10]

참고 문헌

1. [Strohmer09] : A. Herman and Thomas Strohmer, "High-Resolution Radar via Compressed Sensing", IEEE Trans. On Signal Processing, 2009, 인용회수 : 468회
2. [Pi11] : J. Xu and X. Pi, "Compressive Sensing in Radar high Resolution Range Imaging", Journal of Computational Information System, 2011, 인용회수: 7회
3. [Ender10] J. H. G. Ender, "On compressive sensing applied to radar", Signal Processing, 2010, 인용회수 : 204번

Literature survey on compressed sensing radar (1/6)

출 처	내 용	실 험 결 과		
[Baraniuk07]	• 압축센싱 기반의 레이더의 개념을 **처음으로 제안함** • 제안된 레이더는 **정합필터 (matched filter)가 필요 없음**.	 원본	 정합필터 기반 레이더 결과	 압축센싱 기반 레이더 결과
[Strohmer09]	• 압축센싱 기반 레이더의 해상도 (range resolution, Doppler resolution)의 한계에 대해 분석 • **압축센싱 기반 레이더의 해상도**가 정합필터 기반 레이더의 해상도보다 **우수한 것을 보임**.	 원본	 정합필터 기반 레이더 결과	 압축센싱 기반 레이더 결과

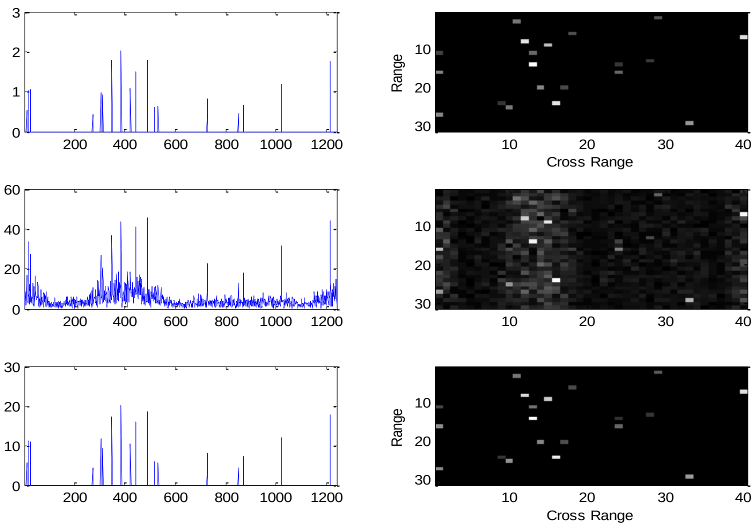
특히, [Strohmer09] 논문에서 아래와 같이 직접적으로 압축센싱 기반 레이더의 우수성을 강조

1. "our stylized compressed sensing radar which under appropriate conditions **can "beat" the classical radar uncertainty principle**!"
2. "the benefit of employing compressed sensing recovery manifests itself as **a dramatic increase in resolution**."
3. "Experimentally confirm that **compressed sensing radar can achieve much higher resolution** than traditional techniques

참고 문헌

1. [Strohmer09] : A. Herman and Thomas Strohmer, "High-Resolution Radar via Compressed Sensing", IEEE Trans. On Signal Processing, 2009, 인용회수: 468 회
2. [Baraniuk07] : R. Baraniuk and P. Steeghs, "Compressive Radar Imaging", IEEE Radar Conference, 2007, 인용회수: 498회

Literature survey on compressed sensing radar (2/6)

출 처	내 용	실 험 결 과
[Chen08]	• 압축센싱 기반 MIMO 레이더를 **처음으로 제안함** • 표적의 거리, 방위각, 속도를 탐색하고자 함 • 실험을 통해 압축센싱 기반 레이더의 해상도가 정합 필터 기반 레이더의 것 보다 우수한 것을 보임	원본 정합 필터 기반 레이더 압축센싱 기반 레이더 

참고 문헌

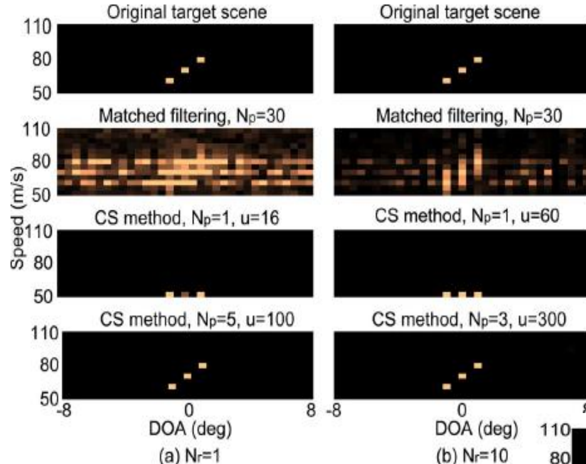
1. [Chen08] : Chun-Yang Chen and P. P. Vaidyanathan, “Compressed Sensing in MIMO Radar”, Signals, systems and computers Asilomar, 2008 인용회수: 67회

Literature survey on compressed sensing radar (3/6)

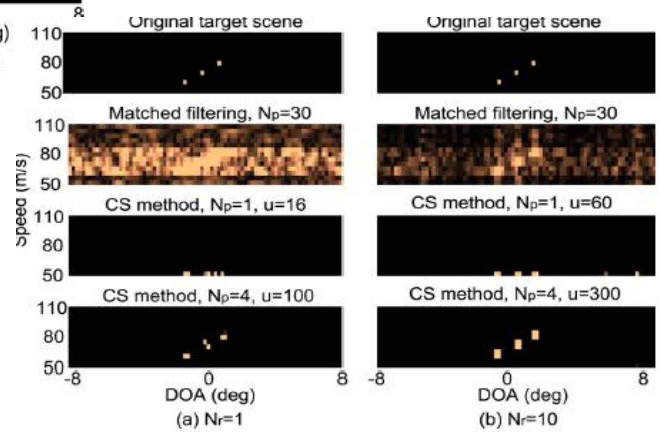
<table>
<tr><th>출 처</th><th>내 용</th><th>실 험 결 과</th></tr>
<tr>
<td>[Poor10]</td>
<td>

- 압축센싱 기반 MIMO 레이더 제안
- 표적의 위치와 방위각을 추정하고자 함
- 탐색 공간을 방위각-도플러 평면상에 표시 [Appendix 1 참조]
- Np: Tx의 개수
- Nr: Rx의 개수

</td>
<td>



- 표적이 격자 점에 존재하는 경우(왼쪽)



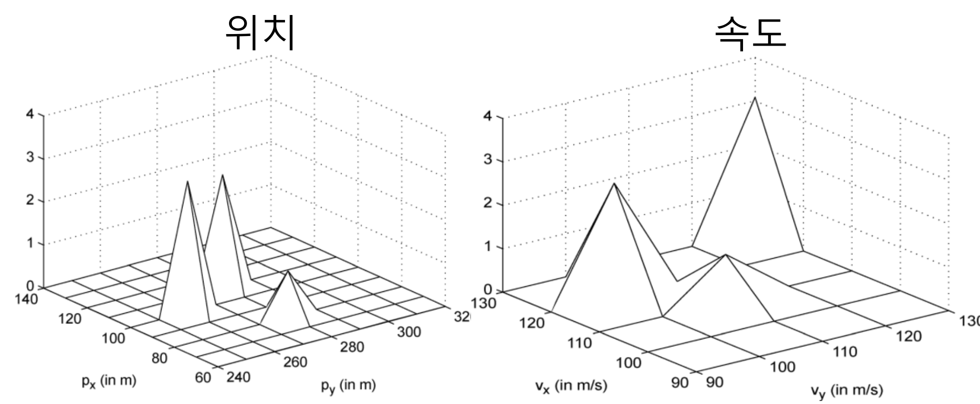
- 표적이 격자 점에서 어긋나있는 경우(오른쪽) [Appendix2 참조]

</td>
</tr>
</table>

참고 문헌

1. [Poor10] : Y. Yu, A. P. Petropulu, H. V. Poor, “MIMO Radar Using Compressive Sampling”, IEEE Journal of Selected Topics in Signal Processing, 2010, 인용회수: 109회

Literature survey on compressed sensing radar (4/6)

출 처	내 용	실 험 결 과
[Nehorai11]	• Distributed MIMO Radar 제안 • 복원하는 신호-즉 표적들의 **위치**와 **속도**-를 **block sparse signals**로 모델 • 동일한 source에 대하여, Rx의 위치에 따라 **Radar Cross Section 값이 다르다**라고 가정 [그림 7참조]	

위 실험 결과 환경

위치 단위 meter, 속도 단위 meter/sec

– 2개의 Tx와 2개의 Rx 위치는 다음과 같음

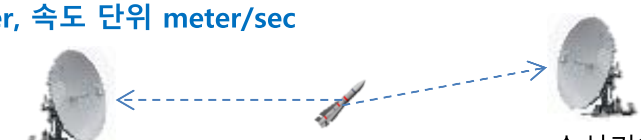
$$Tx_1 = [100,0], Tx_2 = [200,0], Rx_1 = [0,200], Rx_1 = [0,100]$$

– 1GHz의 carrier frequency, 3개의 pulse

– 3개의 표적들의 위치 및 속도는 다음과 같음

$$p_1 = [110,280], p_2 = [80,280], p_3 = [100,260]$$

$$v_1 = [120,100], v_2 = [110,110], v_3 = [130,130]$$



수신기 위치 Vs. RCS

동일한 표적일지라도 수신기의 위치에 따라 RCS가 다를 수 있다고 가정

Block Sparse signals은 Joint sparse signals을 concatenated한 것

참고 문헌

1. [Nehorai11] : Sandeep Gogineni, Arye Nehorai, "Target Estimation Using Sparse Modeling for Distributed MIMO Radar", IEEE Trans. On Signal Processing, 2010, 인용회수: 49회

Literature survey on compressed sensing radar (5/6)

<table>
<tr><th>출 처</th><th>내 용</th><th colspan="2">실 험 결 과</th></tr>
<tr>
<td>[Zhang12]</td>
<td>

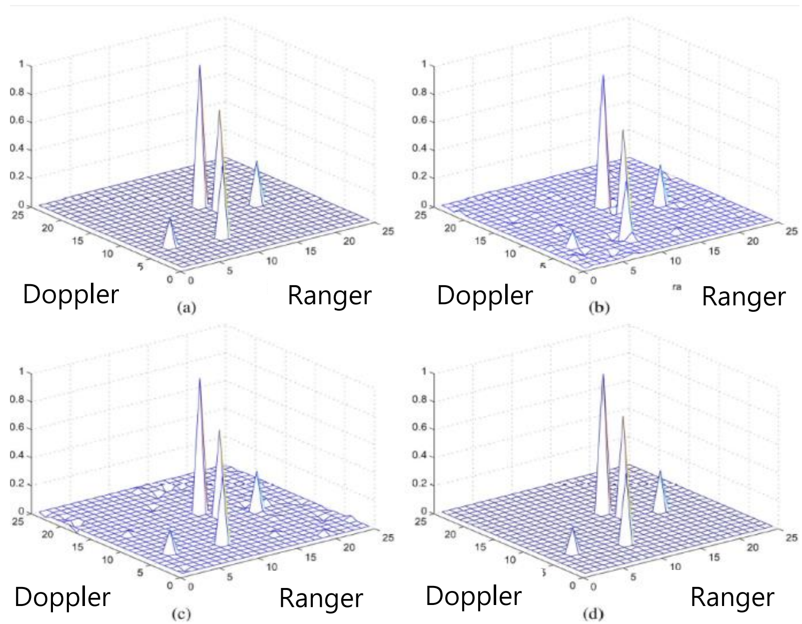
- **피드백 정보**를 이용하여 레이더의 전송 신호 파형과 **측정 행렬을 갱신**
- 갱신된 전송 신호 파형과 측정 행렬은 표적들 간의 상관도를 줄임으로써 성능을 향상시킴

</td>
<td>

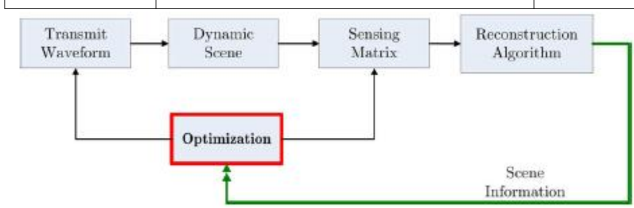
Range-Doppler 평면 상의 이미지 복원실험

(a) 실제 표적
(b) 랜덤 신호파형과 센싱 행렬 사용
(c) 피드백 없이 최적화한 신호파형과 센싱 행렬 사용
(d) 피드백 된 정보를 사용하여 최적화한 신호파형과 센싱 행렬 사용

</td>
<td>



</td>
</tr>
</table>



적응형 레이더 시스템

Scene Information

– 여기서 피드백 되는 정보는 복원 알고리즘에 의해 **복원된 Range-Doppler 정보**임

참고 문헌

1. [Zhang12] : J. Zhang, D. Zhu, G. Zhang, “Adaptive Compressed Sensing Radar Oriented Toward Cognitive Detection in Dynamic Sparse Target Scene”, IEEE Trans. on Signal Processing, 2012, 인용회수: 34 회

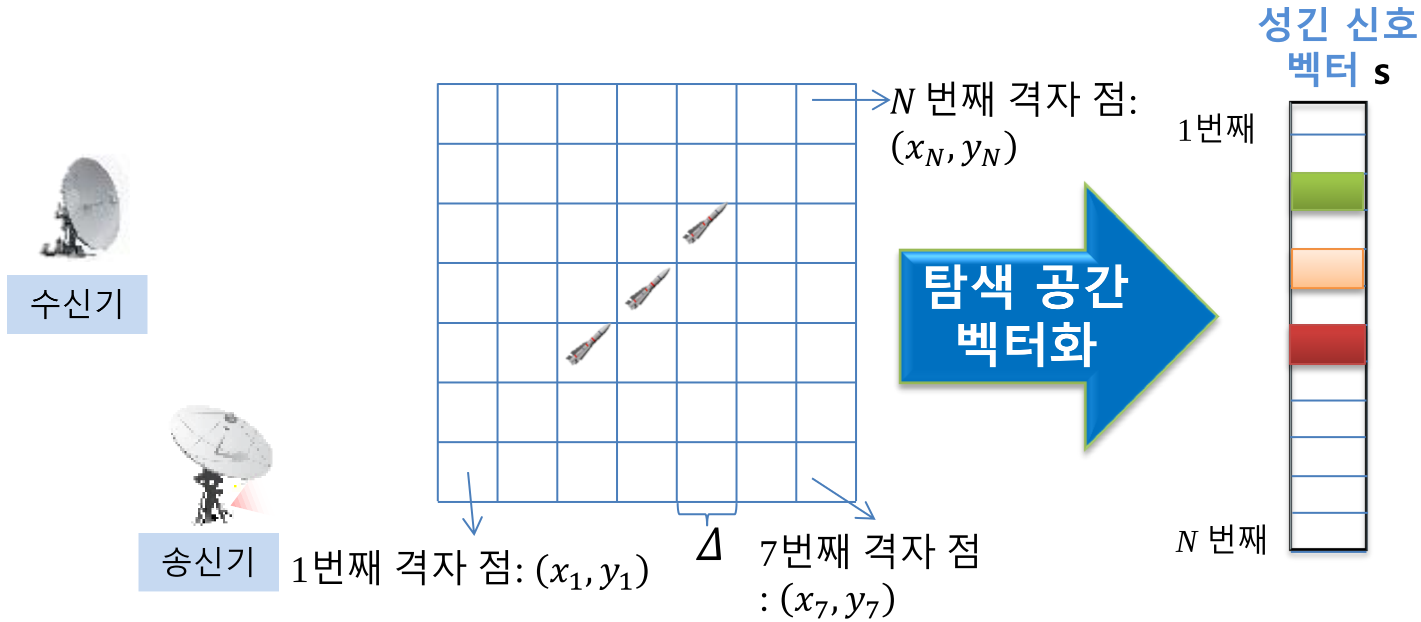
Literature survey on compressed sensing radar (6/6)

출 처	내 용	이 론 결 과
[Strohmer14]	• 압축센싱 MIMO radar 성능 분석 • 표적들의 위치,방위각 및 속도를 올바르게 복원하기 위한 이론적 툴을 제시 • 제시된 툴을 numerical simulation을 통해 올바른 것을 확인	• Tx의 개수, Rx의 개수, 방위각 해상도, 속도 해상도 등이 결정 되면, **탐지 가능한 표적의 최대 개수의 바운드를 이론적으로 제시** [해당 논문, 이론 1 참조] • Tx의 개수, Rx의 개수, 방위각 해상도, 거리 해상도, 속도 해상도 등이 결정 되면, **탐지 가능한 표적의 최대 개수의 바운드를 이론적으로 제시** [해당 논문, 이론 5 참조]

참고 문헌

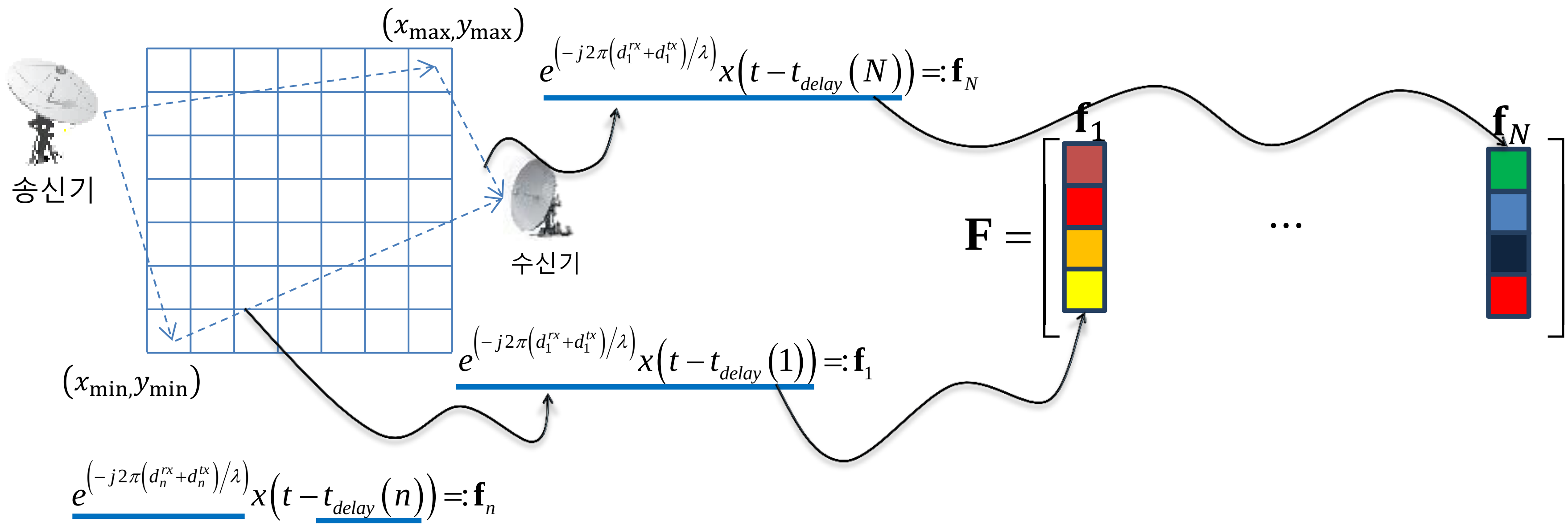
1. [Strohmer14] : Thomas Strohmer and Benjamin Friedlander, “Analysis of sparse MIMO radar”, Applied and Computational Harmonic Analysis, 2014, 인용회수: 8회

System model – sparse signal



- 색칠해진 부분은 자탄이 해당 탐색 공간 안에 존재 함을 의미
- 공란은 표적이 없음을 뜻함
- n번째 격자에 자탄이 있으면 벡터의 n번 째 원소 값은 0이 아닌 것으로 모델

System model – sensing matrix



송신기가 보낸 신호가 n번째 격자 공간을 지나서 수신기에 도착 했을 때의 위상 지연 값

송신기가 보낸 신호가 n번째 격자 공간을 지나서 수신기에 도착하는데 까지 걸린 시간

- 열 벡터 $\mathbf{f}_n$은 송/수신기와 n번째 격자 공간 사이의 물리적 반응(physical response) 관계
- 이 물리적 반응은 **wave propagation**에 의해 제시

Summary of compressed sensing radar

- Compressed sensing radar yields better resolutions that cannot be achieved by matched filter radar
  - It break out fundamental limits on the rage resolution of matched filter radar

$$\Delta r \geq \frac{c}{2B}, c \approx 3 \times 10^{8}$$

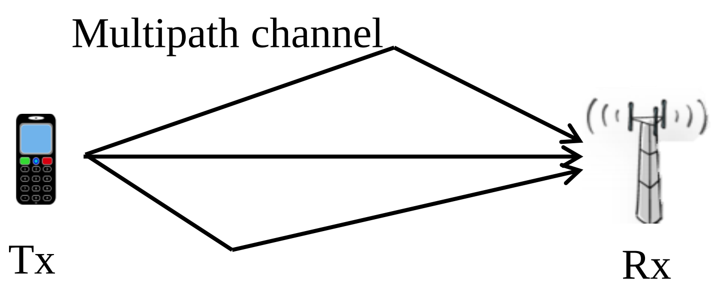
  - where B is the bandwidth

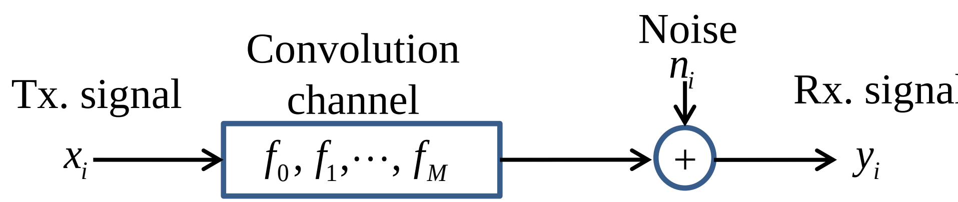
- A few papers analyzed the performance of compressed sensing radar
  - It gives bounds on the number of Tx and Rx to achieve a proper resolution…
    - [Strohmer14] : Thomas Strohmer and Benjamin Friedlander, "Analysis of sparse MIMO radar", Applied and Computational Harmonic Analysis, 2014,

Compressed Sensing Channel Estimation and Multi-user Detection

Applications in Communication (1)

- In the problem of channel estimation,
  - Tx transmits its signal vector $\mathbf{x}$ to Rx through a multipath channel.
  - The aim is to estimate an unknown channel impulse response based on the transmitted signal vector $\mathbf{x}$ and the received signal vector $\mathbf{y}$ at Rx.





$$y_i = \sum_{m=0}^{M} f_m x_{i-m} + n_i$$

- The received signal vector is represented by

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N-M} \end{bmatrix} = \underbrace{\begin{bmatrix} f_M & \cdots & f_0 & 0 & \cdots & 0 \\ 0 & f_M & & f_0 & & 0 \\ \vdots & & \ddots & & \ddots & \vdots \\ 0 & \cdots & 0 & f_M & \cdots & f_0 \end{bmatrix}}_{\triangleq \mathbf{F} \in \mathbb{C}^{(N-M)\times N}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}}_{\triangleq \mathbf{x} \in \mathbb{C}^{N\times 1}} + \mathbf{n}$$

Applications in Communication (2)

- The received signal vector $\mathbf{y}$ can be represented by

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N-M} \end{bmatrix} = \underbrace{\begin{bmatrix} x_{M+1} & x_M & \cdots & x_1 \\ x_{M+2} & x_{M+1} & \cdots & x_2 \\ \vdots & \vdots & & \vdots \\ x_N & x_{N-1} & \cdots & x_{N-M} \end{bmatrix}}_{\triangleq \mathbf{X} \in \mathbb{C}^{(N-M)\times(M+1)}} \underbrace{\begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_M \end{bmatrix}}_{\triangleq \mathbf{f} \in \mathbb{C}^{(M+1)\times 1}} + \mathbf{n}$$

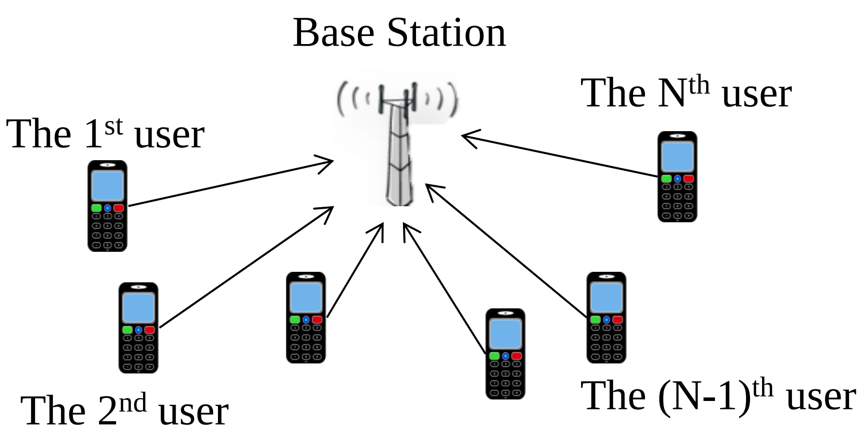
- An unknown sparse channel impulse response vector $\mathbf{f}$ is estimated by

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \|\mathbf{f}\|_1 \text{ subject to } \|\mathbf{X}\mathbf{f} - \mathbf{y}\|_2^2 \le \varepsilon$$

- The following papers also consider channel estimation problems
  - Ultra-wideband channel estimation based on compressed sensing [Wang07]
  - An optimization of the pilot placement for sparse channel estimation in OFDM system [Wu11]
  - To get the benefits such as small phases, low PAPR and low-rate sampling, a compressed sensing framework for OFDM channel estimation was proposed in [Yu14]

Applications in Communication (3)

- In the problem of multi-user detection,
  - Each user has its own signature vector.
  - The number of users sending its own signature to the base station is sparse.
  - The active users transmit their signatures with their messages to a base station.
  - The aim is to estimate the active users and their transmitted signals based on the received vector **y**.



- The received vector **y** at the base station is

$$\mathbf{y} = \sum_{i \in \mathcal{P}} x_i \mathbf{f}_i + \mathbf{n} = \mathbf{F}\mathbf{x} + \mathbf{n}$$

- $\mathcal{P}$ is an index set of active users.
- **n** is the additive white Gaussian noise vector.
- x_i is the message by the ith user.
- $\mathbf{f}_i$ is the signature vector by the i^{th} user.

Applications in Communication (4)

- The active users and their messages are identified and estimated by solving

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{y} - \mathbf{F}\mathbf{x}\|_2^2 \leq \varepsilon.$$

- Suppose that the *i*th coefficient of $\hat{\mathbf{x}}$ is non-zero. Then,
  - the *i*th user is active, and
  - the nonzero coefficients is the transmitted message by the *i*th user.

- The following papers also consider multi-user detection problems.
  - A simple on-off random multiuser detection was analyzed in [Goyal09]
  - The hybrid method that combines the OMP algorithm and the LMMSE (linear minimum mean square error) estimation was proposed in [Shim12]

- Tanaka [Tanaka13] introduces communication applications based on compressed sensing.

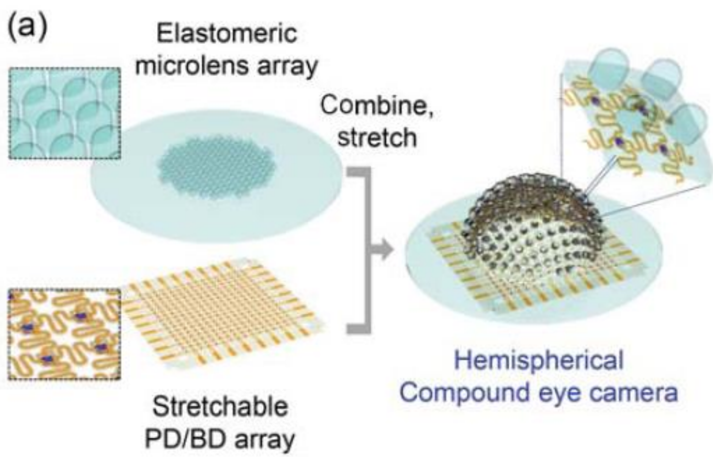
References

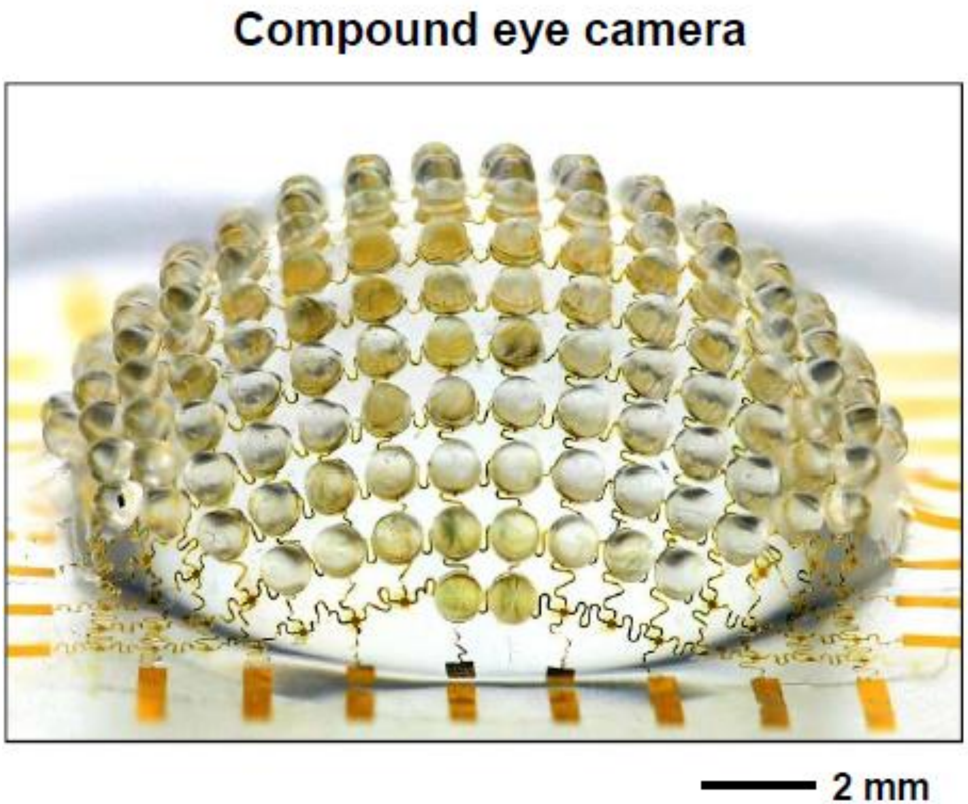
- [Tanka13] K. Hayashi, M. Nagahara and T. Tanaka, “A user’s guide to compressed sensing for communications systems”, IEICE Trans. Commun., vol. E96 – B, no. 3, pp. 685 – 712, Mar. 2013.
- [Wang07] J. L. Paredes, G. R. Arce, and Z. Wang, “Ultra-wideband compressed sensing: Channel estimation” IEEE J. Sel. Top. Signal Process., vol. 1, no. 3, pp. 383 – 395, Oct. 20007
- [Yu11] C. Qi and L. Wu, “Optimized pilot placement for sparse channel estimation in OFDM Systems”, IEEE Signal Process. Lett., vol. 18, no. 12, pp. 749 – 752, Dec. 2011.
- [Yu14] N. Y. Yu and L. Gan, “Convolutional compressed sensing using decimated Sidelnikov sequences,” IEEE Signal Processing Letters, vol. 21, no. 5, pp. 591 – 594, May 2014.
- [Goyal09] A. K. Fletcher, S. Rangan, V. K. Goyal, “A sparsity detection framework for on-off random access channels,” IEEE International Symposium on Information Theory, 2009, Seoul.
- [Shim12] B. Shim and B. Song, “Multiuser detection via compressed sensing,” IEEE Communications Letters, vol. 16, no. 7, pp. 972 – 974, Jul, 2012

COMPutational Compound EYE (COMP-EYE) imaging system

Hemispherical Apposition Compound Eyes

- Implemented by stretchable microlens array and photodiodes
- Limitation: 180 pixels (16x16 photo diodes)

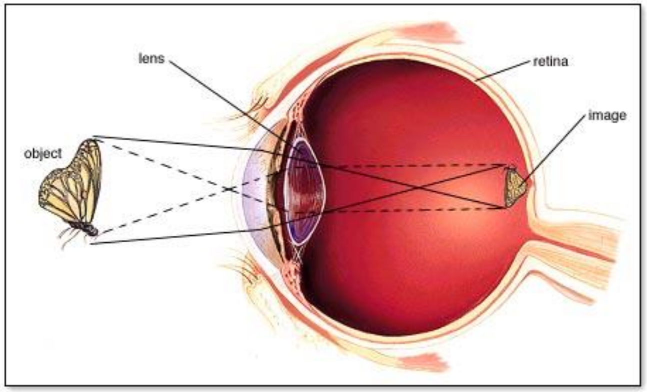




Computational Compound EYE imaging system

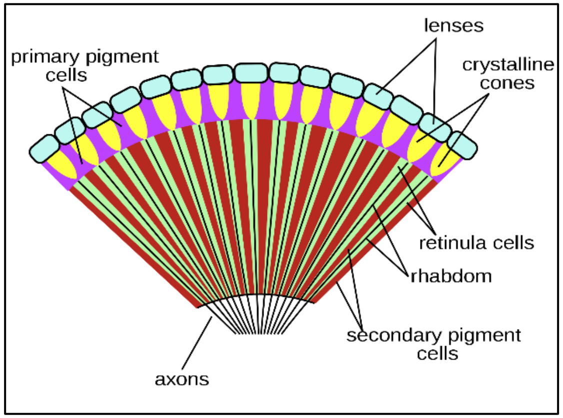
- Eyes in nature
  - Camera-type eye vs. Compound eye

Camera-type Eye



- Single lens system
- High resolution
- Pattern recognition

Compound Eye

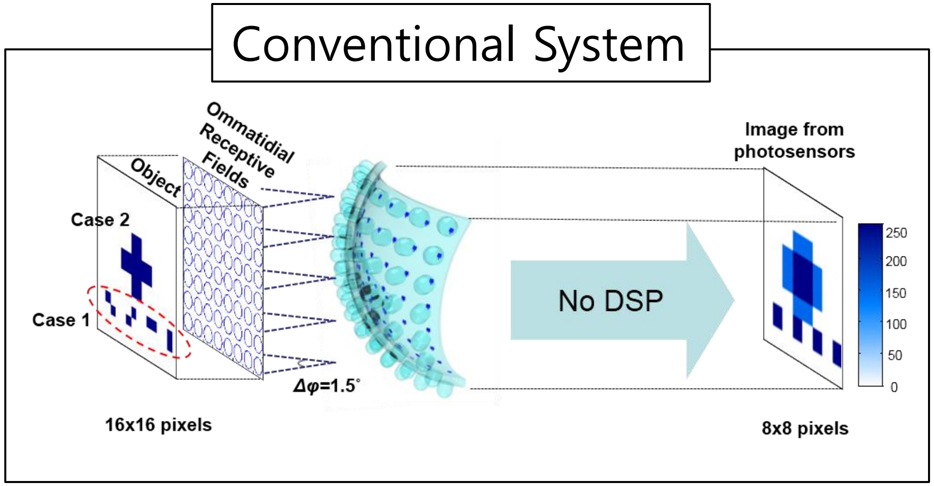


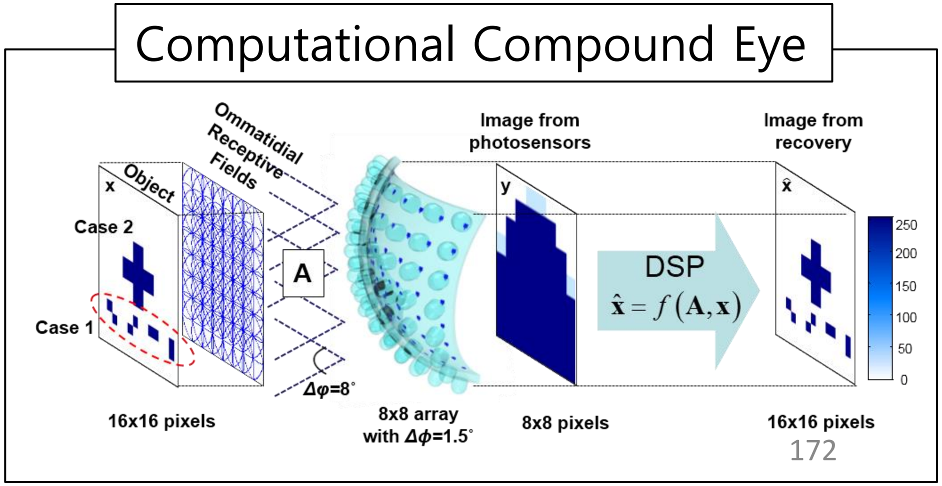
- Multi-lens system
- Wide field of view (FOV), Infinite depth of field (DOF)
- Motion detection

- Due to diffraction limit and low density of photoreceptors, the resolution of compound eyes is limited.

Computational Compound EYE imaging system

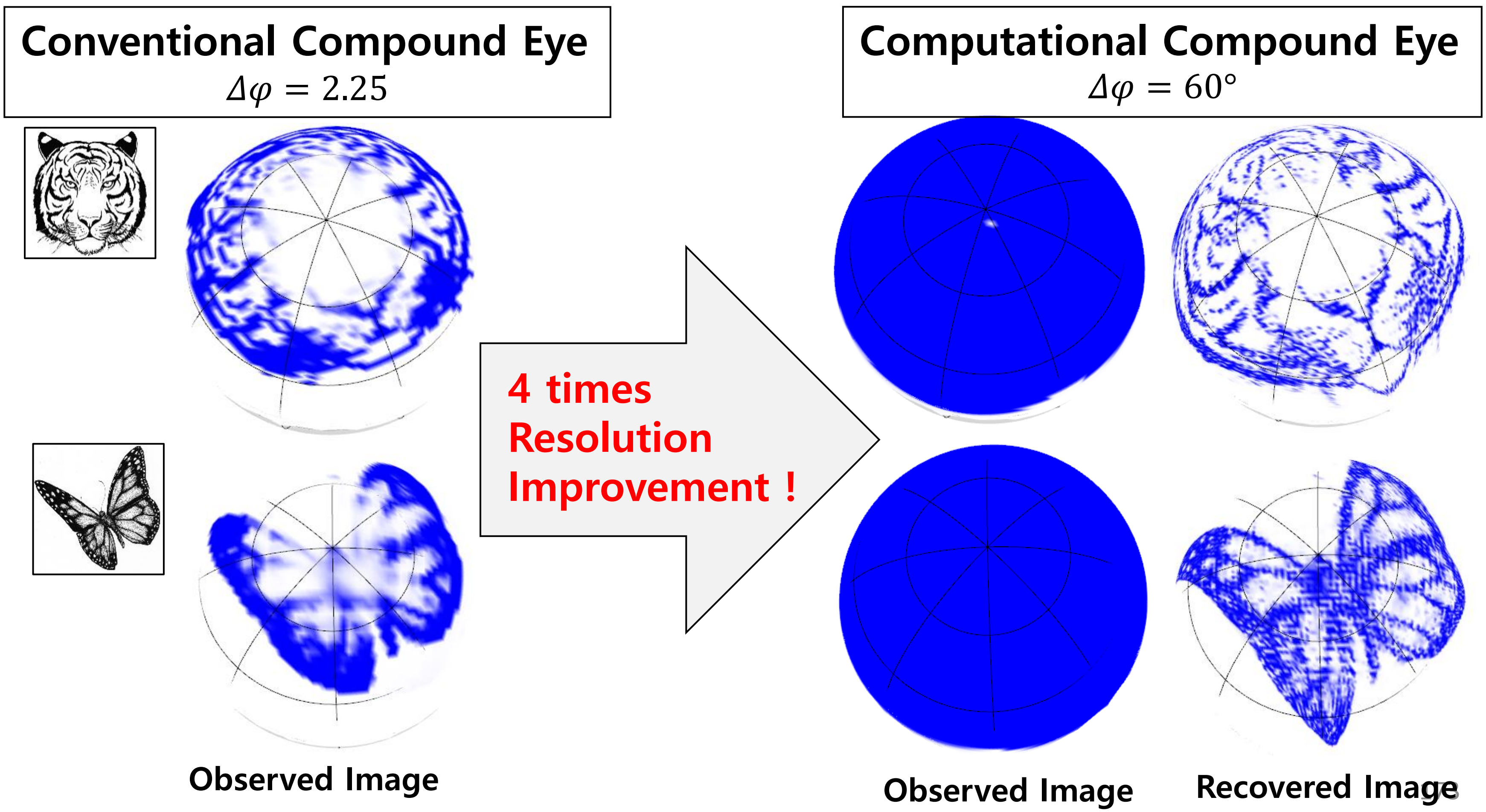
- COMP-EYE
  - We aim to improve the resolution of the compound eye imaging system by designing larger acceptance angles of ommatidia and using a digital signal processing (DSP) technique
  - Larger acceptance angles enable each ommatidium to observe multiple pieces of information all at once.
  - Each piece of information is observed multiple times by multiple ommatida each with different perspectives.
  - By exploiting this, the DSP technique recovers the object image with high resolution.





Computational Compound EYE imaging system

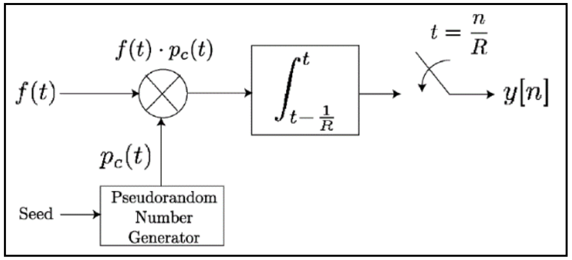
- Simulation results
  - M=80 x 80 ommatidia, N= 160 x 160 pixels



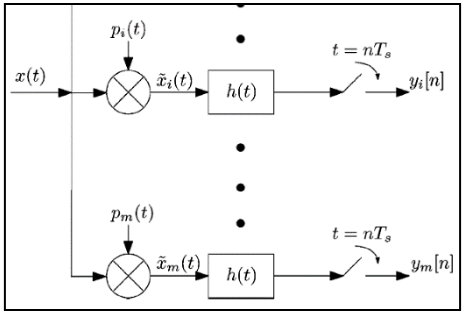
Sub-Nyquist Sampling/Xampling

Sub-Nyquist Sampling

- Compressed Sensing 이론을 적용하여 광대역 신호를 Sub-Nyquist 샘플링 할 수 있음
- 대표적인 CS based Sub-Nyquist 샘플링 시스템은, Random demodulator[Tropp10]와 Modulated Wideband Converter [Eldar 10]가 있음



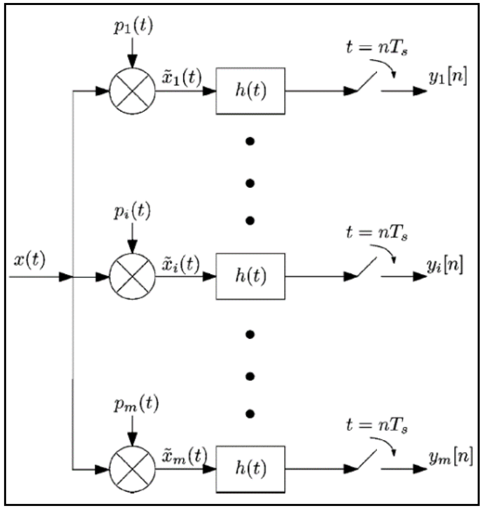
Random Demodulator

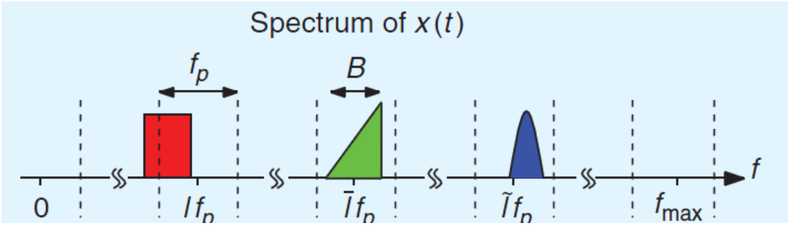


Modulated Wideband Converter

- 두 시스템 모두 아날로그 광대역 신호를 고속 Pseudo Random Binary Sequence(PRBS)와 혼합(Mix) 압축(Integrator or LPF) 한 후 저속 샘플링 함
- CS복원 알고리즘을 사용하여 압축 저속 샘플로 부터 Original 샘플을 복원
- 두 시스템의 핵심적인 차이는 구조적으로는 PRBS의 주기성에 있으며, 분석적으로는 CS모델에서 신호의 희소화 방식에 있음
- MWC에서 발생 할 수 있는 문제는 CS recovery를 보장하는 고속 PRBS의 디자인과 제어에 관한 문제, 그리고 광대역 신호간의 비선형적 혼합에 따른 시스템의 non-ideal 동작에 관한 문제 등이 있음

Modulated Wideband Converter(MWC) [Eldar10]





- 대상 신호: Blind & Multi-band sparse signal
  - Multi-band sparse: 소수의 협대역 spectral band가 넓은 주파수 영역에서 희소하게 산재함
  - Blindness: 각 협대역 스펙트럼들의 carrier frequency가 알려져 있지 않음
  - 입력 신호는 대역 제한되어있음: $F = [-f_{NYQ}/2, f_{NYQ}/2]$
  - 이때, $f_{NYQ} = 2 f_{\max}$
  - 각 협대역 spectral band의 대역폭은 B[Hz]보다 작음

Modulated Wideband Converter(MWC) [Eldar10]

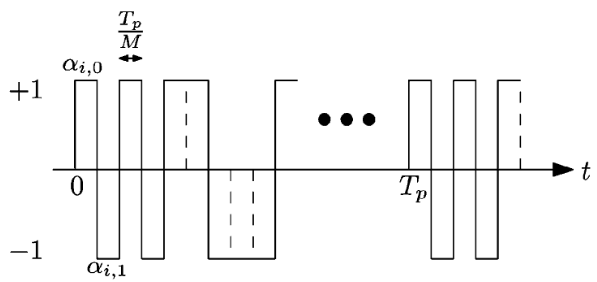
- Motivation: Random Demodulator [Kirolos06]

  – 각 밴드의 주파수 위치를 안다면 IF frequency로 각 밴드를 down convert 한 후 Filtering and sampling을 수행 할 수 있지만, 이 경우에는 각 밴드의 주파수 위치를 모르기 때문에, 일정 주기를 갖는 신호를 시간영역에서 곱해줌으로써 협대역 spectral band들의 aliasing을 유도함

  – 주기가 T_p 인 신호 $p_i(t)$ 는 다음과 같은 Fourier expansion으로 표현 가능

$$p_i(t) = \sum_{l=-\infty}^{\infty} c_{il} e^{j\frac{2\pi}{T_p}lt}$$

  – 이때, 주기 신호 $p_i(t)$를 PRBS로 선택 할 경우 Fourier coefficient는 간단하게 계산됨



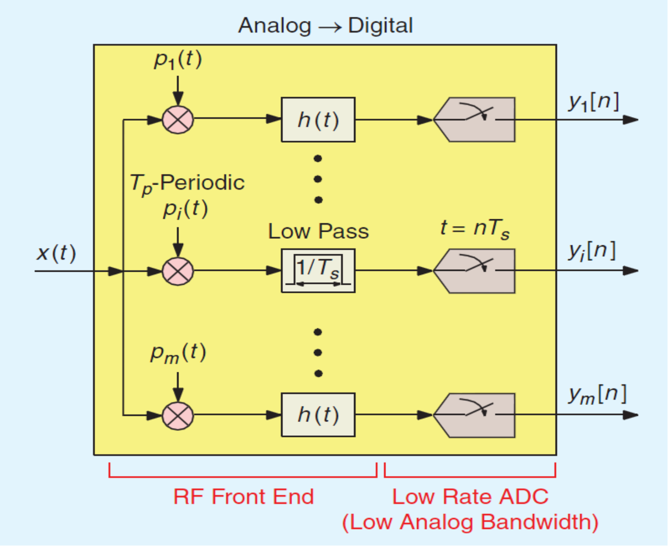
$$c_{i,l} \triangleq \frac{1}{T_p}\int_0^{T_p} p_i(t) e^{-j2\pi l f_p t} dt$$

$$= \frac{1}{T_p}\int_0^{\frac{T_p}{M}} \sum_{k=0}^{M-1} a_{ik} e^{-j2\pi l f_p \left(t + k\frac{T_p}{M}\right)} dt$$

$$= \frac{1}{T_p}\sum_{k=0}^{M-1} \underbrace{a_{ik}}_{a_{i,l}} \underbrace{e^{-j2\pi \frac{lk}{M}}}_{F_{i,l}} \underbrace{\int_0^{\frac{T_p}{M}} e^{-j2\pi l f_p t} dt}_{d_l}$$

Modulated Wideband Converter(MWC) [Eldar10]

❖ Sampling scheme



- 시스템은 m개의 channel로 구성되어 있음
- PRBS의 주기와 Lowpass filter의 cut-off frequecy는 다음과 같은 관계를 고려하여 결정되어야 함
- $f_s \geq f_p \geq B$
- 여기서 $f_s = 1/T_s$, $f_p = 1/T_p$
- 먼저, 혼합단을 통과한 신호 $s_i(t)$는,

$$s_i(t) = x(t)p_i(t) = \sum_{l=-\infty}^{\infty} c_{i,l} \exp(j2\pi l f_p t) x(t)$$

- 혼합신호의 주파수 표현은, $S_i(f) = \sum_{l=-\infty}^{\infty} c_{il} X(f - lf_p)$
- Lowpass filter를 통과 한 신호 $y_i(t)$의 퓨리에 변환은,

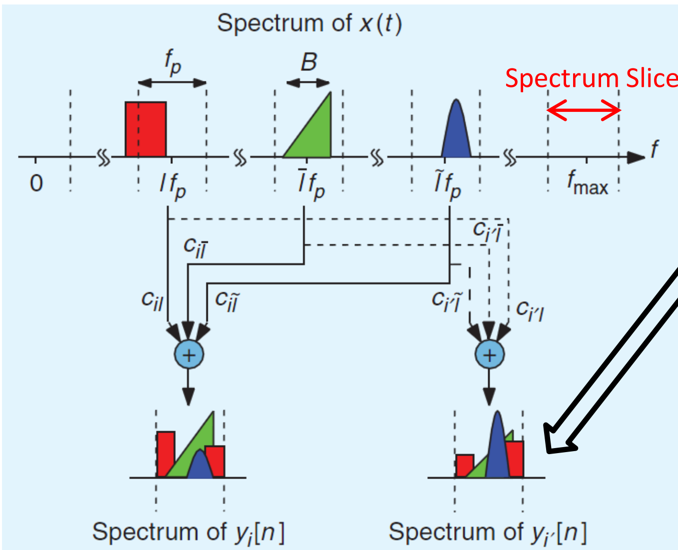
$$Y_i(f) = \begin{cases} \sum_{l=-\infty}^{\infty} c_{il} X(f - lf_p) & f \in [-f_s/2, f_s/2] \\ 0 & f \notin [-f_s/2, f_s/2] \end{cases}$$

- 이때, 입력 신호 $X(f)$의 대역은 $\pm f_{max}$로 제한되어 있으므로 $Y_i(f)$의 무한 항은 유한 항으로 감소됨. 즉,

$$Y_i(f) = \sum_{l=-L_0}^{L_0} c_{il} X(f - lf_p), \; f \in [-f_s/2, f_s/2]$$

Modulated Wideband Converter(MWC) [Eldar10]

❖ Linear CS model



$$Y_i(f)=\sum_{l=-L_0}^{L_0} c_{il}X(f-lf_p),\ f\in F_s$$

$$\Downarrow$$

$$\begin{bmatrix} Y_1(e^{j2\pi f}) \\ Y_2(e^{j2\pi f}) \\ \vdots \\ Y_m(e^{j2\pi f}) \end{bmatrix} = \overbrace{\begin{bmatrix} c_{1,L_0} & \cdots & c_{1,0} & \cdots & c_{1,-L_0} \\ c_{2,L_0} & \cdots & c_{2,0} & \cdots & c_{2,-L_0} \\ \vdots & & \vdots & & \vdots \\ c_{m,L_0} & \cdots & c_{m,0} & \cdots & c_{m,-L_0} \end{bmatrix}}^{L \text{ colums}} \begin{bmatrix} \mathrm{LPF}[X(f-L_0)] \\ \vdots \\ \mathrm{LPF}[X(f)] \\ \vdots \\ \mathrm{LPF}[X(f+L_0)] \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} y_1[n] \\ y_2[n] \\ \vdots \\ y_m[n] \end{bmatrix} = \overbrace{\begin{bmatrix} c_{1,L_0} & \cdots & c_{1,0} & \cdots & c_{1,-L_0} \\ c_{2,L_0} & \cdots & c_{2,0} & \cdots & c_{2,-L_0} \\ \vdots & & \vdots & & \vdots \\ c_{m,L_0} & \cdots & c_{m,0} & \cdots & c_{m,-L_0} \end{bmatrix}}^{L \text{ colums}} \begin{bmatrix} \mathrm{LPF}\left[x(t)\exp(j2\pi L_0 f_p t)\right]_{t=nT_s} \\ \vdots \\ \mathrm{LPF}\left[x(t)\right]_{t=nT_s} \\ \vdots \\ \mathrm{LPF}\left[x(t)\exp(-j2\pi L_0 f_p t)\right]_{t=nT_s} \end{bmatrix}$$

$$\Leftrightarrow \mathbf{Y}=\mathbf{CZ}$$

– Sampling time T_s마다 matrix $\mathbf{Y}$와 $\mathbf{Z}$ 의 column이 누적됨 -> MMV Problem
– Matrix $\mathbf{Z}$의 각 row의 성분은 $X(f)$의 각 스펙트럼 슬라이스별 샘플에 대응됨
– 즉, $\mathbf{Z}$의 support set은 $X(f)$ 를 구성하는 협대역 spectral bands의 위치이며, $\mathbf{Z}$의 성분을 획득 한 후 $x(t)$ 를 복원 가능

References

- [Eldar10] Moshe Mishali, Yonina C. Eldar, “From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals,” IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, VOL. 4, NO. 2, APRIL 2010
- [Tropp10] Joel A. Tropp, Jason N. Laska, Marco F. Duarte, Justin K. Romberg, Richard G. Baraniuk, “Beyond Nyquist: Efficient Sampling of Sparse Bandlimited Signals,” IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 56, NO. 1, JANUARY 2010
- [Kirolos06] S. Kirolos, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. G. Baraniuk, “Analog to information conversion via random demodulation,” in Proc. 2006 IEEE Dallas/CAS workshop on Design, Applications, Integration and Software, Dallas, TX, Oct. 2006 pp. 71-74

Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems [Heath14]

- Millimeter wave (mmWave) cellular systems ~ gigabit-per-second data rates ~ the large bandwidth available at mmWave frequencies.

- To realize sufficient link margin, directional beamforming with large antenna arrays is used at both the transmitter and receiver.

- Due to the high cost and power consumption of gigasample mixed-signal devices, mmWave precoding will likely be divided among the analog and digital domains.

- The large number of antennas and the presence of analog beamforming requires the development of mmWave-specific channel estimation and precoding algorithms.

- This paper develops an adaptive algorithm to estimate the mmWave channel parameters that exploits the poor scattering nature of the channel.

- To enable the efficient operation of this algorithm, a novel hierarchical multi-resolution codebook is designed to construct training beamforming vectors with different beamwidths.

Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems [Heath14]

- The adaptive channel estimation algorithm is then extended to the multi-path case relying on the sparse nature of the channel.

- Using the estimated channel, this paper proposes a new hybrid analog/digital precoding algorithm that overcomes the hardware constraints on the analog-only beamforming, and approaches the performance of digital solutions.

- Simulation results show that the proposed low-complexity channel estimation algorithm achieves comparable precoding gains compared to exhaustive channel training algorithms.

- The results also illustrate that the proposed channel estimation and precoding algorithms can approach the coverage probability achieved by perfect channel knowledge even in the presence of interference.

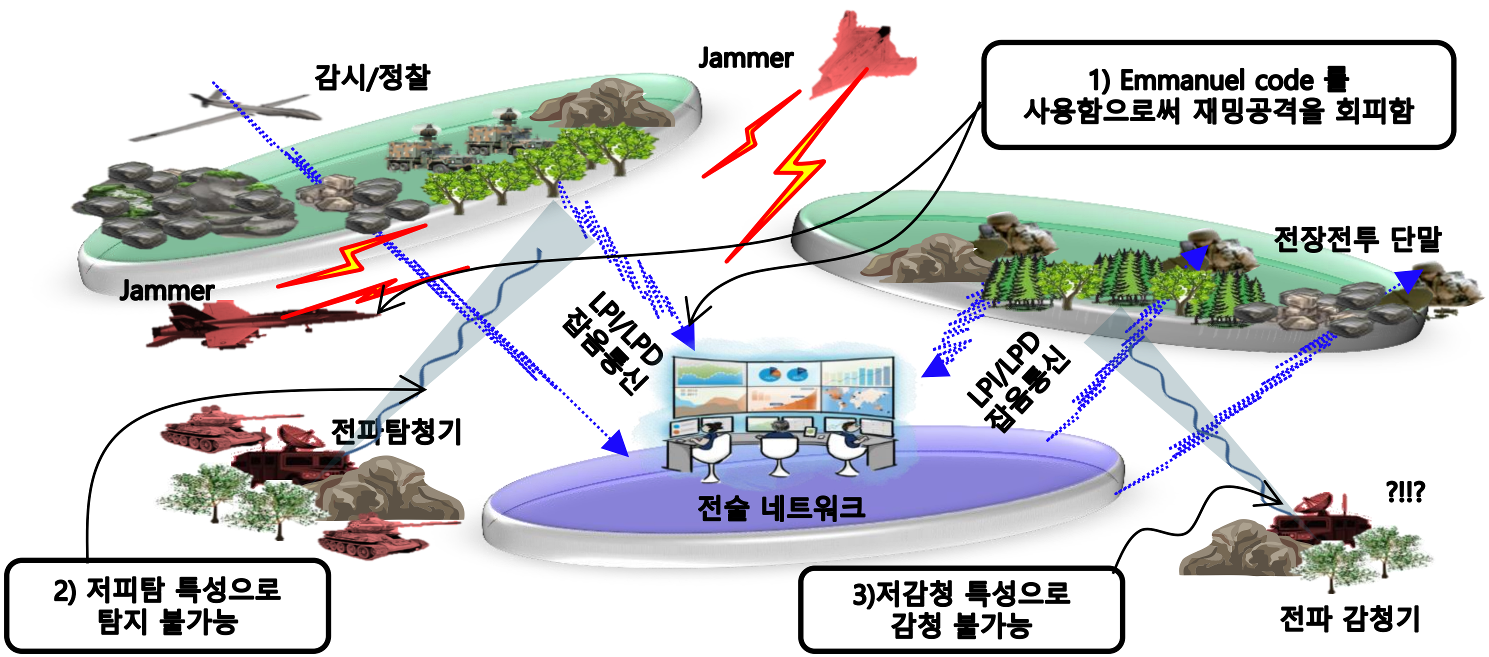
References

- [Heath14] A. Alkhateeb, O. El Ayach, G. Leus, R. W. Heath, "**Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems**" *IEEE J. Selected Topics in Sig. Proc.*, 8(5), pp. 831-846, Oct. 2014.
- [Heath15] A. Alkhateeb, G. Leus, R. W. Heath, "**Compressed Sensing based Multi-User Millimeter Wave Systems: How Many Measurements are Needed?**" *IEEE ICASSP 2015*, Apr. 2015.
- [Choi15] J. Choi, "**Beam Selection in mm-Wave multiuser MIMO Systems using Compressive Sensing**", *IEEE Trans. on Comm.*, 63(8), pp. 2936-2947, Aug. 2015.

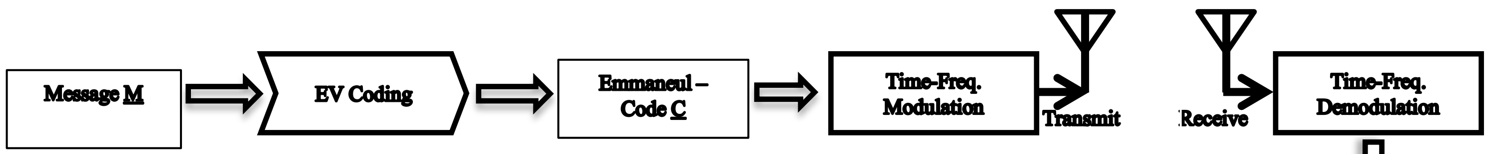
Noise Communication

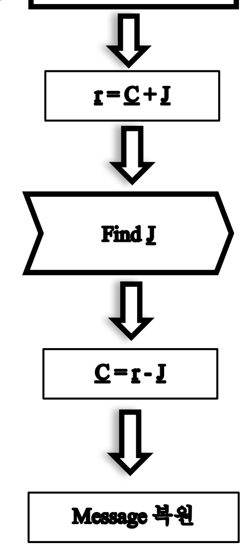
❖ 다양한 재밍 공격을 한정된 자원(소비전력, 채널 대역폭 등)을 사용하여 방어하는 적응형 다중모드 항재밍 통신 기술 개발 및 아군의 통신보안(Communication Security)을 위한 저피탐/저감청 통신 기술 개발

- 한정된 자원(소비전력, 채널 대역폭 등)을 최대한 활용하여 적군의 다양한 재밍 공격에 방어 할 수 있는 적응형 다중모드 항재밍 통신 기술 필요
- 적군의 전파 탐청기를 회피 할 수 있는 저피탐 통신 기술 필요
- 탐지된 통신을 적군이 감청할 수 없는 저감청 통신 기술 필요



Noise Communication



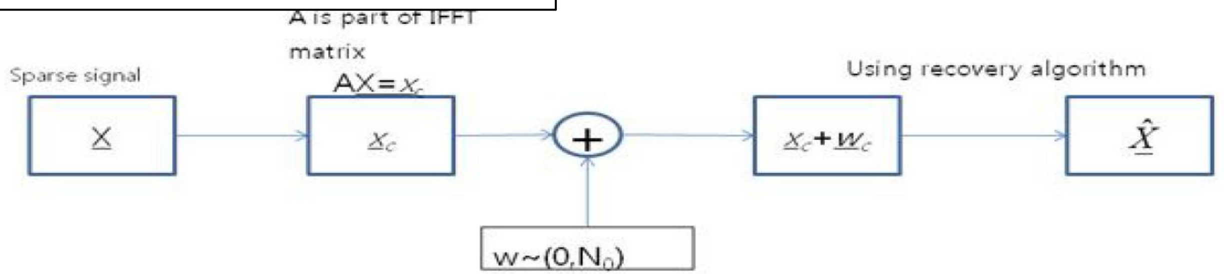


- EV 코드는 수신 단에서 받은 EV 코드 워드 벡터 C에 포함된 에러 벡터 e가 희소하게 표현이 될때 사용 되는 채널 코딩의 한가지 방법임.

- EV 코딩 행렬 G를 Gaussian들로 구성 함.
- 생성된 EV 코드 워드 벡터 c는 Gaussian 벡터로 표현 됨.
- Emmanuel 코드 워드 벡터 c는 시간/주파수 영역에서 항상 Gaussian 벡터로 표현 됨.
- 주파수 전 대역에 넓게 확산된 EV 코드 워드 벡터 c와 무선 채널에 존재하는 Gaussian 잡음은 서로 혼합 됨. 적군은 확산된 EV 코드 워드 벡터 c를 검출하기 어려움.

Compressive sensing via OFDM system

CS based OFDM system diagram



❖ 목표: OFDM 시스템에서 PAPR의 감소

❖ 기존의 문제점: 기존 OFDM 시스템은 다중 반송파를 이용하기 때문에 높은 PAPR로 갖을 수 있고, 이로인해 amplifier의 선형 동작 범위를 벗어나고 시스템의 성능이 저하

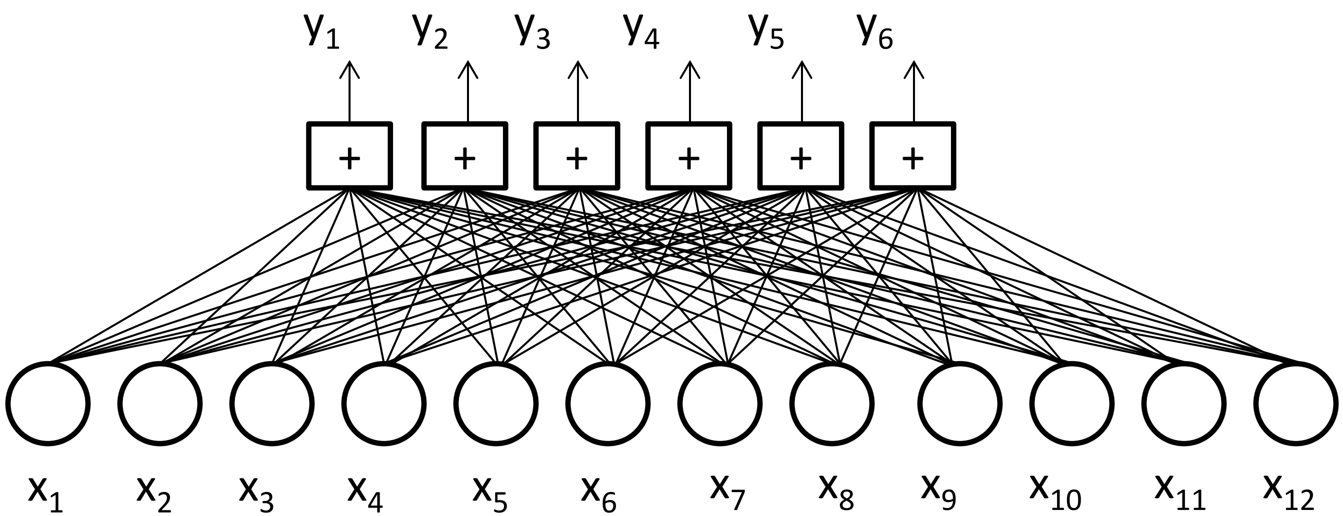
❖ 방법: 희소 신호와 스프레딩 기술을 OFDM 에 적용하여 PAPR 감소 효과 획득.
- 시간 도메인 신호는 압축 센싱된 희소 신호로서, PAPR 감소에 대한 노이즈의 영향을 적게 받을 수 있다.

❖ 결과 및 기대효과
- 일반적인 OFDM 방식 보다 뛰어난 BER 감소 성능
- 릴레이 통신, 네트워크 코드가 사용되거나, 아날로그 네트워크 시스템

Overall Summary of This Talk

- CS Theory can be recast as a parity check problem in coding theory.
  - LDPC codes
  - Channel coding theorem
- The CS narrative has shown useful in many applications including medical imaging, ADCs, spectrum sensing, super resolution areas.
- 혁신적이고 새로운 application들이 속속 등장 하고 있습니다.
  - Lensfree camera, lensfree spectrometers, 고 해상도 휴대용 imaging 기기, multi-target tracking radars, single-pixel camera, 초 고속 MRI 등등
- Application 뿐 아니라 정보이론 및 신호처리 연구도 절실히 필요한 상황입니다.
- IoT, Big Data: 널려있는 부정확한 센서자원은 많은 Data를 생산해 내고 있는데, 그것들을 통해 세상을 보고자 한다. 어떻게 고해상도, 유용한 정보, features, 혹은 주요 정보를 추출해 낼 것인가?

Have you seen this before?



- We can draw a graph for y = Fx, shown above.
- Yes!!!
- Channel codes—Syndrome decoding!

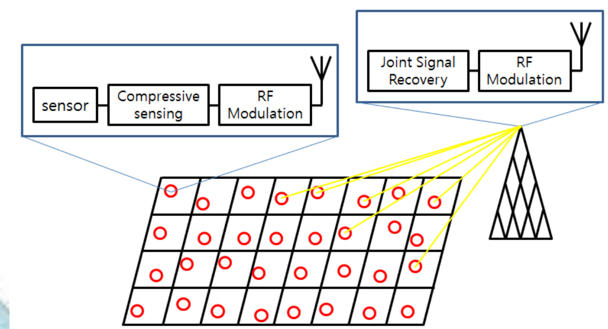
- Two applications:
  - Compressed measurement **y** and recovery of uncompressed data **x**
  - Super-resolution of hidden **x** from limited measurement **y**

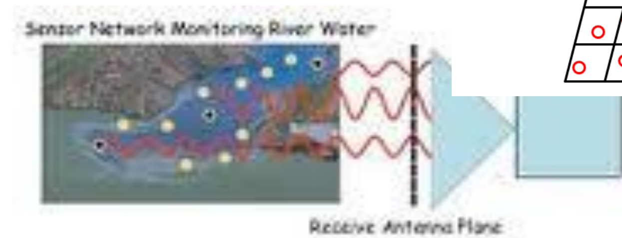
Many Applications

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

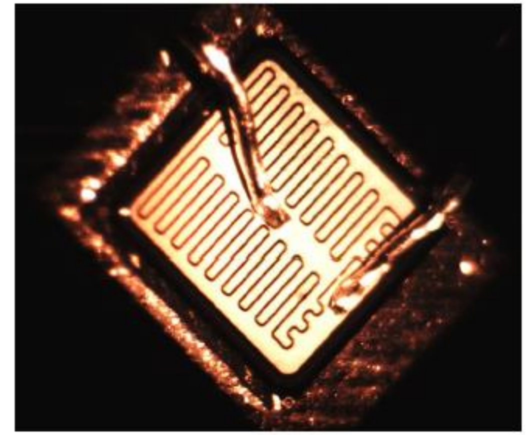
- Compressive Imaging
- Medical Imaging
- Analog-to-Information Conversion
- Ultra-wideband radios
- Compressive Spectrum Sensing
- Classification using Sparse Rep
- Super Resolution imaging



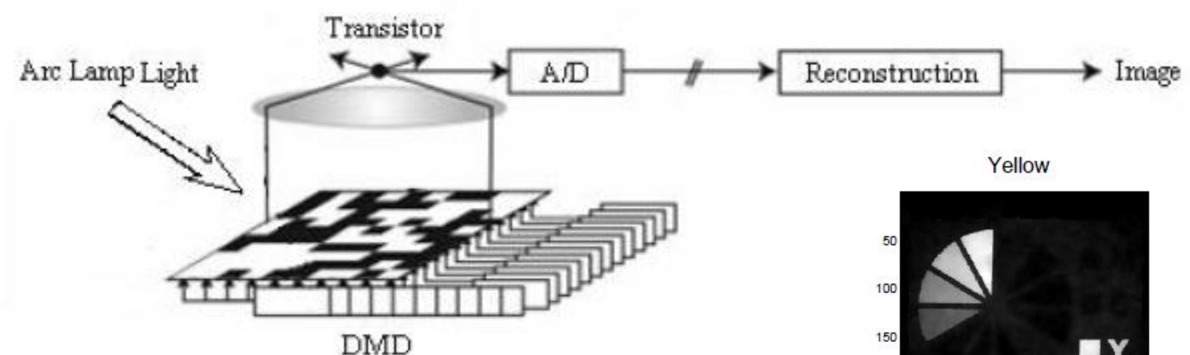


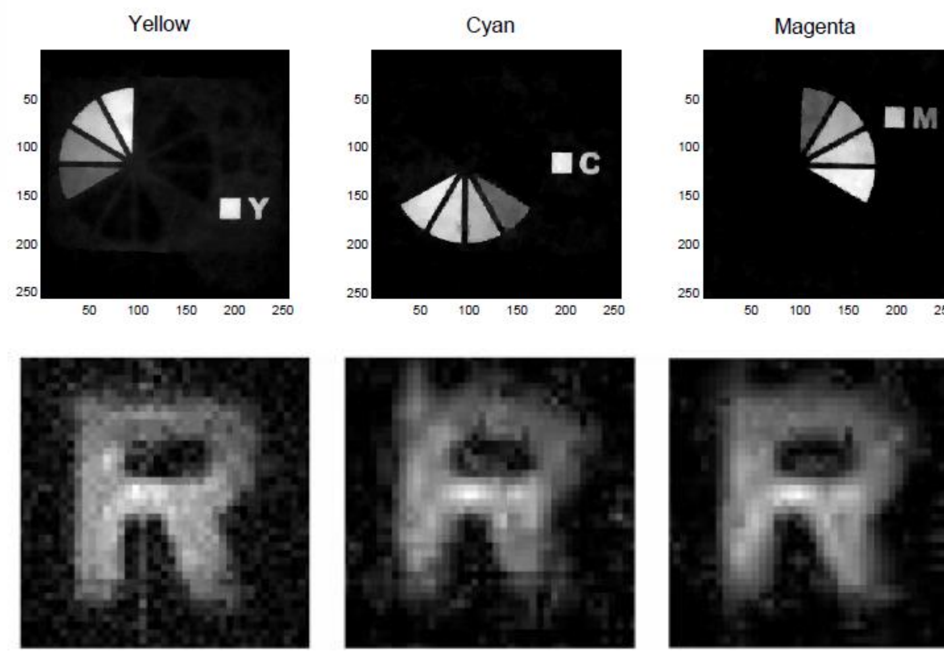



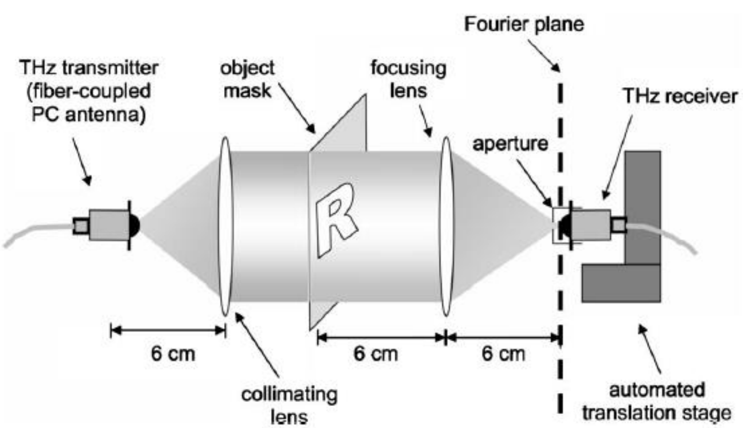
Many Exciting New Applications



- Circuit Failure Analysis[Kevin09]
- Compressive Hyperspectral Imaging [Yin12]
- Terahertz Imaging [Mittleman08]







Questions & Answers

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

References

- [Delgado05] S. F. Cotter, B. D. Rao and K. Kreutz-Delgado, “Sparse Solutions to Linear Inverse Problems With Multiple Measurement Vectors,” IEEE Trans. on Signal Processing, 2005.
- [Huo05] J. Chen and X. Huo, “Theoretical Results on Sparse Representations of Multiple-Measurement Vectors,” IEEE Trans. on Signal Processing, 2006.
- [Baraniuk05] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, “Distributed Compressed Sensing of Jointly Sparse Signals,” Asilomar Conference on Signals, Systems & Computers, 2005.
- [Nehorai10] G. Tang and A. Nehorai, “Performance Analysis for Sparse Support Recovery,” IEEE Trans. on Information Theory, 2010.
- [Davies12] J. D. Blanchard and M. E. Davies, “Recovery Guarantees for Rank Aware Pursuits,” IEEE Signal Processing Letters, 2012.
- [Rao13] Y. Jin, and B. D. Rao, “Support Recovery of Sparse Signals in the Presence of Multiple Measurement Vectors”, IEEE Transaction on Information Theory, 2013
- [Lee12] S. Park and H. Lee, “Number of compressed measurements needed for noisy distributed compressed sensing,” IEEE International Symposium on Information Theory Proceedings, 2012.
- [Lee13] S. J. Park, and H. Lee, “A study for establishing a sufficient condition for successful joint support set reconstruction”, ITC-CSCC 2013.

References

- [Lee13s] J. Oliver, WoongBi Lee, and Heung-No Lee, “Filters with random transmittance for improving resolution in filter-array-based spectrometers,” Optics Express, Vol. 21, No. 4, pp. 3969–3989, Feb. 2013.
- [Lee12s] J. Oliver, WoongBi Lee, SangJun Park , and Heung-No Lee, “Improving resolution of miniature spectrometers by exploiting sparse nature of signals,” Optics Express, vol. 20, no. 3, pp. 2613-2625, Jan. 2012.
- [Vetterli07] Ali Hormati and Martin Vetterli, “Annihilation Filter-based decoding in compressed sensing framework, proceeding of SPIE, CA, 2007
- [Vetterli02] Martin Vetterli, Pina Marziliano, and Thierry Blu, “Sampling signals with finite rate of innovation,” IEEE Trans. On Sig. Proc. Vol. 50, no.6, June 2002
- [Prony1795] Gaspard Riche de Prony, "Cours d'Analyse appliquée à la mécanique," Journal de l'École Polytechnique, an 111, tome 1, cahier 1, 92-119, 1795.
- [Candes05] E.J. Candes and T. Tao, “Decoding by Linaer Programming,”, vol. 51, no.12, 2005
- [Lee11a] J. Oliver and Heung-No Lee, “A Realistic Distributed Compressed Sensing Framework for multiple Wireless Sensor Networks”, SPARS 2011, Scotland, June 27-30, 2011.
- [Lee13b] Jin-Taek Seong and Heung-No Lee*, “Necessary and Sufficient Conditions for Recovery Performance of Sparse Signals over Finite Fields,” Accepted for IEEE Communications Letters, Extended version.

References

- [Liu13] Xinji Liu and Shutao Xia, “Constructions of Quasi-Cyclic Measurement Matrices Based on Array Codes,” ISIT 2013.
- [Kevin09] Ting Sun, Gary L. Woods, Chengbo Li, Yin Zhang, Kevin Kelly, “OBIC Measurements without Lasers or Raster-Scanning Based on Compressive Sensing”, The 35th International Symposium for Testing and Failure Analysis (ISTFA), Nov. 2009.
- [Yin12] Chengbo Li, Ting Sun, Kevin Kelly, Yin Zhang, “A Compressive Sensing and Unmixing Scheme for Hyperspectral Data Processing”, IEEE Trans. On Image Processing, Mar. 2012.
- [Mittleman08] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, ”A single-pixel terahertz imaging system based on compressive sensing”, Applied Physics Letters, 2008.
- [Hyder09] M. M. hyder and K. Mahata, “A robust Algorithm for Joint Sparse Recovery”, IEEE Signal Processing Letter, 2009
- [Nehorai10] G. Tang and A. Nehorai, “Performance Analysis for Sparse Support Recovery,” IEEE Trans. on Information Theory, 2010.
- [Rao13] Y. Jin, and B. D. Rao, “Support Recovery of Sparse Signals in the Presence of Multiple Measurement Vectors”, IEEE Transaction on Information Theory, 2013

References

- [Lee12] S. Park and H. Lee, “Number of compressed measurements needed for noisy distributed compressed sensing,” IEEE International Symposium on Information Theory Proceedings, 2012.
- [Lee13] S. J. Park, and H. Lee, “A study for establishing a sufficient condition for successful joint support set reconstruction”, ITC-CSCC, 2013.
- [Duarte13] M. F. Duarte, M. B. Wakin, D. Baron, S. Sarvotham, and R. G. Baraniuk, “Measurement Bounds for Sparse Signal Ensembles via Graphical Models”, IEEE Transaction on Information Theory, 2013.
- [Davenport10] Mark A. Davenport and Michael B. Wakin, “Analysis of Orthogonal Matching Pursuit Using the Restricted Isometry Property,” IEEE Trans. On Info. Theory, vol. 56, no. 9, September 2010.\

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]

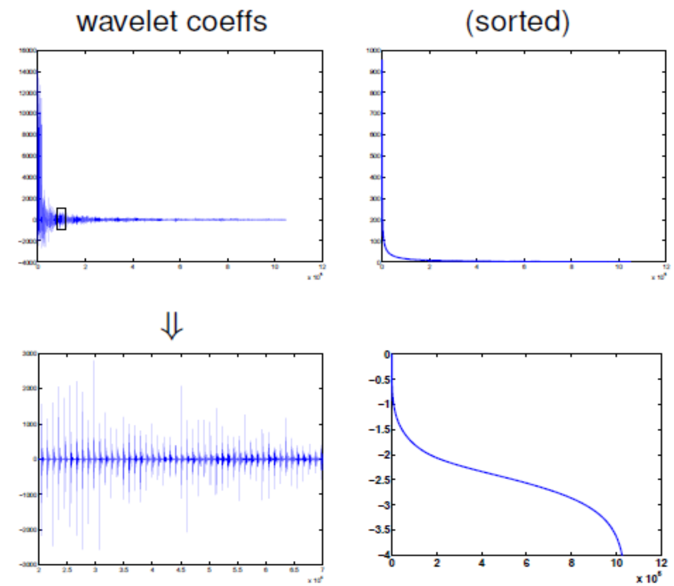
- in another one of his paper, “The sampling theorem of Shannon-Nyquist-Kotelnikov-Whittaker has been of tremendous importance in engineering theory and practice. Straightforward and precise, it sets forth the number of measurements required to reconstruct any bandlimited signal. However, the sampling 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]

CS Tutorials

- Many CS Tu torials on li ne show res ults verifyin g what was said in [Don oho06]
- Charts from Romberg-W akin's CS tu torial, 2007.

Wavelets and Images

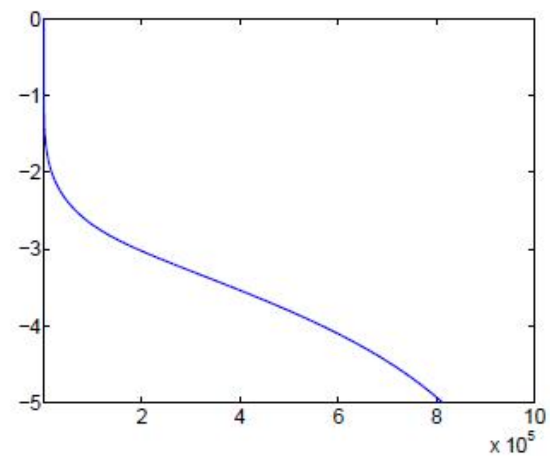
1 megapixel image



zoom in

($\log_{10}$ sorted)

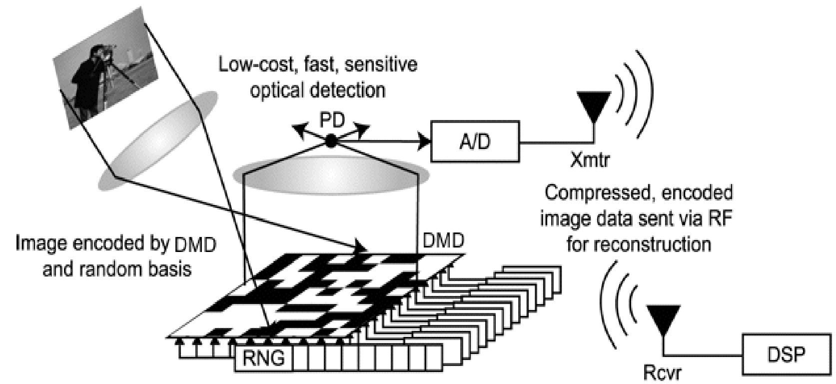
25k term approx



B-term approx error

Background

- 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, …



$$M < N$$

$$y = F\,x$$

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

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

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 com giving its ordinates at a series of poin*

Signal Dimension $\sim 2TW$

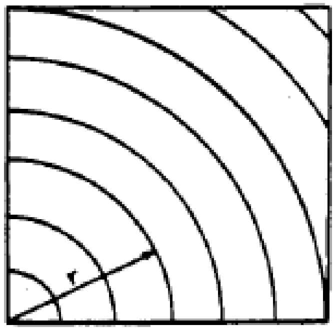


Fig. 2—Reduction of dimensionality through equivalence classes.

- 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 frequenc y would not need to be specified independently, but only ; that is, the total amplitude for this frequency. T he 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 spe ech 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}

Key Issues in CS

- Good sensing matrix *F*
  - *Random (Gaussian, Bernoulli, Fourier, …)*
  - *Vandermonde frames*
  - *Low density frames*
- Good recovery algorithms
  - L1 min recovery is good for sparse signal recovery
- How many measurements are good enough, *M*?
  - There are many
  - Depends on recovery algorithms, L0, L1, L2, …
  - In general, $M > K$ and $M < N$.
- Performance guarantee, in what form:
  - probabilistic vs. deterministic