Applications of Wishart Matrices in Compressive Sensing

Oliver James

INFONET Lab., GIST, KOREA

2014 INFONET WORKSHOP

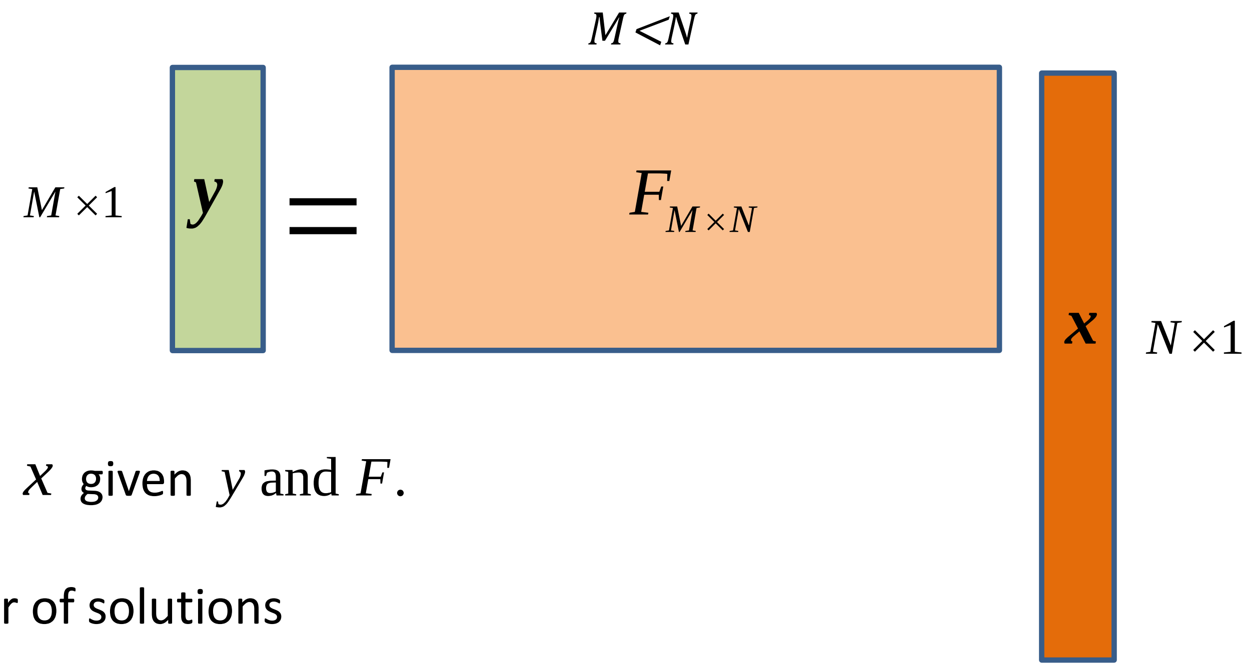
July 8th, 2014

Outline

- Compressive sensing (CS) basics
- Performance evaluation: Restricted Isometry constant
- Random matrices in CS : Gaussian
- Wishart matrices: New probability distributions
- Undersampling Analysis in CS

CS Basics

- Under-determined System: $y = Fx$
- Equations < unknowns



- Problem: Find x given y and F.
- Infinite number of solutions
- Approx. solution: $\widehat{x} = F^T \left(FF^T\right)^{-1} y.$
- Can we find an unique solution? No. ☹ and YES ☺
- YES! says CS provided the signal or the vector x is sparse

CS Basics (contd.)

- A sparse signal has only a fewer non-zero components than zero components.
  - *x*= [0 1 0 0 0 -2 -1 0 0 7 0 0]
  - Length *N*=12
  - No. non-zero components *K*=4 (Sparsity is 4, 4-sparse signal)
  - Support set: Locations of non-zero components $\mathcal{K} = \{2,6,7,10\}$
  - No. of support sets is $\binom{N}{K}$

- x= [0 1 0 0 1] 2-sparse signal

- x= [-1 1 1 -2 1] Not a sparse signal

- Non-sparse signal can be converted into a sparse signal

$$\text{Ex.} \quad \begin{bmatrix} -1 \\ 1 \\ 1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 & 2 & -3 \\ -2 & 3 & -3 & 4 & -2 \\ 3 & -1 & 1 & 2 & 2 \\ 0 & -1 & 0 & 1 & -1 \\ 1 & 1 & -2 & -2 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{aligned} y &= Fx \\ &= \underbrace{FB}_{A}s \qquad B - N \times N \ \text{basis matrix} \\ &= As \end{aligned}$$

CS Basics (contd.)

$$\boldsymbol{y} = F\boldsymbol{x}$$

$M \times 1$
Measurement

$$\boldsymbol{y} = F_{M \times N}\,\boldsymbol{x}$$

$N \times 1$
$K-$sparse signal

- F is a sensing matrix
  - Deterministic matrices
  - Random matrices: Gaussian and Bernoulli
  - Structured matrices: Circulant matrices, Vandermonde

- Two broad research questions in CS
  - What is a good sensing matrix? (Such as Coherence & RIC)
  - What is a good recovery algorithm?

What is a good sensing matrix?

- To measure the goodness of sensing matrix for sensing and recovery of sparse signals [Candes05] introduced Restricted Isometry Constant (RIC) of a sensing matrix, denoted by δ_K.

- δ_K takes values between 0 to 1.

- If δ_K is close to zero, then the matrix is good.

- Restricted Isometry Constant (RIC): The RIC of a sensing matrix F is the smallest quantity such that following equation holds

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

($F_{\mathcal{K}}$: $M \times K$)

for all support sets $\mathcal{K}$ and for all values of $\boldsymbol{x}_{\mathcal{K}}$.

- *RIP of order K:* If a sensing matrix F satisfies above equation with $\delta_K < 1$, then F is said to satisfy *RIP of order K* (Good matrix)

Finding RIC

- Two ways to compute RIC
  - By direct numerical evaluation
  - Through eigenvalues of sensing matrices.

<u>Direct numerical evaluation</u>

- Evaluate the ratio $\frac{\|F_{\mathcal{K}}\mathbf{x}_{\mathcal{K}}\|^2}{\|\mathbf{x}_{\mathcal{K}}\|^2}$ for $\binom{N}{K}$ support sets and for all signal values and find minimum and maximum of the ratio. (Not possible for large values of N and K)

- For a given (deterministic) sensing matrix finding RIC is an NP- hard problem.

- Constructing a matrix F for a given RIC is still an open problem.

Finding RIC (Contd.) –An easy way

- Definition by Candes [Candes05]

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

- From Rayleigh quotient we know that

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

- Observation:

  Eigenvalues of the matrix $F_{\mathcal{K}}^T F_{\mathcal{K}}$ does not depend on signal. RIC is a property of a sensing matrix and not of the sparse signal or the recovery algorithm

- How eigenvalues related to δ_K ?

RIC and Eigenvalues

- [Candes 05] showed that for "large size" Gaussian ensemble

$$(1-\delta_K) \leq \lambda_{\min}\left(F_{\mathcal{K}}^T F_{\mathcal{K}}\right) \leq \lambda_{\max}\left(F_{\mathcal{K}}^T F_{\mathcal{K}}\right) \leq (1+\delta_K)$$

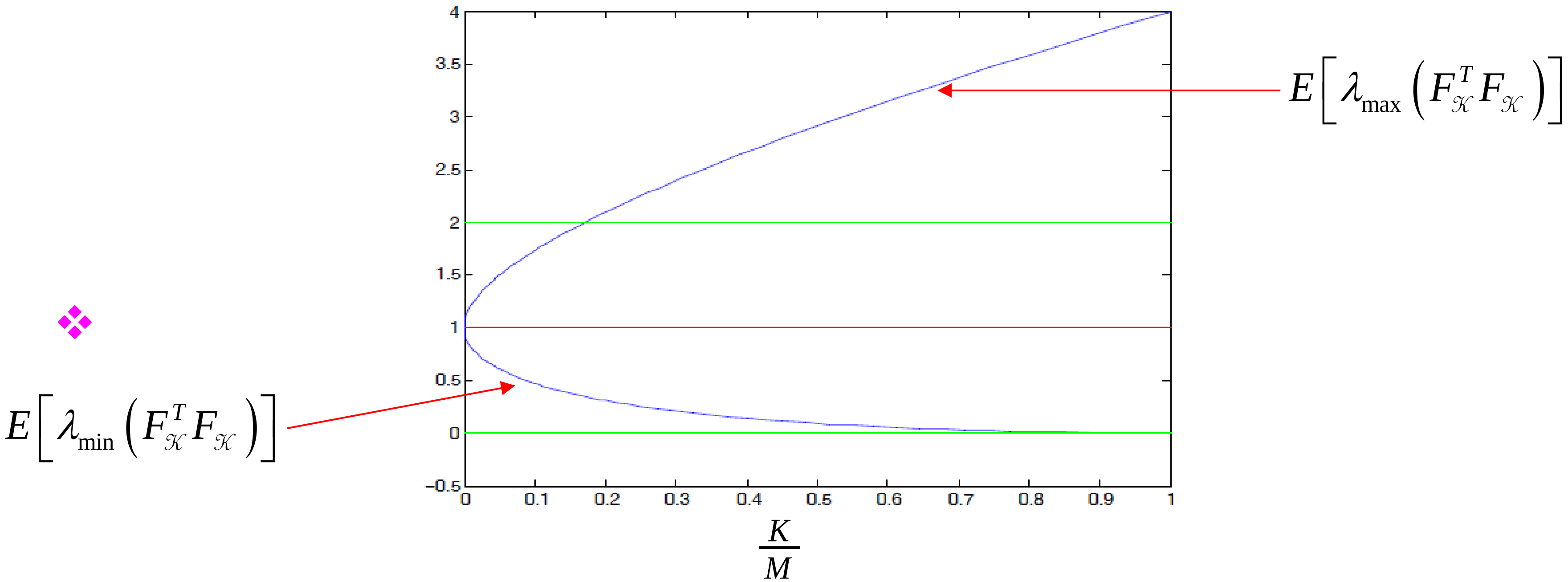
- When $F_{\mathcal{K}}$ is an Gaussian matrix, then $F_{\mathcal{K}}^T F_{\mathcal{K}}$ is called Wishart matrix (popular in multivariate statistics [Murihead 1982] and MIMO communications [Verdu 2004])

- Knowing eigenvalues of Wishart matrix guides us to say probabilistic statements about RIC [Candes 05] [Baraniuk08]

- How does the above relation hold for Gaussian matrices of various sizes?

RIC and Eigenvalues

- [Blanchand11a, Bah10] studied $E\left[\lambda_{\min}\left(F_{\mathcal{K}}^T F_{\mathcal{K}}\right)\right]$ and $E\left[\lambda_{\max}\left(F_{\mathcal{K}}^T F_{\mathcal{K}}\right)\right]$



- $$\left(1-\delta_K^L\right) \le \lambda_{\min}\left(F_{\mathcal{K}}^T F_{\mathcal{K}}\right) \le \lambda_{\max}\left(F_{\mathcal{K}}^T F_{\mathcal{K}}\right) \le \left(1+\delta_K^R\right)$$

- Observation 1. Values of RIC depends on the size of the sensing matrix!
- 2. Minimum eigenvalue still confined between 0 and 1

RIC and Eigenvalues

- An easy way to study performance of sensing matrix is via eigenvalues

$$\left(1-\delta_K^L\right) \le \lambda_{\min}\left(F_{\mathcal{K}}^T F_{\mathcal{K}}\right) \le \lambda_{\max}\left(F_{\mathcal{K}}^T F_{\mathcal{K}}\right) \le \left(1+\delta_K^R\right)$$

- For Gaussian sensing matrices, $\delta_K^L \ne \delta_K^R$

- Using minimum eigenvalues we evaluate δ_K^L as

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

- For a Gaussian matrix $F_{\mathcal{K}}$, we can evaluate δ_K^L via the minimum eigenvalue of the Wishart matrix $F_{\mathcal{K}}^T F_{\mathcal{K}}$

- We can state: $\Pr\left\{\min_{\mathcal{K}} \lambda_{\min}\left(F_{\mathcal{K}}^T F_{\mathcal{K}}\right) > a\right\}$

Eigenvalue distributions of Wishart Matrices

- If $F_{\mathcal{K}}$ is an $M \times K$ Gaussian matrix, $F_{\mathcal{K}}^T F_{\mathcal{K}}$ is a $K \times K$ Wishart matrix (real or complex)

- The Wishart matrix has K eigenvalues

- Eigenvalues of Complex Wishart matrices
  - [Khatri 1967] - in terms of product of beta integrals.
  - [Krishnaiah 1971] - Zonal polynomials (available up to $K = 6$)
  - [Alouini 2004] - CDF for $K = 2$ case
  - [Chinai 2009] - in terms of matrix determinants

- Eigenvalues of real Wishart matrices
  - [Sugiyama 1964] - in terms of Zonal polynomials.
  - [Edelman 1989] - Tricomi functions (no closed-form for $K > 25$)

Eigenvalue distributions of Wishart Matrices

- Aim: To determine new eigenvalue distributions, compact and tractable.
- Approach: Start with the joint eigenvalue distribution $f(\boldsymbol{\lambda})$, and find $f(\lambda_{\max})$ & $f(\lambda_{\min})$.
- For real Wishart matrix [Murihead 1982]

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

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

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

- Two key step in order to perform $\left(K-1\right)-\mathrm{dimensional\ integration}$
  1. Expansion of the Vandermonde determinant along the desired eigenvalue.
  2. Multiple integration of sub-determinants using the theory of skew-symmetric matrices [Bruijn1955].

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

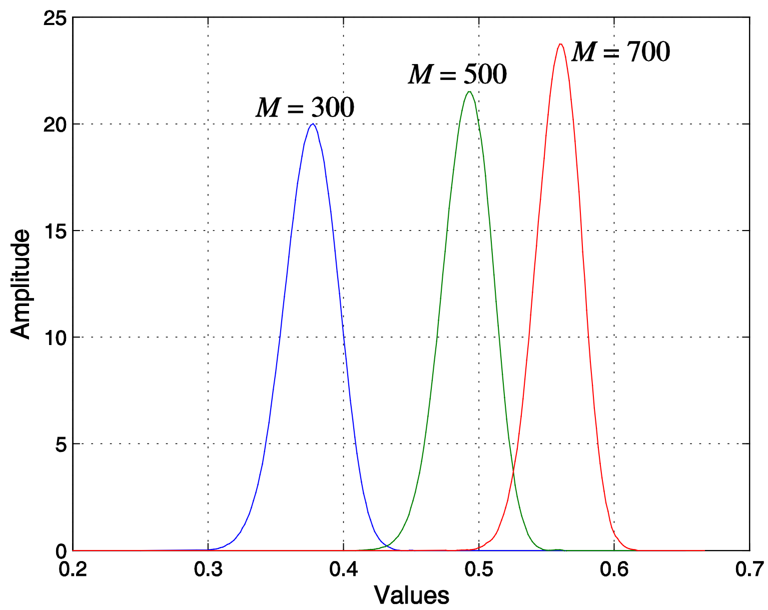
- Minimum eigenvalue distribution

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

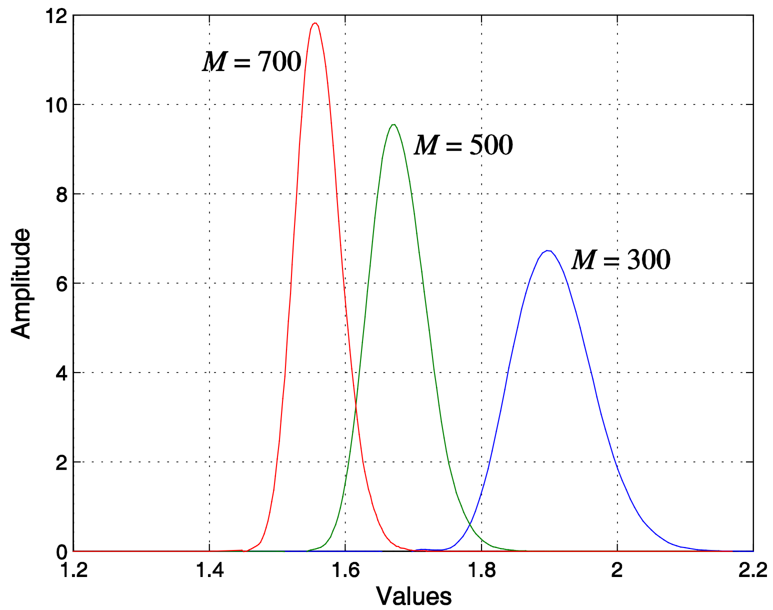
Plots of eigenvalues

- Simulation results

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



Minimum eigenvalue of Wishart matrix



Maximum eigenvalue of Wishart matrix

- Both eigenvalues move close to 1 as M increases

Undersampling Analysis in CS

- Aim: To determine the sufficient number of measurements needed for the exact recovery of K-sparse signal.

- What literature says? If RIC of a sensing matrix satisfies a certain theoretical guarantee, then a recovery algorithm exactly recovers sparse signals

Ex:

$$\delta_{2K}^{L} < 0.421 \quad L_1 \text{ algorithm [Candes11]}$$

$$\delta_{K+1}^{L} < \frac{1}{3\sqrt{K}} \quad \text{OMP algorithm [Davenport10]}$$

- See [Mo11] for summary of theoretical guarantees

- In practice, what does the theoretical guarantee translates to?

- Way to interpret theoretical guarantees in practical terms is termed as undersampling analysis (UA).

Undersampling Analysis in CS (Contd.)

- Undersampling analysis suggest the number of measurements that are needed in order to satisfy a particular theoretical guarantee.

Undersampling Ratio: $\theta = \frac{M}{N}; \quad M < N$

$$\theta = \frac{\text{No. of measurements}}{\text{Length of sparse signal}} = \frac{\text{No. of rows of } F}{\text{No. of cols. of } F}$$

- <u>Our result summary:</u>

For an RIC condition of the form $\delta_K^L < \delta$ there exists a number $\theta_{\text{th}}^{\infty}$ so that $M > \theta_{\text{th}}^{\infty} N$ measurements are sufficient for any matrix chosen at random from a Gaussian ensemble of size $M \times N$.

Eigenvalue-based Undersampling Analysis

- Our approach
  - Start with an arbitrary condition $\delta_K^L < \delta$ (Good matrix)
  - Evaluate $\Pr\left\{\delta_K^L < \delta\right\}$
  - Find conditions at which $\Pr\left\{\delta_K^L < \delta\right\} \to 1$
- RIC is related to minimum eigenvalue as

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

- Definition: 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 as a well-conditioned matrix

Probability of Well-Conditioned Matrix

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

$$= \left[\Pr\left\{\lambda_{\min}\left(K, M, \Sigma\right) > a\right\}\right]^{\binom{N}{K}}$$

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

$$\overset{(a)}{\leq} \binom{N}{K} \Pr\left\{\lambda_{\min}\left(K, M, \Sigma\right) \leq a\right\}$$

$$\leq \binom{N}{K} \mathrm{P}_U$$

- (a) is the union bound, P_U an upper bound

$$\Pr\{\text{Well-conditioned matrix}\} = 1 - \Pr\{\text{Ill-conditioned matrix}\}$$

$$\geq 1 - \binom{N}{K} \mathrm{P}_U$$

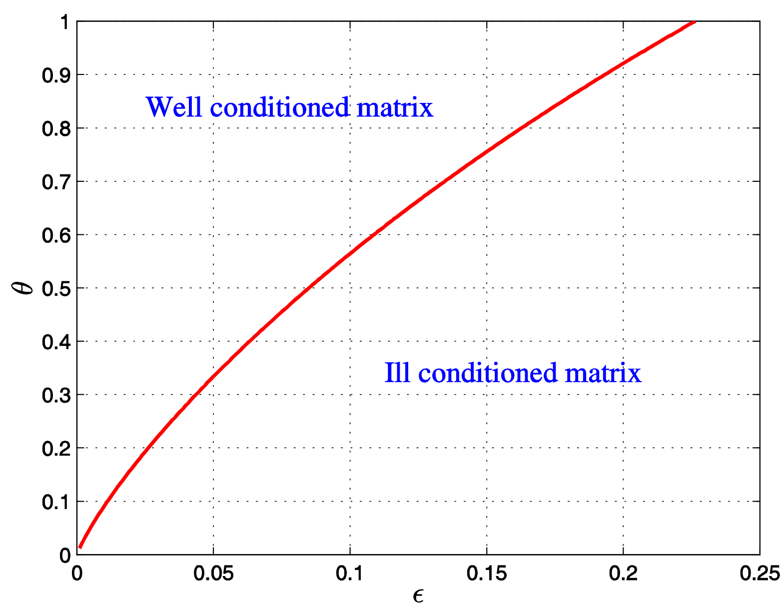
Probability of Well-Conditioned Matrix

$$\Pr\{\text{Well-conditioned matrix}\} = \Pr\left\{1-\delta_K^L > a\right\}$$

$$\geq 1-\binom{N}{K} ? P_U$$

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

- The exponent $E_K = \frac{1}{c_1}\frac{M}{N} - \frac{K}{N}\log\frac{N}{K} - c_2\frac{K}{N} + o(N)$ is a function sparsity ratio $\varepsilon := \frac{K}{N}$ and undersampling ratio $\theta := \frac{M}{N}$



Undersampling Analysis

- Under what condition Pr{ Well-conditioned matrix}→ 1 ?

$$\Pr\{\delta_K < \delta\} \to 1 \text{ when } E_K > 0.$$

E_K is a function of $\varepsilon = \frac{K}{N}$ and $\theta = \frac{M}{N}$.

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

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

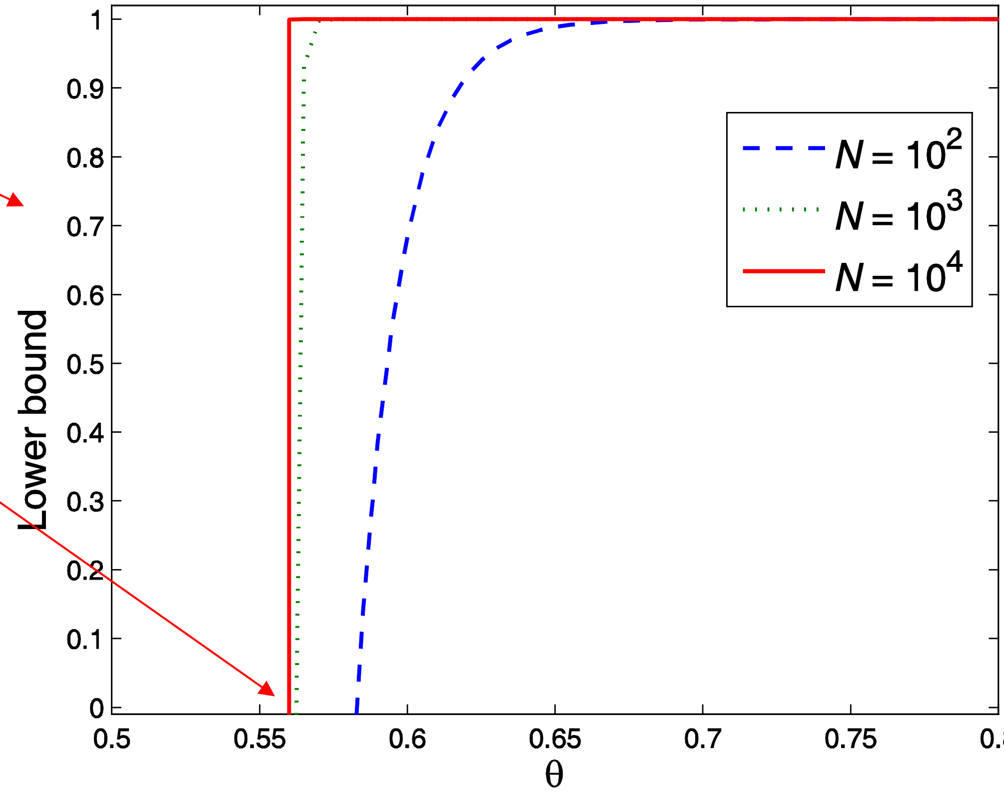
- We aim to find
  - For fixed ε, what values of θ possible?
  - What happens as $N \to \infty$; $\lim_{N\to\infty} \Pr\left\{\delta_{K+1} < \frac{1}{3\sqrt{K}}\right\} \to 1 \quad \theta \to ?$

Undersampling Analysis (Contd.)

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

$$\lim_{N\to\infty} \Pr\left\{\delta_{K+1} < \tfrac{1}{3\sqrt{K}}\right\} \to 1 \quad \theta \to \theta_{\mathrm{th}}^{\infty}$$

$$\theta_{\mathrm{th}}^{\infty} = 0.565$$



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

Summary

- We have derived new eigenvalue distributions of Wishart matrices using the theory of skew-symmetric matrices.

- Our distributions are exact, compact and are useful for the eigenanalysis of small and large systems

- We have found a lower bound on the existence of a good Gaussian sensing matrix for the purpose of undersampling analysis

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

References

- [Candes05] E.J. Candes and T. Tao, "Decoding by Linear Programming,", vol. 51, no.12, 2005.
- [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.
- [Bah 10] B. Bah and J. Tanner, "Improved bounds on restricted isometry constants for Gaussian matrices," SIAM Journal on Matrix Analysis, vol. 31, no. 5, pp. 2882–2898, 2010.
- [Blanchard 11a] J. D. Blanchard, C. Cartis, and J. Tanner, "Compressed Sensing: How sharp is the restricted isometry property?" SIAM Review, vol. 53, no. 1, pp. 105–125, 2011.
- [Blanchard 11b] J. D. Blanchard, C. Cartis, J. Tanner, and A. Thompson, "Phase transitions for greedy sparse approximation algorithms," Appl. Comput. Harmon. Anal., vol. 30, no. 2, pp. 188–203, 2011.
- [Baraniuk08] R. Baraniuk, M. Davenport, R. Devore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," Construct. Approx., vol. 29, no. 3, pp. 253-263, 2008.
- [Bruijn1955] N. G. De Bruijn, "On some mltiple integrals involving determinants," Journal of Indian Mathematical Society, vol. 19, pp. 133-151, 1955.
- [Mo11]Qun Mo and Song Li, "New bounds on the restricted isometry constant," Appl. Comput. Harmon. Anal., vol. 31, pp. 460-468, 2011.

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.

Predicting the Performance of Cooperative Wireless Networking Schemes with Random Network Coding

Jin-Taek Seong

J.-T. Seong and H.-N. Lee, "Predicting the Performance of Cooperative Wireless Networking Schemes with Random Network Coding," Early Access, *IEEE Trans. Commun.*

Outline

1. **Introduction**
   - **Diversity Techniques**
   - **Dynamic Network Cods**
   - **Motivation**

2. **Cooperative Network**
   - **Transmission Strategy**
   - **Received Signal**
   - **Outage Probability**

3. **Modeling of Transmission Matrices**
   - **Random Transmission Matrix**
   - **Probability Distribution**

4. **Upper Bound**
   - **Nullspace**
   - **Expectation of Nullity**
   - **Decoding Failure Probability**
   - **Homogeneous/Heterogeneous Connectivity**
   - **General Connectivity**
   - **Asymptotic Nullity**

5. **Conclusions**

6. **References**

Introduction – Diversity Technique

- **Channel fading** is one of the underlying causes of performance degradation in wireless networks
- To combat fading, **diversity techniques** have been proposed and employed in the time, frequency, and space domains
- **Cooperative networking** is one of the current approaches that aim to utilize **spatial diversity** via user cooperation
- Each user participates collaboratively, and shares the benefit of **a virtual antenna array** in transceiver messages that are available through another user's antenna

Introduction – Dynamic Network Codes (1/2)

- **Dynamic Network Codes (DNC)** proposed by Xiao and Skoglund [Xiao10] is to handle a dynamic network topology
- Wireless links are **unreliable**, and then **link failures will occur randomly** in the inter-user channels
- In the DNC scheme, multiple network code matrices are used; each one is designed to handle particular link outage occurrences
- A certain occurrence of link outages results in **a particular restriction** to the elements of the network code matrix
- Reblatto *et al.* [Rebelatto12] extended the two-phase transmission framework of the DNC to multiple phases in the **Generalized Dynamic Network Code (GDNC)** scheme to further enhance the transmission rate and the diversity order

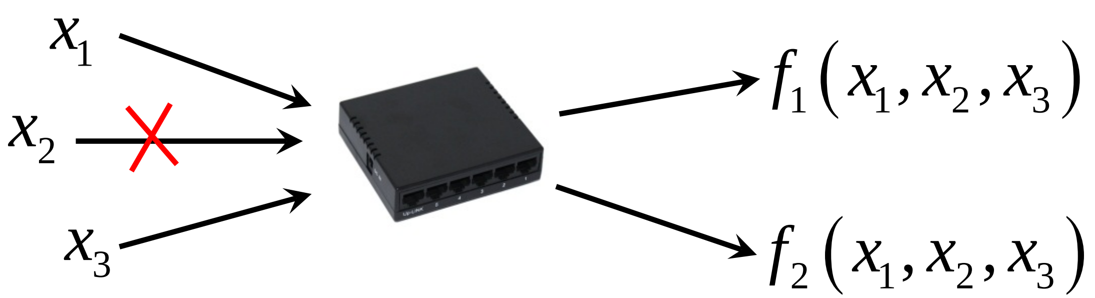
[Xiao10] M. Xiao, M. Skoglund, "Multiple-User Cooperative Communications Based on Linear Network Coding," *IEEE Trans. Commu.*, vol. 58, no. 12, pp. 3345–3351, 2010.

[Rebelatto12] J. L. Rebelatto, B. F. U-Filho, Y. Li, B. Vucetic, "Multiuser Cooperative Diversity through Network Coding Based on Classical Coding Theory," *IEEE Trans. Sig. Proc.*, vol. 60, no. 2, pp. 916–926, 2012.

Introduction – Dynamic Network Codes (2/2)

- An intermediate node fails to decode some of the messages received from the other nodes, and **creates a linear combination of the messages it could successfully decode**, and forwards it to the base station

Link Failure



Linear combinations for outgoing messages

$$f_1(x_1, x_2, x_3) = \alpha x_1 + 0 \cdot x_2 + \gamma x_3$$

$$f_2(x_1, x_2, x_3) = p x_1 + 0 \cdot x_2 + r x_3$$

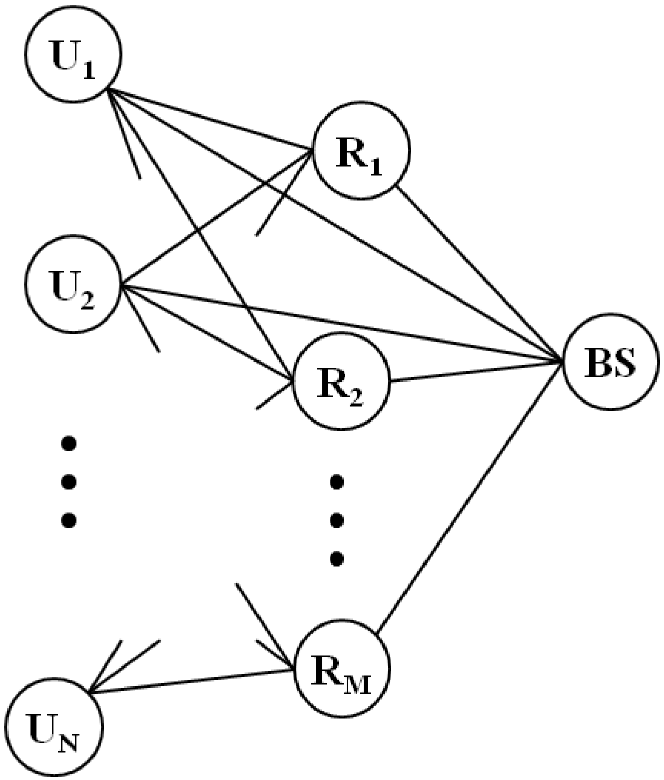
Introduction – Motivation

- While a series of performance analyses for DNC and GDNC are provided, the authors **rely on the exhaustive investigation** of all the individual network code matrices to determine if the resultant transmission matrix is sufficiently able to decode the source messages
- The performance analyses to determine the probability of successful decoding are performed only for **small and non-general networks**
- A successful decoding is assumed to be achieved when the network code matrix at the base station has a sufficient number of linearly independent vectors which at least equals the number of unknown source messages
- Then, the success probability is obtained by adding all the individual probabilities of such events over all the possible link failures.
- To this end, the authors followed the approach of tracking down each network code matrix individually, and determining if each was full in rank.
- **This is an exhaustive process.**

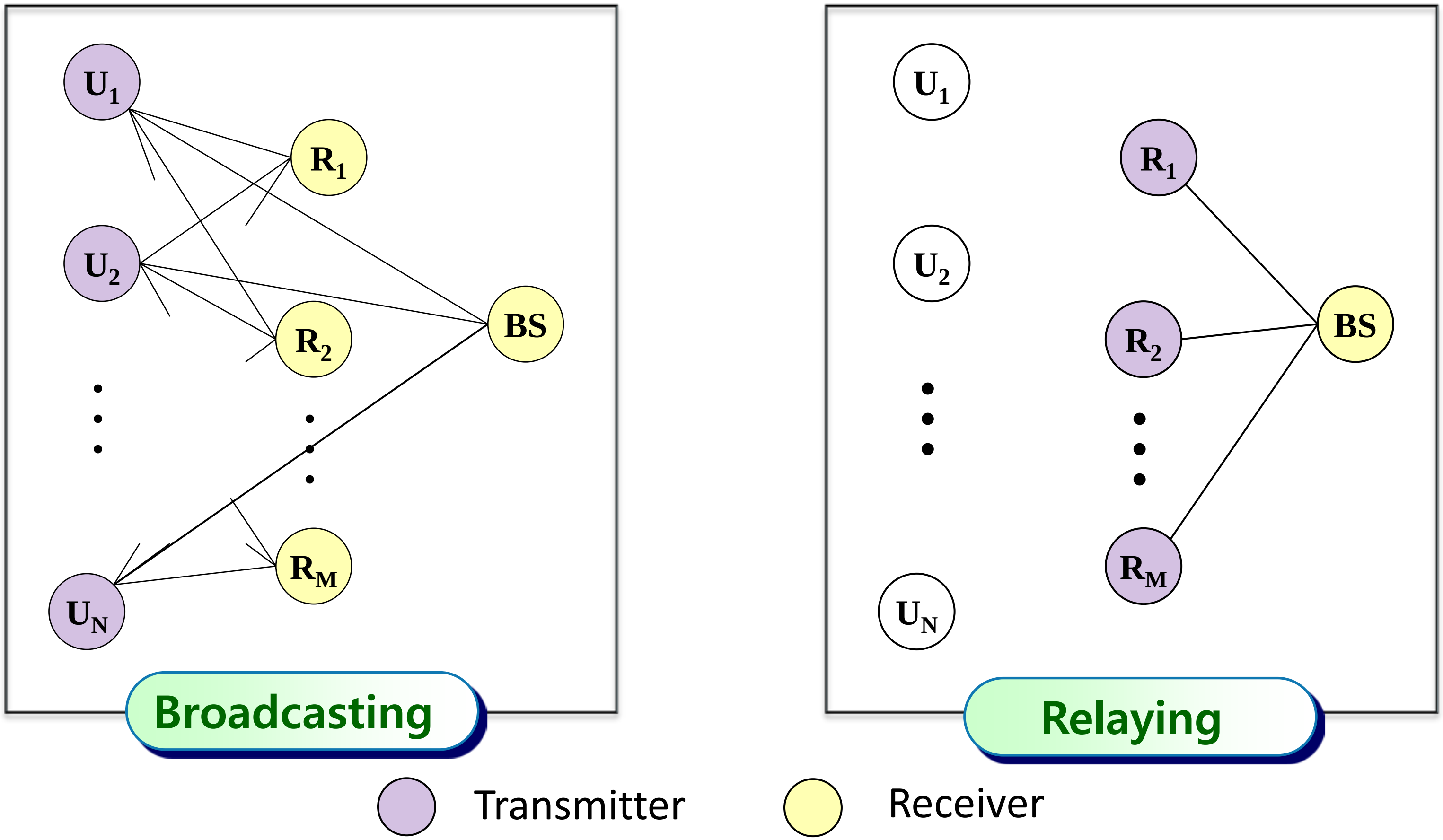
Introduction – Research Goal

Aim

- Proposing an **efficient evaluation framework** of the system performance in cooperative wireless networking schemes
- **Investigating impacts** of the number of relays and the field size of network coding on the system performance



Cooperative Network – Transmission Strategy



- An (N, M) cooperation scheme for wireless networks
- Two phase transmissions : *broadcasting* and *relaying*

Cooperative Network – Received Signal

Broadcasting phase

$$y_{u,d_1} = \sqrt{P_u} h_{u,d_1} x_{u,d_1} + n_{u,d_1}$$

$$u \in \{U_1, U_2, \ldots, U_N\}, \;\; d_1 \in \{R_1, R_2, \ldots, R_M, BS\}$$

Relaying phase

$$y_{r,d_2} = \sqrt{P_r} h_{r,d_2} x_{r,d_2} + n_{r,d_2}$$

$$r \in \{R_1, R_2, \ldots, R_M\}, \quad d_2 \in \{BS\}$$

Rayleigh fading channel

- $h_{r,d_2} \sim \mathcal{CN}\left(0, \sigma^2_{r,d_2}\right), \;\; h_{u,d_1} \sim \mathcal{CN}\left(0, \sigma^2_{u,d_1}\right)$
  - $\sigma^2_{u,d_1} := \rho^{-\eta}_{u,d_1}, \quad \sigma^2_{r,d_2} := \rho^{-\eta}_{r,d_2}$
  - ρ_{u,d_1} and ρ_{r,d_2} are the distances, u-to-d_1 and r-to-d_2

Cooperative Network – Outage Probability

- All the channels are **spatially and temporally independent**
- In Rayleigh fading channels, two outage probabilities are used as

$$\delta_{u,d_1} = \Pr\left\{\log\left(1+\gamma_{u,d_1}\right) < R_{th}\right\},$$

$$\delta_{r,d_2} = \Pr\left\{\log\left(1+\gamma_{r,d_2}\right) < R_{th}\right\}.$$

  – where R_{th} is the predefined threshold of the spectral efficiency

- The instantaneous SNRs of the two channels are denoted as

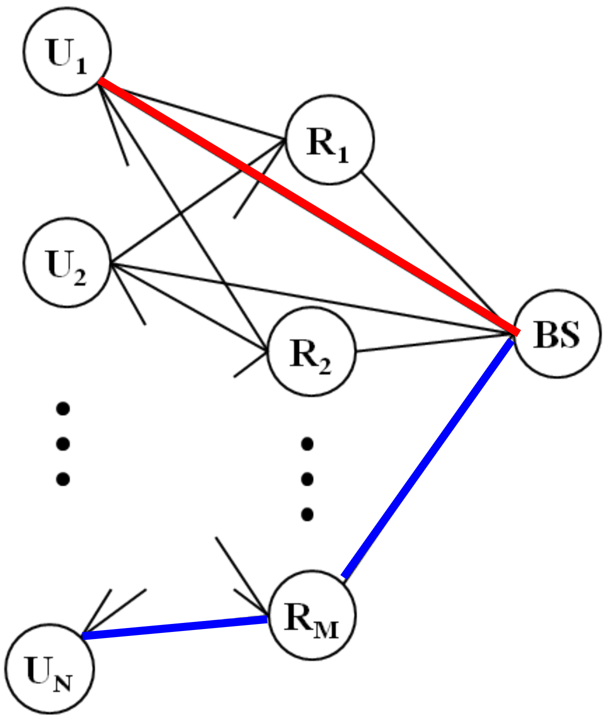
$$\gamma_{u,d_1} := \left|h_{u,d_1}\right|^2 P_u / N_0, \quad \gamma_{r,d_2} := \left|h_{r,d_2}\right|^2 P_r / N_0$$

- **Each of the outage probabilities is a function of the instantaneous SNR and the distance between two nodes**

Modeling – Random Transmission Matrix

- A random transmission matrix can be used to represent a family of network coding matrices for an (*N*, *M*) cooperative scheme
- The vector $\mathbf{y}$ received at the BS is then given by

$$\mathbf{y}^{(N+M)\times 1} = \mathbf{A}^{(N+M)\times N}\mathbf{x}^{(N\times 1)} = \begin{bmatrix} \mathbf{D}^{N\times N} \\ \mathbf{P}^{M\times N} \end{bmatrix}\mathbf{x}^{(N\times 1)}$$



$$\begin{bmatrix} y_{1,1} \\ \vdots \\ y_{N,1} \\ y_{1,2} \\ \vdots \\ y_{M,2} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_{NN} \\ \beta_{11} & \cdots & \beta_{1N} \\ \vdots & \ddots & \vdots \\ \beta_{M1} & \cdots & \beta_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

WHY NEW MODEL IS NEEDED?

- **Performance evaluation** : a transmission matrix has full rank or not
- For $N = 2$ and $M = 1$, there are 8 transmission matrices having full rank

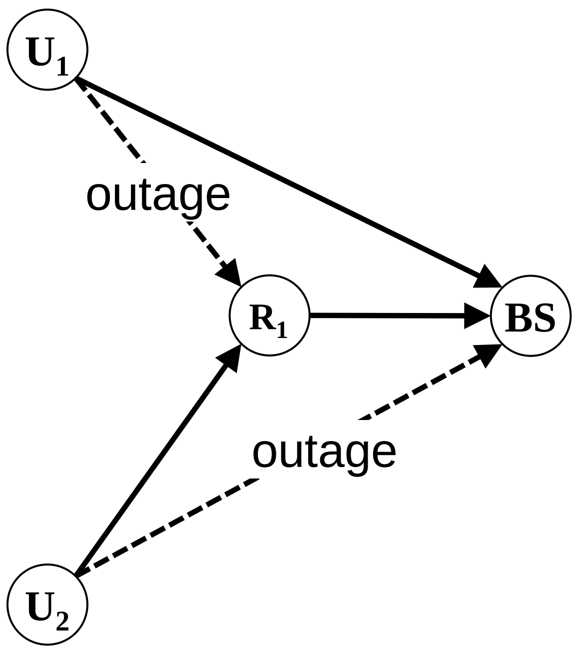
$$\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}\quad\begin{bmatrix}1&0\\0&1\\1&0\end{bmatrix}\quad\begin{bmatrix}1&0\\0&1\\0&1\end{bmatrix}\quad\begin{bmatrix}1&0\\0&1\\1&1\end{bmatrix}\quad\begin{bmatrix}1&0\\0&0\\1&1\end{bmatrix}\quad\begin{bmatrix}1&0\\0&0\\0&1\end{bmatrix}\quad\begin{bmatrix}0&0\\0&1\\1&1\end{bmatrix}\quad\begin{bmatrix}0&0\\0&1\\1&0\end{bmatrix}$$

$\sum$ Sum of individual probabilities

- The success probability is obtained by adding all the individual probabilities of such events over all the possible link failures
- This approach is exhaustive !!! This is intractable as the size of networks grows [Xiao10], [Rebelatto12]
  - # of random matrices with full rank for the binary field : $\prod_{i=1}^{N}\left(2^N - 2^{i-1}\right)$
- **We need an efficient and systematic performance evaluation**

HOW DOES THE TRANSMISSION MATRIX BECOME RANDOM (1/2)

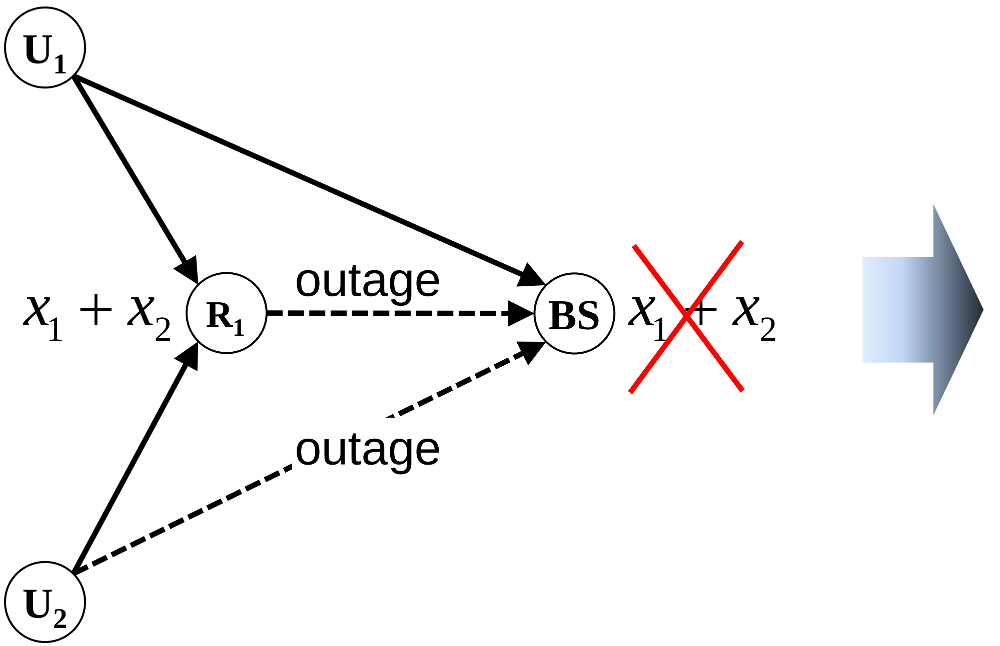
- Consider an element of a transmission matrix as a random variable
- The main idea of the determination of random elements
  - A nonzero coefficient of network coding is chosen if the wireless link between two nodes is successful
  - Otherwise, the coefficient must be zero
- In this work, we use an outage probability as a link failure



$$\begin{bmatrix} y_{1,1} \\ y_{2,1} \\ y_{1,2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

HOW DOES THE TRANSMISSION MATRIX BECOME RANDOM (2/2)

- The relay combines successful decoded messages
- If the outage between the relay and the base station occurs, the corresponding row of a random transmission matrix becomes all zero elements
- The reason is that the relay cannot forward user’s messages to the base station

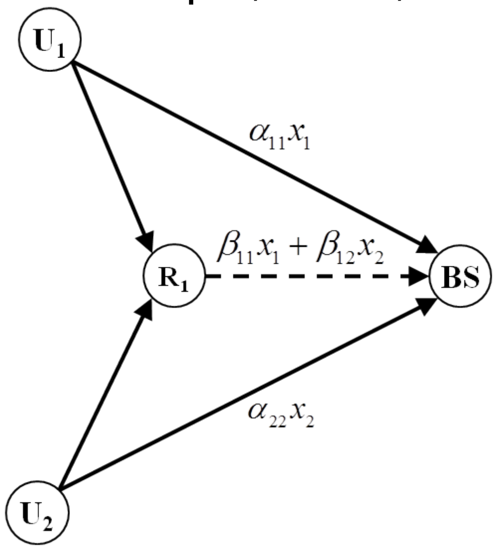


$$x_1 + x_2 \qquad x_1 + x_2$$

$$\begin{bmatrix} y_{1,1} \\ y_{2,1} \\ y_{1,2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Modeling – Example 1

- In a small example, $N = 2$, $M = 1$, $q = 2$,



$$\begin{bmatrix} y_{1,1} \\ y_{2,1} \\ y_{1,2} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & 0 \\ 0 & \alpha_{22} \\ \beta_{11} & \beta_{12} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- Determination of the transmission matrix (O : success, X : failure)

U_1-BS	U_2-BS	**D**	U_1-R_1	U_2-R_2	R_1-BS	**P**
O	O	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	O	O	O	$\begin{bmatrix} 1 & 1 \end{bmatrix}$
O	X	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	O	X	O	$\begin{bmatrix} 1 & 0 \end{bmatrix}$
X	O	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$	X	O	O	$\begin{bmatrix} 0 & 1 \end{bmatrix}$
X	X	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	X/-	X/-	O/X	$\begin{bmatrix} 0 & 0 \end{bmatrix}$

Conditional Probability Distribution

- Assume that all wireless links are independent
- Let $\bar{\mathcal{E}}_j$ and $\mathcal{E}_j$ be the nonoccurrence and occurrence of an outage event, relay R_j –to-BS
- If the outage event $\mathcal{E}_j$ occur, we set the conditional probability without the conditions of channel links between users and the relay as

$$\Pr\left\{\beta_{ji}=0\middle|\mathcal{E}_j\right\}=1$$

- For the event $\bar{\mathcal{E}}_j$, each element β_{ji} of $\mathbf{P}^{M\times N}$ can be independently defined,

$$\Pr\left\{\beta_{ji}=\theta\middle|\bar{\mathcal{E}}_j\right\},\quad \theta\in\mathbb{F}_q$$

- Using the outage probability, we set the conditional probability having zero value as

$$\Pr\left\{\beta_{ji}=0\middle|\bar{\mathcal{E}}_j\right\}=\delta_{U_i,R_j}$$

- For nonzero values, $\theta\in\mathbb{F}_q\setminus\{0\}$, we consider the uniform and MDS distributions

Modeling – Probability Distribution (1/2)

- Modeling each diagonal element of **D**

$$\Pr\{\alpha_{ii} = \theta\} = \begin{cases} \delta_{U_i,BS}, & \text{if } \theta = 0 \\ 1-\delta_{U_i,BS}, & \text{if } \theta = 1 \end{cases}$$

- Consider two types for modeling the elements of **P**
  - **Uniform distribution**

$$\Pr\{\beta_{ji} = \theta \,|\, \bar{\mathcal{E}}_j\} = \begin{cases} \delta_{U_i,R_j}, & \text{if } \theta = 0 \\ \left(1-\delta_{U_i,R_j}\right)/(q-1), & \text{if } \theta \neq 0 \end{cases}$$

  - **MDS distribution**

$$\Pr\{\beta_{ji} = \theta \,|\, \bar{\mathcal{E}}_j\} = \begin{cases} \delta_{U_i,R_j}, & \text{if } \theta = 0 \\ 1-\delta_{U_i,R_j}, & \text{if } \theta = \chi \\ 0\ , & \text{otherwise} \end{cases}$$

- Note that $\Pr\{\beta_{ji} = 0 \,|\, \mathcal{E}_j\} = 1$ without the conditions of outages, (U_i-R_j)

Modeling – Probability Distribution (2/2)

- For $N = 8$ and $M = 4$, the (4 x 8) submatrix of the systematic MDS code is

$$\begin{bmatrix} 9 & 13 & 14 & 7 & 2 & 15 & 13 & 12 \\ 15 & 3 & 9 & 12 & 12 & 10 & 12 & 2 \\ 14 & 9 & 12 & 7 & 8 & 1 & 3 & 7 \\ 4 & 5 & 5 & 10 & 9 & 3 & 4 & 1 \end{bmatrix}$$

- The conditional probability is defined as

$$\Pr\left\{\beta_{11} = 9 \middle| \overline{\mathcal{E}}_1\right\} = 1 - \delta_{U_1,R_1}, \qquad \Pr\left\{\beta_{11} = \theta \middle| \overline{\mathcal{E}}_1\right\} = 0, \theta \in \mathbb{F}_{16} \setminus \{0,9\}$$

- Investigating how much amount of improvement on the reconstruction performance is provided when using MDS codes in a cooperative wireless network

Upper Bound – Nullspace

- If a transmission matrix for a dynamic network topology has full rank, the BS can uniquely decode all messages from all sources

$$\mathbf{y}^{(N+M)\times 1} = \mathbf{A}^{(N+M)\times N}\mathbf{x}^{(N\times 1)}$$

- Let $\mathbf{A}$ be an $(N+M)\times N$ matrix over the finite field with size *q* as $\mathbb{F}_q$.

- Columns $A_1,\ldots,A_N$ of $\mathbf{A}$ are **linearly dependent** if and only if a vector $\mathbf{c}=(c_1,\ldots,c_N)\in\mathbb{F}_q^N$ exists, with at least one nonzero c_i, such that

$$\sum_{i=1}^{N} c_i A_i = 0$$

Definition 1. (***Number of nonzero coefficient vectors***) Let $L(\mathbf{A})$ be the number of all such nonzero vectors which belongs to the nullspace of the given matrix. Let the rank of a realized transmission matrix be $R(\mathbf{A})$. Thus, $L(\mathbf{A})$ can be represented as $L(\mathbf{A}) = q^{N-R(\mathbf{A})} - 1$

Upper Bound – Expectation of Nullity

Definition 2. (***Nullity***) Let $nullity(\mathbf{A})$ be the dimension of the nullspace in the column space of the matrix $\mathbf{A}$.

Proposition 3. For a random matrix $\mathbf{A}$, the expectation of nullity of $\mathbf{A}$ is upper bounded by $\mathbb{E}\left[nullity(\mathbf{A})\right] \le \log_q\left(\mathbb{E}\left[L(\mathbf{A})\right]+1\right)$.

- The rank-nullity theorem in linear algebra theory

$$nullity(\mathrm{A}) = N - R(\mathrm{A})$$

- Using Jensen's inequality and Definition 1,

$$\begin{aligned} \mathbb{E}\left[nullity(\mathbf{A})\right] &:= N - \mathbb{E}\left[R(\mathbf{A})\right] \\ &= \mathbb{E}\left[\log_q\left(L(\mathbf{A})+1\right)\right] \\ &\le \log_q\left(\mathbb{E}\left[L(\mathbf{A})\right]+1\right) \end{aligned}$$

Upper Bound – Decoding Failure Probability

Theorem 4. Let P_{fail} be the decoding failure probability for the reconstruction of source messages. Then, $P_{fail} \le \frac{1}{q-1}\mathbb{E}\left[L(\mathbf{A})\right]$

- The decoding failure probability is upper bounded by

$$\begin{aligned} P_{fail} &:= \Pr\left\{R(\mathbf{A}) < N\right\} \\ &= \Pr\left\{\exists \mathbf{c} : \sum_{i=1}^{N} c_i A_i = 0\right\} \\ &\le \sum_{\mathbf{c}\in\mathbb{F}_q^N \setminus \{0^T\}} \Pr\left\{\mathbf{A}\mathbf{c} = 0^T\right\} \\ &= \mathbb{E}\left[L(\mathbf{A})\right]. \end{aligned}$$

- The upper bound can be tighten as follows

$$P_{fail} \le \frac{1}{q-1}\mathbb{E}\left[L(\mathbf{A})\right],$$

  – using $\bigcup_{\mathbf{c}_1 \in \{\mathbf{c}, \theta\mathbf{c}, \ldots, \theta^{q-2}\mathbf{c}\}} \left\{\mathbf{A} : \mathbf{A}\mathbf{c}_1 = 0^T\right\} = \left\{\mathbf{A} : \mathbf{A}\mathbf{c} = 0^T\right\}$

Upper Bound – Performance Evaluation

Expectation of Nullity

$$\mathbb{E}\left[nullity\left(\mathbf{A}\right)\right] \le \log_q\left(\mathbb{E}\left[L\left(\mathbf{A}\right)\right]+1\right)$$

Decoding Failure Probability

$$P_{fail} \le \frac{1}{q-1}\mathbb{E}\left[L\left(\mathbf{A}\right)\right]$$

Three Types of Network Connectivity

Link	Homogeneous	Heterogeneous	General
User-Relay	δ	δ_1	δ_{U_i,R_j}
User-BS	δ	δ_1	$\delta_{U_i,BS}$
Relay-BS	δ	δ_2	$\delta_{R_j,BS}$

Upper Bound – Homogeneous Connectivity (1/3)

- Assuming all the outage probabilities are the same as δ
- Consider a vector $\mathbf{c} = (c_1, \ldots, c_N) \in \mathbb{F}_q^N$ with the first *k* nonzero elements

$$\mathbb{E}\left[L(\mathbf{A})\right] = \sum_{\mathbf{c}\in\mathbb{F}_q^N \setminus \{0^T\}} \Pr\left\{\mathbf{A}\mathbf{c} = 0^T\right\}$$

$$= \sum_{k=1}^{N} \binom{N}{k} (q-1)^k P_k$$

  - where $P_k := \Pr\left\{\sum_{i=1}^{k} c_i A_i = 0\right\}$

- The probability P_k is given by

$$P_k = \prod_{i=1}^{k} \Pr\{\alpha_{ii} = 0\} \prod_{j=1}^{M} \Pr\left\{\sum_{i=1}^{k} \beta_{ji} = 0\right\}$$

$$= \delta^k S_k^M$$

  - where $S_k := \Pr\left\{\sum_{i=1}^{k} \beta_{ji} = 0\right\}$

Upper Bound – Homogeneous Connectivity (2/3)

Lemma 5. For the homogeneous connectivity with the distributions, the probability S_k is given by

$$S_k = \delta + (1-\delta)\left(q^{-1} + \left(1-q^{-1}\right)\left(1-\frac{1-\delta}{1-q^{-1}}\right)^k \right)$$

- The probability can be decomposed by the condition of the outage event

$$S_k = \Pr\{\varepsilon_j\}\Pr\left\{\sum_{i=1}^{k}\beta_{ji}=0 \middle| \mathcal{E}_j\right\} + \Pr\{\overline{\varepsilon}_j\}\Pr\left\{\sum_{i=1}^{k}\beta_{ji}=0 \middle| \overline{\mathcal{E}}_j\right\}$$

$$= \delta + (1-\delta)\Pr\left\{\sum_{i=1}^{k}\beta_{ji}=0 \middle| \overline{\mathcal{E}}_j\right\}$$

  – using $\Pr\left\{\beta_{ji}=0 \middle| \mathcal{E}_j\right\}=1$

Upper Bound – Homogeneous Connectivity (3/3)

- Let f_k be the probability, $f_k := \Pr\left\{\sum_{i=1}^{k} \beta_{ji} = 0 \middle| \overline{\mathcal{E}}_j\right\}$

- The probability can be rewritten

$$f_k = \Pr\left\{\sum_{i=1}^{k-1} \beta_{ji} = 0 \middle| \overline{\mathcal{E}}_j\right\} \Pr\left\{\beta_{jk} = 0 \middle| \overline{\mathcal{E}}_j\right\}$$

$$+ \sum_{\theta \in \mathbb{F}_q \setminus \{0\}} \Pr\left\{\sum_{i=1}^{k-1} \beta_{ji} = \theta \middle| \overline{\mathcal{E}}_j\right\} \Pr\left\{\beta_{jk} = -\theta \middle| \overline{\mathcal{E}}_j\right\}$$

$$= f_{k-1}\delta + \left(1 - f_{k-1}\right)\frac{1-\delta}{q-1}.$$

- Let $g_k := f_k - q^{-1}$, substituting , we obtain,

$$g_k = g_{k-1}\left(1 - \frac{1-\delta}{1-q^{-1}}\right).$$

$$f_k = q^{-1} + \left(1 - q^{-1}\right)\left(1 - \frac{1-\delta}{1-q^{-1}}\right)^k$$

Upper Bound – Heterogeneous Connectivity

Proposition 6. Given an (N, M) cooperative network with the homogeneous connectivity based on some outage probability δ, $\mathbb{E}\left[L(\mathbf{A})\right]$ of a random transmission matrix $\mathbf{A}$ is

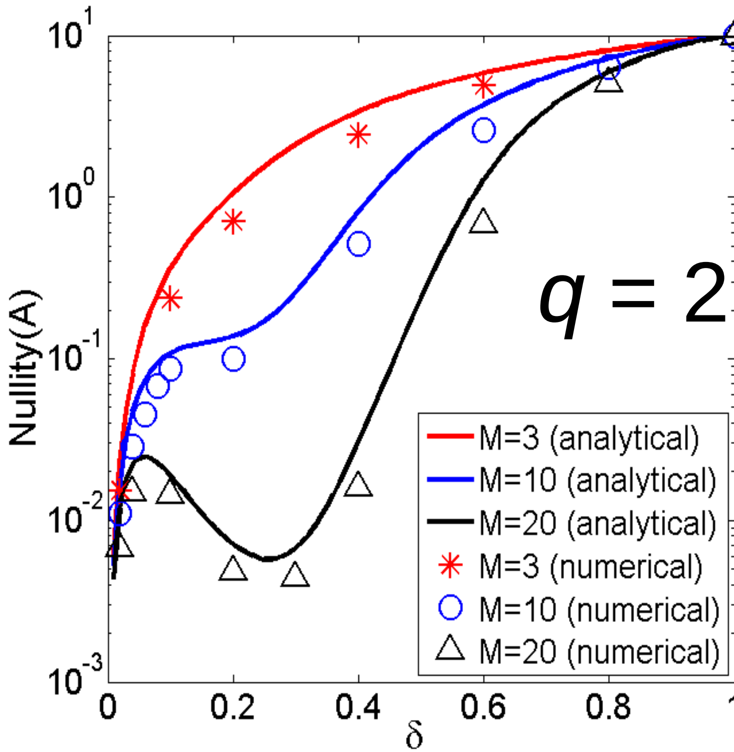
$$\mathbb{E}\left[L(\mathbf{A})\right]=\sum_{k=1}^{N}\binom{N}{k}(q-1)^k\,\delta^k\left[\delta+(1-\delta)\left(q^{-1}+\left(1-q^{-1}\right)\left(1-\frac{1-\delta}{1-q^{-1}}\right)^k\right)\right]^M.$$

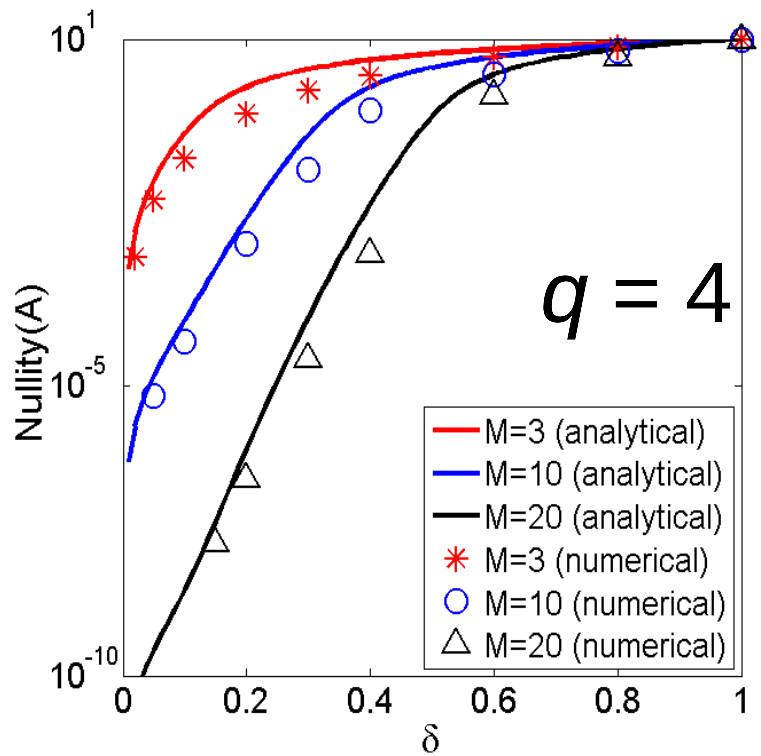
- Consider the **heterogeneous connectivity**, $\delta_1=\delta_{U_i,BS}=\delta_{U_i,R_j}$, $\delta_2=\delta_{R_j,BS}$

Proposition 7. Given the heterogeneous (N, M) cooperative network defined by the two outage probabilities δ_1 and δ_2, $\mathbb{E}\left[L(\mathbf{A})\right]$ of a random transmission matrix $\mathbf{A}$ is

$$\mathbb{E}\left[L(\mathbf{A})\right]=\sum_{k=1}^{N}\binom{N}{k}(q-1)^k\,\delta_1^k\left[\delta_2+(1-\delta_2)\left(q^{-1}+\left(1-q^{-1}\right)\left(1-\frac{1-\delta_1}{1-q^{-1}}\right)^k\right)\right]^M.$$

Upper Bound – Results of Homogeneous Connectivity





- The nullity of a random matrix for a homogeneous (10, *M*) cooperative wireless network with *N* = 10 and *M* = 3, 10, and 20.
- For *q* = 2, identical rows appear in random matrices from 0.1 to 0.3 of the outage probability
- The increase of field sizes more generates independent columns

Upper Bound – General Connectivity (1/3)

- Consider the general connectivity in which all outage probabilities are different

Proposition 8. Given an (N, M) cooperative network with the general connectivity, $\mathbb{E}\left[L(\mathbf{A})\right]$ of a random transmission matrix $\mathbf{A}$ is

$$\mathbb{E}\left[L(\mathbf{A})\right]=\sum_{k=1}^{N}(q-1)^{k}\,Q_{k},$$

where $Q_k := \sum_{l=1}^{|\mathcal{L}_k|} Q_{k,l},\ l\in\left\{1,2,..,|\mathcal{L}_k|\right\},\ |\mathcal{L}_k| := \binom{N}{k}$, and $\mathcal{L}_{k,l}$ is the lth entry of a set $\mathcal{L}_k$. Let $\mathcal{L}_k$ denote the collection of the sets of k distinct indices among $[N]:=\{1,2,...,N\}$, i.e., $\mathcal{L}_k := \left\{\{\lambda_1,\lambda_2,...,\lambda_k\} : \lambda_i\in\{1,2,...,N\}, \lambda_i\neq\lambda_j, i\neq j\right\}$.

Let $Q_{k,l} := \Pr\left\{\sum_{i\in\mathcal{L}_{k,l}} c_i A_i = 0\right\}$.

Upper Bound – General Connectivity (2/3)

- Consider a (2, 1) cooperative wireless network for q = 2, N = 2, and M = 2
- There are three nonzero vector $\mathbf{c}$ in $\mathbb{F}_2^2$: (10), (01), and (11)
- The probability $Q_{1,1}$:

$$\begin{aligned} Q_{1,1} &= \Pr\{c_1 A_1 = 0\} \\ &= \Pr\{\alpha_{11} = 0\} \Pr\{\beta_{11} = 0\} \\ &= \delta_{U_1,BS}\left(\delta_{R_1,BS} + \left(1-\delta_{R_1,BS}\right)\delta_{U_1,R_1}\right). \end{aligned}$$

- The probability $Q_{1,2}$:

$$\begin{aligned} Q_{1,2} &= \Pr\{c_2 A_2 = 0\} \\ &= \Pr\{\alpha_{22} = 0\} \Pr\{\beta_{12} = 0\} \\ &= \delta_{U_2,BS}\left(\delta_{R_1,BS} + \left(1-\delta_{R_1,BS}\right)\delta_{U_2,R_1}\right). \end{aligned}$$

- The probability $Q_{2,1}$:

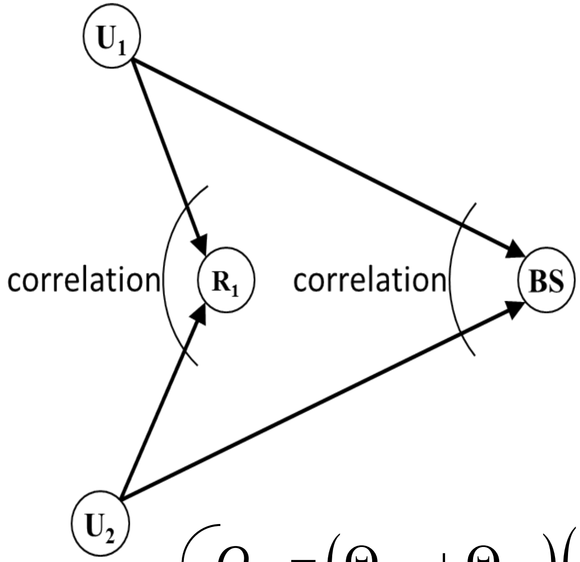
$$\begin{aligned} Q_{2,1} &= \Pr\{c_1 A_1 + c_2 A_2 = 0\} \\ &= \Pr\{\alpha_{11} = 0\} \Pr\{\alpha_{22} = 0\} \Pr\{\beta_{11} + \beta_{12} = 0\} \\ &= \delta_{U_1,BS}\delta_{U_2,BS}\left(\delta_{R_1,BS} + \left(1-\delta_{R_1,BS}\right)\Pr\{\beta_{11} + \beta_{12} = 0 \,|\, \overline{\mathcal{E}_1}\}\right). \end{aligned}$$

- Finally, $\mathbb{E}\left[L(\mathrm{A})\right] = Q_{1,1} + Q_{1,2} + Q_{2,1}$

Upper Bound – General Connectivity (3/3)

- The **proposed evaluation framework** can be extended to cases where the outages between different links are not independent, but **correlated**
- A pair of two outage events, U_1-BS and U_2-BS, makes a joint probability

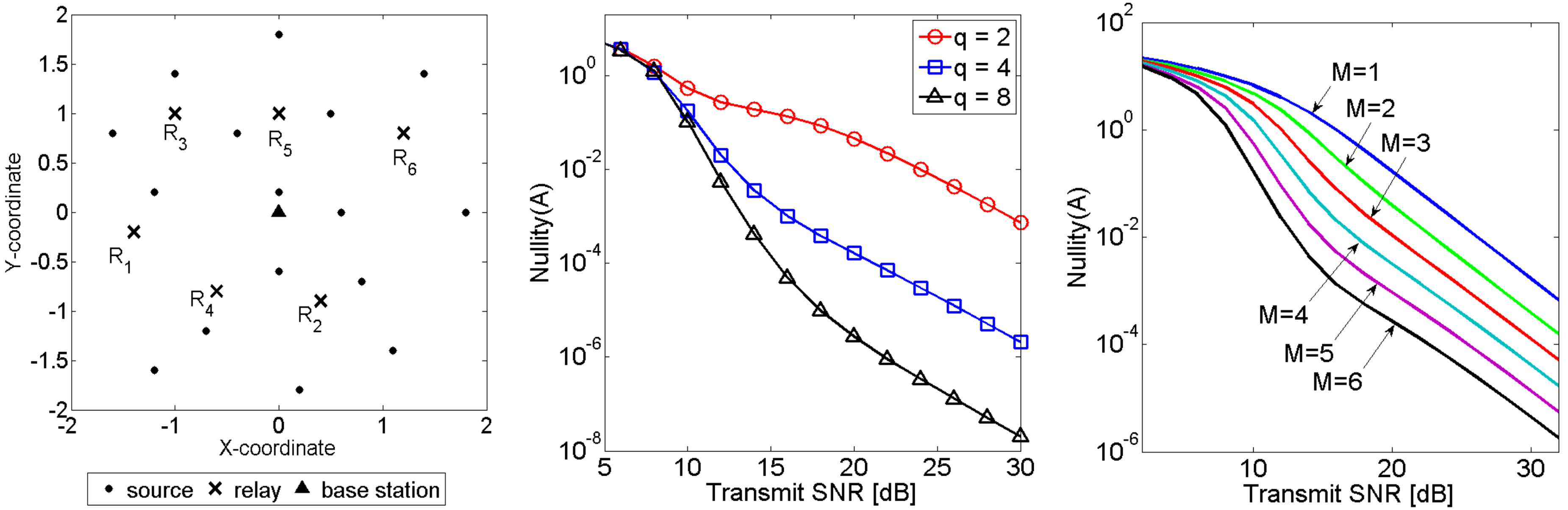
$$\Pr\left\{\alpha_{11}=\theta_1,\alpha_{22}=\theta_2\right\}=\Theta_{\theta_1,\theta_2}$$



(U_1-R_1, U_2-R_1)	(β_{12}, β_{11})	$\Pr\left\{\beta_{11}=\gamma_1,\beta_{12}=\gamma_2 \middle| \bar{\mathcal{E}}_1\right\}$
(O, O)	(1, 1)	$\Gamma_{0,0}$
(O, x)	(1, 0)	$\Gamma_{0,1}$
(x, O)	(0, 1)	$\Gamma_{1,0}$
(x, x)	(0, 0)	$\Gamma_{1,1}$

$$\begin{cases} Q_{1,1}=\left(\Theta_{0,0}+\Theta_{0,1}\right)\left(\delta_{R_1,BS}+\left(1-\delta_{R_1,BS}\right)\left(\Gamma_{0,0}+\Gamma_{0,1}\right)\right) \\ Q_{1,2}=\left(\Theta_{0,0}+\Theta_{1,0}\right)\left(\delta_{R_1,BS}+\left(1-\delta_{R_1,BS}\right)\left(\Gamma_{0,0}+\Gamma_{1,0}\right)\right) \\ Q_{2,1}=\Theta_{0,0}\left(\delta_{R_1,BS}+\left(\Gamma_{0,0}+\Gamma_{1,1}\right)\left(1-\delta_{R_1,BS}\right)\right) \end{cases}$$

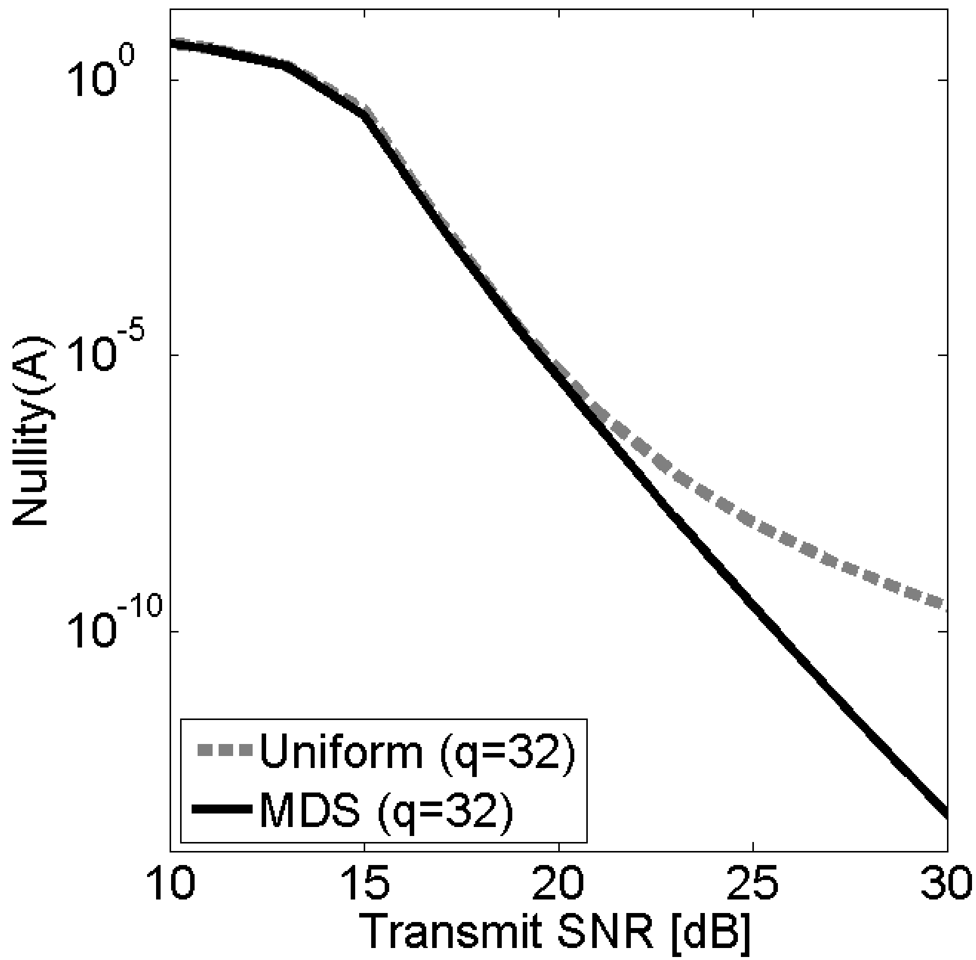
Upper Bound – Results of General Connectivity (1/3)



(a) (b) (c)

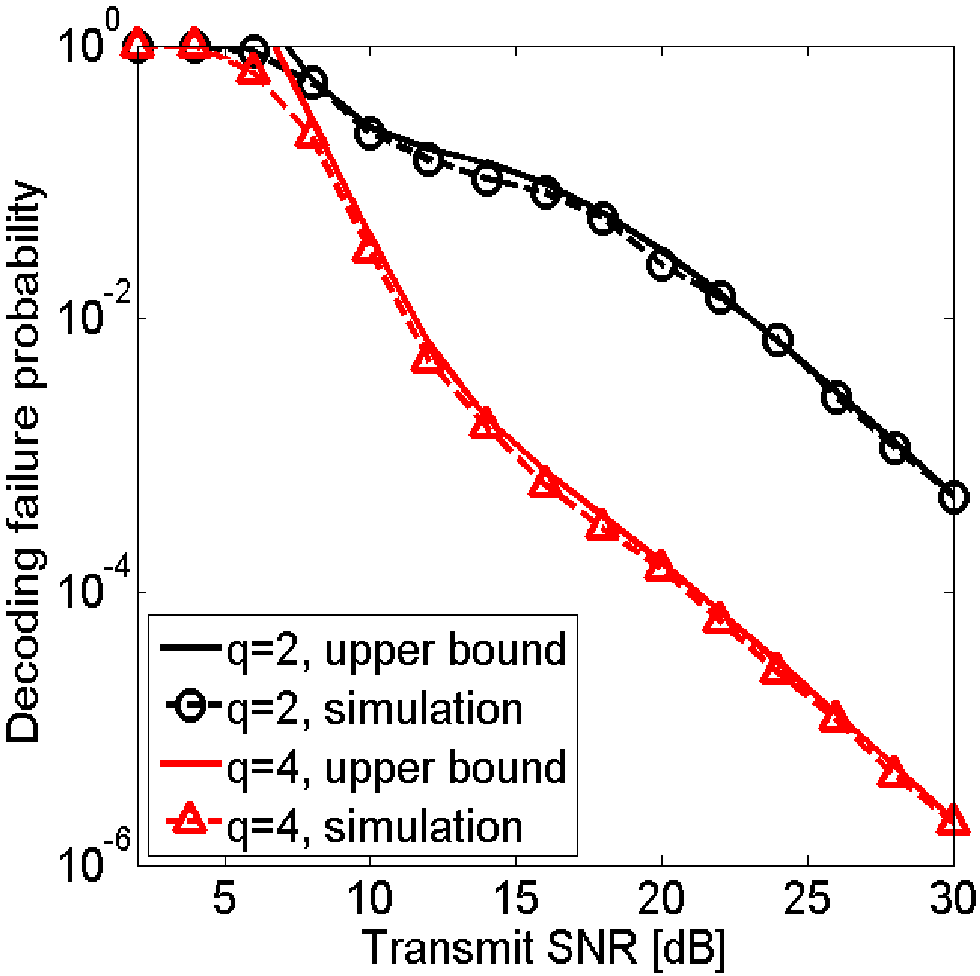
- (a) Location of 16 sources and 6 relays in a 2D space for a (16, 6) cooperative wireless network. (b) Results of upper bounds on with differing network coding field sizes *q* = 2, 4, and 8, (c) varying the number of relays at *q* = 4

Upper Bound – Results of General Connectivity (2/3)



- Comparison of upper bounds on $\mathbb{E}\left[nullity\left(\mathbf{A} \right) \right]$ for the uniform and MDS distributions

Upper Bound – Results of General Connectivity (3/3)



- Comparison of the decoding failure probabilities with the upper bound using Proposition 8 and the numerical simulation for $q = 2$ and 4

Upper Bound – Asymptotic Nullity (1/2)

- The homogeneous connectivity scheme is a specific case among general connectivity schemes
- Consider $\mathbb{E}\left[L(\mathbf{A})\right]$ for q = 2 in the homogeneous connectivity
- The general form of Q_k is given by

$$Q_k = \binom{N}{k} \delta^k S_k$$

- In this case, $\mathbb{E}\left[L(\mathbf{A})\right]$ can be obtained

$$\mathbb{E}\left[L(\mathbf{A})\right] = \sum_{k=1}^{N} \binom{N}{k} \delta^k S_k$$

- In high SNR regions , assuming δ is minimal, an approximation of $\mathbb{E}\left[L(\mathbf{A})\right]$ is obtained as

$$\mathbb{E}\left[L(\mathbf{A})\right] = \sum_{k=1}^{N} Q_k$$

$$\overset{(a)}{\approx} \binom{N}{1} \delta S_1 + \binom{N}{2} \delta^2 S_2$$

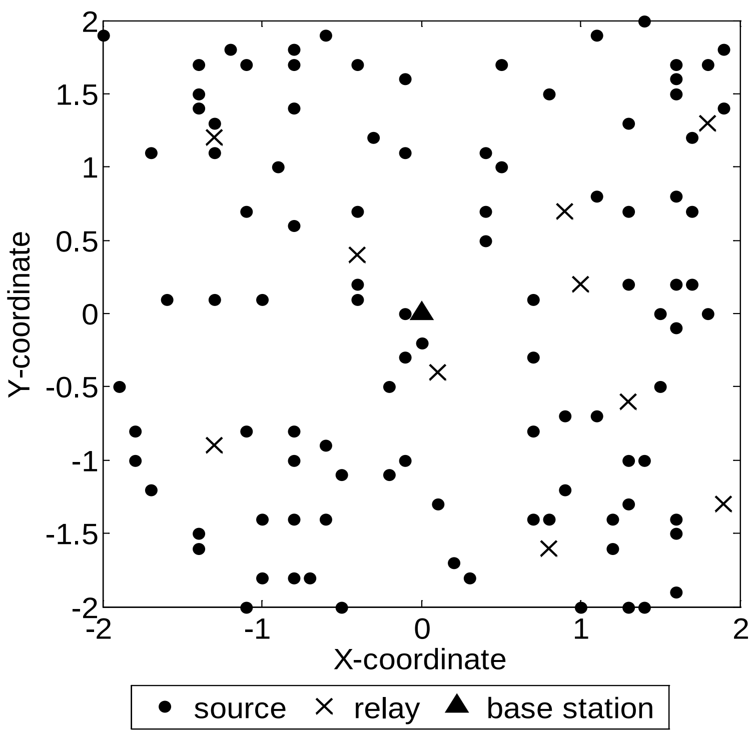
Upper Bound – Asymptotic Nullity (2/2)

Corollary 9. Given an (N, M) cooperative network with general connectivity, $\mathbb{E}\left[L(\mathbf{A})\right]$ is simplified in the high SNR regime

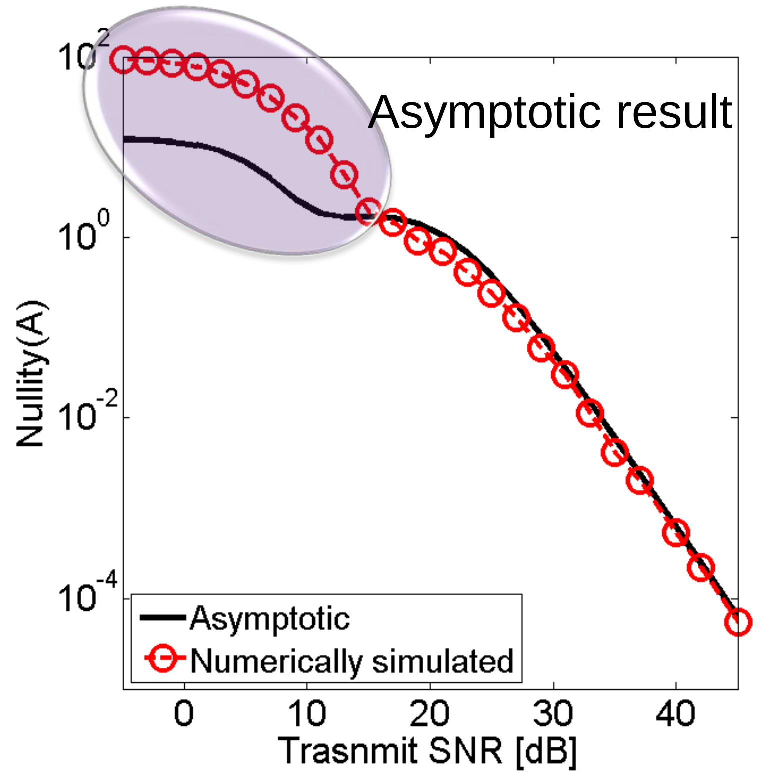
$$\mathbb{E}\left[L(\mathbf{A})\right] \approx (q-1)Q_1 + (q-1)^2 Q_2.$$

- For the computation of $\mathbb{E}\left[L(\mathbf{A})\right]$, two terms Q_1 and Q_2 are sufficient in high SNR regions
- Therefore, in high SNR regions, $\mathbb{E}\left[L(\mathbf{A})\right]$ **converges to the second order of the transmit SNR**

Upper Bound – Results of Asymptotic Nullity (1/2)



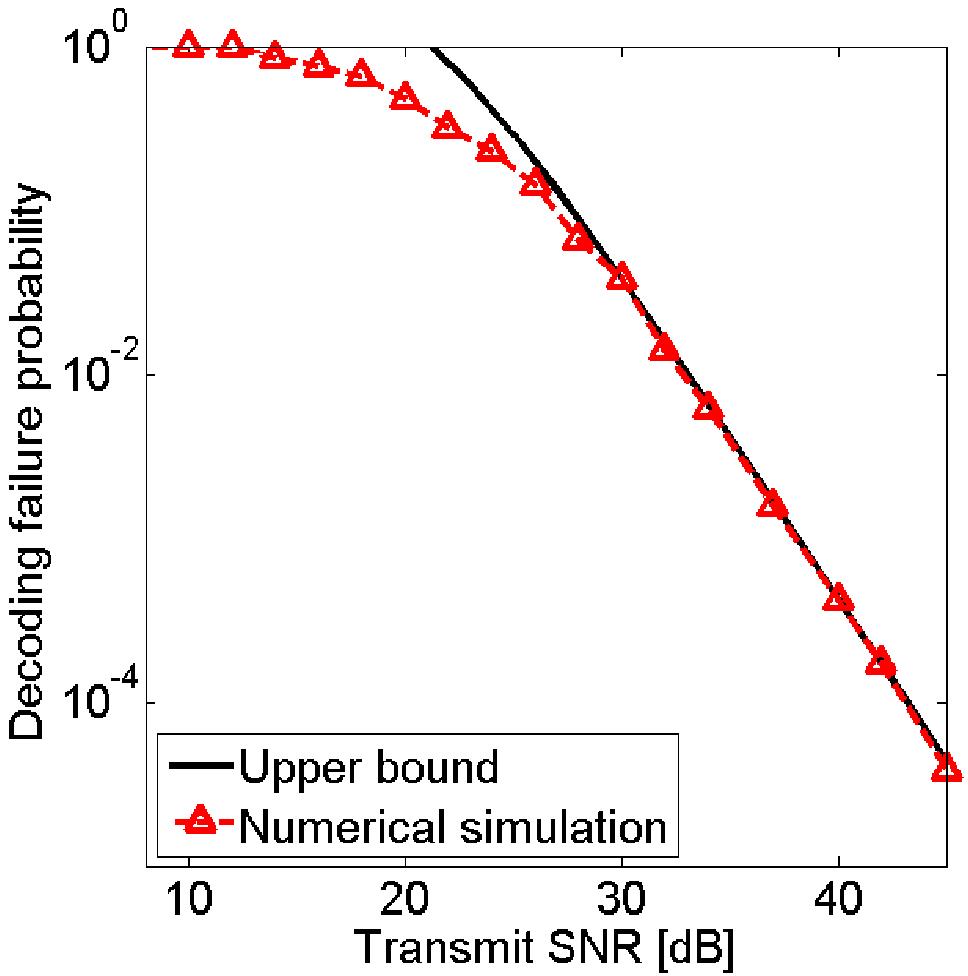
(a)



(b)

- (a) Locations of 100 sources and 10 relays in a 2D space for a (100, 10) cooperative wireless network. (b) Comparison of with the numerically simulated result and the upper bound using Corollary 9 for *q* = 2 and the uniform distribution

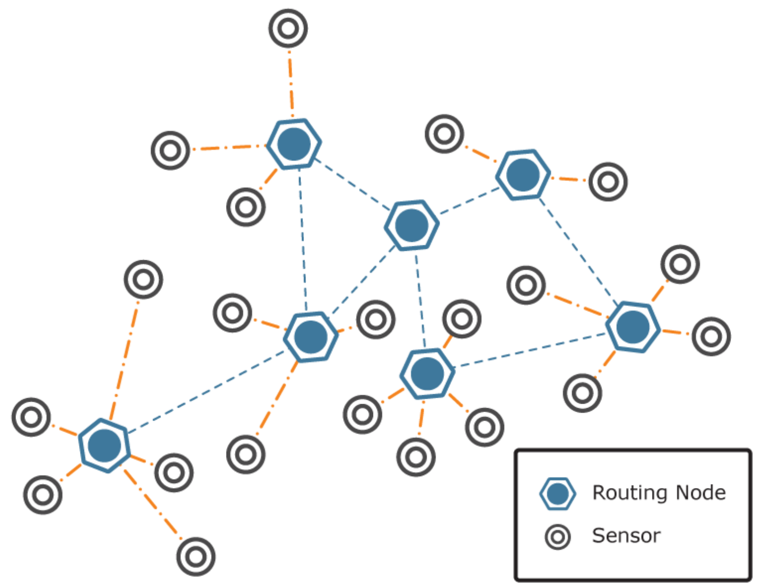
Upper Bound – Results of Asymptotic Nullity (1/2)



- Comparison of the decoding failure probabilities with the numerical simulation and the upper bound using Corollary 9

Who can use this work? (1/2)

- This work can be utilized in **wireless sensor networks** which aim to collect measured data using network coding [Wang10], [Yang13]

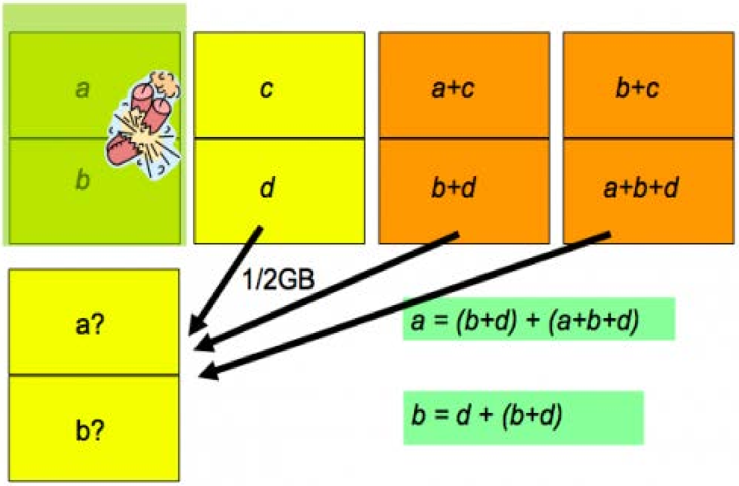


< Wireless sensor networks >

- Using our proposed framework, a system performance can be analyzed in terms of the number of sensors, the total number of transmissions, and power consumption, etc.

Who can use this work? (2/2)

- The application of **distributed data storage** based on network coding can analyze its system performance [Dimakis10]
- Dimakis *et al.* addressed how to generate encoded packets while transferring as little data as possible across the network



< Distributed data storages >

- In this example, three packets, *d*, *b*+*d*, and *a*+*b*+*d*, are needed to recover failed data *a* and *b*
- The proposed framework can be connected to the performance evaluation of distributed data storage systems

Conclusions

- We considered a cooperative wireless network where *N* sources are assisted with *M* relays in two phase transmissions.
- Our main goal was to propose a new performance analysis framework for evaluating the reconstruction performance of source messages at the BS.
- We modeled the elements of the transmission matrix as random variables.
- We derived two tight upper bounds on the expected nullspace dimension of the random transmission matrix, as well as the decoding failure probability.
- The result is a framework that is more effective than the rank-based method proposed in the previous literature.
- Three types of connectivity schemes are considered in this paper, as they enhance the framework’s scalability and suitability for general-purpose installations.

References

[Xiao10] M. Xiao, M. Skoglund, “Multiple-User Cooperative Communications Based on Linear Network Coding,” *IEEE Trans. Commu.*, vol. 58, no. 12, pp. 3345–3351, 2010.

[Rebelatto12] J. L. Rebelatto, B. F. U-Filho, Y. Li, B. Vucetic, “Multiuser Cooperative Diversity through Network Coding Based on Classical Coding Theory,” *IEEE Trans. Sig. Proc.*, vol. 60, no. 2, pp. 916–926, 2012.

[Seong14] J.-T. Seong and H.-N. Lee, “Predicting the Performance of Cooperative Wireless Networking Schemes with Random Network Coding,” Accepted for Publication in *IEEE Trans. Commun.*

[Fiore11] O. T.-Cruces, J. M. B.-Ordinas, M. Fiore, “Exact Decoding ProbabilityUnder Random Linear Network Coding,” *IEEE Comm. Lett.*, vol. 15, no. 1, pp. 67–69, 2011.

[Wang10] X. Wang, J. Wang, and Y. Xu, “Data Dissemination with Wireless Sensor Networks with Network Coding,” *EURASIP J. Wireless Comm. Networking*, 2010.

[Yang13] X. Yang, X. Tao, E. Dutkiewicz, X. Huang, Y. J. Guo, and Q. Cui, “Energy-Efficient Distributed Data Storage for Wireless Sensor Networks Based on Compressed Sensing and Network Coding”, *IEEE Trans. Wirel. Comm.*, vol. 12, no. 10, 2013.

[Dimakis10] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, K. Ramchandran, “Network Coding for Distributed Storage Systems,” *IEEE Trans. Info. Theory*, vol. 56, no. 9, 2010.

IT Bounds and Applications to Radar Performance Analysis

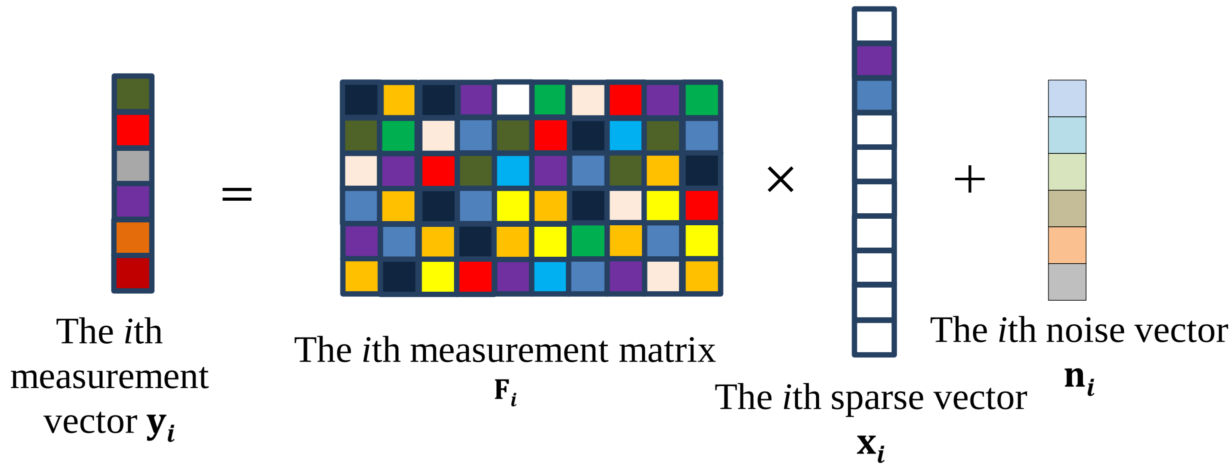
Sangjun Park

INFONET Lab., GIST, KOREA

2014 INFONET WORKSHOP

July 8th, 2014

Performance limits on reconstructing the multiple sparse vectors in the joint sparsity model



- ## Research Goal
  - Theoretical performance limits of compressive sensing problems in the joint sparsity model [Braniuk09].
- ## Approaches
  - Defining the failure probability that we cannot reconstruct the joint support set of the multiple sparse vectors.
  - Derive an upper bound on the failure probability
  - Finding sufficient conditions for vanishing the upper bound.

A failure event & a joint typical decoder

❖ A failure event & a joint typical decoder

- We define the failure probability based on a joint typical decoder

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

- This decoder all possible subsets to find the joint support set.
- If the output of the decoder satisfies the following inequality, then joint support set reconstruction is successful.

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

where $\mathbf{Q}(\mathbf{F}) := \mathbf{I} - \mathbf{F}\mathbf{F}^{\dagger}$ and $\mathbf{F}_{s,\mathcal{J}}$ is a matrix constructed by collecting column vectors of $\mathbf{F}_s$ corresponding to the indices of the output of the decoder.

A failure probability

- **A failure probability**
  - Let $\mathrm{E}(\mathbf{Y},\mathcal{I},\delta)$ be an event that the output of the decoder is declared as the joint support set.
  - With this event, the failure event can be defined by

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

where $\mathrm{E}(\mathbf{Y},\mathcal{I},\delta)^c$ denotes the joint support set is not declared as the joint support set and denotes the incorrect support set is declared as the joint support set.

  - The probability of the event is $\mathrm{E}(\mathbf{Y},\mathcal{J},\delta)$

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

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

where the inequality appears due to the Union bound.

Random variable(1/4)

- **The properties of the random variable**
  - The random variable in the event is

$$\sum_{s=1}^{S} \mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right)\mathbf{y}_s = \sum_{s=1}^{S} \mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right)\left(\mathbf{n}_s + \sum_{u\in\mathcal{I}\setminus\mathcal{J}} \mathbf{f}_{s,u} x_s(u)\right)$$
$$= \sum_{s=1}^{S} \mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right)\mathbf{c}_s$$

where $\mathbf{c}_s$ is the multivariate Gaussian with

$$\mathbb{E}\left[\mathbf{c}_{s,\mathcal{J}}\right] = \mathbf{0}_M \text{ and } \mathbb{V}\left[\mathbf{c}_{s,\mathcal{J}}\right] = \left(\sigma_n^2 + \left\|\mathbf{x}_{s,\mathcal{I}\setminus\mathcal{J}}\right\|_2^2\right)\mathbf{I}_M .$$

  - Also, the projection matrix $\mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right)$ can be decomposed

$$\mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right) = \mathbf{U}_{s,\mathcal{J}}\mathbf{S}_{s,\mathcal{J}}\mathbf{U}_{s,\mathcal{J}}^T$$

where $\mathbf{U}_{s,\mathcal{J}}$ is an $M \times M$ unitary matrix and $\mathbf{S}_{s,\mathcal{J}}$ is an $M \times M$ diagonal matrix.

Random variable(2/4)

- Continuously, we define a multivariate Gaussian vector

$$\mathbf{w}_{s,\mathcal{J}} = \mathbf{U}_{s,\mathcal{J}}^{T}\mathbf{c}_{s,\mathcal{J}}.$$

where $\mathbb{E}\left[\mathbf{w}_{s,\mathcal{J}}\right] = \mathbf{0}_M$ and $\mathbb{V}\left[\mathbf{w}_{s,\mathcal{J}}\right] = \mathbb{V}\left[\mathbf{c}_{s,\mathcal{J}}\right]$.

- Assume that the $M - K$ non-zero diagonal elements of $\mathbf{S}_{s,\mathcal{J}}$ are on the first $M - K$ diagonals.
- The random variable can be represented in terms of this multivariate Gaussian vector

$$\sum_{s=1}^{S}\left\|\mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right)\mathbf{y}_s\right\|_2^2 = \sum_{s=1}^{S}\left\|\mathbf{S}_{s,\mathcal{J}}\mathbf{w}_{s,\mathcal{J}}\right\|_2^2$$

$$= \sum_{s=1}^{S}\sum_{i=1}^{M-K}\left|w_{s,\mathcal{J}}\left(i\right)\right|^2.$$

Random variable(3/4)

– Last, we define a $S(M-K)\times 1$ vector

$$\mathbf{b}_{\mathcal{J}} = \begin{bmatrix} \mathbf{b}_{1,\mathcal{J}}^T & \mathbf{b}_{2,\mathcal{J}}^T & \cdots & \mathbf{b}_{S,\mathcal{J}}^T \end{bmatrix}^T$$

– where $\mathbf{b}_{s,\mathcal{J}} = \begin{bmatrix} w_{s,\mathcal{J}}(1) & w_{s,\mathcal{J}}(2) & \cdots & w_{s,\mathcal{J}}(M-K) \end{bmatrix}^T.$

– Then, the random variable finally can be rewritten as

$$\sum_{s=1}^{S} \left\| \mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right)\mathbf{y}_s \right\|_2^2 = \mathbf{b}_{\mathcal{J}}^T \mathbf{b}_{\mathcal{J}}.$$

– It states that the random variable is *a quadratic random variable* [Scharf91].

Random variable(4/4)

- **Lemma 1:** Let $M > K$, $Z_{\mathcal{I}} := \sum_{s=1}^{S} \left\| \mathbf{Q}\left(\mathbf{F}_{s,\mathcal{I}}\right)\mathbf{y}_s \right\|^2 \big/ \sigma_n^2$ and $\mathcal{I}$ be the joint support set. For any real number t, we have

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

- **Lemma 2:** Let $M > K$, and $Z_{\mathcal{J}} := \sum_{s=1}^{S} \left\| \mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right)\mathbf{y}_s \right\|_2^2$ for any $\mathcal{J} \in \mathcal{S} \setminus \{\mathcal{I}\}$. Then, for any real number t, we have

$$\mathbb{E}\left[Z_{\mathcal{J}}\right] = \mathrm{tr}\left(\mathbf{R}_{\mathcal{J}}\right), \mathbb{V}\left[Z_{\mathcal{J}}\right] = 2\,\mathrm{tr}\left(\mathbf{R}_{\mathcal{J}}^T \mathbf{R}_{\mathcal{J}}\right) \text{ and}$$
$$\mathbb{E}\left[\exp\left(tZ_{\mathcal{J}}\right)\right] = \prod_{n=1}^{S(M-K)} \left(1-2t\lambda_n\left(\mathbf{R}_{\mathcal{J}}\right)\right)^{-1/2}.$$

where

$$\mathbf{R}_{\mathcal{J}} = \begin{bmatrix} \left(\sigma_n^2 + \left\|\mathbf{x}_{1,\mathcal{I}\setminus\mathcal{J}}\right\|_2^2\right)\mathbf{I}_{M-K} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \left(\sigma_n^2 + \left\|\mathbf{x}_{2,\mathcal{I}\setminus\mathcal{J}}\right\|_2^2\right)\mathbf{I}_{M-K} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \left(\sigma_n^2 + \left\|\mathbf{x}_{S,\mathcal{I}\setminus\mathcal{J}}\right\|_2^2\right)\mathbf{I}_{M-K} \end{bmatrix}.$$

Upper bounds on the failure probabilities(1/5)

- **Upper bounds**
  - Remind

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

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

  - The properties the random variables of the events are given in Lemma 1 and Lemma 2 respectively.
  - Hence, we can compute upper bounds on the failure probabilities.

Upper bounds on the failure probabilities(2/5)

- **An upper bound on** $\mathbb{P}\left\{\mathrm{E}\left(\mathbf{Y},\mathcal{J},\delta\right)\right\}$.

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

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

$$=\mathbb{P}\left\{Z_{\mathcal{J}}<SM\delta+S\left(M-K\right)\sigma_n^2\right\}$$

$$\overset{(b)}{\leq}\mathbb{E}\left[\exp\left(tZ_{\mathcal{J}}\right)\right]\exp\left(-t\left(SM\delta+S\left(M-K\right)\sigma_n^2\right)\right)$$

$$=\prod_{i=1}^{S(M-K)}\left(1-2t\lambda_i\left(\mathbf{R}_{\mathcal{J}}\right)\right)^{-1/2}\exp\left(-t\left(SM\delta+S\left(M-K\right)\sigma_n^2\right)\right)$$

$$\overset{(c)}{\leq}\left(1-2t\lambda\left(\mathcal{J}\right)\right)^{-\frac{S(M-K)}{2}}\exp\left(-t\left(SM\delta+S\left(M-K\right)\sigma_n^2\right)\right)$$

  - **(a) : Union bound, (b) Chernoff bound, (c) due to the inequality**

$$\prod_{i=1}^{S(M-K)}\left(1-2t\lambda_i\left(\mathbf{R}_{\mathcal{J}}\right)\right)^{-1/2}\leq\left(1-2t\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right)\right)^{-S(M-K)/2}.$$

Upper bounds on the failure probabilities(3/5)

- **An upper bound on** $\mathbb{P}\{\mathrm{E}(\mathbf{Y},\mathcal{J},\delta)\}$.

$$\mathbb{P}\{\mathrm{E}(\mathbf{Y},\mathcal{J},\delta)\} \le \left(1-2t\lambda(\mathcal{J})\right)^{-\frac{S(M-K)}{2}} \exp\left(-t\left(SM\delta + S(M-K)\sigma_n^2\right)\right)$$
$$:= g(t)$$

  - Note that the function $g(t)$ is convex.
  - Thus, we can find its minimum value by solving

$$t^* = \arg\min_t g(t) \text{ subject to } t<0.$$

  - Finally, the upper bound is

$$\mathbb{P}\{\mathrm{E}(\mathbf{Y},\mathcal{J},\delta)\} \le g(t^*;W_2)$$
$$= \left(\frac{\sigma_n^2 + \frac{M\delta}{M-K}}{\lambda_{\min}(\mathbf{R}_{\mathcal{J}})}\right)^{\frac{S(M-K)}{2}} \exp\left(-\frac{S(M-K)}{2}\left(\frac{\sigma_n^2 - \lambda_{\min}(\mathbf{R}_{\mathcal{J}}) + \frac{M\delta}{M-K}}{\lambda_{\min}(\mathbf{R}_{\mathcal{J}})}\right)\right)$$

Upper bounds on the failure probabilities(4/5)

- **Lemma 3:** Let $\mathcal{I}$ be the joint support set and $M > K$. Then, we have for any $\delta > 0$,

$$\mathbb{P}\left\{\mathrm{E}\left(\mathbf{Y},\mathcal{I},\delta\right)^c\right\} \le 2\left(1+\frac{M\delta}{\left(M-K\right)\sigma_n^2}\right)^{\frac{S(M-K)}{2}} \exp\left(-\frac{S\left(M-K\right)}{2}\frac{M\delta}{\left(M-K\right)\sigma_n^2}\right)$$
$$:= p_1\left(S,M,K,\sigma_n^2,\delta\right)$$

- **Lemma 4:** Let $\mathcal{J}$ be not the joint support set and $M > K$. Then, we have

$$\mathbb{P}\left\{\mathrm{E}\left(\mathbf{Y},\mathcal{I},\delta\right)^c\right\} \le \left(\frac{\sigma_n^2+\frac{M\delta}{M-K}}{\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right)}\right)^{\frac{S(M-K)}{2}} \exp\left(-\frac{S\left(M-K\right)}{2}\left(\frac{\sigma_n^2-\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right)+\frac{M\delta}{M-K}}{\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right)}\right)\right)$$
$$:= p_2\left(S,M,K,\sigma_n^2,\delta,\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right)\right).$$

for any $0<\delta<\frac{M-K}{M}\left(\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right)-\sigma_n^2\right).$

Upper bounds on the failure probabilities(5/5)

- We have presented the upper bounds on the failure probabilities.
- With these upper bounds, we finally have

$$\mathbb{P}\{failure\} \le p_1\left(S, M, K, \sigma_n^2, \delta\right) + \sum_{\forall \mathcal{J} \ne \mathcal{I}, |\mathcal{J}| = K} p_2\left(S, M, K, \sigma_n^2, \delta, \lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right)\right)$$

$$\le p_1\left(S, M, K, \sigma_n^2, \delta\right) + \binom{N}{K} \max_{\forall \mathcal{J} \ne \mathcal{I}, |\mathcal{J}| = K} \left(p_2\left(S, M, K, \sigma_n^2, \delta, \lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right)\right)\right).$$

- The main results in this talk were obtained by finding conditions for vanishing the upper bound on $\mathbb{P}\{failure\}$.

Main results

- **Theorem 1:** Suppose that both $M > K$ and

$$0 < \delta < \min_{\forall \mathcal{J} \neq \mathcal{I}, |\mathcal{J}| = K} \left(\left(1 - K/M\right) \left(\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right) - \sigma_n^2 \right) \right)$$

are met. Then, $\mathbb{P}\{\mathit{failure}\}$ linearly converges to zero as S goes to infinity.

  – a) It states that taking more measurement vectors can reduce effects of noises.

- **Theorem 2:** Suppose that both $M = \Omega\left(\frac{K}{S}\log\frac{N}{K} + K\right)$ and

$$0 < \delta < \min_{\forall \mathcal{J} \neq \mathcal{I}, |\mathcal{J}| = K} \left(\left(1 - K/M\right) \left(\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right) - \sigma_n^2 \right) \right)$$

are met. Then, $\mathbb{P}\{\mathit{failure}\}$ converges to zero as N goes to infinity.

  – a) It shows an inversion relation between M and S.

Comparisons with other leading results

	Model	Random variables	Sufficient conditions	Necessary conditions
ours	$\mathbf{y}_i = \mathbf{F}_i\mathbf{x}_i + \mathbf{n}_i$, $i = 1, 2, \cdots, S$	$\mathbf{F}_i$ and $\mathbf{n}_i$	$M = \Omega\left(\frac{K}{S}\log\frac{N}{K} + K\right)$ $M > K$	?
[Baraniuk13]	$\mathbf{y}_i = \mathbf{F}_i\mathbf{x}_i$, $i = 1, 2, \cdots, S$	$\mathbf{F}_i$	$M > K$	$M > K$
[Nehorai09]	$\mathbf{y}_i = \mathbf{F}\mathbf{x}_i + \mathbf{n}_i$, $i = 1, 2, \cdots, S$	$\mathbf{F}, \mathbf{x}_i$ and $\mathbf{n}_i$	$M = \Omega\left(K\log\frac{N}{K}\right)$	$M = \Omega\left(K\log\frac{N}{K}\right)$
[Rao13]	$\mathbf{y}_i = \mathbf{F}\mathbf{x}_i + \mathbf{n}_i$, $i = 1, 2, \cdots, S$	$\mathbf{F}$ and $\mathbf{n}_i$	$M > \frac{\log(N)}{c(\mathbf{W})}$	$M < \frac{\log(N)}{c(\mathbf{W})}$

where $\mathbf{W}$ is a matrix whose elements are correspoding to the non-zero coefficients of $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_S]$,

$$c(\mathbf{W}) := \min_{\mathcal{J}\subseteq[K],\mathcal{J}\neq\varnothing} \frac{\log\left(\det\left(\mathbf{I}_S + \mathbf{W}_{\mathcal{J}}^T\mathbf{W}_{\mathcal{J}}\,\sigma_\sigma^2/\sigma_n^2\right)\right)}{2|\mathcal{J}|}.$$

a. We currently do not provide a necessary condition.

b. The other sufficient conditions explicitly do not provide relations between M and S.

c. The necessary and sufficient conditions by Baraniuk et al. are not asymptotical.

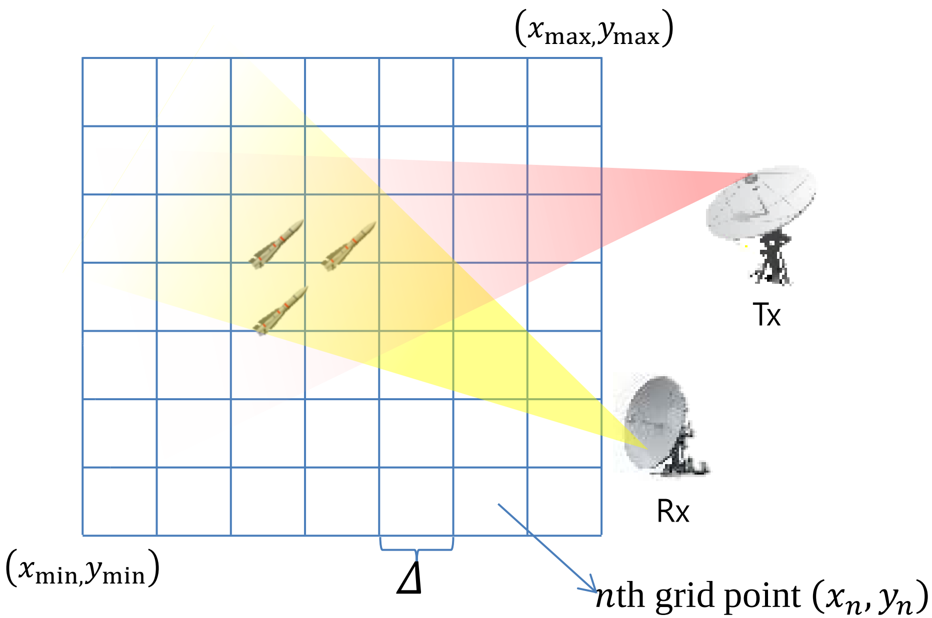
d. The necessary and sufficient conditions by Nehorai et al. and Rao et al. are asymptotical with respect to N.

Summary

- Taking more measurement vectors can reduce effects of noises.
  - $M > K$ measurements per each measurement vector are sufficient for joint support set reconstruction.
  - In particular, $K + 1$ measurements per each measurement vector are enough when the number of measurement vectors is sufficiently large.

- The number of measurements can be inversely decreased as taking more measurement vectors.
  - $M = \Omega\left(\frac{K}{S}\log\frac{N}{K} + K\right)$ measurements per each measurement vector are sufficient for joint support set reconstruction.
  - In particular, $M = \Omega\left(K\right)$ measurements per each measurement vector are enough when $S > K\log\frac{N}{K}$.

Compressive Sensing Radar



- Δ denotes a range resolution
- The number of targets is assumed to be less than the number of grid points

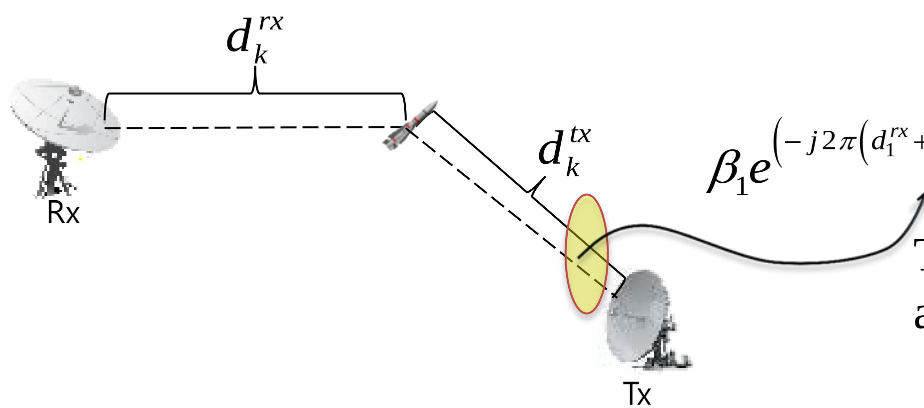
- **Research Goal**
  - Apply the theory of compressive sensing to radar problems where we aim to estimate the position, velocity, azimuth angle of multiple targets.
- **Approaches**
  - Discretize a radar model.
  - Transformed this radar model to a linear model.
  - Use compressive sensing algorithms to this linear model.

Compressive Sensing Radar model (1/5)

- **System model**
  - Let $x(t)$ be the transmitted signal by Tx.
  - Let $z_k(t)$ be the reflected signal by the kth target.
  - The received signal by Rx can be expressed as

$$\tilde{y}(t) = \sum_{k=1}^{K} \beta_k z_k\left(t - d_k^{rx}/c\right)$$
$$= \sum_{k=1}^{K} \beta_k x\left(t - \left(d_k^{rx} + d_k^{tx}\right)/c\right) \exp\left(j 2\pi f_c \left(t - \left(d_k^{rx} + d_k^{tx}\right)/c\right)\right)$$

where β_k is a constant proportional to the radar cross section value of the kth target.



$$\beta_1 e^{\left(-j 2\pi \left(d_1^{rx} + d_1^{tx}\right)/\lambda\right)} x\left(t - \left(d_k^{rx} + d_k^{tx}\right)/c\right)$$

The equation gives relations among Tx, Rx and the kth target.

Compressive Sensing Radar model (2/5)

– Rx demodulates the received signal.

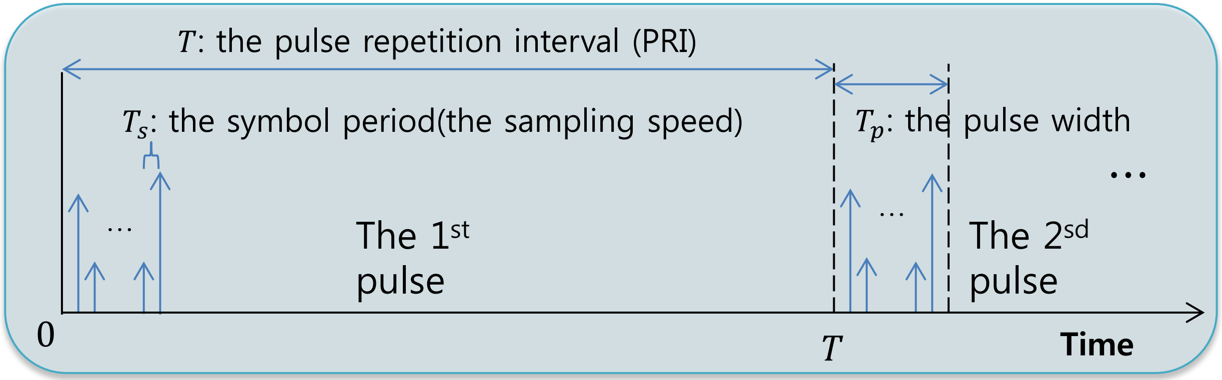
– Then, the demodulated signal can be expressed as

$$\begin{aligned} y(t) &= \tilde{y}(t)\exp(-j2\pi f_c t) \\ &= \sum_{k=1}^{K} \beta_k x\left(t-\left(d_k^{rx}+d_k^{tx}\right)/c\right) e^{\left(-j2\pi\left(d_k^{rx}+d_k^{tx}\right)/\lambda\right)} e^{j2\pi f_k t} \end{aligned}$$

Notations	Meaning
$t_{delay} = (d_k^{rx} + d_k^{tx})/c$	The delay of the transmitted signal
$e^{(-j2\pi(d_k^{rx}+d_k^{tx})/\lambda)}$	The phase delay
$e^{j2\pi f_k t}$	The Doppler term

Compressive Sensing Radar model (3/5)

– Assume that the pulse width of the transmitted signal is T_p and the transmitted signal consists of L sequences.



– Let $\tau(n) = \lceil t_{delay}(n) \rceil$ be the time delay of the demodulated signal through the nth grid point.

– Then, the discretized demodulated signal can be expressed as

$$\mathbf{y} = \begin{bmatrix} y(t=0) \\ \vdots \\ y(t=(L+\tau_{\max}-1)T_s) \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{K} \beta_k x\left(-\lceil t_{delay}(k) \rceil\right) e^{\left(-j2\pi\left(d_k^{rx}+d_k^{tx}\right)/\lambda\right)} \\ \vdots \\ \sum_{k=1}^{K} \beta_k x\left((L+\tau_{\max}-1)T_s - \lceil t_{delay}(k) \rceil\right) e^{\left(-j2\pi\left(d_k^{rx}+d_k^{tx}\right)/\lambda\right)} \end{bmatrix}$$

where $\tau_{\max} = \max_n \tau_n$.

Compressive Sensing Radar model (4/5)

- To use the theory of compressive sensing, the discretized demodulated forms a matrix-vector multiplication.
- At time t, the discretized demodulated signal can be expressed as

$$y(t) = \sum_{n=1}^{N} s_n x\left(T_s - t_{delay}(n)\right)\left(-j2\pi\left(d_n^{rx} + d_n^{tx}\right)/\lambda\right)$$

$$= \left[x\left(t - t_{delay}(1)\right) e^{\left(-j2\pi\left(d_1^{rx}+d_1^{tx}\right)/\lambda\right)} \quad \cdots \quad x\left(t - t_{delay}(N)\right) e^{\left(-j2\pi\left(d_N^{rx}+d_N^{tx}\right)/\lambda\right)} \right] \begin{bmatrix} s_1 \\ \vdots \\ s_N \end{bmatrix}$$

$$= \mathbf{f}_t \mathbf{s}$$

where $s_n = \begin{cases} \beta_k & \text{if the } k\text{th target is at } (x_n, y_n) \\ 0 & o.w. \end{cases}$.

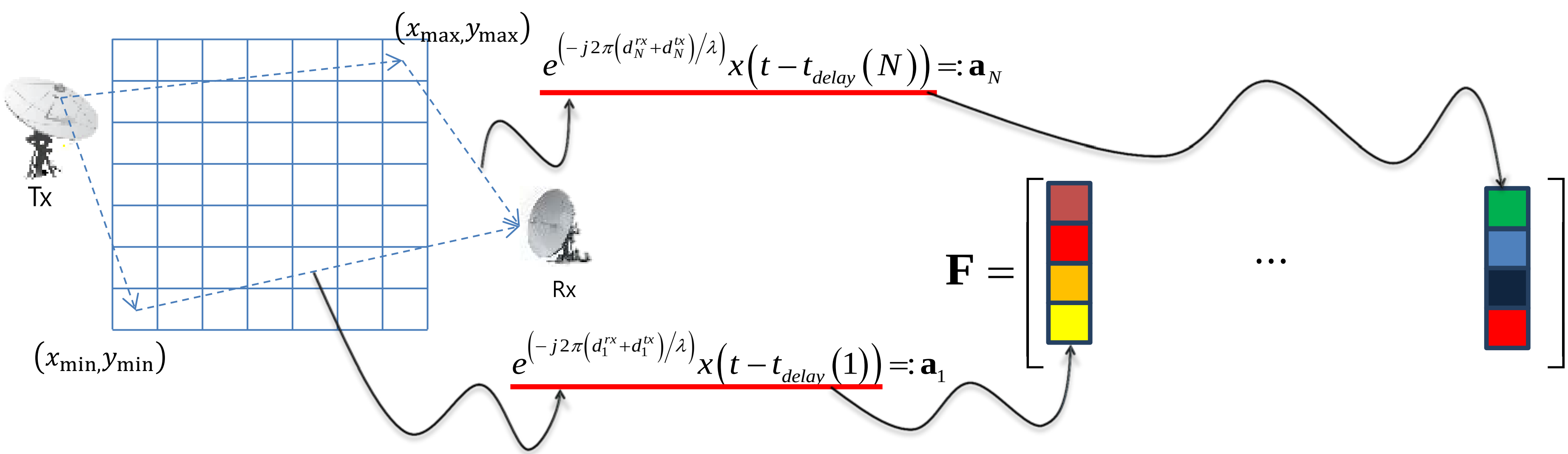
	Size	Meaning
$\mathbf{s}$	$N \times 1$	1. The number of non-zero elements is equivalent to the number of targets. 2. Suppose that the nth element is non-zero. It means that there is a target in the nth grid point.
$\mathbf{f}_t$	$1 \times N$	1. Physical responses among Tx, Rx and the gird points at time t. 2. In particular, the ith element is physical response among Tx, Rx and the ith grid point at time t.

Compressive Sensing Radar model (5/5)

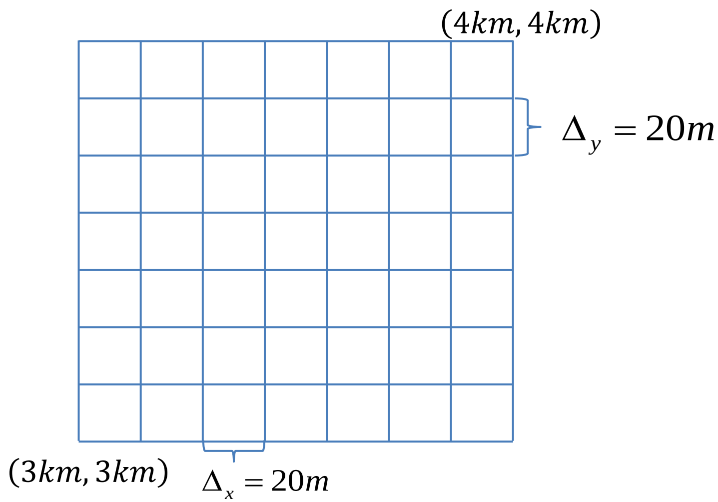
– Thus, the discretized demodulated signal at Rx can be expressed as

$$\mathbf{y} = \begin{bmatrix} \mathbf{f}_0 \\ \vdots \\ \mathbf{f}_{L+\tau_{\max}-1} \end{bmatrix} \mathbf{s} = \mathbf{F}\mathbf{s}.$$

where $\mathbf{F}$ is a $(L + \tau_{\max}) \times N$ matrix.



Computer Simulation (1/2)

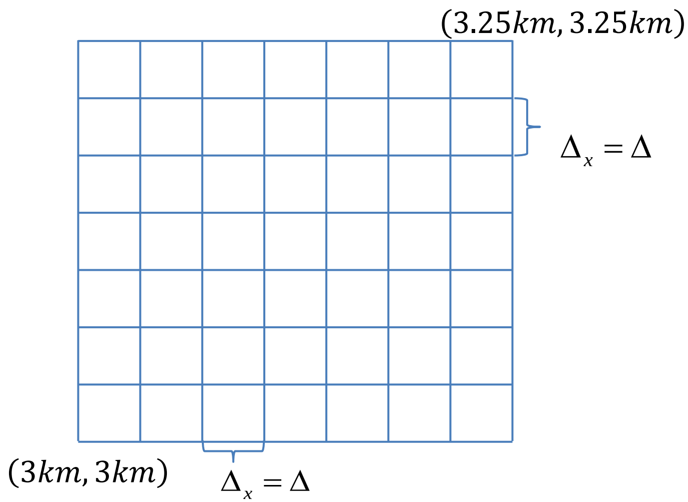


1. The number of targets is 20
2. There is no noise
3. Each pulse consists of 128 sequences
4. $T_s = 5 \times 10^{-8}$(bandwidth is 20MHz)
5. $f_c = 10GHz$

	$M_r = M_t = 1$	$M_r = M_t = 2$	$M_r = M_t = 3$	$M_r = M_t = 4$
Results	failure	failure	**Success**	**Success**
the total number of samples	128	256	384	512

1. M_r : the number of Rx
2. M_t : the number of Tx

Computer Simulation (2/2)



1. The number of targets is 20
2. There is no noise
3. Each pulse consists of 128 sequences
4. $T_s = 5 \times 10^{-8}$(bandwidth is 20MHz)
5. $f_c = 10GHz$
6. 4 Tx and 4 Rx.

The range resolution	$\Delta = 7.5m$	$\Delta = 5m$	$\Delta = 3m$	$\Delta = 1m$
성공 여부	Success	Success	Success	**Failure**

– The range resolution by the matched filtered radar is $\Delta r = \frac{c}{2B} = 7.5m$.

– The above results shows the superiority of the compressive sensing radar.

References

[Braniuk09] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarotham and R. G. Baraniuk, "Distributed Compressive Sensing", arXiv.org/abs/0901.3403

[Scharf91] L. L. Scharf, Statistical Signal Processing: Detection, Estimation, and Time Series Analysis, Prentice Hall, 1991

[Baraniuk13] M. F. Duarte, M. B. Wakin, D. Braon, S. Sarvotham, and R. G. Baraniuk, "Measurement Bounds for Sparse Signal Ensembles via Graphical Models," IEEE Transaction on Information Theory, 2013

[Nehorai10] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," IEEE Transaction 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

Compressive Sensing for Imaging through turbid media

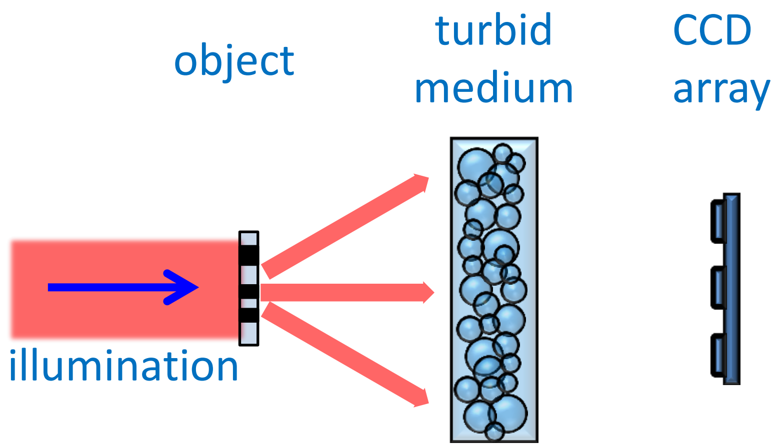
Hwanchol Jang

INFONET Lab., GIST, KOREA

2014 INFONET WORKSHOP
July 8th, 2014

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



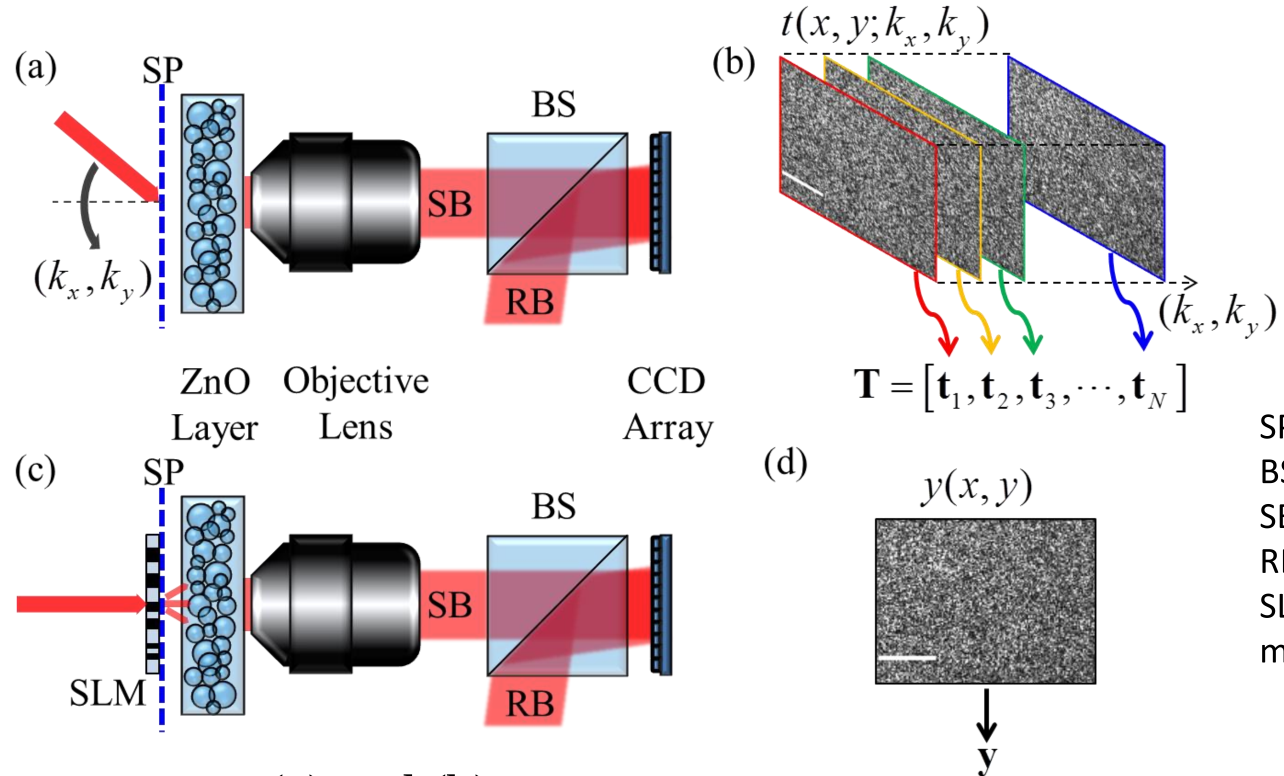
- Due to the multiple scattering, the outgoing object waves are spatially scrambled and become a speckle field (SF) at an observation plane
- For image recovery, the multiple scattering should be suppressed; the object image should be recovered

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

- Object: $o(x,y) = \sum_{k_x} \sum_{k_y} a(k_x,k_y)\, p(x,y;k_x,k_y)$

  where $o(x,y)$ is the object wave, $a(k_x,k_y)$ and $p(x,y;k_x,k_y)$ are the *angular spectrum* and the *plane wave*, respectively, of the object wave with the propagation angle with k_x and k_y

  - k_x and k_y are the wave vector components of the wave in the x and y directions ($k_x/2\pi = \sin\theta_x/\lambda$).
- TM
  - $t(x,y;k_x,k_y)$, the response of the medium to $p(x,y;k_x,k_y)$, for all (k_x,k_y) are measured and recorded

Compressive Sensing for Imaging through turbid media

- Object SF: $y(x,y)=\sum_{k_x}\sum_{k_y} a(k_x,k_y)t(x,y;k_x,k_y)+n(x,y)$

  where $n(x,y)$ is the additive noise
  - The response of the medium to the object wave
- Estimate: $\hat{a}(k_x,k_y)$
  - With $t(x,y;k_x,k_y)$ for all (k_x,k_y) considered, the estimate of the angular spectrum, $\hat{a}(k_x,k_y)$, is made for a given $y(x,y)$.
- Recovery: $\hat{o}(x,y)$
  - Using the angular spectrum estimate, the recovery of the object image is made

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

- The recovered object wave

$$\hat{o}(x,y) = \sum_{k_x}\sum_{k_y}\hat{a}(k_x,k_y)\,p(x,y;k_x,k_y)$$
$$= o(x,y) + \sum_{k_x}\sum_{k_y}e(k_x,k_y)\,p(x,y;k_x,k_y)$$

  where $e(k_x,k_y) := \hat{a}(k_x,k_y) - a(k_x,k_y)$ is the angular spectrum estimation (ASE) error for a given angle (k_x,k_y)

  – the speckle in the recovered object image is made directly from the ASE error.

- With this finding, we come up with a new speckle suppression approach *via* reducing the ASE error rather than the ensemble averaging approach.

Compressive Sensing for Imaging through turbid media

- CS has received great deal of interests recently for it enables correct estimation of a signal even for underdetermined measurement systems ($M \leq N$) [Bruckstein2009], [Candès2011] .
- For successful applications of CS, there are two key conditions to be met
  - Compressibility: the signal is well approximated with a small number of nonzero elements, say K where $K \ll N$.
  - Isometry: the measurement matrix preserves the energy of a signal well after the signal is transformed by the matrix.

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 K largest element approximation (K-LEA)
  - Optimal among all K element approximations in terms of the L1 norm and the L2 norm error senses [Candès2011]
    - The K -LEA can be made by an *oracle* estimator which is assumed to know all the coefficients of elements of the original signal exactly (by setting all elements to zero while keeping the K largest elements intact)
- Sparse signal estimation (SSE) in the CS framework
  - It behaves like the oracle estimator [Candès2011]
  - For a Gaussian TM, about $M = O(K\log(N/K))$ number of measurements shall be good enough [Candès2011]
    - PC requires at least $M=N$ for this estimation
  - This means that the SSE offers the perfect ASE in the TM-based recovery when the angular spectrum of object wave has K or less nonzero elements; with $M = O(K\log(N/K))$
  - For object waves with more than K nonzero angular spectrum elements, if the angular spectrum elements other than the K largest elements are insignificant, the ASE error of SSE is negligible.

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]

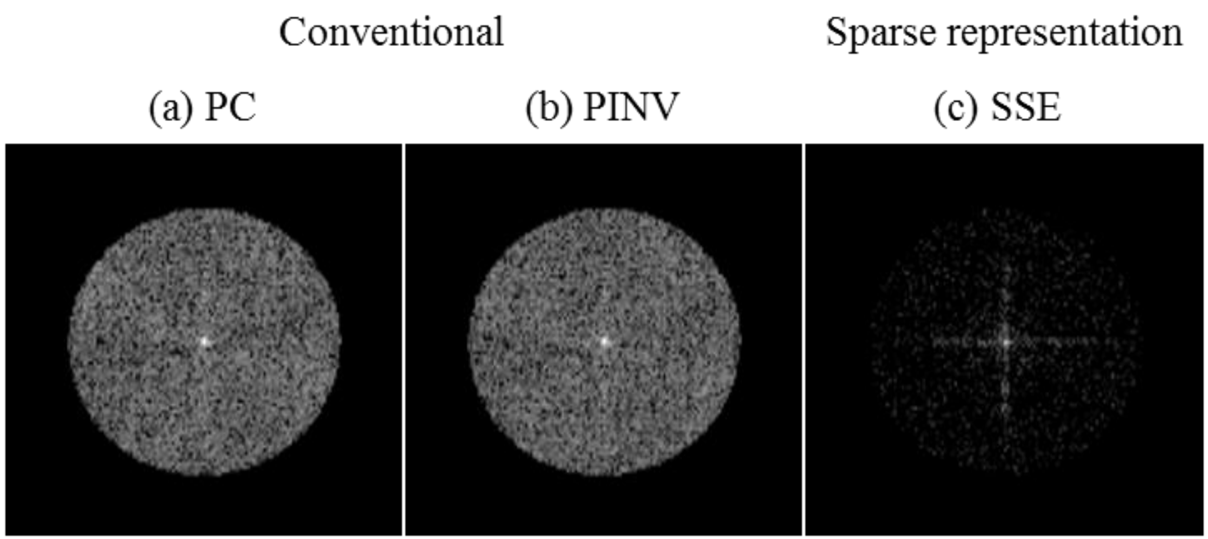
$$\hat{\mathbf{a}}_{\mathrm{SSE}} = \arg\min_{\mathbf{a}} \|\mathbf{a}\|_1 \ \text{ subject to } \mathbf{y} = \mathbf{T}\mathbf{a}.$$

  - 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

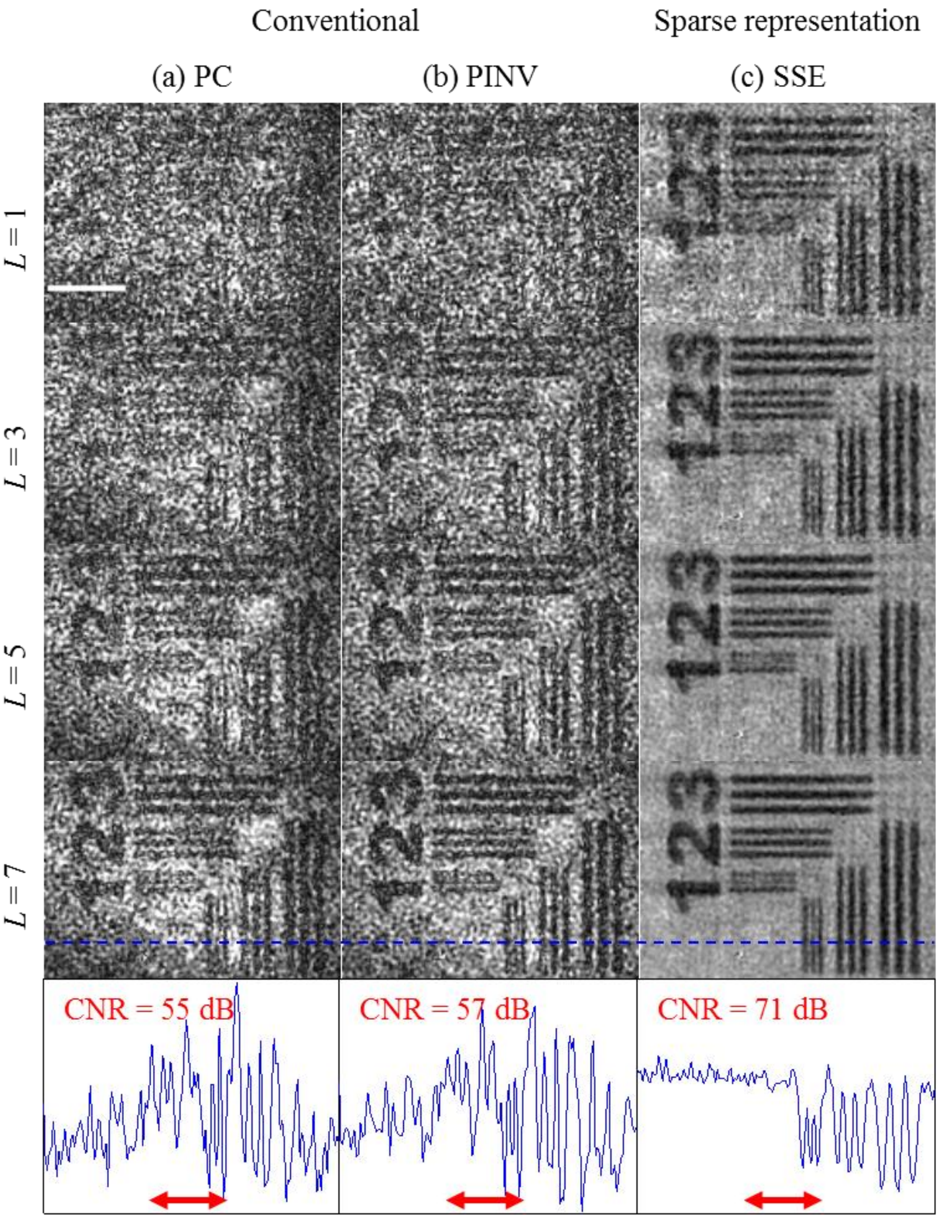


Estimated angular spectra using (a) PC, (b) PINV, and (c) SSE, respectively. Here, M = 4389, N = 20000. All angular spectra are represented in log scale for better visibility.

- 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.



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.

Resolution improvements on

1. miniature spectrometers
2. compound eye type of cameras

by compressive sensing

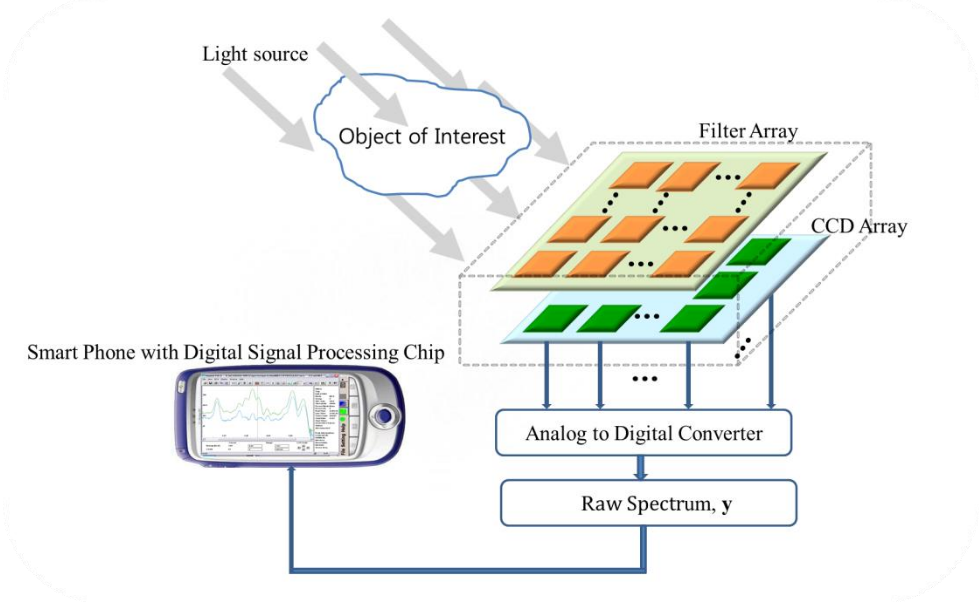
Woongbi Lee

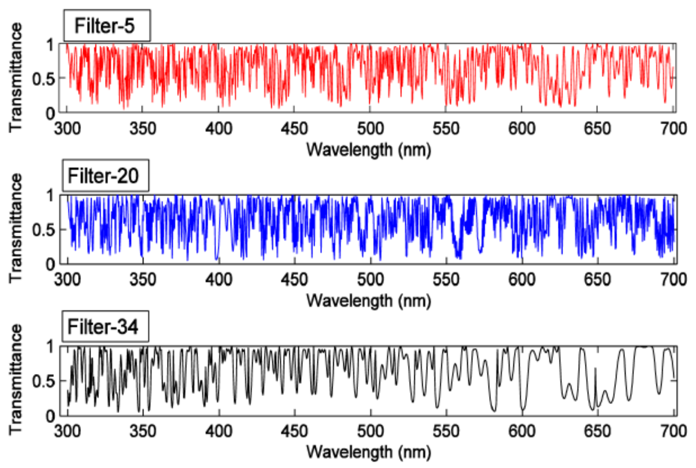
INFONET WORKSHOP, 2014

July 10th, 2014

Miniature Spectrometers

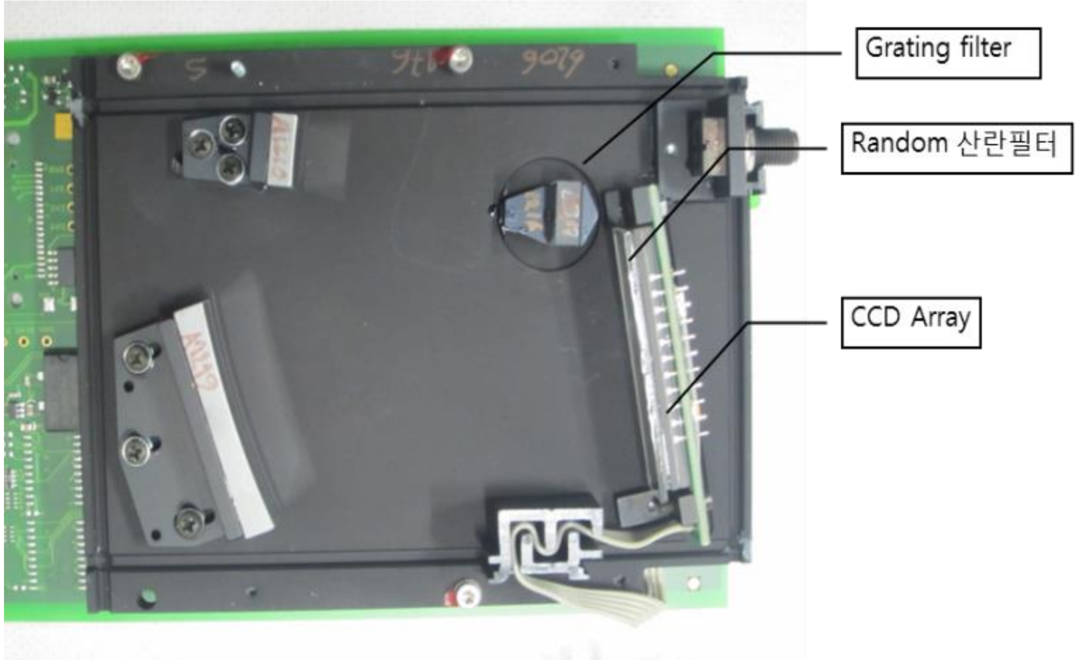
- We have achieved resolution improvements beyond the resolution limit by
  - Digital Signal Processing (2012)
  - Filters having random transmittance (2013)





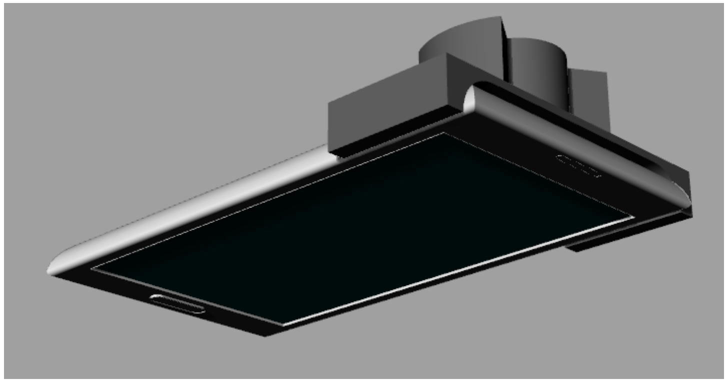
Prototype spectrometer

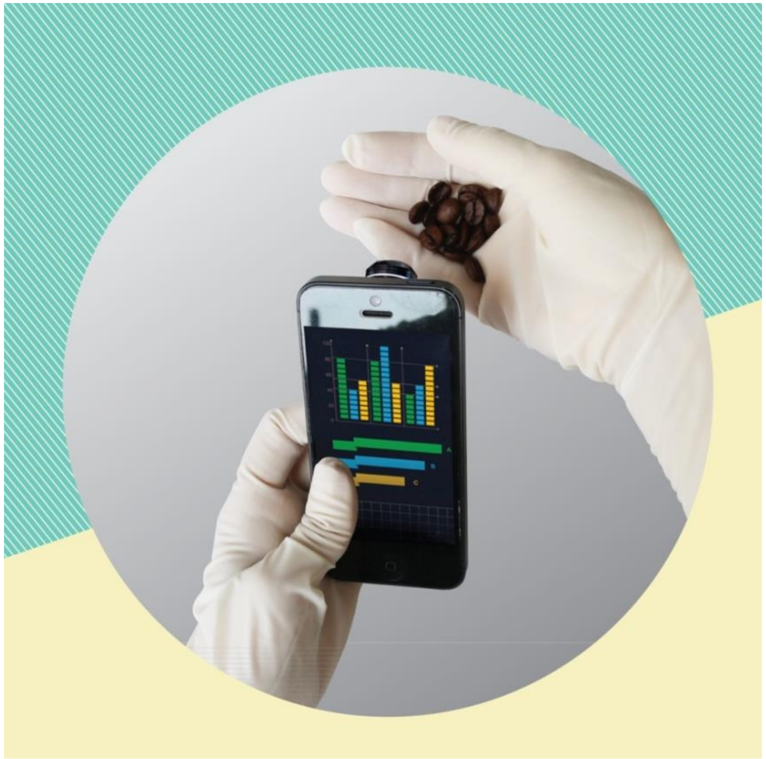
❖ Prototype using random scattering filter (2012~2013)



Future works

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

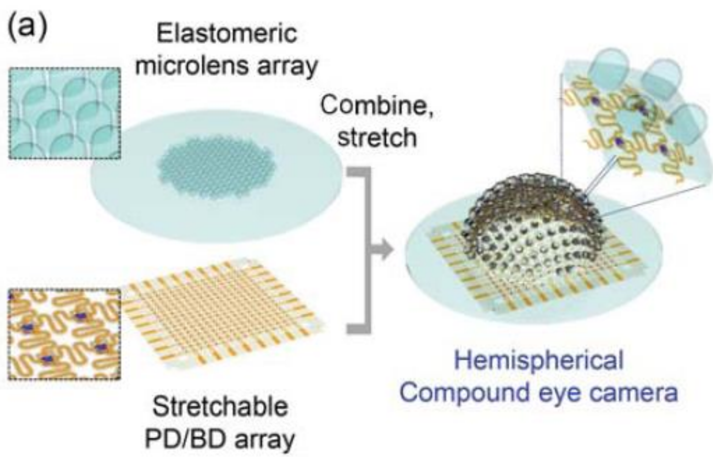


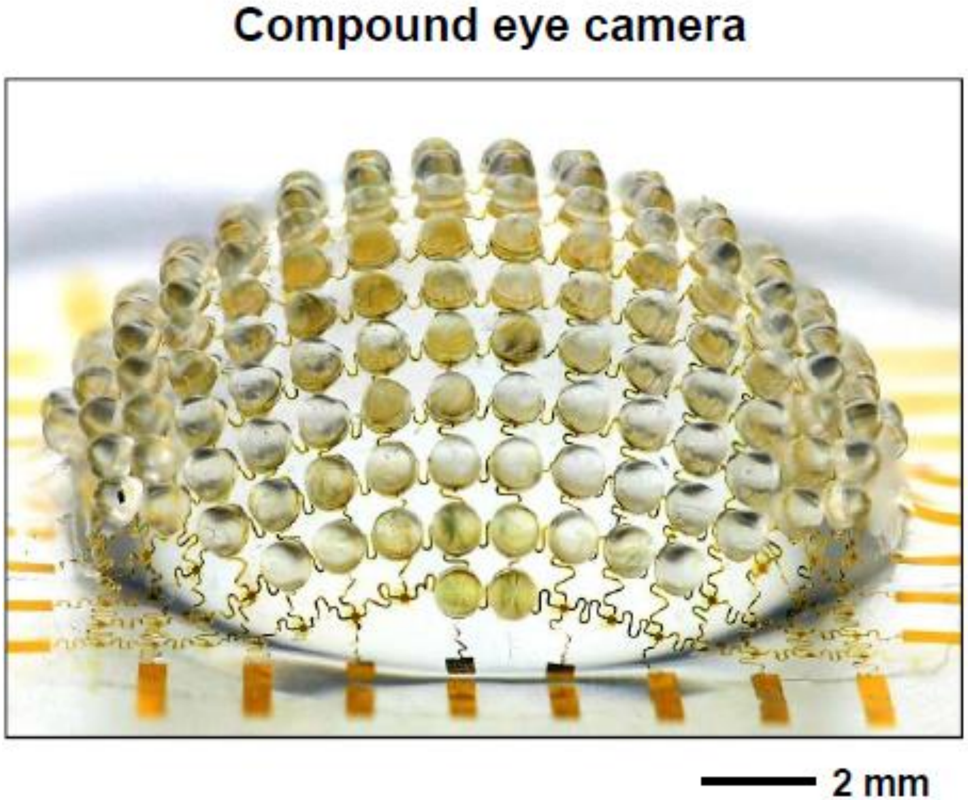


Hemispherical Apposition Compound Eyes

- Implemented by strechable microlens array and photodiodes

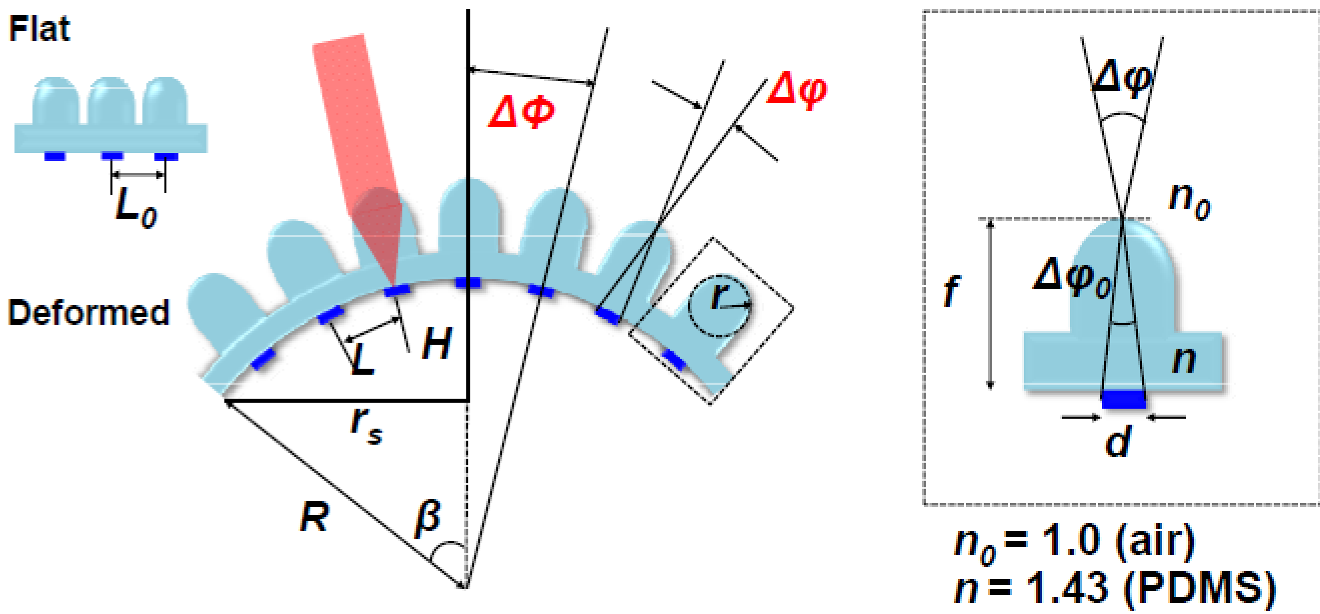
- Limitation: 180 pixels (16x16 photo diodes)





Optical Design

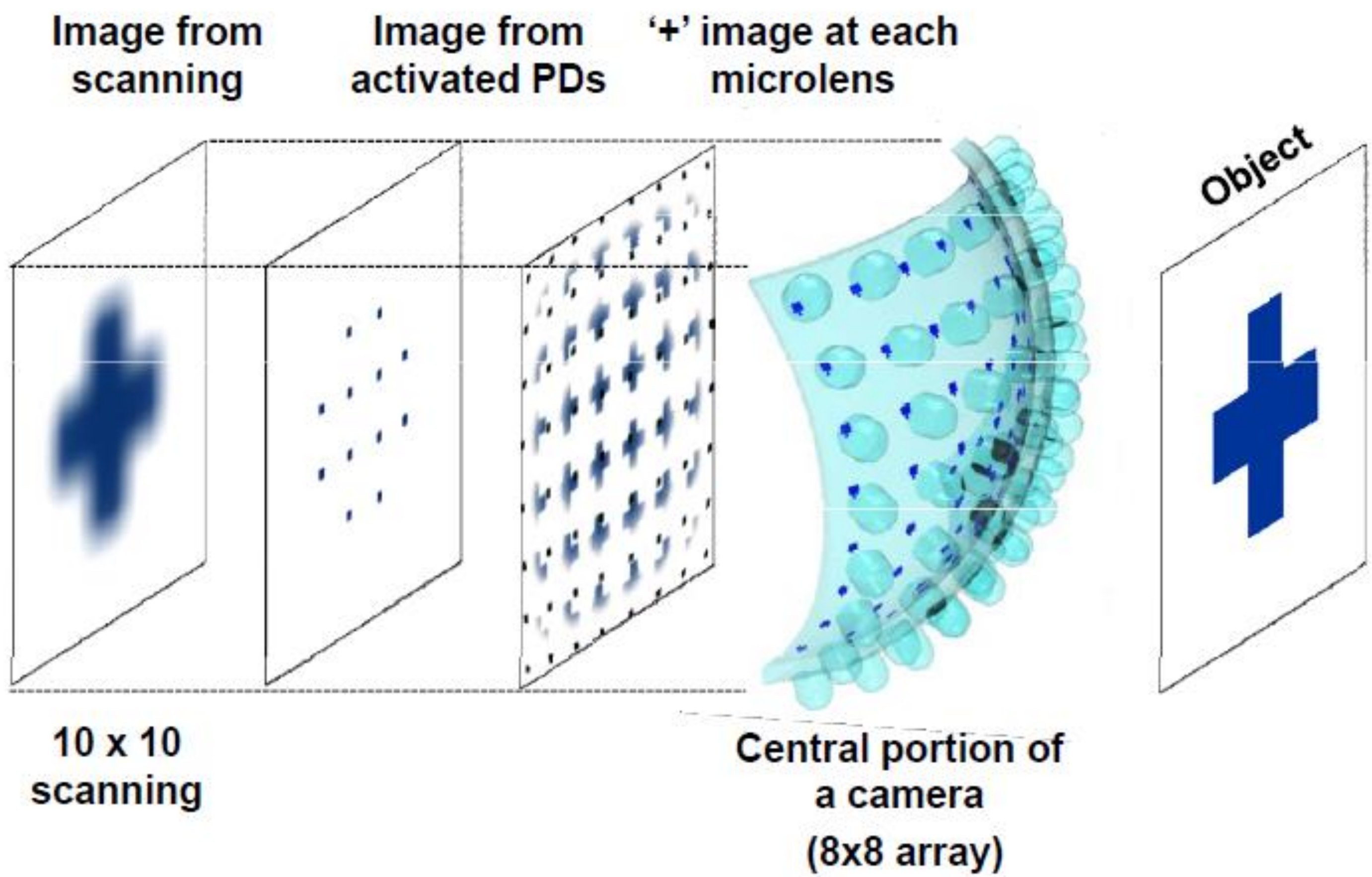
- Key parameters:
  - Acceptance angle
  - Inter-ommatidial angle



Inter-ommatidial angle (*ΔΦ*) **>** Acceptance angle (*Δφ*)

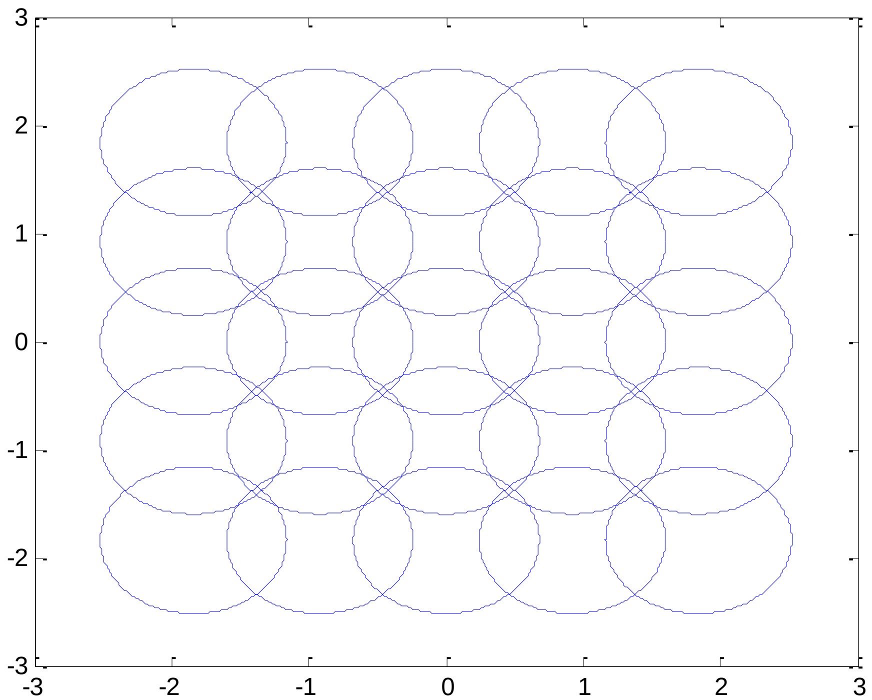
$$\Delta\Phi = \frac{\rho L_0}{R}, \ \rho = \frac{2R\beta}{2r_s} \qquad \Delta\varphi = \frac{d}{f}, \ f = \frac{rn}{n-1}$$

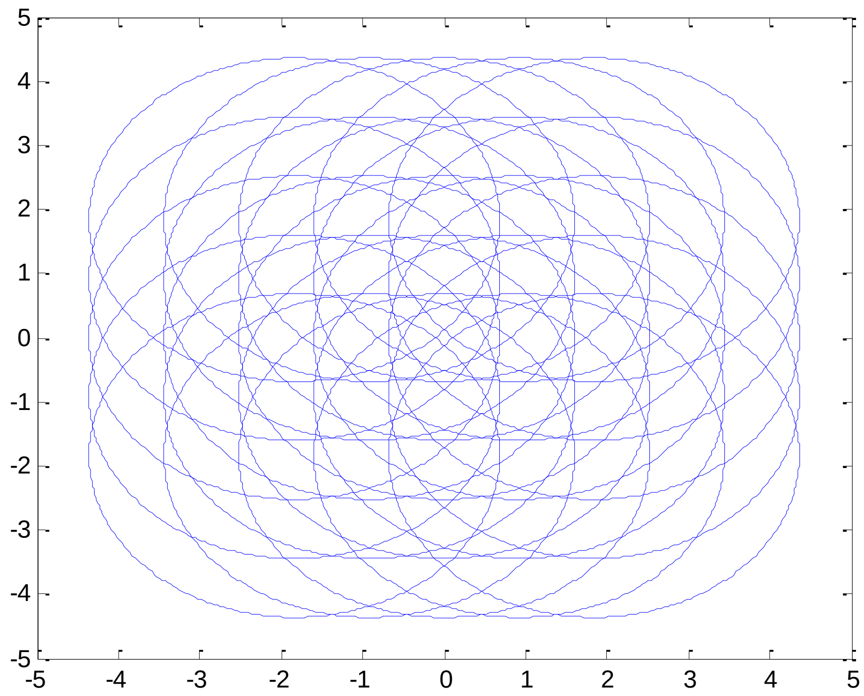
Operating Principle



Sensors on the plane

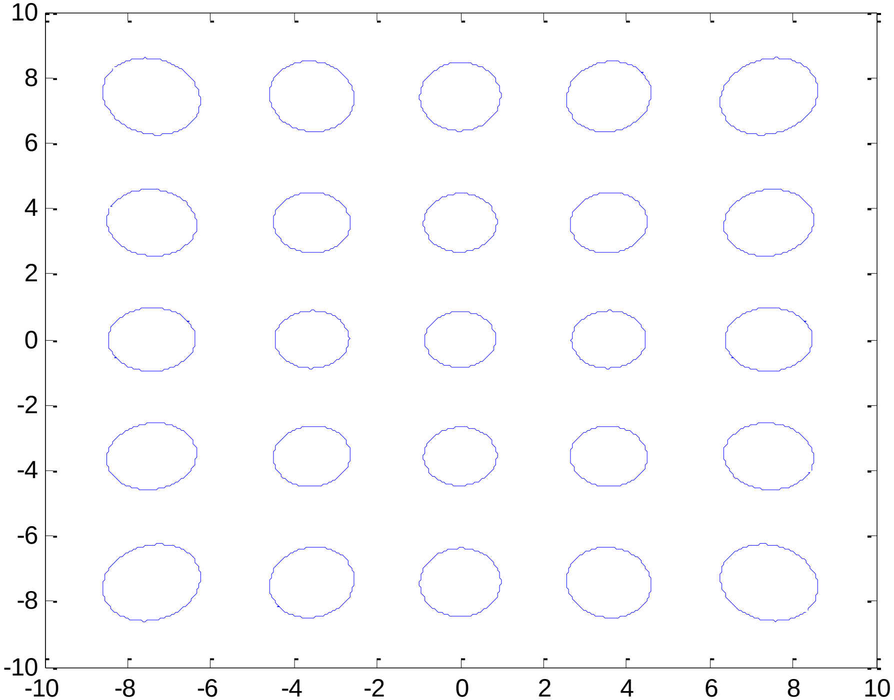
- 5x5 Sensors with acceptance angle =9.7 and 35
- Distance from Image plane and sensor plane is 10mm

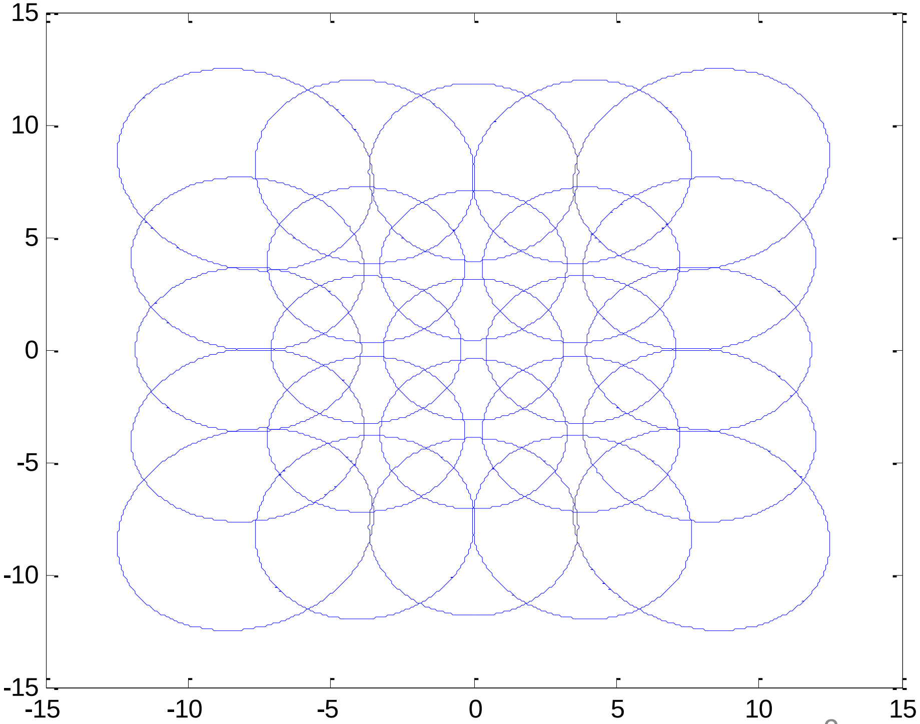




Sensors on the hemisphere

- 5x5 sensors with acceptance angle =9.7 and 35
- Distance between Image plane and center of the hemisphere is 10mm





Future works

- Resolution Improvements
- Depth estimation

Development and Evaluation of Sparse Representation based Classification Method for Motor Imagery based BCI systems

Younghak Shin

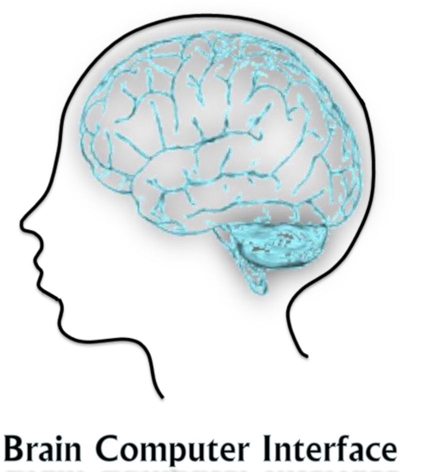
INFONET Lab., GIST, KOREA

2014 INFONET WORKSHOP

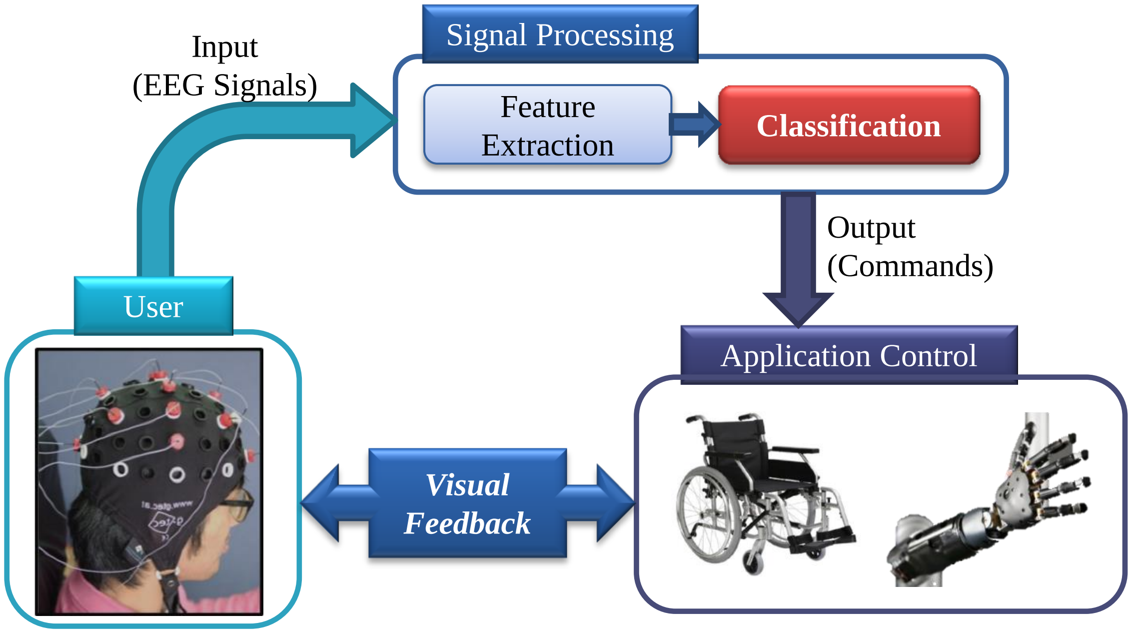
July 8th, 2014

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

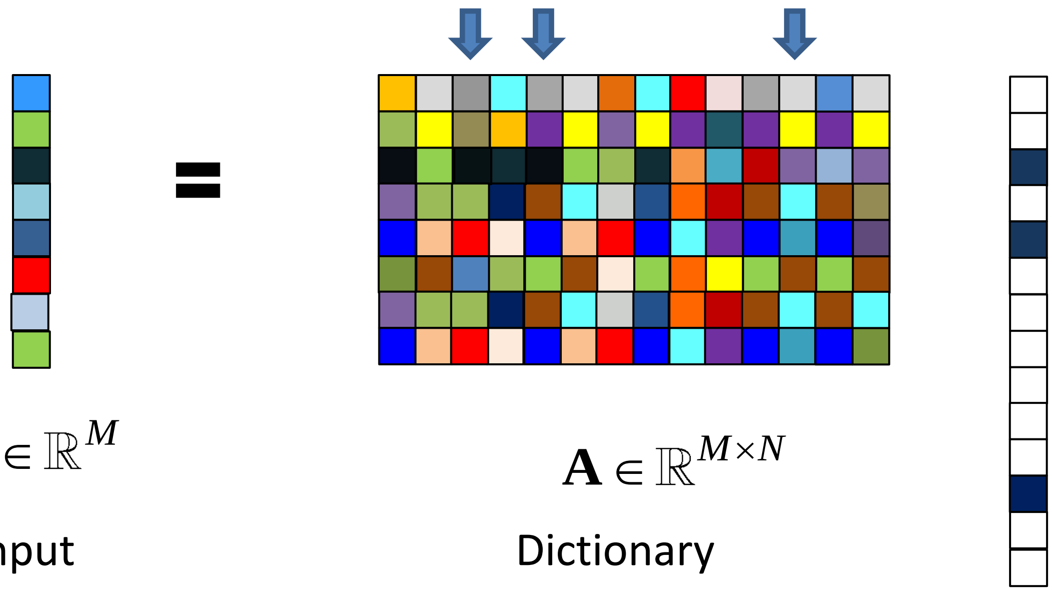


- BCI is a novel communication and control channel between person and external world.
- The BCIs allows communication only using user's intention or imagination instead brain's normal output pathways of peripheral nerves and muscle
- In the BCIs, classification is needed to transform the extracted feature of a user's intention into a computer command to control the external device.
- However, EEG signals are very noisy and have non-stationarity 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].



$\mathbf{y} \in \mathbb{R}^{M}$

Input

$\mathbf{A} \in \mathbb{R}^{M \times N}$

Dictionary

$\mathbf{x} \in \mathbb{R}^{N \times 1}$

Sparse coefficient

y = A x

Sparse Representation (SR)

- The problem of SR is to find the coefficient $\mathbf{x} \in \mathbb{R}^{N\times 1}$:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

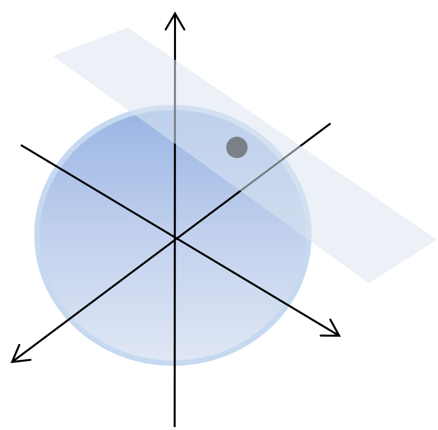
where, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is known over-complete dictionary $(M \ll N)$

$\mathbf{y} \in \mathbb{R}^{M}$ is measured signal $\|\mathbf{x}\|_0$ denotes the L0 norm.

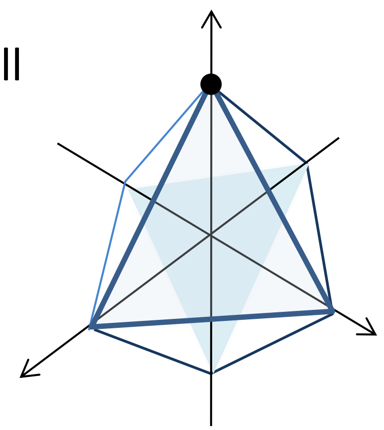
- Solving this under-determined problem is NP hard.
- Recently developed Compressive Sensing theory [Donoho 2006] reveals that if a solution is sparse enough, L1 norm solution is equivalent to the L0 norm solution.

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

L2 ball

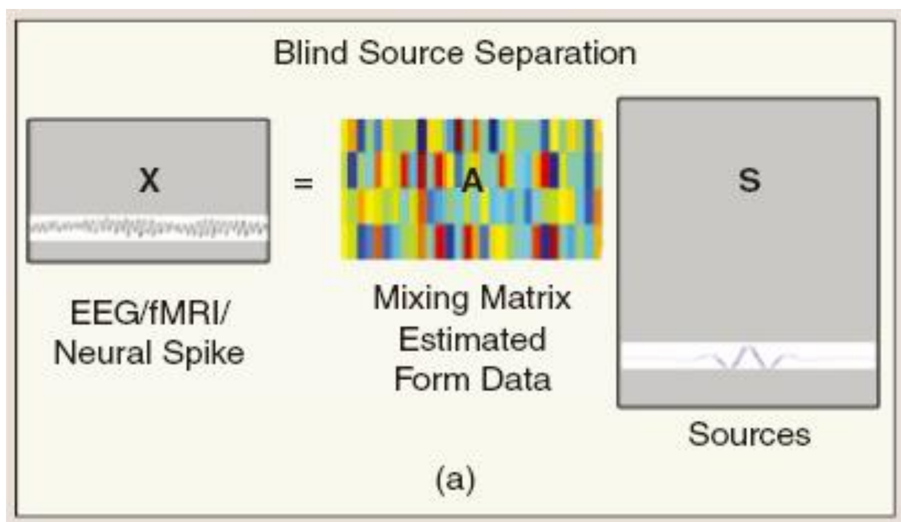


L1 ball



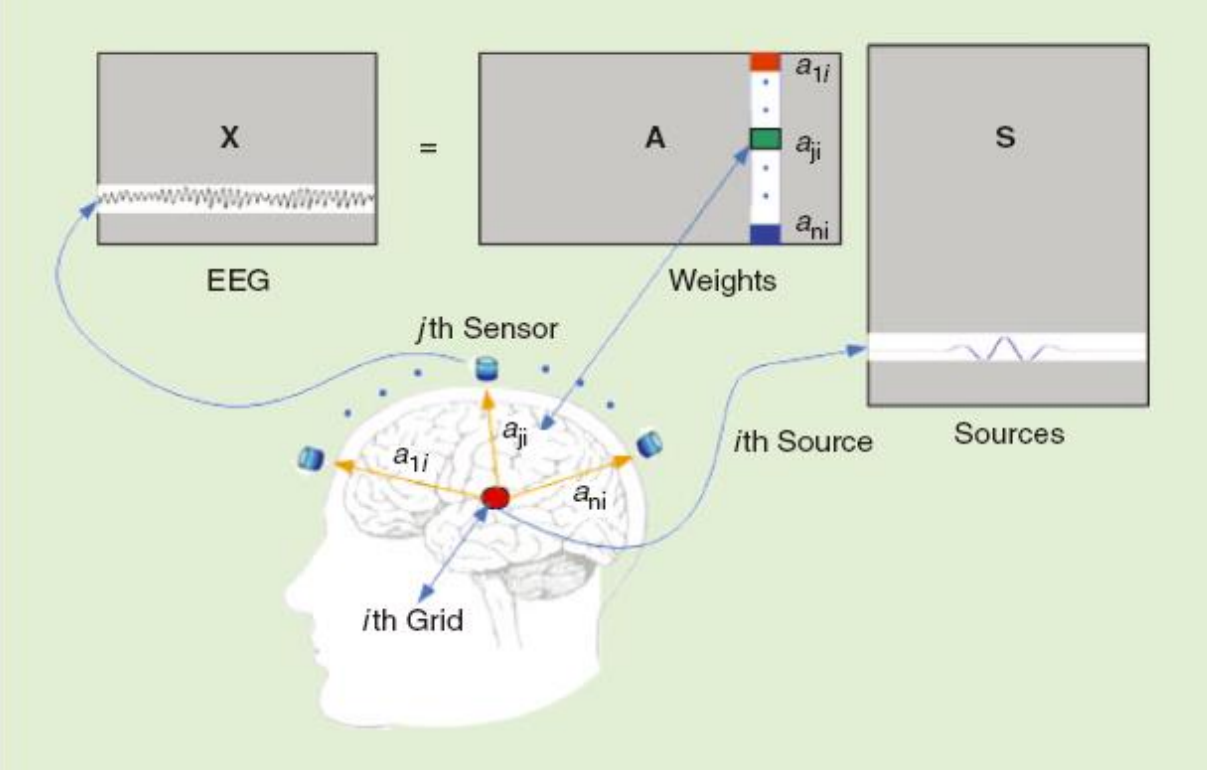
Sparse representation for brain signal processing [Yuanqing 2014, ieee sig. proc. magazine]

- Blind source separation
  - EEG signals can be considered as the linear mixtures of unknown sources with an unknown mixing matrix.
  - The brain sources can be assumed to be sparse in a domain such as the time domain 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.



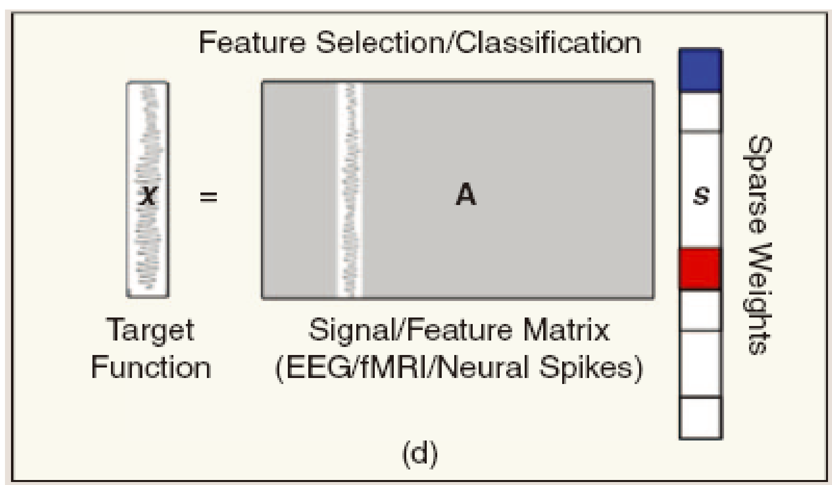
Sparse representation for brain signal processing [Yuanqing 2014, ieee sig. proc. magazine]

- 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 [Yuanqing 2014, ieee sig. proc. magazine]

- 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.

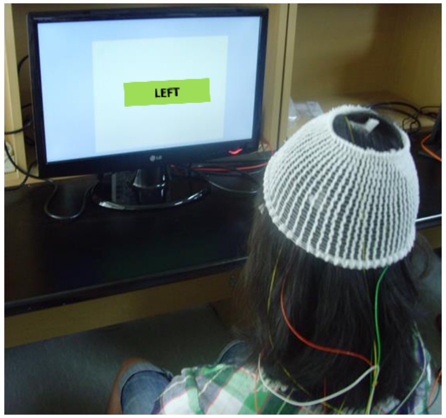


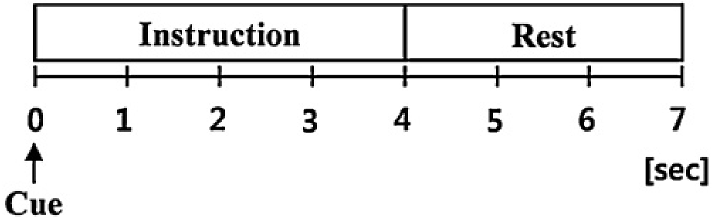
Motivation and Purpose

- 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.

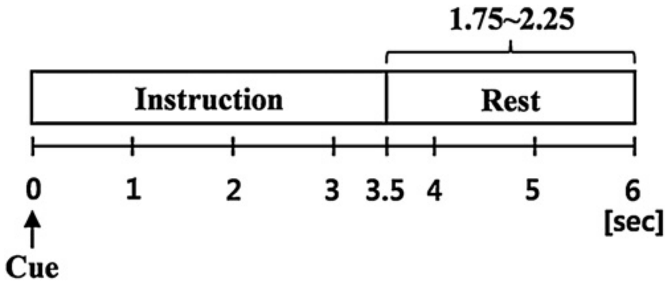
Data acquisition

- 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

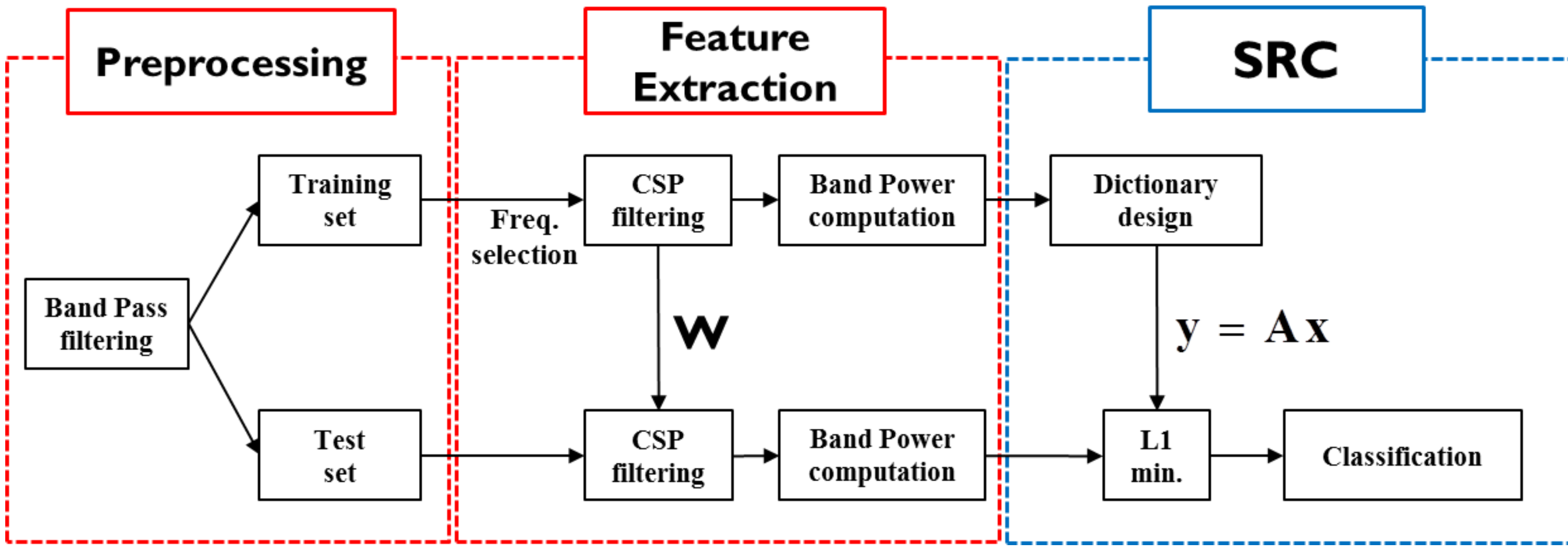




  - Berlin dataset
    - BCI competition dataset (Data set IVa)
    - Five healthy subjects
    - Right hand and right foot imaginations
    - 118 EEG channels
    - 140 trials per class

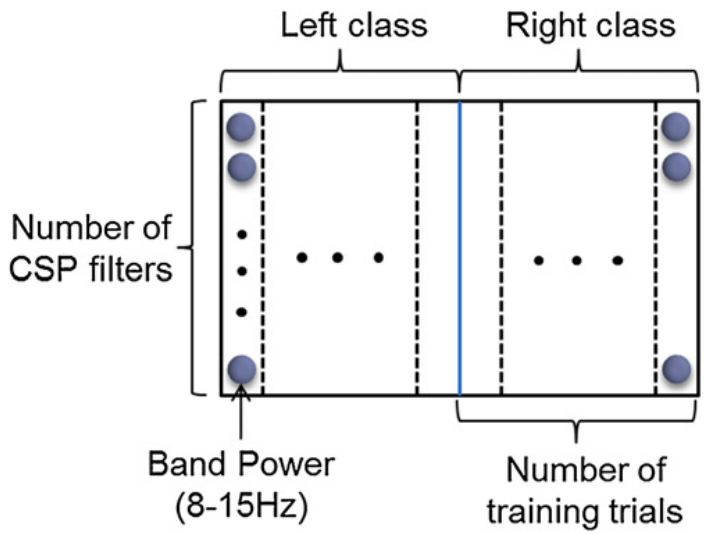


Proposed SRC scheme



- 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



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

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

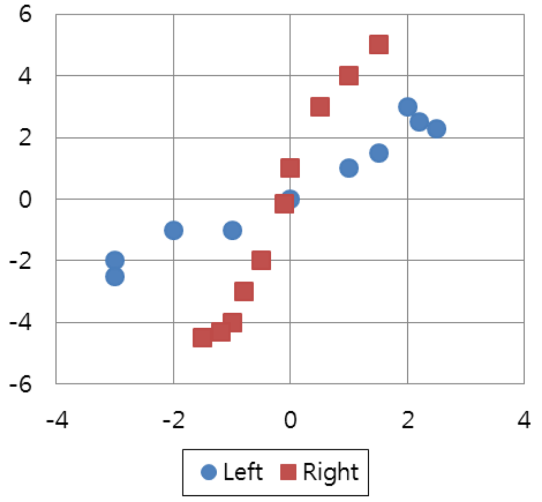
where, i is class, N_t is total trials

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

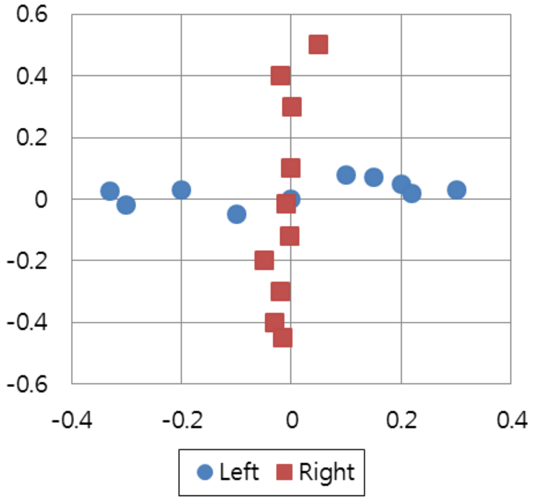
- *M* is the measure of mutual coherence of two component dictionaries; when *M* is small, we say that the dictionary is incoherent .
- The incoherent dictionary promotes the sparse representation of the test signal under the L1 minimization [Donoho 2003].
- We use the CSP filtering to design an *incoherent*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

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.

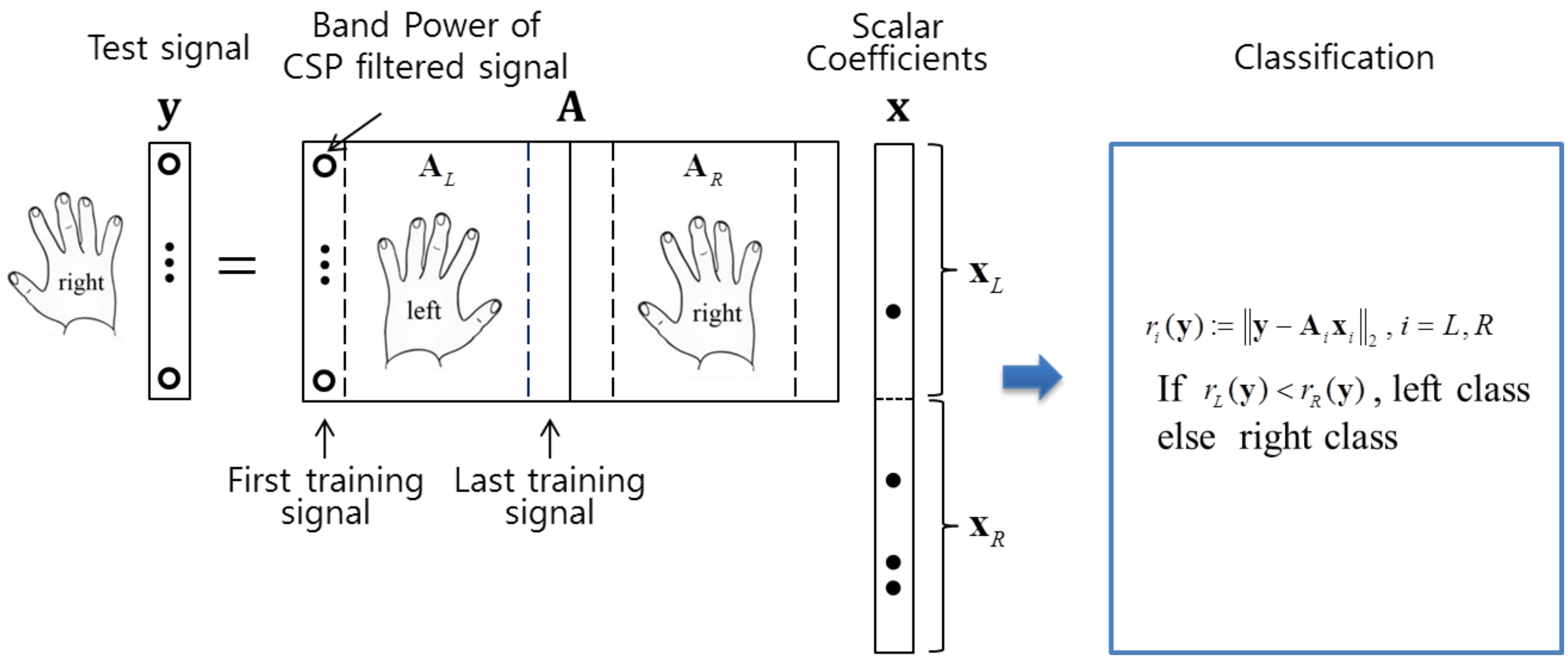


[before CSP filtering]



[after CSP filtering]

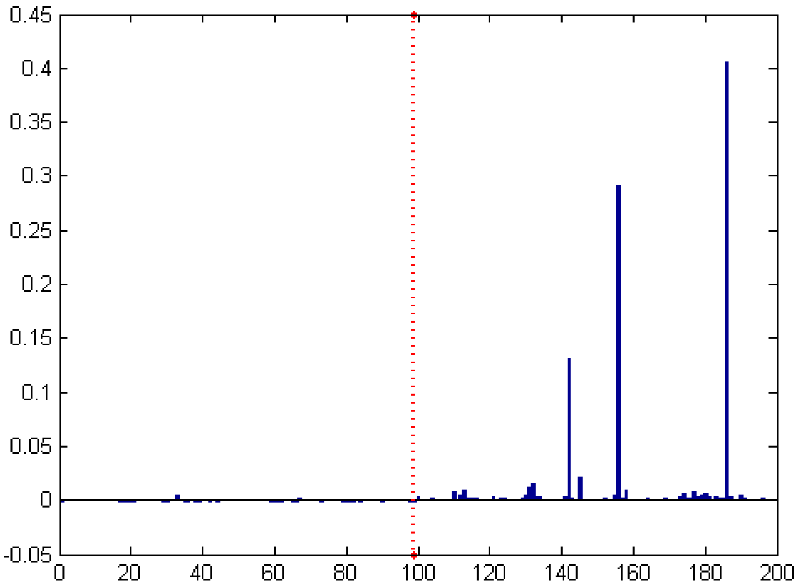
Sparse Representation and Classification



- 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 **x** in $\mathbf{y} = \mathbf{A}\mathbf{x}$.
  - The class of the test trial is the right hand imagery
  - The test signal of right class sparsely represented with some 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.

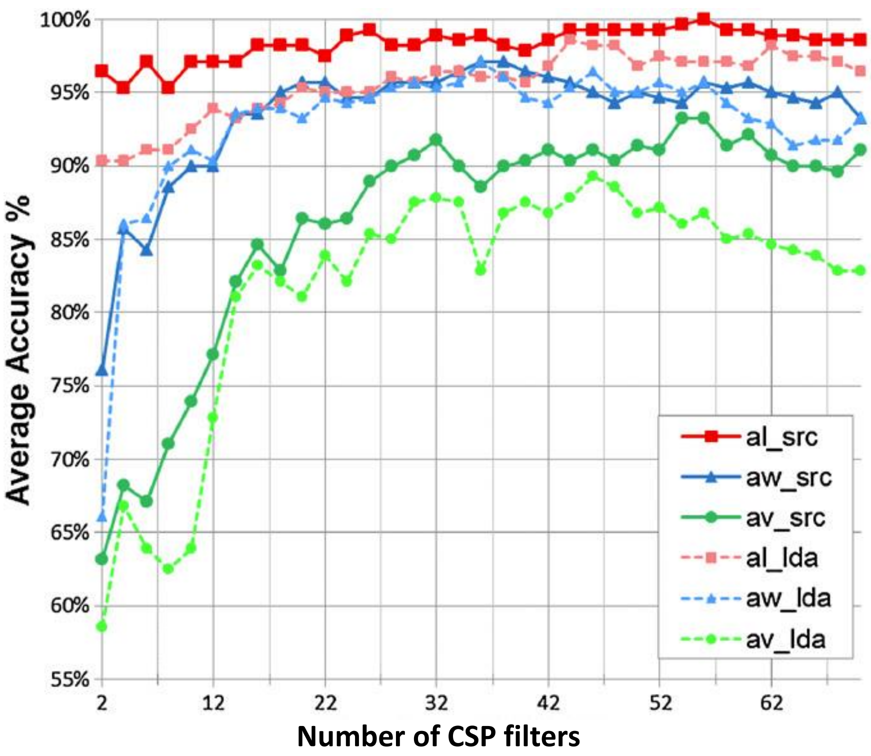
Classification accuracy of Berlin dataset

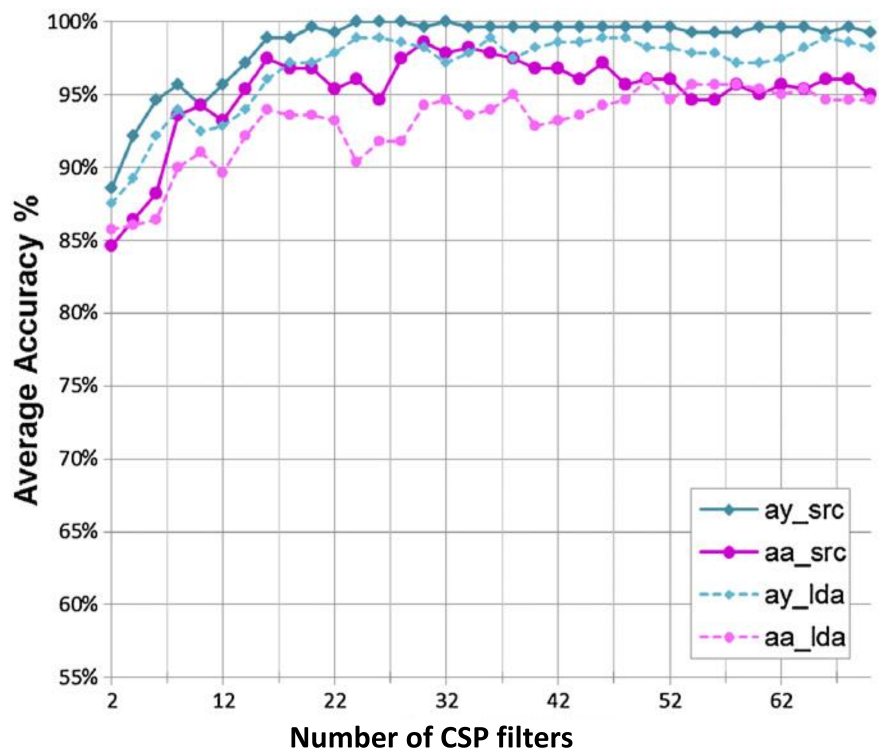
Subject	SRC Accuracy [%]	LDA Accuracy [%]
al	98.93	96.43
ay	100	97.14
aw	95.71	95.36
aa	97.86	94.64
av	91.79	87.86
Mean (SD)	96.85 (3.25)	94.29 (3.72)

- 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.

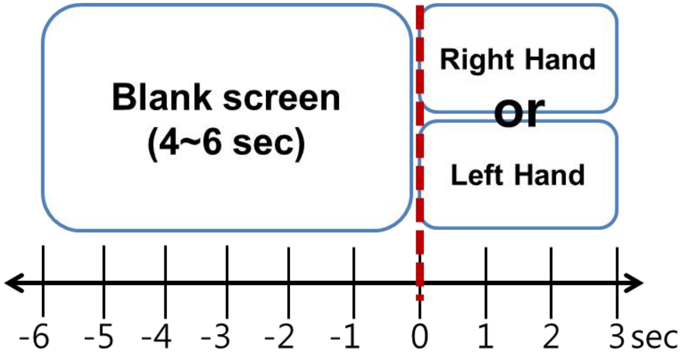
Evaluation of SRC method

Motivation and purpose

- Compare classification performance of SRC with that of SVM, which is a widely known for providing the best performance in many studies.
  - We use large BCI data set (20 subjects)
- Evaluate noise robustness of SRC method
  - Test signal will be changed and contaminated by noise.
  - Therefore, position of test feature is shifted from that of training feature.
  - We compare the noise robustness results of SRC and that of SVM.
  - We use two types of noise such as white Gaussian and resting state noise.
- Data classification using training data
  - Classifier will be self tested with training data which is used for designing classifier.
  - If classifier perfectly classify the training data itself, it will be powerful for testing data.

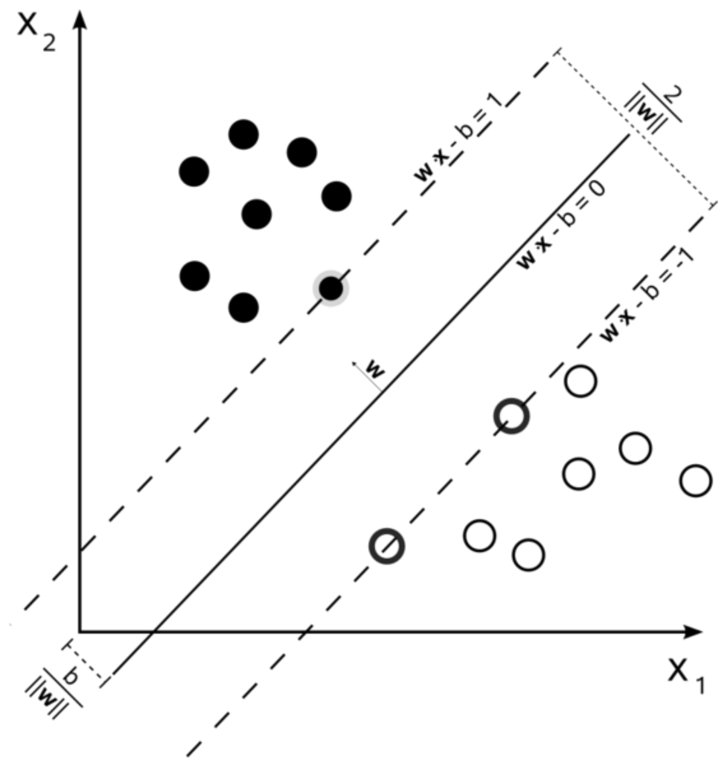
Data acquisition

- We use two-class MI dataset obtained from 20 subjects.
- Right hand and Left hand of motor imagery movements
- 64 EEG channels and 512 sampling rate
- 100 trials per class
- We also record the resting state for each subject to estimate the resting noise.
- For the resting state, subject just open their eyes.



Support vector machine (SVM)

- The idea of SVM is proposed by Vapnik aimed to find decision hyperplane with maximum margin which is the distance between the hyperplane and the nearest training feature vectors (support vectors).



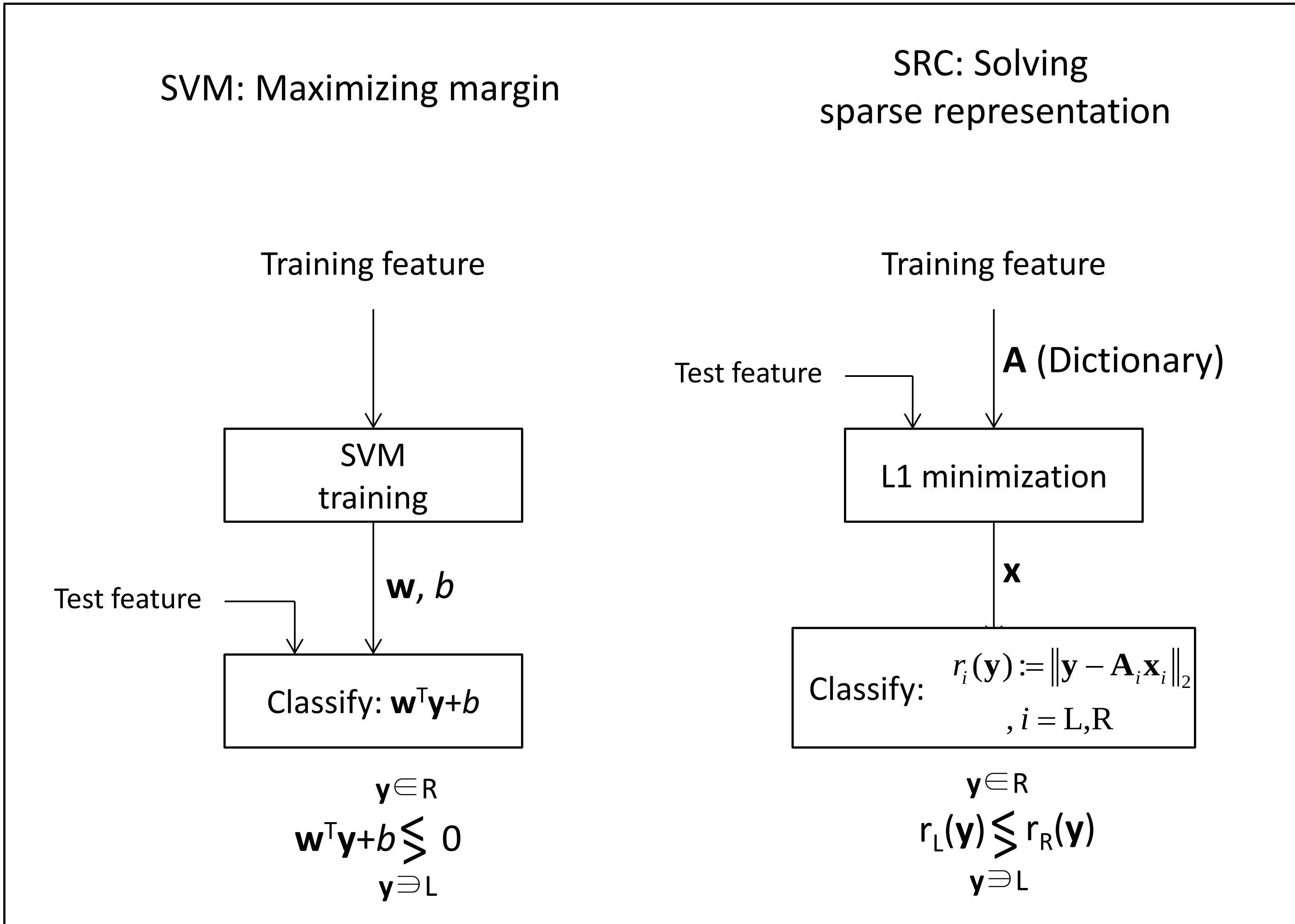
$$\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2,$$

$$\text{subject to} \quad y_n(\mathbf{w}^T\mathbf{x}_n + b) \geq 1$$

$$n = 1, \ldots, N$$

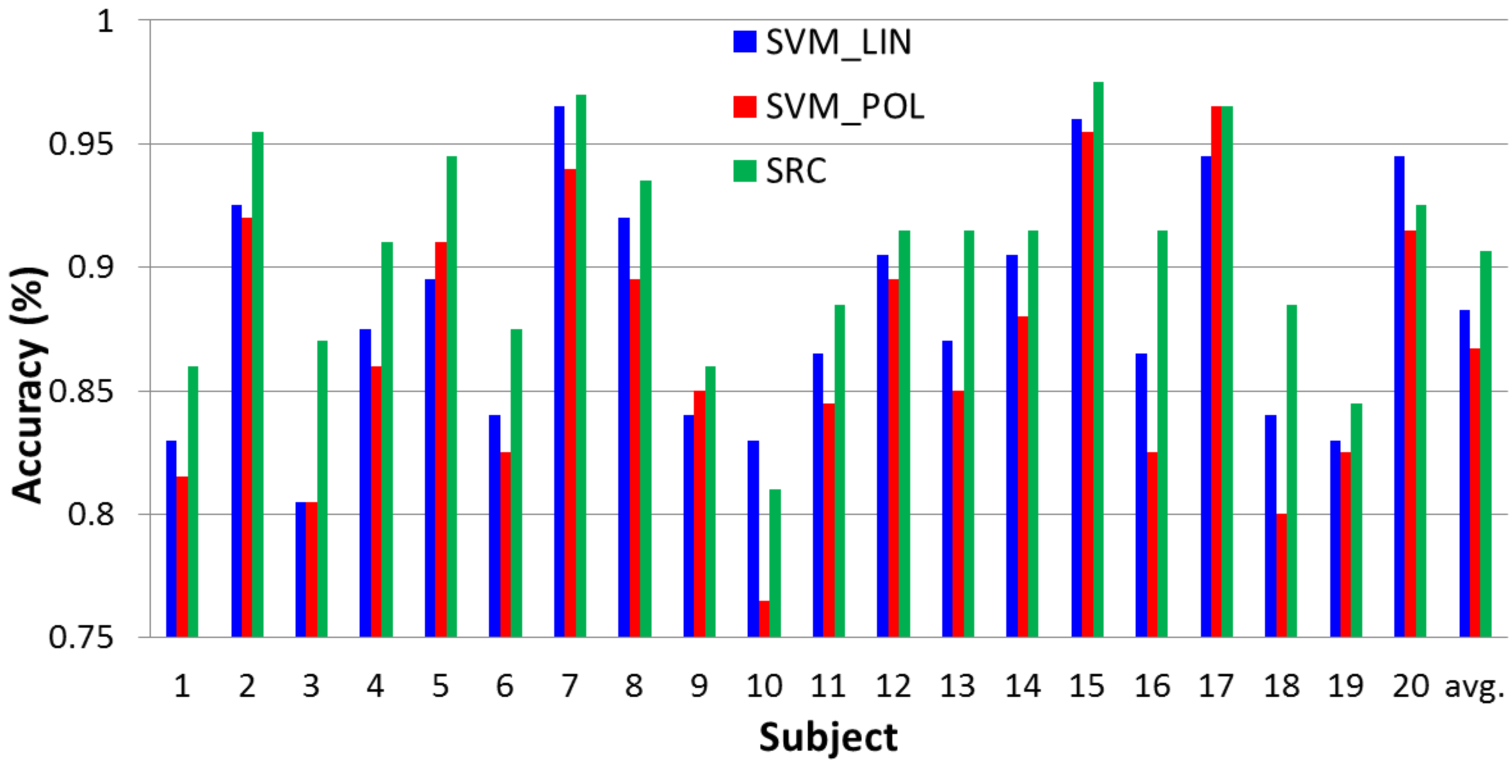
- In the BCI field, SVM has shown the robust classification performance in many experiments.

Classifier algorithm comparison



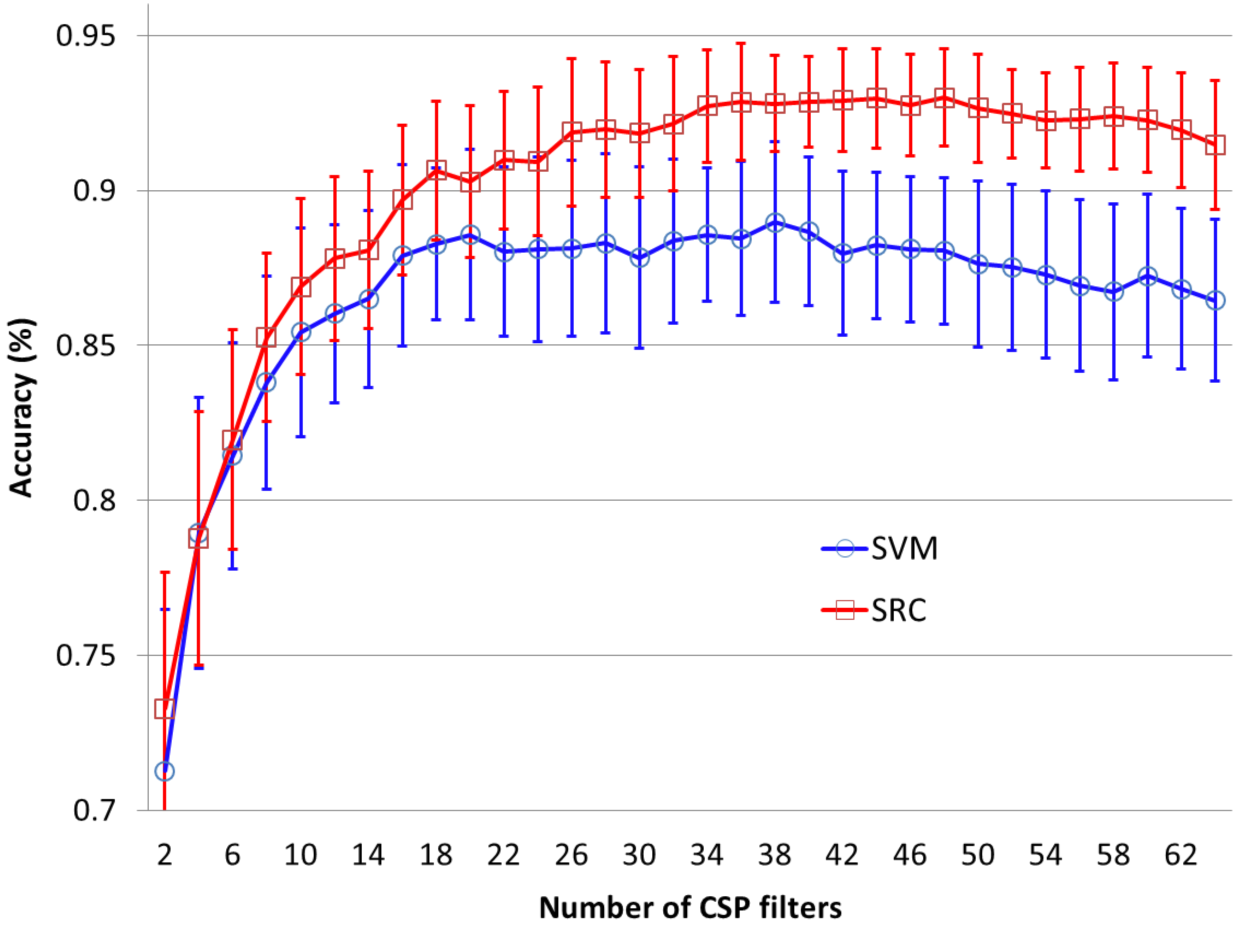
Comparison of classification results

- We use 9 CSP filters for all subjects.
- SRC shows higher classification accuracy than SVM for 18 subjects.
- Average difference between SRC and SVM is statistically significant ($p < 0.001$).



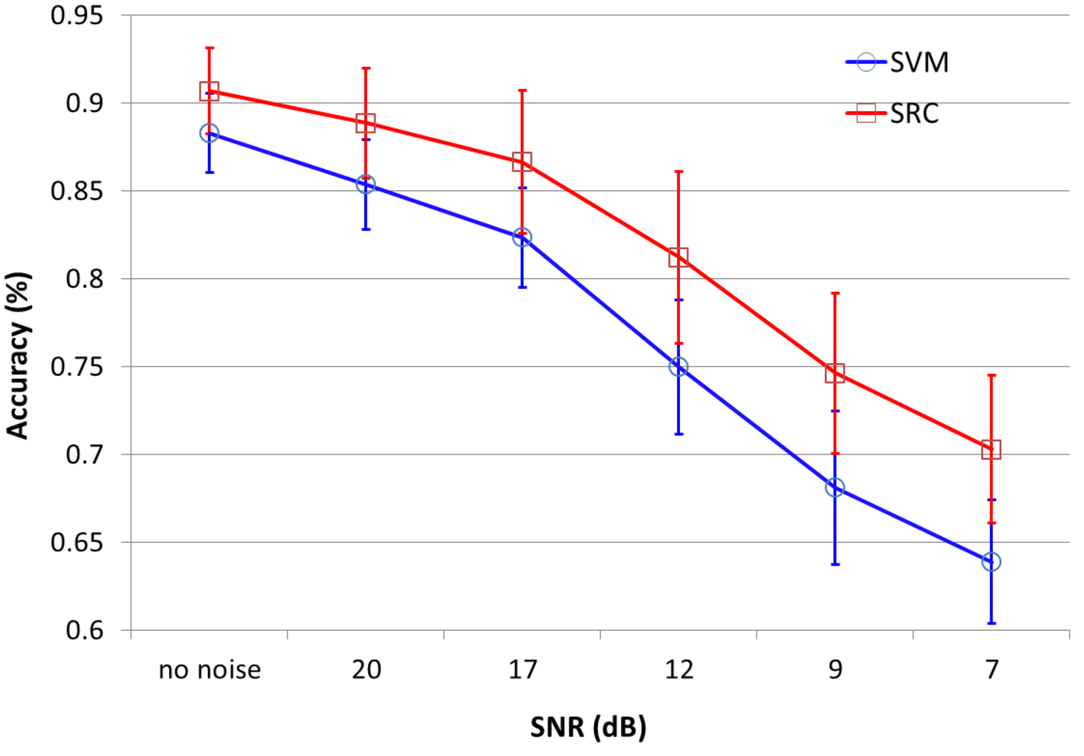
Comparison of classification results

- Classification accuracy as a function of the number of CSP filters.
- Regardless of feature dimension, SRC outperforms SVM.



Noise robustness (Gaussian noise)

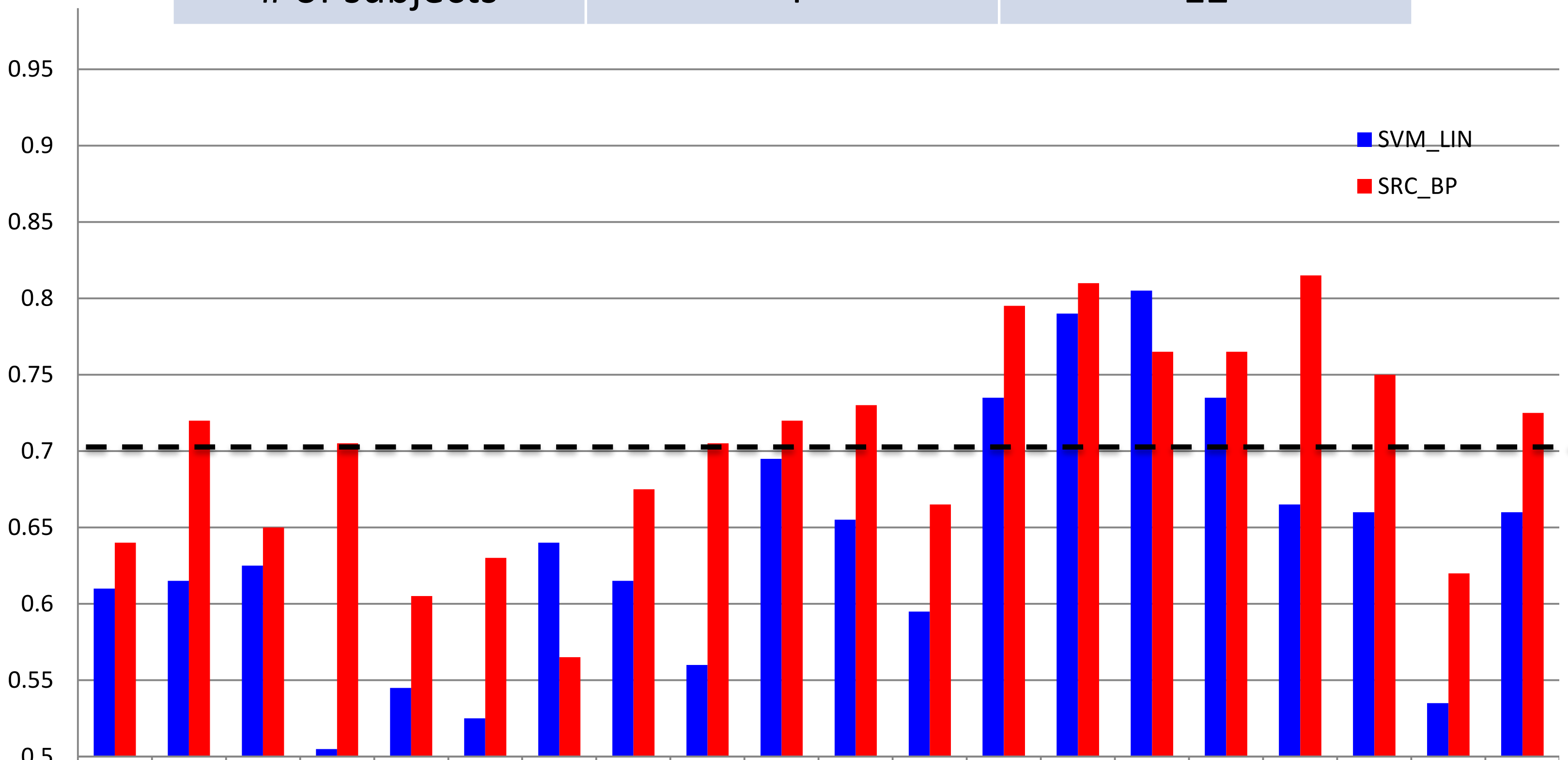
- We generate white Gaussian noise based on mean and standard deviation.
- mu=0; std=100; noise=random('norm',mu,std,64,512);
- We add Gaussian noise to the original test signal while increasing noise level.



Noise robustness (Gaussian noise)

- Accuracy more than 70% allows communication and device control [Kübler 2001].

Over 70%	SVM	SRC
# of subjects	4	12



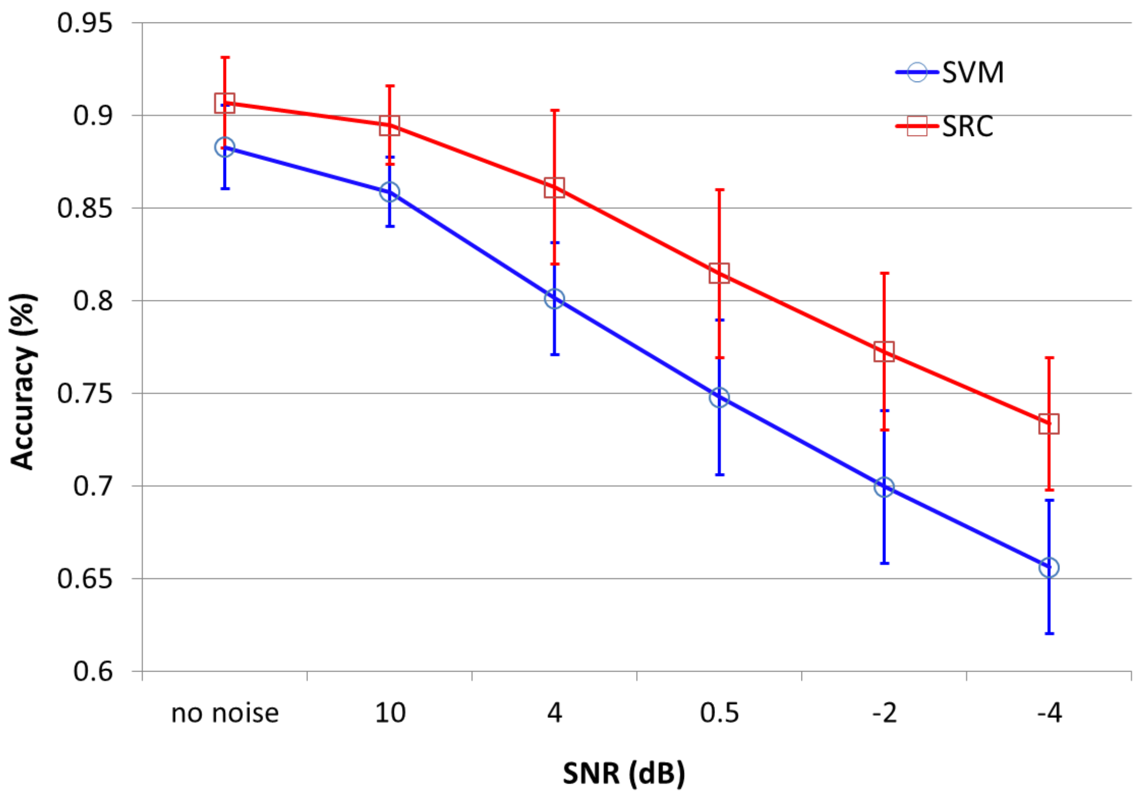
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
SVM_LIN	0.61	0.615	0.625	0.505	0.545	0.525	0.64	0.615	0.56	0.695	0.655	0.595	0.735	0.79	0.805	0.735	0.665	0.66	0.535	0.66
SRC_BP	0.64	0.72	0.65	0.705	0.605	0.63	0.565	0.675	0.705	0.72	0.73	0.665	0.795	0.81	0.765	0.765	0.815	0.75	0.62	0.725

Noise robustness (resting noise)

- We generate noisy signal using resting state signal:

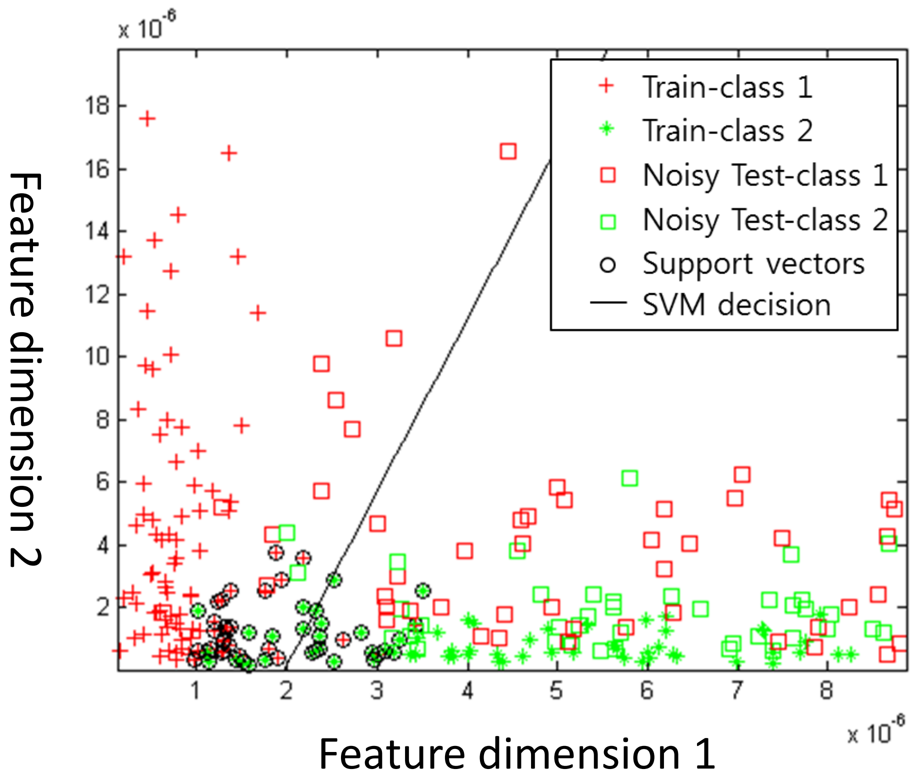
  Noisy test signal = test signal + (noise level * resting noise)

- We compute classification accuracy while increasing noise level.



Noise robustness (resting noise)

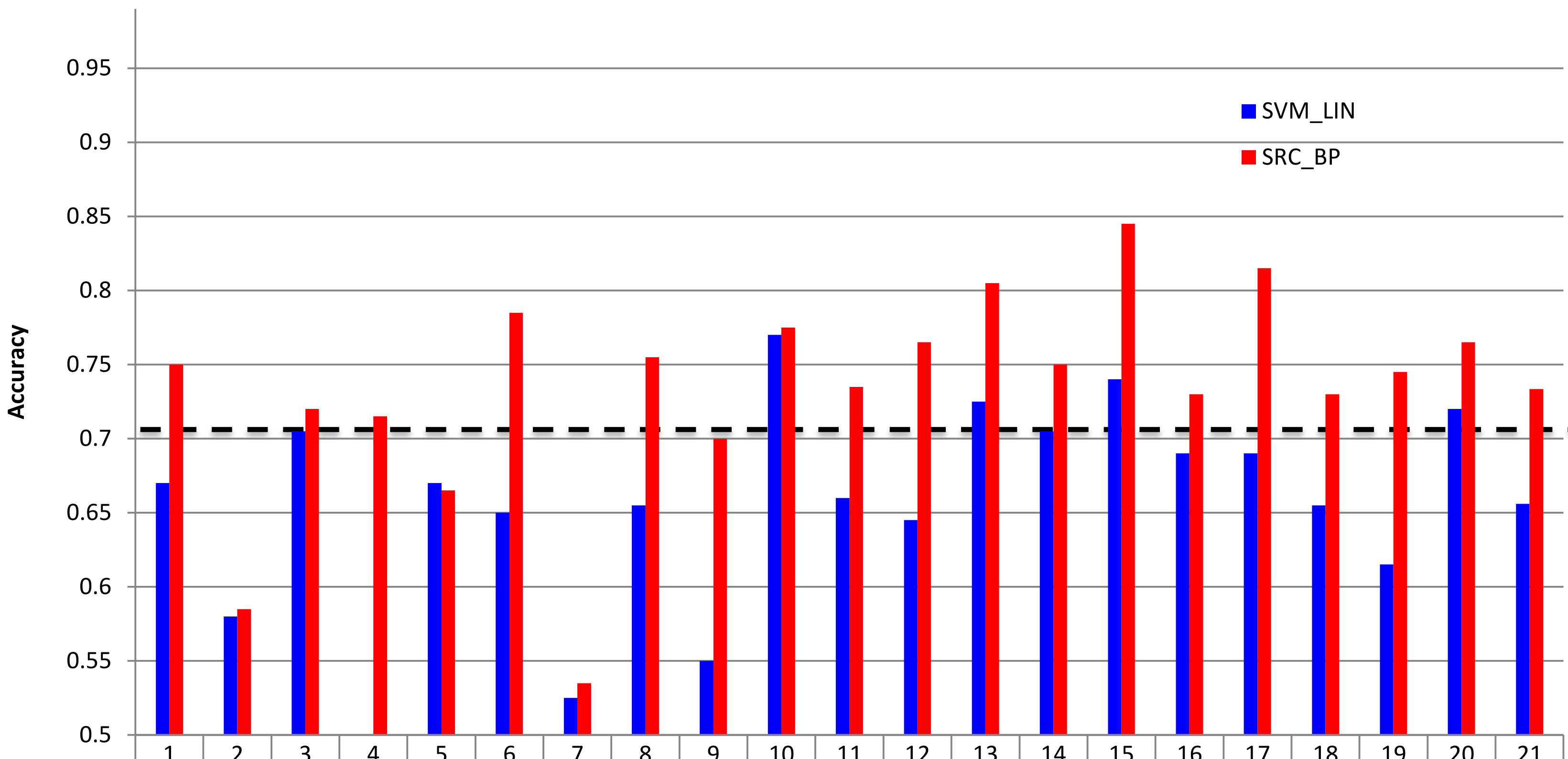
- 2D example, SNR = 4dB



	SVM	SRC
SNR=4dB	55.5%	64.5%

Noise robustness (resting noise)

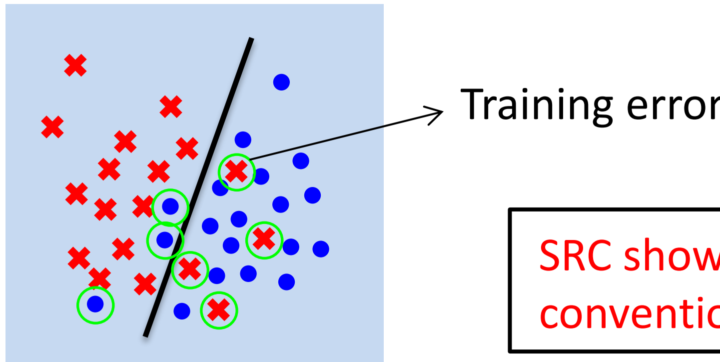
Over 70%	SVM	SRC
# of subjects	6	17



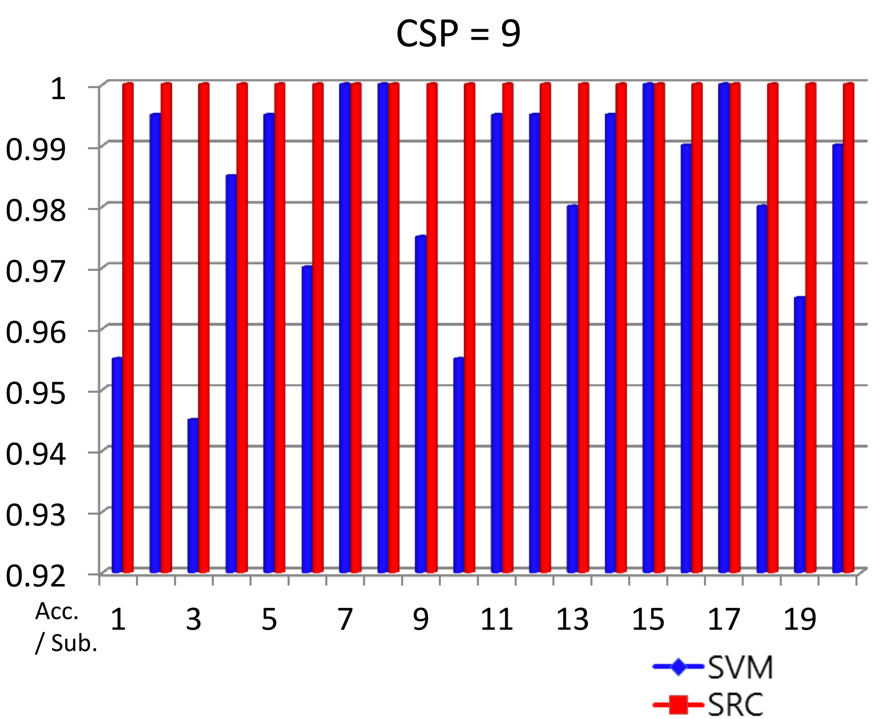
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
SVM_LIN	0.67	0.58	0.705	0.5	0.67	0.65	0.525	0.655	0.55	0.77	0.66	0.645	0.725	0.705	0.74	0.69	0.69	0.655	0.615	0.72	0.656
SRC_BP	0.75	0.585	0.72	0.715	0.665	0.785	0.535	0.755	0.7	0.775	0.735	0.765	0.805	0.75	0.845	0.73	0.815	0.73	0.745	0.765	0.734

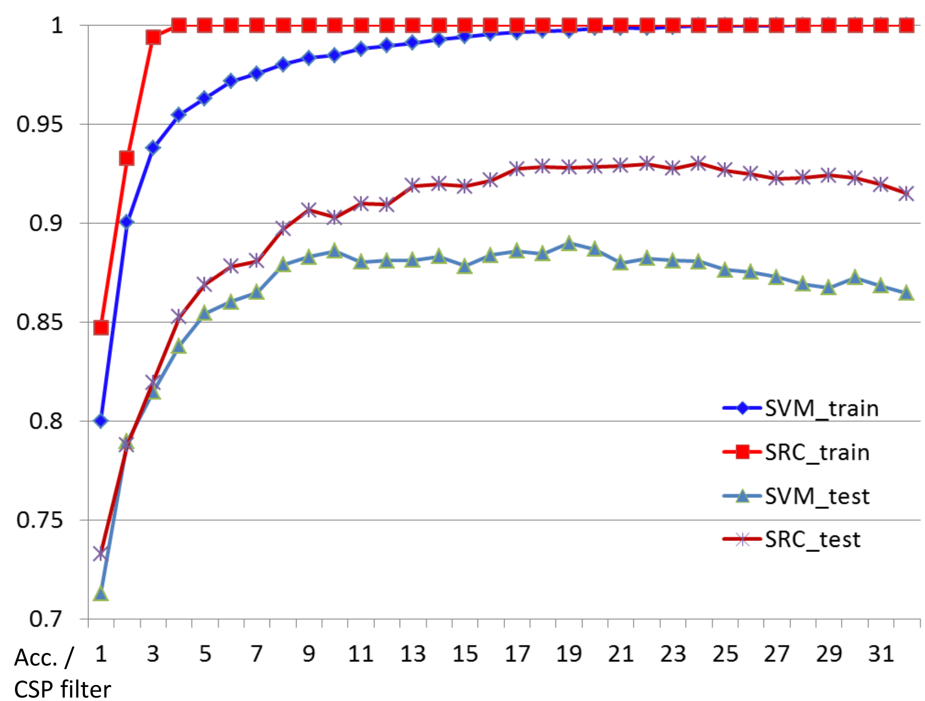
Training data classification

- Data classification using only training data which is used for classifier design.



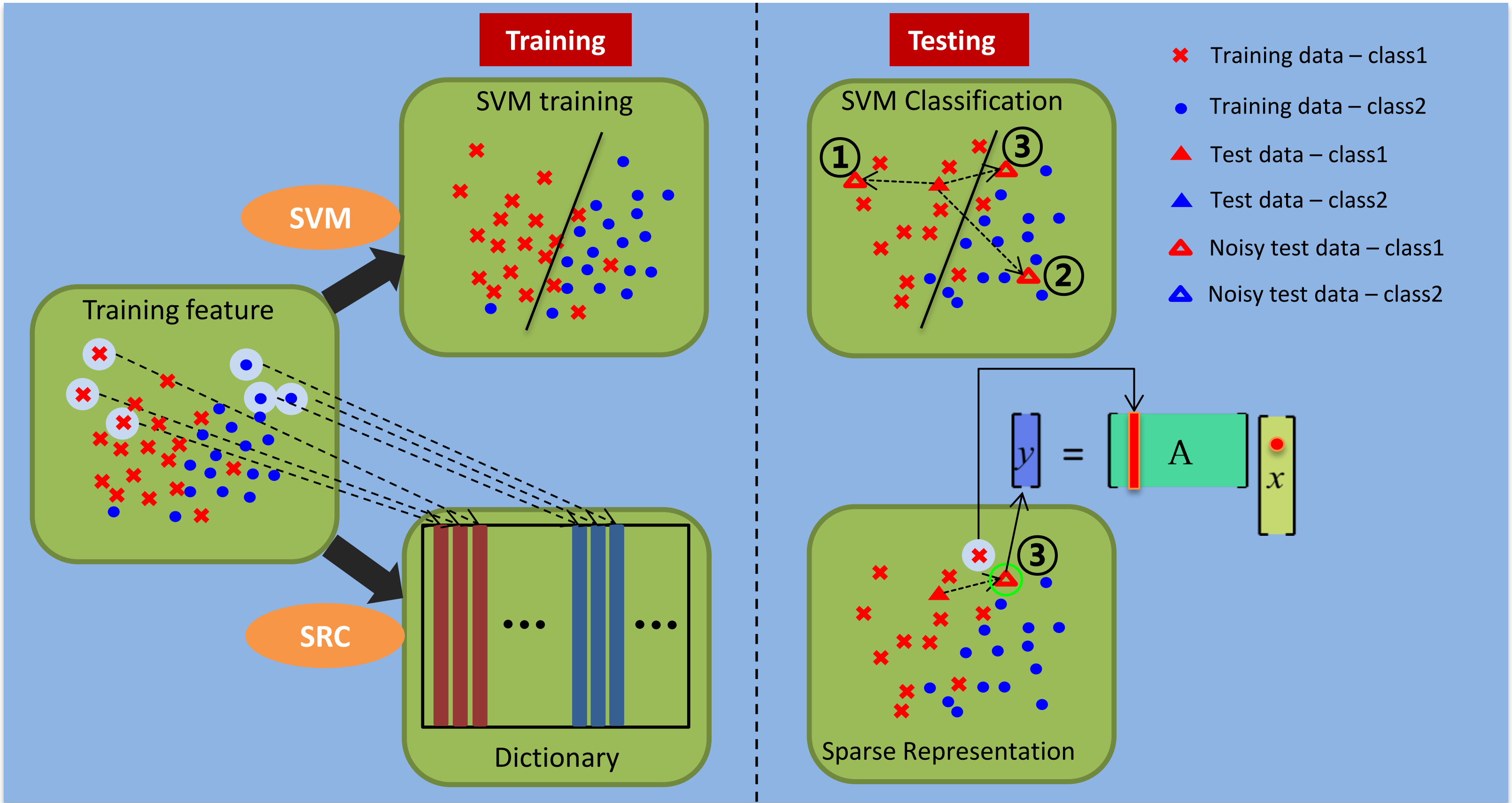
SRC shows superior performance than conventional classifiers for training data!





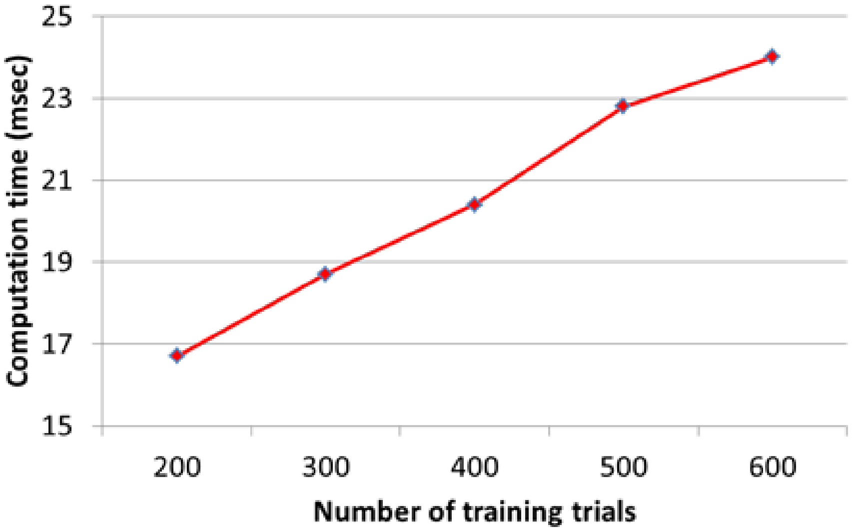
A new concept of classification

- In the SVM, the classifier is trained (learned) using training data. Then, this fixed classification rule is applied to test data.
- However, in SRC, the design of classification rule is not needed. For each test signal, using direct training signal, independent classification task(i.e., sparse representation) is performed.
- In third case, based on the decision rule obtained from training data, SVM always wrongly classify. On the other hand, SRC has still chance to correct sparse representation with same class training data



Computation time

- The computation time of SRC method increase by the number of test trials. Thus, the robust classification performance of the SRC involves the cost of computation time at each test trial.
- For a single test trial, SVM and SRC take 12.1 and 16.7msec average computation time respectively.
- We compute the average computation time of SRC as the function of number of training trials.
- The difference of computation time is just few milliseconds.



Summary

- We compare the classification accuracy of SRC with that of SVM using large datasets.
- SRC shows better classification accuracy than SVM.
- We evaluate the noise robustness of SRC using Gaussian and resting state noise.
- SRC shows more robust performance than SVM for both Gaussian and resting noise.
- The improved performance of SRC might be caused by the different classification approach with conventional classifiers.
- Computation time for each testing is cost of the robust classification accuracy.

Future work

- Real time online classification
  - Simulate the SRC using online feedback dataset.
  - Apply SRC to online BCI experiment.
- Adaptive classifier
  - Develop dictionary adaptation technique for long time use.
  - Apply dictionary learning techniques
- Multi-class performance
  - Apply SRC scheme to the multi-class BCI applications.

Thank you for attention!

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.

Approximate Message-Passing for High-Dimensional Piecewise-Constant Recovery

Jaewook Kang, Hyoyoung Jung, Heung-No Lee, Kiseon Kim*

Speaker: Jaewook Kang
CSNL laboratory at GIST-DIC
jwkkang@gist.ac.kr

Infonet Workshop 2014 Summer July 8th, 2014

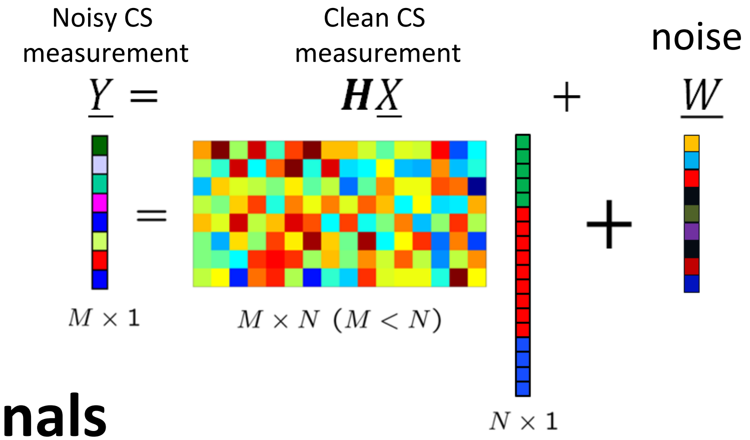
List of Contents

- Background
- Motivations
- Contributions
- The Spike-and-slab AMP (ssAMP) algorithm
- Performance validation
- Conclusion and further works

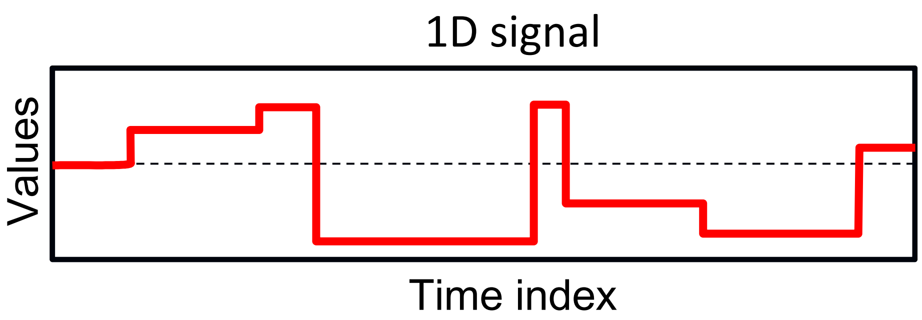
Background (1/3)

Piecewise-constant Recovery (PcR) Problem

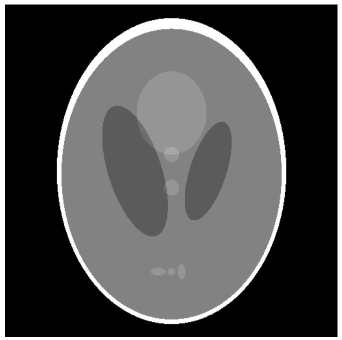
- The signal $\underline{X} \in \mathbb{R}^N$ consists of K+1(<<N) different constant values.
- Our aim is to "fast reconstruction of $\underline{X}$ from the compressive measurements $\underline{Y} \in \mathbb{R}^M$ under a noisy CS framework".



Examplary piecewise-constant signals



2D image signal



Background (2/3)

❖ Fixed-point solvers for the PcR Problem

	Year	Name	Complexity	Description	Publication
Milestones	1992	TV/ROF denoising	-	-First work to proposed TV-penalty for the PcR image denoising	Physica D nonlinear phen. 1992
	1996	LS-TV	-	- Solving TV-penalized least square using Newton-method	SIMA J. Sci. Comput 1996
	2005	Fused Lasso	-	- Solving Lasso problem including the TV-penalty term	J.R. statist. 2005
Recent	2008	TVAL3	O(MN)	-Variable splitting approach (alternating-direction approach) - Accelerating convergence - Not good PTC	SIMA J. imag. 2008
	2009	Split Bregman	O(MN)	- Further developing the variable splitting approach - Not good PTC	SIMA J. imag. 2009
	2011	Champolle-Pock (CP)	O(MN)	- Solving the primal-dual problem simultaneously -Reducing the duality gap, convergence guarantee	J. Math. Imag. 2011

Background (3/3)

- **AMP solver has got attention for high-dimensional CS problems !!**
  - **Reason 1)** LASSO-optimal performance in high-dimensional setting
    - Showing phase transition curve equivalent to Lasso. [PNAS'09:Donoho et.al] [TIT'13:Donoho et.al]
  - **Reason 2)** Low-computational nature of AMP iteration.
    - Handling only O(M+N) message per iteration. [PNAS'09:Donoho et.al],[PhysRev'12:Krzakala et.al]
  - **Reason 3)** MSE prediction via state evolution.
    - MSE of the AMP estimate is deterministically predictable over iteration [PNAS'09:Donoho et.al],[TIT'11:Bayati et.al]

- **Several extension of AMP to various types of signals**

Year	Name	Description	Publication
2011	GAMP	Generalized input/output distribution	ISIT 2011, ArXiv2010
2013	BAMP	Block sparsity (Group Lasso)	TIT 2013, ArXiv2013
2013	CAMP	Complex sparsity	TIT 2013
2013	TV-AMP	Piecewise-constancy (TV Lasso)	TIT 2013
2013	GrAMPA	A variation of GAMP to the analysis-CS framework (Generalized Lasso)	ArXiv 2013

CSNL Communications and Sensor Networks Lab.
INFONET INFOrmation sensing, processing, controlling, and NETworking

Motivations (1/2)

❖ Existing AMP for the PcR problem (1): TV-AMP [TIT13:Donoho et. al]

- TV-AMP operates iteratively according to:

$$\underline{\mu}^{(t+1)} = \eta_{\mathrm{TV}}(\mathbf{H}^T \underline{r}^{(t)} + \underline{\mu}^{(t)})$$

$$\underline{r}^{(t)} = \underline{y} - \mathbf{H}\underline{\mu}^{(t)} + \underline{r}^{(t-1)} \frac{N}{M} \left\langle \eta'_{\mathrm{TV}}(\mathbf{H}^T \underline{r}^{(t-1)} + \underline{\mu}^{(t-1)}) \right\rangle$$

➢ where $\underline{\mu}^{(t)} \in \mathbb{R}^N$ and $\underline{r}^{(t)} \in \mathbb{R}^M$ are a signal estimate and residual at the t-th iteration respectively.

- **Contribution)** TV-AMP is the first extension of AMP to the PcR problem.
- **Limitation 1)** TV-AMP has shown limited success in practice due to its denoiser which is solved by an external optimization package.

$$\eta_{\mathrm{TV}}(\underline{\rho}) \equiv \arg\min_{\underline{X}} \tfrac{1}{2} \left\| \underline{\rho} - \underline{X} \right\|_2^2 + \lambda \sum_{i=2}^{N} |X_i - X_{i-1}|,$$

  ➢ Complexity of TV-AMP highly depend upon that of the external package.

- **Limitation 2)** The TV-AMP denoiser is not scalarwise, leading to difficulty in the MSE prediction via the SE framework of [TIT'11:Bayati et.al].

CSNL Communications and Sensor Networks Lab.
INFONET INFOrmation sensing, processing, controlling, and NETworking

Motivations (2/2)

❖ Existing AMP for the PcR problem (2): GrAMPA [ArXiv13:Schniter et al.]

- GrAMPA is a variation of GAMP to the analysis-CS problem [ACHA'13:Nam et.al], solving

$$(P_{\text{analysis-CS}}): \hat{\underline{x}} = \arg\min_{\underline{X}} \tfrac{1}{2} \| \underline{Y} - \mathbf{H}\underline{X} \|_2^2 + \lambda g(\mathbf{\Omega}\underline{X})$$

  - Where the function $g(.)$ is called the analysis function.
  - When $\mathbf{\Omega} = \mathbf{D}$(a finite-difference matrix), $(P_{\text{analysis-CS}})$ becomes the PcR problem.
  - In Bayesian viewpoint, the choice of the function $g(.)$ is determined by the prior for $\mathbf{\Omega}\underline{X}$.

- **Contribution)** GrAMPA operates with a scalarwise MMSE denoiser for the PcR problem, which is given as

$$\eta_{\text{GrAMPA}}(\rho_i; v_i, \tau) \equiv \mathbf{E}[X_i - X_{i-1} \mid \rho_i, v_i, \tau] = \frac{\rho_i}{1 + \tau \mathcal{N}(0; \rho_i, v_i)}$$

  - The denoiser, $\eta_{GrAMPA}(\cdot)$, selects a sparse support of difference $X_i - X_{i-1}$ with a Bernoulli-Uniform prior, where the input ρ_i is a noisy observation of $X_i - X_{i-1}$.
- **Limitation)** In the GrAMPA iteration, the residual calculation for $X_i - X_{i-1}$ is based on not real measurements but the estimate by $\eta_{GrAMPA}(\cdot)$.
  - This may lead to ineffective convergence of the recovery when the sampling rate M/N is very low.

Contributions

Contributions

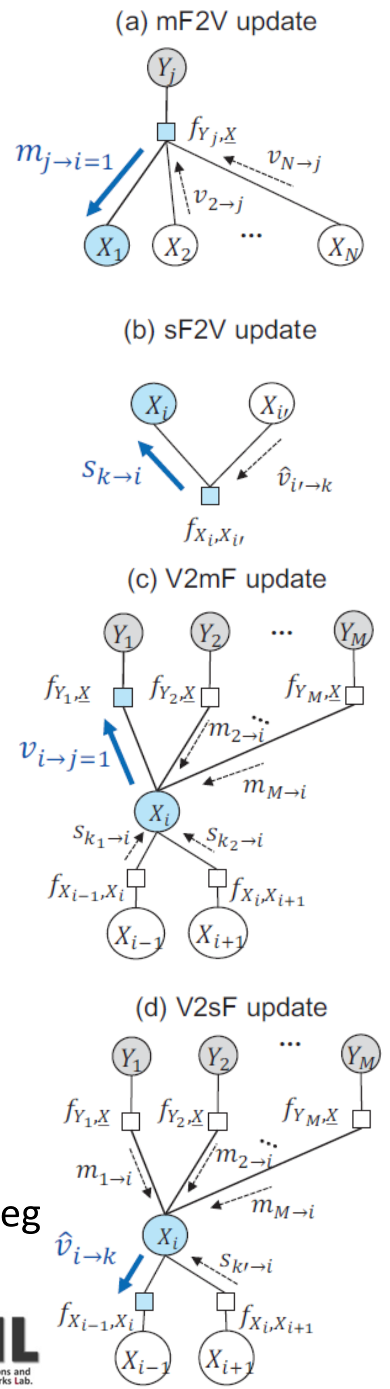
- The present work proposes an alternative AMP solver for the PcR problem called, **“spike-and-slab Approximate Message-Passing (ssAMP)”**.

- **Contribution 1)** ssAMP includes a novel scalarwise denoiser satisfying the Lipschitz continuity.
  - Which can be an alternative of the TV-AMP denoiser using the external package.
  - Which can be applized to the SE framework of [TIT’11:Bayati et.al] for MSE prediction.
- **Contribution 2)** ssAMP shows phase transition curve (PTC) covering that of the two existing AMPs for the PcR problem: TV-AMP and GrAMPA.
- **Contribution 3)** Computational advantage of ssAMP is remarkable compared to the other solvers for the PcR problem in a high-dimensional setting.

Related Publications

- 1) **Jaewook Kang**, Hyoyoung Jung, Heoun-No Lee, Kiseon Kim, "Spike-and-Slab Approximate Message-Passing for High-Dimensional Piecewise-Constant Recovery," submitted to *IEEE Journal of selected topics in Signal processing* July 1st.
- 2) **Jaewook Kang**, Heung-No Lee, and Kiseon Kim, "Piecewise-Constant Recovery via Spike-and-Slab Approximate Message-Passing using a Scalarwise Denoiser," to appear in proc. of the 48th Asilomar Conference (Asilomar, CA), Nov. 2014.

The ssAMP algorithm (1/2)



❖ Algorithm construction approach

- **Step I)** Drawing a joint PDF from a factor graph model

$$f_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) = \frac{1}{Z}\prod_{i=2}^{N} f_{X_{i-1},X_i}(x_{i-1},x_i)\prod_{j=1}^{M} f_{Y_j|\underline{X}}(y_j|\underline{X})$$

- **Step II)** Assigning potential function for the piecewise-constancy
  - We go with Spike-and-slab potential function.

$$f_{X_{i-1},X_i}(x_{i-1},x_i) = (1-q)\delta(x_i - x_{i-1}) + q\mathcal{N}(x_i - x_{i-1}; 0, \sigma_0^2)$$

- **Step III)** Developing classical MP rule
  - Measurement fidelity: the mF2V and V2mF update
  - Piecewise-constant pursuit: the sF2V and V2sF update

- **Step IV)** MP rule to AMP rule
  - **1) Parameterization step**: Density-passing to parameter-passing
  - **2) First-order approximation step**: handling O(MN) meg. to O(M+N) meg
  - **3) Simplification step**: handling the sF2V and V2sF upate

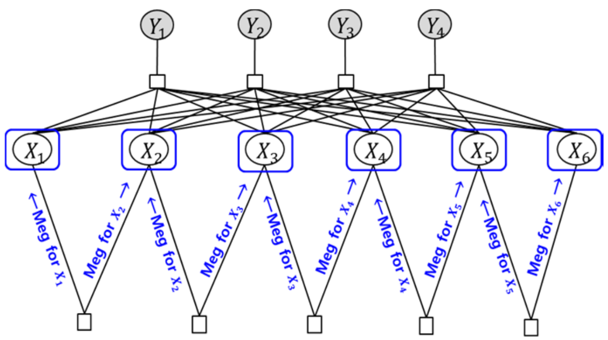
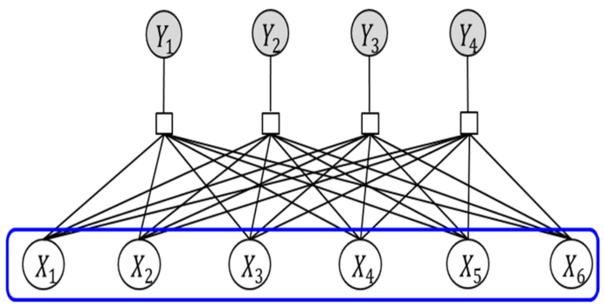
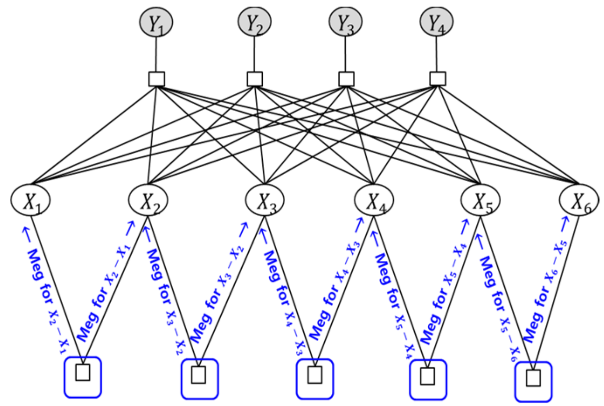
The ssAMP algorithm (2/2)

❖ ssAMP Iterations [JSTSP14:Kang et. al]

mF2V update	set $\underline{r}^{(t)} = \underline{y} - \mathbf{H}\underline{\mu}^{(t-1)} + \underline{r}^{(t-1)} \frac{N}{M} \left\langle \eta'(\underline{\rho}; \theta, \{\underline{a}_l\}^{(t-1)}, \{\underline{b}_l\}^{(t-1)}, \{\underline{c}_l\}^{(t-1)}) \right\rangle_N$ set $\underline{\rho}^{(t)} = \mathbf{H}^T \underline{r}^{(t)} + \underline{\mu}^{(t-1)}$ set $\theta^{(t)} = \Delta + \frac{N}{M} \left\langle (\underline{\sigma}^2)^{(t-1)} \right\rangle_N$	▪ Measurement fidelity task - Producing a residual $\underline{r}^{(t)}$ from $\underline{y}$, - Generating a noisy estimate $\underline{\rho}^{(t)}$ of the signal $\underline{X}$. → Conventional
R2P/L2P update	set $\underline{\mu}_{\text{L2P}}^{(t)} = \phi(\underline{\rho}^{(t)}; \theta, \underline{\mu}_{\text{L2P}}^{(t-1)}, \underline{\sigma}_{\text{L2P}}^{2(t-1)})$, set $\underline{\mu}_{\text{R2P}}^{(t)} = \phi(\underline{\rho}^{(t)}; \theta, \underline{\mu}_{\text{R2P}}^{(t-1)}, \underline{\sigma}_{\text{R2P}}^{2(t-1)})$, set $\underline{\sigma}_{\text{L2P}}^{2(t)} = \zeta(\underline{\rho}^{(t)}; \theta, \underline{\mu}_{\text{L2P}}^{(t-1)}, \underline{\sigma}_{\text{L2P}}^{2(t-1)})$, set $\underline{\sigma}_{\text{R2P}}^{2(t)} = \zeta(\underline{\rho}^{(t)}; \theta, \underline{\mu}_{\text{R2P}}^{(t-1)}, \underline{\sigma}_{\text{R2P}}^{2(t-1)})$,	▪ Piecewise-constancy pursuit task Sharing the information on neighboring elements over the signal $\underline{X}$ by the R2P/L2P update → NEW
V2mF update	set $\{\underline{a}_l^{(t)}\}, \{\underline{b}_l^{(t)}\}, \{\underline{c}_l^{(t)}\}$ (which are functions of the R2P/L2P parameters) set $\underline{\mu}^{(t)} = \eta(\underline{\rho}; \theta, \{\underline{a}_l\}^{(t)}, \{\underline{b}_l\}^{(t)}, \{\underline{c}_l\}^{(t)})$ set $(\underline{\sigma}^2)^{(t)} = \gamma(\underline{\rho}; \theta, \{\underline{a}_l\}^{(t)}, \{\underline{b}_l\}^{(t)}, \{\underline{c}_l\}^{(t)})$	▪ ssAMP estimate of $\underline{X}$ Producing an AMMSE estimate of the signal $\underline{X}$ using a scalarwise denoiser $\boldsymbol{\eta}(\cdot)$ → NEW

What's NEW in ssAMP? (1/2)

❖ Comparison of Piecewise-constant pursuit task

ssAMP	TV-AMP	GrAMPA
▪ **The R2P/L2P update** provides information on the two neighbor X_{i-1}, X_{i+1} to the estimation of X_i via the denoiser $\eta(\cdot)$.	▪ **The external denoiser $\eta_{TV}(\cdot)$** performs the picewise-constant pursuit task.	▪ **The denoiser $\eta_{GrAMPA}(\cdot)$** provides an sparse estimate of $X_i - X_{i-1}$. ▪ The info. is given to the estimation of X_i in a form of residual.
 : the ssAMP denoiser $\eta(\cdot)$	 The external denoising function $\eta_{TV}(\cdot)$	 : the GrAMPA denoiser $\eta_{GrAMPA}(\cdot)$

What's NEW in ssAMP? (2/2)

❖ The ssAMP denoiser $\eta(\cdot)$

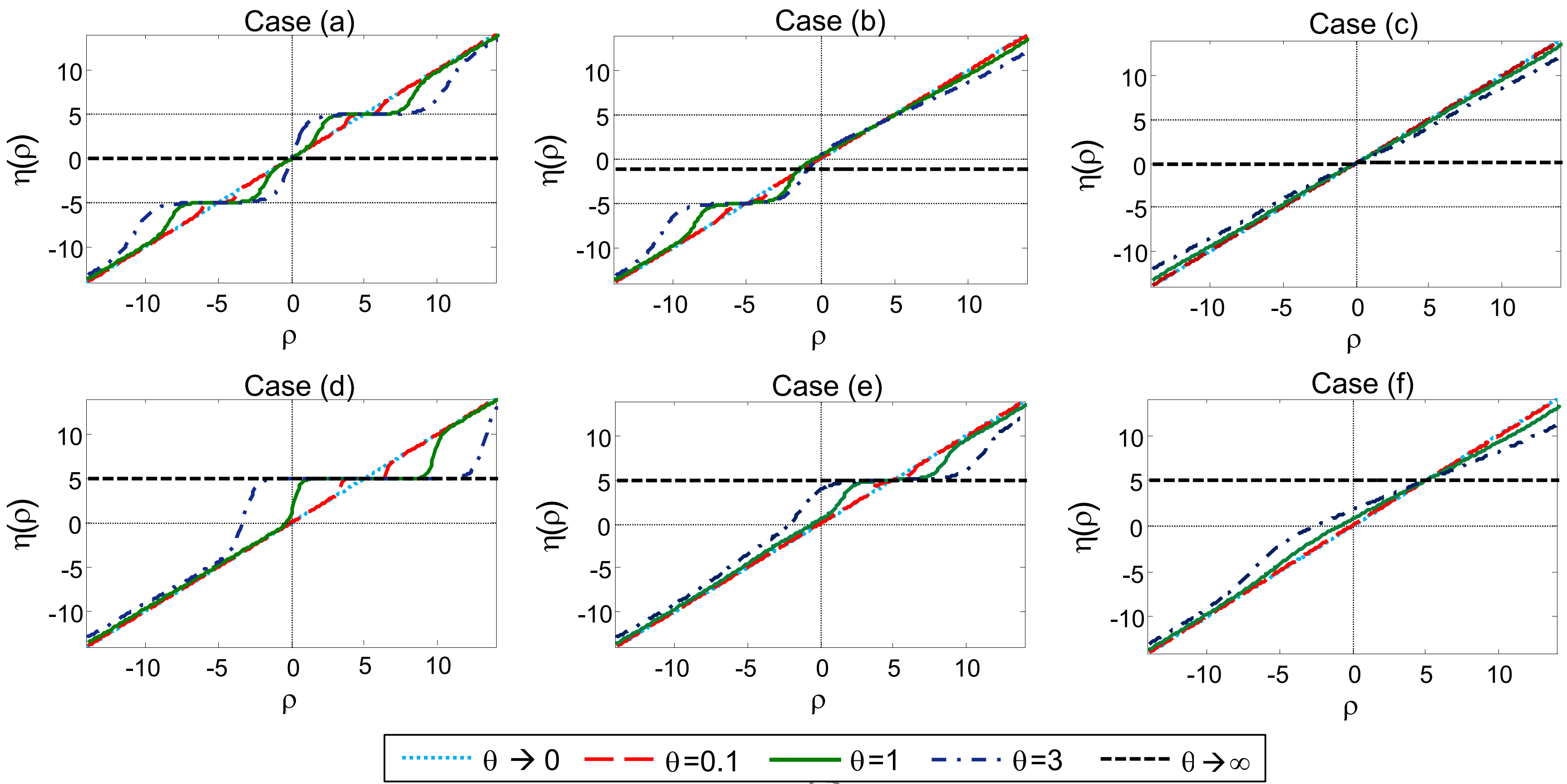
- A scalarwise denoiser produces an AMMSE estimate of X_i in the ssAMP iteration.

$$\mu_i^{(t)} \equiv \mathbb{E}_{f_{X_i}^{(t)}(x_i|\underline{Y})}[X_i|\underline{Y} = \underline{y}]$$
$$= \eta(\rho_i^{(t)}; \theta^{(t)}, \{a_{i,l}\}^{(t)}, \{b_{i,l}\}^{(t)}, \{c_{i,l}\}^{(t)}).$$

	The input parameter $\{\underline{a}_l\}, \{\underline{b}_l\}, \{\underline{c}_l\}$	The input parameter θ
Remarks	▪ Delivering the information on X_{i-1}, X_{i+1} ▪ which are functions of the R2P/L2P parameters ▪ The element index l indicates the four different situations: $1) l=1: X_{i-1} \approx X_i \approx X_{i+1}$, $2) l=2: X_{i-1} \neq X_i \approx X_{i+1}$, $3) l=2: X_{i-1} \approx X_i \neq X_{i+1}$, $4) l=4: X_{i-1} \neq X_i \neq X_{i+1}$	▪ Determining the denoising threshold ▪ When $\theta \to \infty$, the denoiser should generate a constant output for all ρ_i. ▪ When $\theta \to 0$, the denoiser simply passes the input ρ_i such that denoising is not necessary.

- **The ssAMP denoiser** is **Lipschitz continuous** (omitting proof here) such that we can predict MSE of the ssAMP iteration using the Bayati's SE framework [TIT'11:Bayati et.al].

	$\mu_{\mathrm{L2P},i}$	$\mu_{\mathrm{R2P},i}$	$\sigma_{\mathrm{L2P},i}$	$\sigma_{\mathrm{R2P},i}$	$\{a_{i,l}\}$	$\{b_{i,l}\}$	$\{\frac{c_{i,l}}{\sum c_{i,l'}}\}$
Case (a)	5.0	-5.0	0.1	0.1	$\{0.0, 4.99, -4.99, 0.0\}$	$\{0.05, 0.1, 0.1, 50.05\}$	$\{0.0, 0.34, 0.34, 0.31\}$
Case (b)	5.0	-5.0	10.0	0.1	$\{-4.90, 4.09, -4.99, 0.24\}$	$\{0.1, 9.09, 0.1, 52.41\}$	$\{0.0, 0.34, 0.34, 0.31\}$
Case (c)	5.0	-5.0	10.0	10.0	$\{0.0, 4.17, -4.17, 0.0\}$	$\{5.0, 9.17, 9.17, 55.0\}$	$\{0.10, 0.31, 0.31, 0.28\}$
Case (d)	5.0	5.0	0.1	0.1	$\{5.0, 5.0, 5.0, 5.0\}$	$\{0.05, 0.1, 0.1, 50.05\}$	$\{0.90, 0.04, 0.04, 0.02\}$
Case (e)	5.0	5.0	10.0	0.1	$\{5.0, 5.0, 5.0, 5.0\}$	$\{0.1, 9.09, 0.1, 52.41\}$	$\{0.54, 0.17, 0.17, 0.12\}$
Case (f)	5.0	5.0	10.0	10.0	$\{5.0, 5.0, 5.0, 5.0\}$	$\{5.0, 9.17, 9.17, 55.0\}$	$\{0.48, 0.19, 0.19, 0.14\}$



Performance Validation

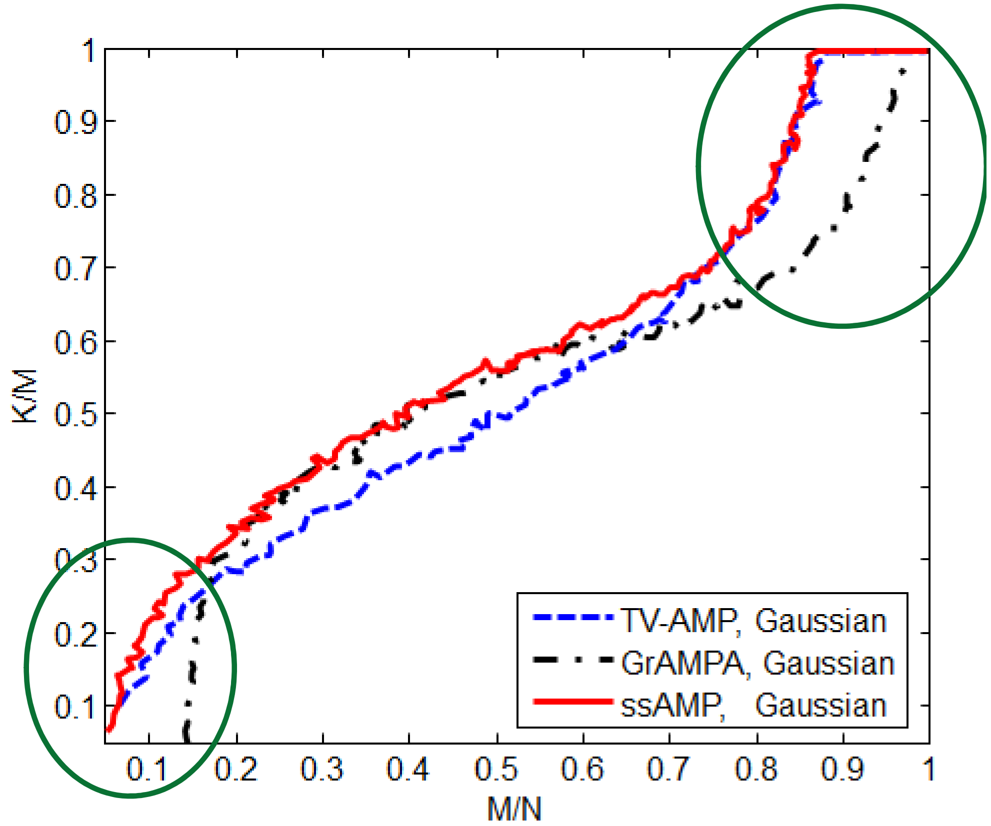
❖ Phase transition Characteristic (PTC)

- **ssAMP outperforms TV-AMP and GrAMPA in terms of PTC.**
  - ➢ In the region of low sampling ratio $(0 < \mathrm{M/N} \leq 0.1)$
  - ➢ In the region of high sampling ratio $(0.6 < \mathrm{M/N} \leq 1.0)$
- **Unsuccessful recovery of GrAMPA due to its residual update**

$$r_k^{(t+1)} = \eta_{\mathrm{GrAMPA}}(\mu_i^{(t)} - \mu_{i-1}^{(t)}) - (\mu_i^{(t)} - \mu_{i-1}^{(t)}) + \mathrm{Onsager},$$

  - ➢ 1) The residual update relies on not real measurements but the estimate of $X_i - X_{i-1}$ by $\eta_{GrAMPA}(\cdot)$.
  - ➢ 2) The onsager term cannot properly cancel cross-interference among the signal elements.

PTC for Gaussian piecewise-constancy: N=500, Noiseless, maxiter=500



Performance Validation

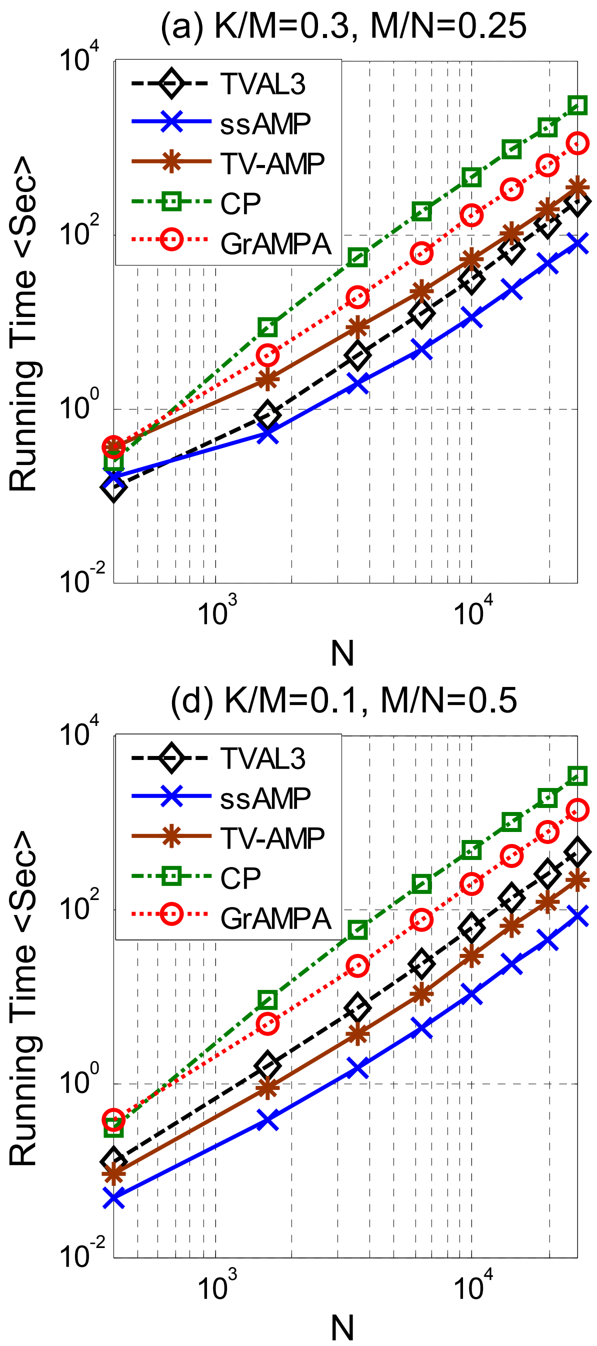
❖ Algorithm runtime over N

- ssAMP has computational advantage over recent TV solvers, TVAL3, TV-AMP, CP, and GrAMPA, in algorithm runtime.

  - ➢ Total complexity of all the algorithms scales as O(MN) since the matrix multiplications, $\mathbf{H}\underline{\boldsymbol{x}}$ or $\mathbf{H}^T\underline{y}$, dominate the complexity.

  - ➢ In such a situation, ssAMP retains its place as the fastest algorithm when $N \geq 10^3$.

  - ➢ When $N = 160^2$, the ssAMP runs more than 10 times faster than CP and GrAMPA.



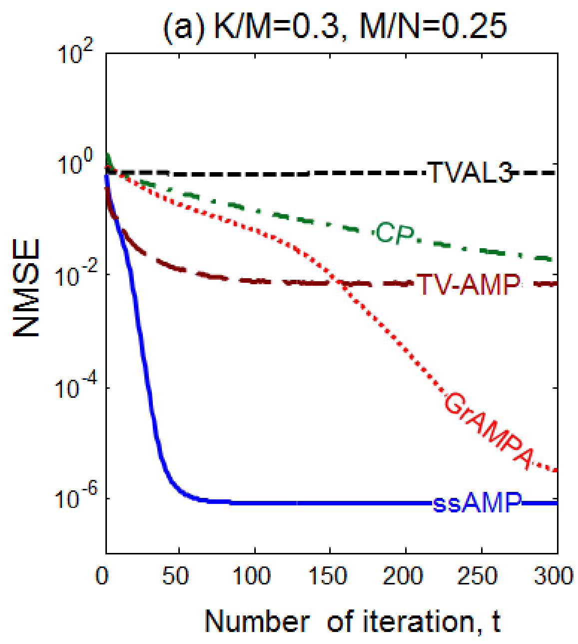
Performance Validation

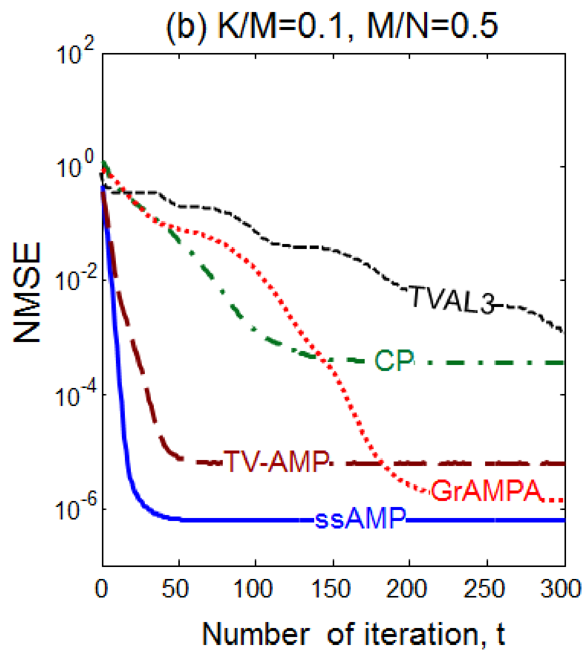
❖ MSE convergence over iterations

- The computational advantage of ssAMP is mainly from its fast MSE convergence.
  - **TV-AMP**: The most competitive wrt. per-iteration runtime, No deep convergence in MSE.
  - **GrAMPA**: Slow convergence, high per-iteration runtime, Deep MSE convergence.
  - **CP**: The most high per-iteration runtime due to many matrix multiplications in its processing.
  - **TVAL3**: variable per-iteration runtime, the worst convergence characteristic

Runtime (in sec) per iteration in the experiment ($N = 100^2, \Delta = 10^{-5}$)

Conditions	TVAL3	ssAMP	TV-AMP	CP	GrAMPA
(a) Severe condition	variable	0.17	0.16	2.23	0.82
(b) Rich condition	variable	0.32	0.31	2.37	0.99





Conclusions and Further works

❖ Conclusions

- The ssAMP algorithm operates with a scalarwise denoiser generating AMMSE estimate and holding the Lipschitz continuity.→ Low complex, Being applicable to the SE framework.
- PTC of ssAMP covers that of two existing AMPs for the PcR problem.
- ssAMP has computational advantage over the recent TV solver in runtime.

❖ Further works

- A simplification work of the denoiser and all embedded functions in the ssAMP iteration (Looking for a co-working master student)
- 2D extension of ssAMP algorithm by applying the tree-reweighted approach
  (An ongoing work with Hyoyoung Jung since June 2014)