Sparse Solutions to Linear Inverse Problems

Yuzhe Jin

Outline

- Intro/Background
- Two types of algorithms
  - Forward Sequential Selection Methods
  - Diversity Minimization Methods
- Experimental results
- Potential application to Speech

Background: Sparseness

- A large vector with only a very small # of non-zero entries

$$\begin{array}{c}
\\
\\
\longrightarrow \\
\\
\longrightarrow \\
\\
\\
\\
\\
\\
\longrightarrow \\
\\
\\
\end{array}
\begin{bmatrix} 0 \\ 0 \\ 2.6 \\ 0 \\ 9 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 3 \\ 0 \\ 0 \end{bmatrix}$$

Why sparse? In what scenarios?

- Bio-magnetic inverse problem
- Band-limited extrapolation, especially for Speech signals
- Direction-of-arrival estimation
- Channel equalization, Echo cancellation
- Image restoration
- …

- **Represent a signal of interest using the minimum number of vectors from an over-completed dictionary.**

Problem Description

$$\begin{bmatrix} | \\ b \\ | \end{bmatrix} = \begin{bmatrix} | & | & | & & | & | & | \\ a_1 & a_2 & a_3 & \dots & a_{n-2} & a_{n-1} & a_n \\ | & | & | & & | & | & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ . \\ x_{n-2} \\ x_{n-1} \\ x_n \end{bmatrix}$$

a measurement

a “dictionary” with each column as a codeword

sparse source, only a few of the entries are non-0

Problem Description

$$\begin{bmatrix} | \\ b \\ | \end{bmatrix} = \begin{bmatrix} | & | & | & & | & | & | \\ a_1 & a_2 & a_3 & \dots & a_{n-2} & a_{n-1} & a_n \\ | & | & | & & | & | & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ x_{n-2} \\ x_{n-1} \\ x_n \end{bmatrix}$$

a measurement

a “dictionary” with each column as a codeword

$$\mathbf{b} = \sum x_i \cdot \mathbf{a}_i$$

sparse source, only a few of the entries are non-0

With multiple measurements

$$\begin{bmatrix} | & \cdots & | \\ b_1 & \cdots & b_m \\ | & \cdots & | \end{bmatrix} = \begin{bmatrix} | & | & | & & | & | & | \\ a_1 & a_2 & a_3 & \cdots & a_{n-2} & a_{n-1} & a_n \\ | & | & | & & | & | & | \end{bmatrix} \begin{bmatrix} x_{11} & \cdots & x_{1,m} \\ x_{21} & \cdots & x_{2,m} \\ x_{31} & \cdots & x_{3,m} \\ . & \cdots & . \\ . & \cdots & . \\ x_{n-2,1} & \cdots & x_{n-2,m} \\ x_{n-1,1} & \cdots & x_{n-1,m} \\ x_{n,1} & \cdots & x_{n,m} \end{bmatrix}$$

Multiple measurements

$$\mathbf{B} = \mathbf{AX}$$

Multiple sources with same sparsity profile

With multiple measurements

$$\begin{bmatrix} | & \ldots & | \\ b_1 & \ldots & b_m \\ | & \ldots & | \end{bmatrix} = \begin{bmatrix} | & | & | & & | & | & | \\ a_1 & a_2 & a_3 & \ldots & a_{n-2} & a_{n-1} & a_n \\ | & | & | & & | & | & | \end{bmatrix} \begin{bmatrix} x_{11} & \ldots & x_{1,m} \\ \boxed{x_{21}} & \boxed{\ldots} & \boxed{x_{2,m}} \\ x_{31} & \ldots & x_{3,m} \\ . & \ldots & . \\ . & \ldots & . \\ \boxed{x_{n-2,1}} & \boxed{\ldots} & \boxed{x_{n-2,m}} \\ x_{n-1,1} & \ldots & x_{n-1,m} \\ x_{n,1} & \ldots & x_{n,m} \end{bmatrix}$$

Multiple measurements

$$\mathbf{B} = \mathbf{AX}$$

Multiple sources with same sparsity profile

Then, add noise to the observations

- New Model:

$$\mathbf{B} = \mathbf{AX} + \underline{\mathbf{N}}$$

- **Modeling error**
- **Noise present**

- Tradeoff between fit and sparsity of solution

Then, add noise to the observations

- New Model:

$$\mathbf{B} = \mathbf{A}\mathbf{X} + \underline{\mathbf{N}}$$

- **Modeling error**
- **Noise present**

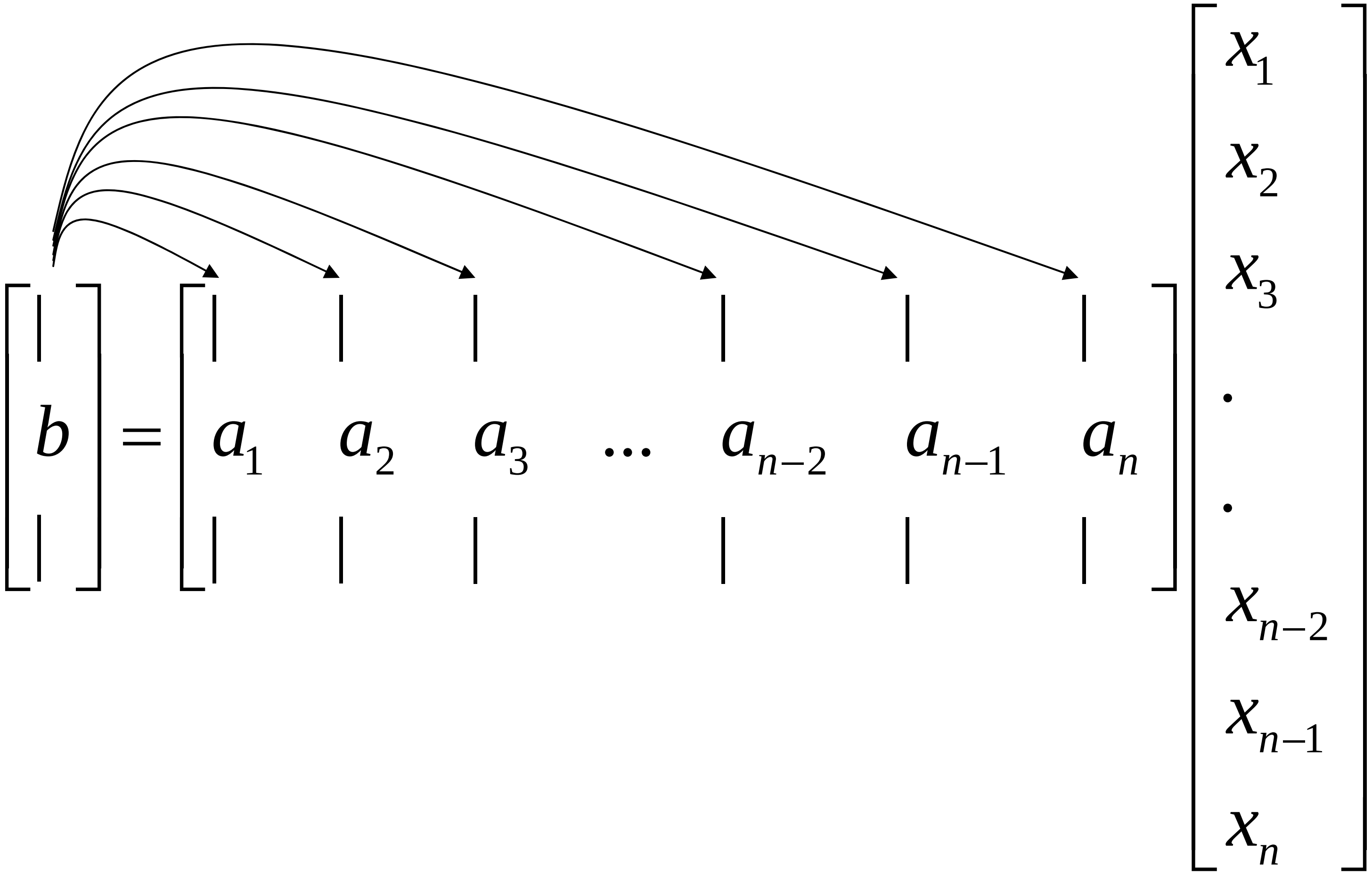
- Tradeoff between fit and sparsity of solution

$$\left\| \mathbf{A}\mathbf{X} - \mathbf{B} \right\|$$

Type I: **Forward Sequential Selection**

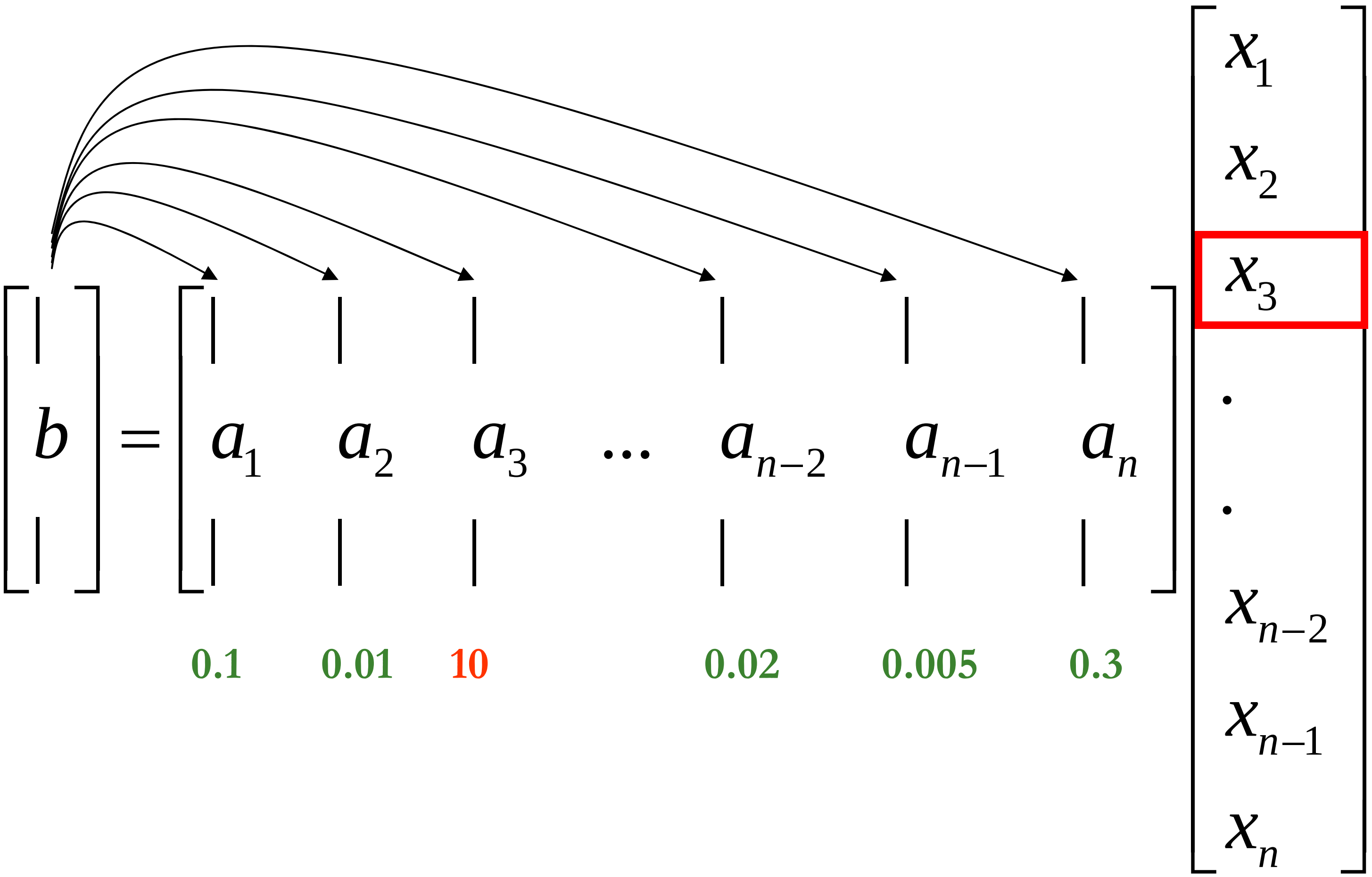
$$\underset{\substack{\uparrow\\ \textbf{Known}}}{\begin{bmatrix} | \\ b \\ | \end{bmatrix}} = \underset{\substack{\uparrow\\ \textbf{Known}}}{\begin{bmatrix} | & | & | & & | & | & | \\ a_1 & a_2 & a_3 & \dots & a_{n-2} & a_{n-1} & a_n \\ | & | & | & & | & | & | \end{bmatrix}} \underset{\substack{\uparrow\\ \textbf{Unknown}}}{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ . \\ x_{n-2} \\ x_{n-1} \\ x_n \end{bmatrix}}$$

Type I: **Forward Sequential Selection**



$$\begin{bmatrix} | \\ b \\ | \end{bmatrix} = \begin{bmatrix} | & | & | & & | & | & | \\ a_1 & a_2 & a_3 & \dots & a_{n-2} & a_{n-1} & a_n \\ | & | & | & & | & | & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ . \\ x_{n-2} \\ x_{n-1} \\ x_n \end{bmatrix}$$

Type I: **Forward Sequential Selection**



$$\begin{bmatrix} | \\ b \\ | \end{bmatrix} = \begin{bmatrix} | & | & | & & | & | & | \\ a_1 & a_2 & a_3 & \cdots & a_{n-2} & a_{n-1} & a_n \\ | & | & | & & | & | & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \boxed{x_3} \\ . \\ . \\ x_{n-2} \\ x_{n-1} \\ x_n \end{bmatrix}$$

How to compute residual?

- **Basic Matching Pursuit**
  - ❑ Remove contribution from the selected vector

Project onto one direction $\longrightarrow$

$$P_{\mathbf{a}_{k_p}}^{\perp}\mathbf{b}_{p-1}$$

- **Orthogonal Matching Pursuit**
  - ❑ Remove contribution from the selected subspace

Project onto a sub-space $\rightarrow$

$$P_{S_p}^{\perp}\mathbf{b}_{p-1} \qquad S_p = [S_{p-1}, \mathbf{a}_{k_p}]$$

Another variation: Order Recursive Matching Pursuit

- Most like Orthogonal Matching Pursuit
- Main Difference:
  - The vector is selected by

$$k_p = \arg\max_k \frac{<\mathbf{b}, \mathbf{a}_k>}{\left\| \mathbf{a}_k^{(p-1)} \right\|^2} \qquad \mathbf{a}_k^{(p-1)} = P_{S_p}^{\perp} \mathbf{a}_k$$

normalization term →

Correction:

- **Remove the contribution from vectors found previously**
- **Normalized inner product (since the codewords are not orthonormal)**

Type II: **Diversity Minimization**

- Metric of Sparseness

$$E^{(p)}(\mathbf{x}) = \sum_{i=1}^{n} \left| x[i] \right|^{p}$$

  - ***p* = 2**: Like 2-norm
    - Commonly used in engineering solutions
    - Nothing about sparsity
  - ***p* = 0**: a count on the non-zero entries in **x**
    - Direct measurement on sparseness
    - Very hard to solve, exhaustive search, NP-hard, combinatorial…
  - **What if *p* takes some value in between?**
    - In practice, setting *p* to some value between 0.8 ~ 1.0 gives good tradeoff between computational complexity and quality of sparse solution

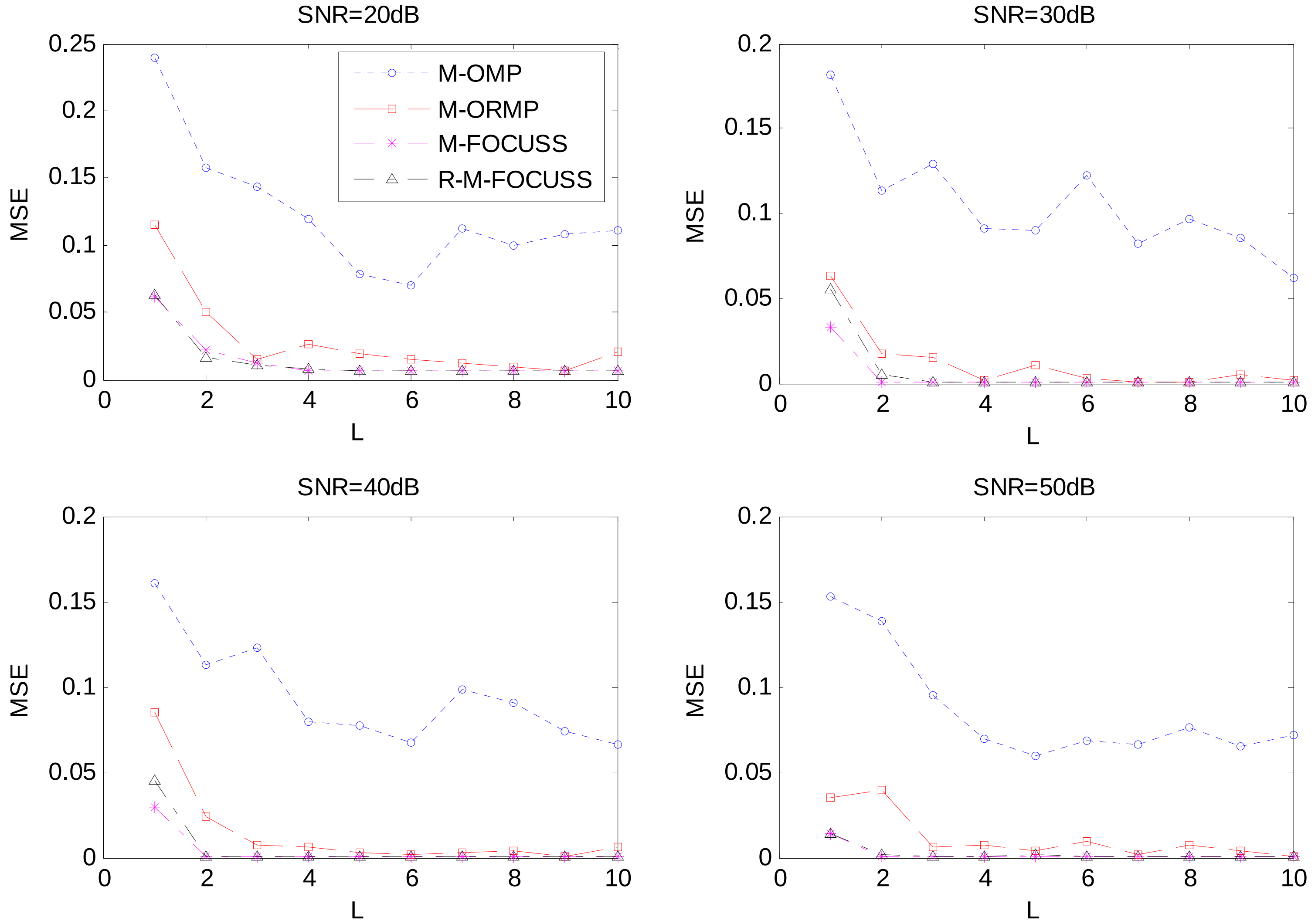
Now, we have a cost function. What's Next?

- Gradient Descent
  - FOCUSS-class Algorithms
    - FOCUSS
    - Regularized FOCUSS
      - Add in a regularization term to improve matrix condition for computing its inverse
      - Introduce bias, tradeoff between bias and quality of convergence

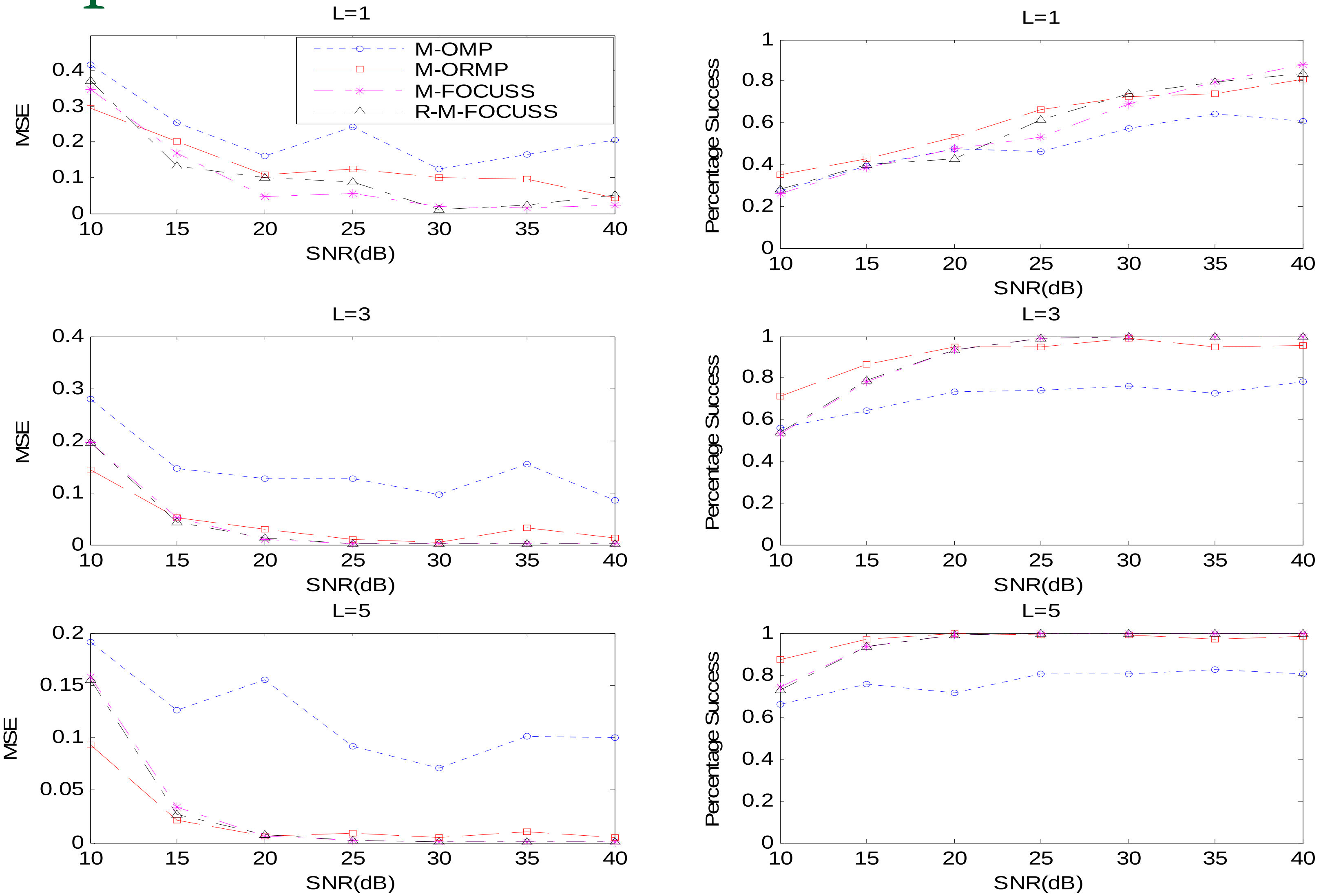
Experiments Setup

- Different number of measurements
- Different SNR
- Compare performances across 4 algorithms
  - M-OMP, M-ORMP, M-FOCUSS, R-M-FOCUSS
- Test:
  - If an algorithm correctly identifies the non-zero positions, **Percentage Success**
  - The MSE between the recovered vector and the ground truth, **MSE**

Experimental results

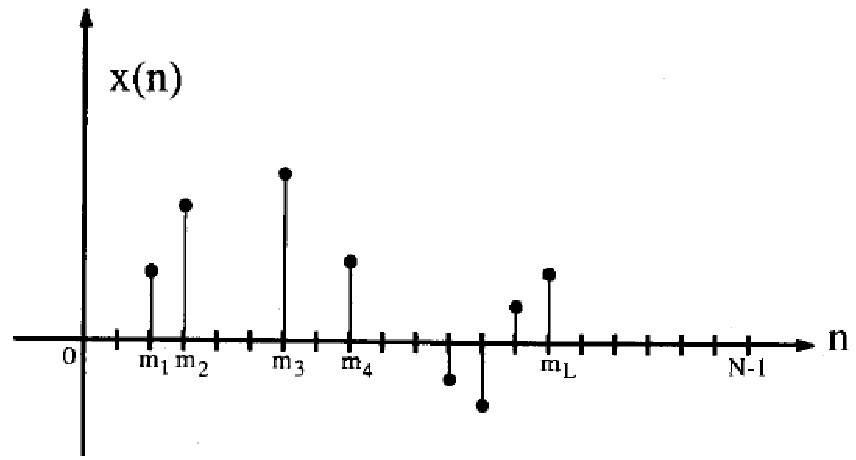


Experimental results



Applications to Speech

- Speech: Sparse in frequency domain
- Extrapolation



- Compression

Conclusion

- These algorithms are able to explore sparse linear inverse problems efficiently. They are robust to signal contaminated by noise.

- Speech signal is sparse in certain transform domain. We can apply this technique to speech.

References

- Sparse solutions to linear inverse problems with multiple measurements vectors, S. F. Cotter, B. D. Rao, K. Engan, K.K-Delgado, IEEE Trans. Sig. Proc., July 2005
- Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm, I. F. Gorodnitsky, B. D. Rao, IEEE Trans. Sig. Proc., March 1997
- Extrapolation and spectral estimation with iterative weighted norm modification, S. D. Cabrera, T. W. Parks, IEEE Trans. Sig. Proc., April 1991
- Forward sequential algorithms for best basis selection, S. F. Cotter, J. Adler, B. D. Rao, K. K-Delgado, IEE Proc. Image Sig. Proc., Oct 1999