Outlier-Robust Greedy Pursuit Algorithms in ℓ_p-Space for Sparse Approximation

Hing Cheung So

http://www.ee.cityu.edu.hk/~hcso

Department of Electronic Engineering
City University of Hong Kong

Outline

- Introduction
- ℓ_p-Correlation and ℓ_p-Orthogonality
- Robust Greedy Pursuit Algorithms
- Numerical Examples
- Summary
- List of References

Introduction

What is Sparse Approximation?

Sparse approximation refers to decomposing a target signal into a linear combination of very few elements drawn from a fixed collection.

Consider the following underdetermined linear system:

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v}$$

where $\boldsymbol{b} \in \mathbb{R}^N$ is the observed vector, $\boldsymbol{A} = [\boldsymbol{a}_1 \ \cdots \ \boldsymbol{a}_M] \in \mathbb{R}^{N\times M}$ with $N < M$ is known coefficient matrix, $\boldsymbol{x} = [x_1 \ \cdots \ x_M]^T \in \mathbb{R}^M$ is the unknown signal-of-interest and $\boldsymbol{v} \in \mathbb{R}^N$ is the noise.

The task is to find the sparsest $\boldsymbol{x}$ or the minimum ℓ_0-norm solution of $\boldsymbol{x}$ such that $\boldsymbol{b} \approx \boldsymbol{A}\boldsymbol{x}$.

Why Sparse Approximation is Important?

In real world, many signals-of-interest have a sparse representation in some basis.

As a result, this problem is a core issue in numerous areas of science and engineering including statistics, signal processing, machine learning, medical imaging and computer vision. It is dual to sparse recovery whose aim is to retrieve a high-dimensional signal based on a small number of linear measurements.

For example, in magnetic resonance imaging, we hope to collect as few observations (i.e., $\boldsymbol{b} \in \mathbb{R}^N$) as possible because scan time reduction means benefits for patients and health care economics.

How to Perform Sparse Approximation?

Finding $\boldsymbol{x}$ with minimum number of nonzero entries or the minimum ℓ_0-norm solution of $\boldsymbol{x}$ is in fact NP hard.

In practice, greedy pursuit and convex optimization are two standard approaches for obtaining an approximate solution.

Formulating the sparse approximation problem as:

$$\min_{\boldsymbol{x} \in \mathbb{R}^M} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2, \quad \text{subject to } \|\boldsymbol{x}\|_0 \leq s$$

where $s \in \mathbb{N}$ is the target sparsity of $\boldsymbol{x}$, the key idea of greedy pursuit is to identify the nonzero components sequentially. At each iteration, one column of $\boldsymbol{A}$ that is best correlated with the residual from the previous iteration is chosen, then its contribution to $\boldsymbol{b}$ is subtracted.

Representative greedy pursuit algorithms include matching pursuit (MP), orthogonal MP (OMP) and weak MP.

On the other hand, the convex optimization approach aims to approximate the ℓ_0-norm by the ℓ_1-norm, and widely-used methods include the least absolute shrinkage and selection operator (LASSO):

$$\min_{\boldsymbol{x}\in\mathbb{R}^M} \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2, \quad \text{subject to } \|\boldsymbol{x}\|_1 \leq \delta, \quad \delta \geq 0$$

basis pursuit (BP):

$$\min_{\boldsymbol{x}\in\mathbb{R}^M} \|\boldsymbol{x}\|_1, \quad \text{subject to } \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2 \leq \epsilon, \quad \epsilon \geq 0$$

and ℓ_1-regularization:

$$\min_{\boldsymbol{x}\in\mathbb{R}^M} \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{x}\|_1, \quad \lambda \geq 0$$

Motivation of ℓ_p-Space Sparse Approximation

Derivation of these conventional techniques is based on the ℓ_2-norm objective function, which implicitly assumes Gaussian data. In spite of providing theoretical and computational convenience, it is generally understood that the validity of the Gaussian distribution is at best approximate in reality.

Non-Gaussian impulsive noise arises in many practical applications. These standard solvers may fail to work properly when the observations contain outliers.

We propose to apply greedy pursuit to solve:

$$\min_{\boldsymbol{x}\in\mathbb{R}^M} \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_p^p, \quad \text{subject to } \|\boldsymbol{x}\|_0 \leq s, \quad 0 < p < 2$$

ℓ_p-Correlation and ℓ_p-Orthogonality

At the kth step of the iterative procedure of ℓ_2-norm based MP, we find a column of $\boldsymbol{A} = [\boldsymbol{a}_1 \ \cdots \ \boldsymbol{a}_M]$ that is most strongly correlated with the residual and the column index i_k is determined as:

$$i_k = \arg \max_{m \in \{1,2,\cdots,M\}} \left| \left\langle \boldsymbol{r}^{k-1}, \boldsymbol{a}_m \right\rangle \right|$$

Updated approximation and residual are then computed as:

$$x_{i_k}^k = x_{i_k}^{k-1} + \frac{\left\langle \boldsymbol{r}^{k-1}, \boldsymbol{a}_{i_k} \right\rangle}{\|\boldsymbol{a}_{i_k}\|_2^2}$$

and

$$\boldsymbol{r}^k = \boldsymbol{r}^{k-1} - \frac{\left\langle \boldsymbol{r}^{k-1}, \boldsymbol{a}_{i_k} \right\rangle}{\|\boldsymbol{a}_{i_k}\|^2} \boldsymbol{a}_{i_k}$$

ℓ_p-Correlation

To derive the robust MP, we first generalize the inner product or correlation, which is based on ℓ_2-norm, to ℓ_p-space.

The ℓ_p-correlation of $\boldsymbol{a} = [a_1 \ \cdots \ a_N]^T$ and $\boldsymbol{b} = [b_1 \ \cdots \ b_N]^T$ with $p > 0$, is defined as:

$$c_p(\boldsymbol{a}, \boldsymbol{b}) \triangleq \|\boldsymbol{b}\|_p^p - \min_{\alpha \in \mathbb{R}} \|\boldsymbol{b} - \alpha \boldsymbol{a}\|_p^p$$

where the ℓ_p-norm is

$$\|\boldsymbol{b}\|_p = \left(\sum_{n=1}^{N} |b_n|^p \right)^{1/p}$$

ℓ_p-correlation has the following properties:

- The case of $\boldsymbol{a} = \boldsymbol{b}$ elicits the definition of auto-ℓ_p-correlation $c_p(\boldsymbol{a}, \boldsymbol{a}) = \|\boldsymbol{a}\|_p^p$.

- If $\boldsymbol{a} = \boldsymbol{0}$ or $\boldsymbol{b} = \boldsymbol{0}$, then $c_p(\boldsymbol{a}, \boldsymbol{b}) = 0$. In other words, any vector has zero ℓ_p-correlation with a zero vector.

- For any $\beta \in \mathbb{R}$, $c_p(\beta\boldsymbol{a}, \boldsymbol{b}) = c_p(\boldsymbol{a}, \boldsymbol{b})$ and $c_p(\boldsymbol{a}, \beta\boldsymbol{b}) = |\beta|^p c_p(\boldsymbol{a}, \boldsymbol{b})$. That is, the ℓ_p-correlation is scale-invariant with respect to the first vector $\boldsymbol{a}$ but homogeneous with respect to the second vector $\boldsymbol{b}$.

- $0 \leq c_p(\boldsymbol{a}, \boldsymbol{b}) \leq \|\boldsymbol{b}\|_p^p$. In addition, $c_p(\boldsymbol{a}, \boldsymbol{b})$ attains its maximum if and only if there exists a scalar β such that $\boldsymbol{b} = \beta\boldsymbol{a}$, that is, $\boldsymbol{a}$ and $\boldsymbol{b}$ are collinear.

To make ℓ_p-correlation scale-invariant with respect to both vectors, we define normalized ℓ_p-correlation coefficient:

$$\theta_p(\boldsymbol{a},\boldsymbol{b}) \triangleq \frac{c_p(\boldsymbol{a},\boldsymbol{b})}{\|\boldsymbol{b}\|_p^p} = 1 - \frac{\min_{\alpha} \|\boldsymbol{b} - \alpha\boldsymbol{a}\|_p^p}{\|\boldsymbol{b}\|_p^p}$$

At $p=2$, $\theta_p(\boldsymbol{a},\boldsymbol{b})$ is reduced to

$$\theta_2(\boldsymbol{a},\boldsymbol{b}) = \frac{|\langle \boldsymbol{a},\boldsymbol{b}\rangle|^2}{\|\boldsymbol{a}\|^2\|\boldsymbol{b}\|^2}$$

That is, $\theta_p(\boldsymbol{a},\boldsymbol{b})$ generalizes the conventional correlation coefficient.

Normalized ℓ_p-correlation has the following properties:

- auto-ℓ_p-correlation coefficient of a nonzero vector $\boldsymbol{a}$ is $\theta_p(\boldsymbol{a},\boldsymbol{a}) = 1$.
- If $\boldsymbol{a} = \mathbf{0}$ or $\boldsymbol{b} = \mathbf{0}$, then $\theta_p(\boldsymbol{a},\boldsymbol{b}) = 0$.
- The ℓ_p-correlation coefficient is scale-invariant. That is to say, for any nonzero scalars $\beta_1 \in \mathbb{R}$ and $\beta_2 \in \mathbb{R}$, we have $\theta_p(\beta_1\boldsymbol{a},\beta_2\boldsymbol{b}) = \theta_p(\boldsymbol{a},\boldsymbol{b})$.
- $0 \leq \theta_p(\boldsymbol{a},\boldsymbol{b}) \leq 1$. In addition, $\theta_p(\boldsymbol{a},\boldsymbol{b})$ attains its maximum 1 if and only if there exists a scalar β such that the nonzero vector $\boldsymbol{b} = \beta\boldsymbol{a}$, that is, $\boldsymbol{a}$ and $\boldsymbol{b}$ are collinear.

ℓ_p-Orthogonality

Two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are ℓ_p-orthogonal if $c_p(\boldsymbol{a},\boldsymbol{b})=0$.

At $p=2$, this reduces to orthogonality in inner product space as $c_2(\boldsymbol{a},\boldsymbol{b})=0$ is equivalent to $\boldsymbol{a}^T\boldsymbol{b}=0$.

The relationship between ℓ_p-orthogonality and the global minimizer of the residual function

$$\alpha^\star = \arg\min_{\alpha} \left\{ f_p(\alpha) = \|\boldsymbol{b}-\alpha\boldsymbol{a}\|_p^p \right\}$$

- If $\alpha^\star=0$, then $c_p(\boldsymbol{a},\boldsymbol{b})=0$, and $\boldsymbol{a}$ and $\boldsymbol{b}$ are ℓ_p-orthogonal for any value $p>0$.
- $c_p(\boldsymbol{a},\boldsymbol{b})=0 \Leftrightarrow \alpha^\star=0, \quad p>1.$

Algorithms for ℓ_p-Correlation Computation

The key step for computing ℓ_p-correlation is to solve

$$\min_{\alpha} \left\{ f_p(\alpha) = \|\boldsymbol{b} - \alpha \boldsymbol{a}\|_p^p \right\}$$

For $p > 1$, $f_p(\alpha)$ is twice differentiable and strictly convex for any α.

The global solution can be easily obtained by Newton's method which has a complexity of $\mathcal{O}(N)$ at each iteration.

For $p = 1$, the problem is:

$$\min_{\alpha} \left\{ f_1(\alpha) \stackrel{\triangle}{=} \sum_{n=1}^{N} |a_n| \left| \alpha - \frac{b_n}{a_n} \right| \right\}$$

where $|a_n| > 0$ is considered as positive weight.

Defining a new sequence

$$d_n = \frac{b_n}{a_n}, \ n = 1, \cdots, N$$

The optimal $\alpha^\star$ is the weighted median of the sequence $\{d_n\}_{n=1}^N$ with weights $\{|a_n|\}_{n=1}^N$:

$$\alpha^\star = \mathrm{WMED}\left(|a_n|, d_n\right)$$

Algorithm for computing weighted median

Input: Weighting coefficients $\{|a_n|\}_{n=1}^N$ and data sequence $\{d_n\}_{n=1}^N$.

1. Determine the threshold $a_0 = (1/2) \sum_{n=1}^N |a_n|$.
2. Sort the data sequence $\{d_n\}_{n=1}^N$ in ascending order with the corresponding concomitant weights $\{|a_n|\}_{n=1}^N$.
3. Sum the concomitant weights, beginning with $|a_1|$ and increasing the order.
4. The weighted median $\alpha^\star$ is d_m whose weight leads to the inequality $\sum_{n=1}^m |a_n| \geq a_0$ hold first.

Output: The weighted median $\alpha^\star = \mathrm{WMED}\left(|a_n|, d_n\right)$.

For $p < 1$, the problem is reformulated as:

$$\min_{\alpha} \left\{ f_p(\alpha) \triangleq \sum_{n=1}^{N} |a_n|^p \left| \alpha - \frac{b_n}{a_n} \right|^p \right\}$$

Assume that $\{d_n\}_{n=1}^N$ has been sorted in ascending order.

The function $f_p(\alpha)$ is piecewise with breakpoints at $\{d_n\}_{n=1}^N$, that is, the domain of $f_p(\alpha)$ can be divided into $N+1$ intervals: $(-\infty, d_1]$, $[d_1, d_2], \cdots$, $[d_{N-1}, d_N]$, and $[d_N, \infty)$.

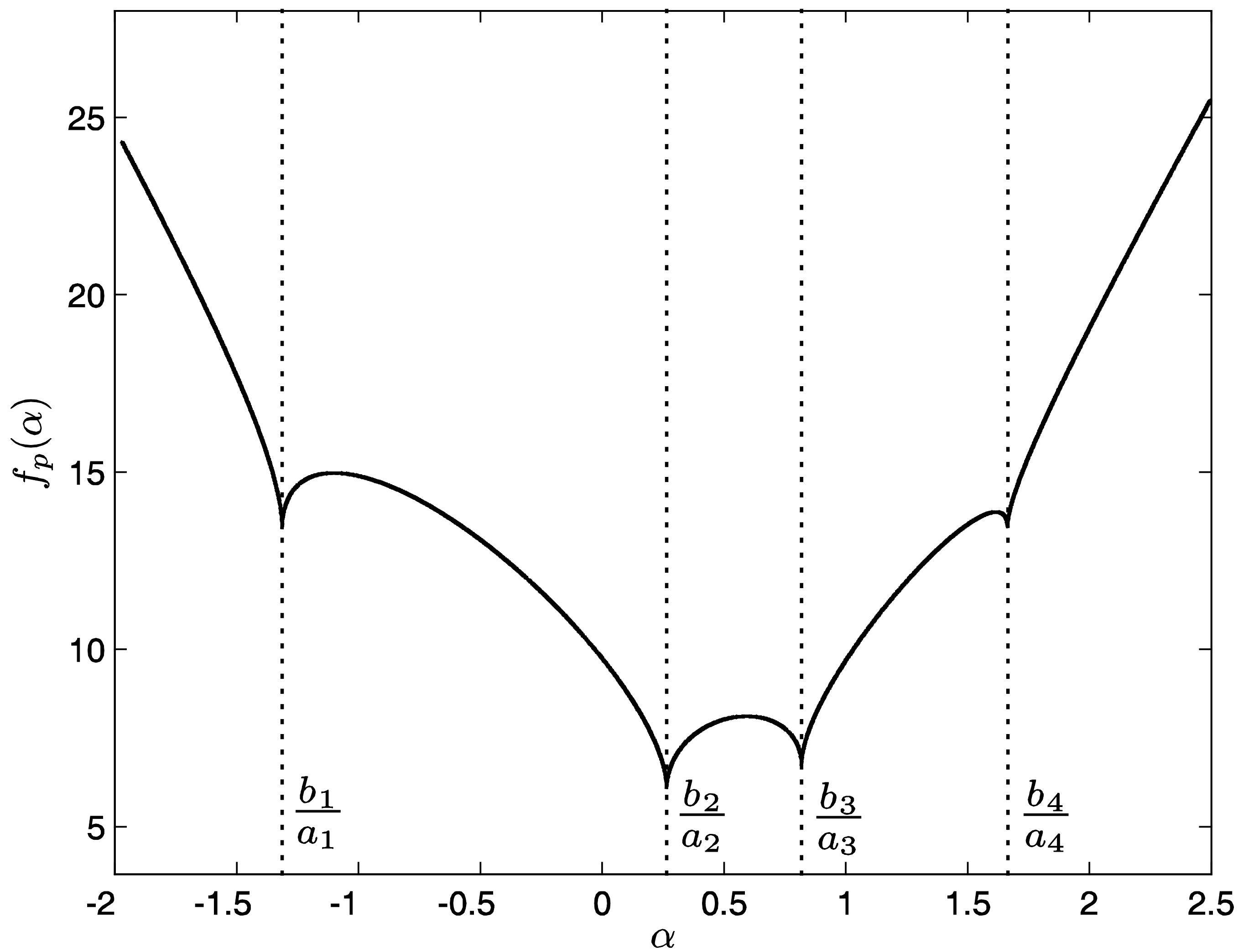
In each interval, the sign of $\{\alpha - d_n\}_{n=1}^N$ is known and the absolute operator $|\cdot|$ can be removed. Noting that $(\alpha - d_n)^p$ or $(d_n - \alpha)^p$ is a concave function due to $p < 1$, $f_p(\alpha)$ is concave because the non-negative combination preserves concavity.

As a concave function attains its minimum at the boundary points, the minimizer of $f_p(\alpha)$ belongs to $\{d_1, \cdots, d_N\}$ since $f_p(-\infty) = f_p(\infty) = \infty$:

$$\min_{\alpha} \left\{ f_p(\alpha) \stackrel{\Delta}{=} \sum_{n=1}^{N} |a_n|^p \left| \alpha - \frac{b_n}{a_n} \right|^p \right\} = \min_{1 \le n \le N} f_p(d_n)$$

The algorithm complexity is $\mathcal{O}(N^2)$.

Take an example: $p = 0.5$, $\boldsymbol{a} = [0.2939 \ -0.7873 \ 0.8884 \ -1.1471]^T$ and $\boldsymbol{b} = [0.4889 \ 1.0347 \ 0.7269 \ -0.3034]^T$. To find the global minimum of $f_p(\alpha)$, we compute the objective function value at the sorted $\boldsymbol{d} = [-1.3143 \ 0.2645 \ 0.8182 \ 1.6636]^T$ where $f_p(\alpha)$ attains its minimum at $d_2 = 0.2645$.



$f_p(\alpha) = \|\boldsymbol{b} - \alpha \boldsymbol{a}\|_p^p$ vs α at $p = 0.5$

Robust Greedy Pursuit Algorithms

ℓ_p-MP

It is the outlier-resistant version of MP.

Let $\boldsymbol{x}^k$, $\boldsymbol{r}^k$ and $\mathcal{I}^k$ be the solution, residual, and index set of the nonzero elements of $\boldsymbol{x}^k$ at the kth iteration.

Algorithm for ℓ_p-MP

Input: $\boldsymbol{A}$, $\boldsymbol{b}$, error tolerance ϵ, and target sparsity s

Initialization: Initialize $k = 0$, and set the initial solution $\boldsymbol{x}^0 = \boldsymbol{0}$, residual $\boldsymbol{r}^0 = \boldsymbol{b}$, and initial index set $\mathcal{I}^0 = \emptyset$.

Repeat

- $k \leftarrow k+1$
- Select the index i_k via

$$i_k = \arg \max_{1 \le i \le M} c_p(\boldsymbol{a}_i, \boldsymbol{r}^{k-1})$$

- Augment the index set $\mathcal{I}^k = \mathcal{I}^{k-1} \cup i_k$.
- Update the solution $\boldsymbol{x}^k \leftarrow \boldsymbol{x}^{k-1}$ and $x_{i_k}^k = x_{i_k}^{k-1} + \alpha_{i_k}^\star$ with

$$\alpha_{i_k}^\star = \arg \min_{\alpha} \|\boldsymbol{r}^{k-1} - \alpha \boldsymbol{a}_{i_k}\|_p^p$$

- Update the residual

$$\boldsymbol{r}^k = \boldsymbol{r}^{k-1} - \alpha_{i_k}^\star \boldsymbol{a}_{i_k}$$

until $\|\boldsymbol{r}^k\|_p^p \le \epsilon$ or $|\mathcal{I}^k| \le s$

Output: $\boldsymbol{x}^k$

Theorem 1

The ℓ_p-norm of the residual of the ℓ_p-MP algorithm decays exponentially with a rate proportional to $I_p(\boldsymbol{A})$:

$$\|\boldsymbol{r}^k\|_p^p \le (1 - I_p(\boldsymbol{A}))^k \, \|\boldsymbol{b}\|_p^p, \quad k = 1, 2, \cdots$$

where

$$I_p(\boldsymbol{A}) = \inf_{\boldsymbol{y} \neq \boldsymbol{0}} \lambda_p(\boldsymbol{A}, \boldsymbol{y}) > 0$$

with

$$\lambda_p(\boldsymbol{A}, \boldsymbol{y}) = \max_{1 \le i \le M} \theta_p(\boldsymbol{a}_i, \boldsymbol{y}) = \max_{1 \le i \le M} \frac{c_p(\boldsymbol{a}_i, \boldsymbol{y})}{\|\boldsymbol{y}\|_p^p} \in (0, 1]$$

being the ℓ_p-decay-factor of $\boldsymbol{y}$ and $\boldsymbol{A}$, which is the maximal normalized ℓ_p-correlation of $\boldsymbol{y}$ and the columns of $\boldsymbol{A}$.

Weak ℓ_p-MP

It is the outlier-resistant version of weak MP.

It does not attempt to find the index associated with the maximal possible ℓ_p-correlation $\max_i c_p(\boldsymbol{a}_i, \boldsymbol{r}^{k-1})$ but chooses the index i_k that satisfies

$$c_p(\boldsymbol{a}_{i_k}, \boldsymbol{r}^{k-1}) \geq \gamma^p \|\boldsymbol{r}^{k-1}\|_p^p, \quad 0 < \gamma < 1$$

Algorithm for Weak ℓ_p-MP

Input: $\boldsymbol{A}$, $\boldsymbol{b}$, error tolerance ϵ, and target sparsity s

Initialization: Initialize $k = 0$, and set the initial solution $\boldsymbol{x}^0 = \mathbf{0}$, residual $\boldsymbol{r}^0 = \boldsymbol{b}$, and initial index set $\mathcal{I}^0 = \emptyset$.

Repeat

- $k \leftarrow k+1$
- Set i_k as first index that satisfies $c_p(\boldsymbol{a}_i, \boldsymbol{r}^{k-1}) \geq \gamma^p \|\boldsymbol{r}^{k-1}\|_p^p$. If there is no such an index, set

$$i_k = \arg \max_{1 \leq i \leq M} c_p(\boldsymbol{a}_i, \boldsymbol{r}^{k-1})$$

- Augment the index set $\mathcal{I}^k = \mathcal{I}^{k-1} \cup i_k$.
- Update the solution $\boldsymbol{x}^k \leftarrow \boldsymbol{x}^{k-1}$ and $x_{i_k}^k = x_{i_k}^{k-1} + \alpha_{i_k}^\star$ with

$$\alpha_{i_k}^\star = \arg \min_{\alpha} \|\boldsymbol{r}^{k-1} - \alpha \boldsymbol{a}_{i_k}\|_p^p$$

- Update the residual

$$\boldsymbol{r}^k = \boldsymbol{r}^{k-1} - \alpha_{i_k}^\star \boldsymbol{a}_{i_k}$$

until $\|\boldsymbol{r}^k\|_p^p \leq \epsilon$ or $|\mathcal{I}^k| \leq s$

Output: $\boldsymbol{x}^k$

Theorem 2

The ℓ_p-norm of the residual of the weak ℓ_p-MP algorithm decays exponentially with a rate proportional to $\zeta_p = \min\left(\gamma^p, I_p(\boldsymbol{A})\right)$:

$$\|\boldsymbol{r}^k\|_p^p \le (1-\zeta_p)^k \|\boldsymbol{b}\|_p^p, \quad k = 1, 2, \cdots$$

ℓ_p-OMP

It is the outlier-resistant version of OMP.

Algorithm for ℓ_p-OMP

Input: $\boldsymbol{A}$, $\boldsymbol{b}$, error tolerance ϵ, and target sparsity s

Initialization: Initialize $k=0$, and set the initial solution $\boldsymbol{x}^0 = \boldsymbol{0}$, residual $\boldsymbol{r}^0 = \boldsymbol{b}$, initial index set $\mathcal{I}^0 = \emptyset$ and $\boldsymbol{A}_{\mathcal{I}^0} = \emptyset$.

Repeat

- $k \leftarrow k + 1$
- Select the index i_k via

$$i_k = \arg \max_{i \notin \mathcal{I}^{k-1}} c_p(\boldsymbol{a}_i, \boldsymbol{r}^{k-1})$$

- Augment the index set and the matrix of chosen atoms as $\mathcal{I}^k = \mathcal{I}^{k-1} \cup i_k$ and $\boldsymbol{A}_{\mathcal{I}^k} = [\boldsymbol{A}_{\mathcal{I}^{k-1}}, \boldsymbol{a}_{i_k}]$.
- Solve the ℓ_p-norm minimization problem:

$$\boldsymbol{x}_{\mathcal{I}^k} = \arg \min_{\boldsymbol{u}} \|\boldsymbol{A}_{\mathcal{I}^k}\boldsymbol{u} - \boldsymbol{b}\|_p^p$$

  to obtain the nonzero coefficients.
- Update the residual

$$\boldsymbol{r}^k = \boldsymbol{b} - \boldsymbol{A}_{\mathcal{I}^k}\boldsymbol{x}_{\mathcal{I}^k}$$

until $\|\boldsymbol{r}^k\|_p^p \leq \epsilon$ or $|\mathcal{I}^k| \leq s$

Output: Index set $\mathcal{I}^k$ and the corresponding coefficients $\boldsymbol{x}_{\mathcal{I}^k}$

Algorithm	Complexity
MP/OMP	$\mathcal{O}(sMN)$
ℓ_p-MP/ℓ_p-OMP ($p > 1$)	$\mathcal{O}(sMN)$
ℓ_p-MP/ℓ_p-OMP ($p = 1$)	$\mathcal{O}(sMN \log N)$
ℓ_p-MP/ℓ_p-OMP ($p < 1$)	$\mathcal{O}(sMN^2)$

Complexity of Index Selection

Algorithm	Complexity
OMP	$\mathcal{O}(s^2N^2)$
ℓ_p-OMP ($p \geq 1$)	$\mathcal{O}(s^2N^2N_{\rm IRLS})$
ℓ_p-OMP ($p < 1$)	$\mathcal{O}(s^2N^2N_{\rm IRLS})$ (local minimum)

Complexity of Orthogonalization

$N_{\rm IRLS}$ is number of iterations used in iteratively reweighted least squares in solving ℓ_p-norm minimization problem

Numerical Examples

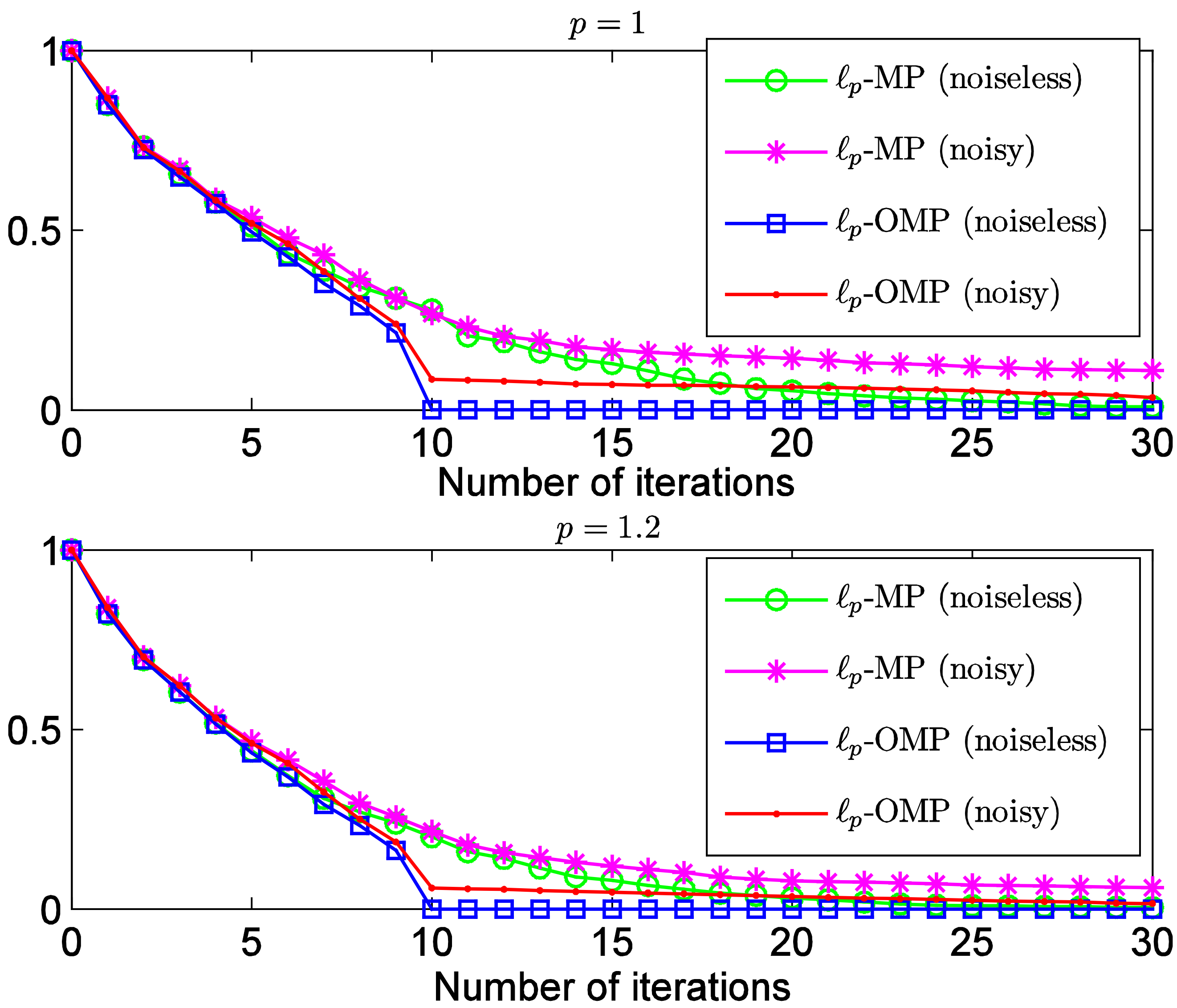
Sparse Recovery

Unknown $\boldsymbol{x} \in \mathbb{R}^{100}$ with $K = 10$ nonzero entries whose magnitudes are uniformly drawn in $[1, 2]$.

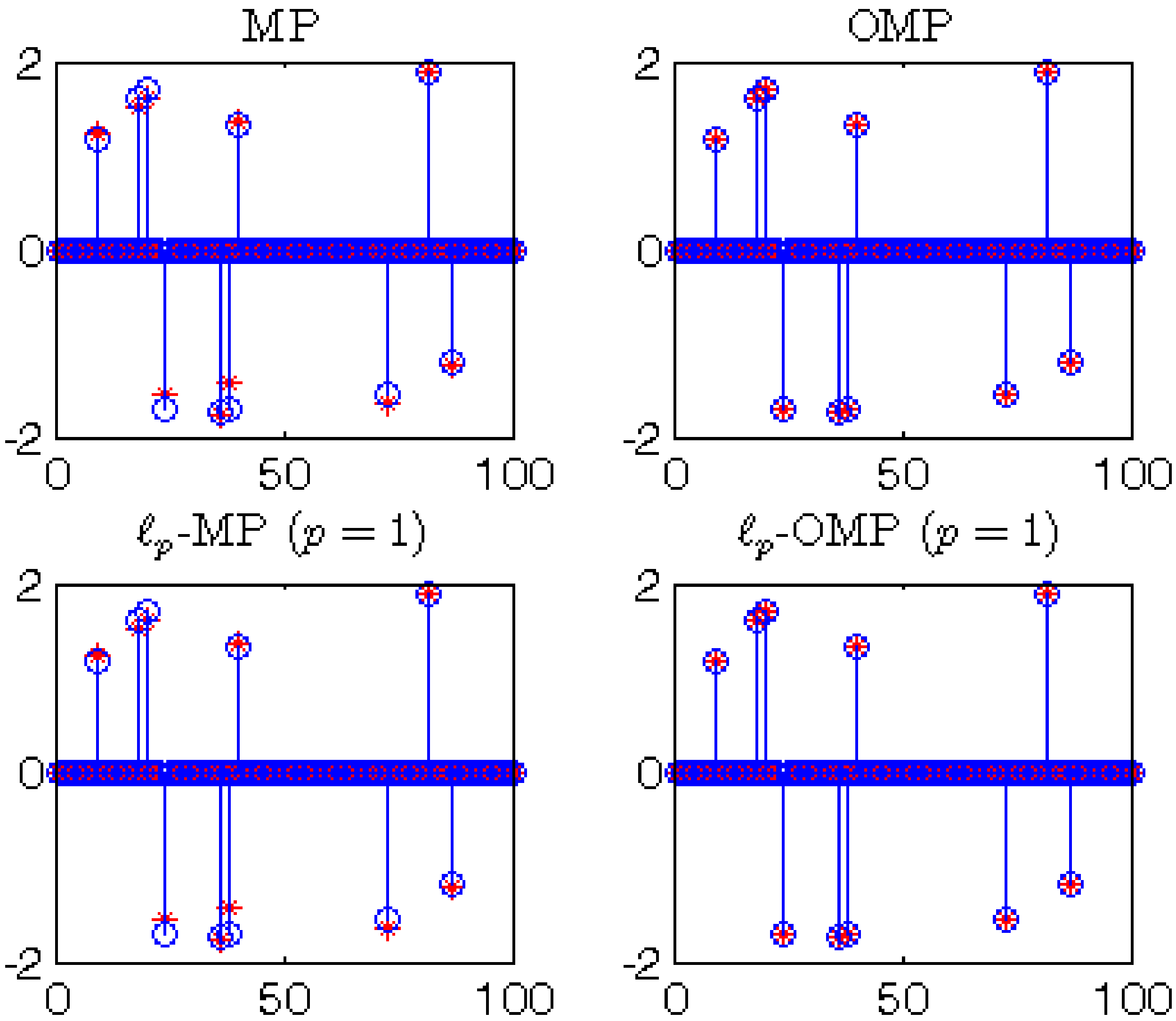
Known $\boldsymbol{A} \in \mathbb{R}^{60 \times 100}$ contains random entries.

Noise $\boldsymbol{v} \in \mathbb{R}^{60}$ contains Gaussian mixture model (GMM) or salt-and-pepper variables.

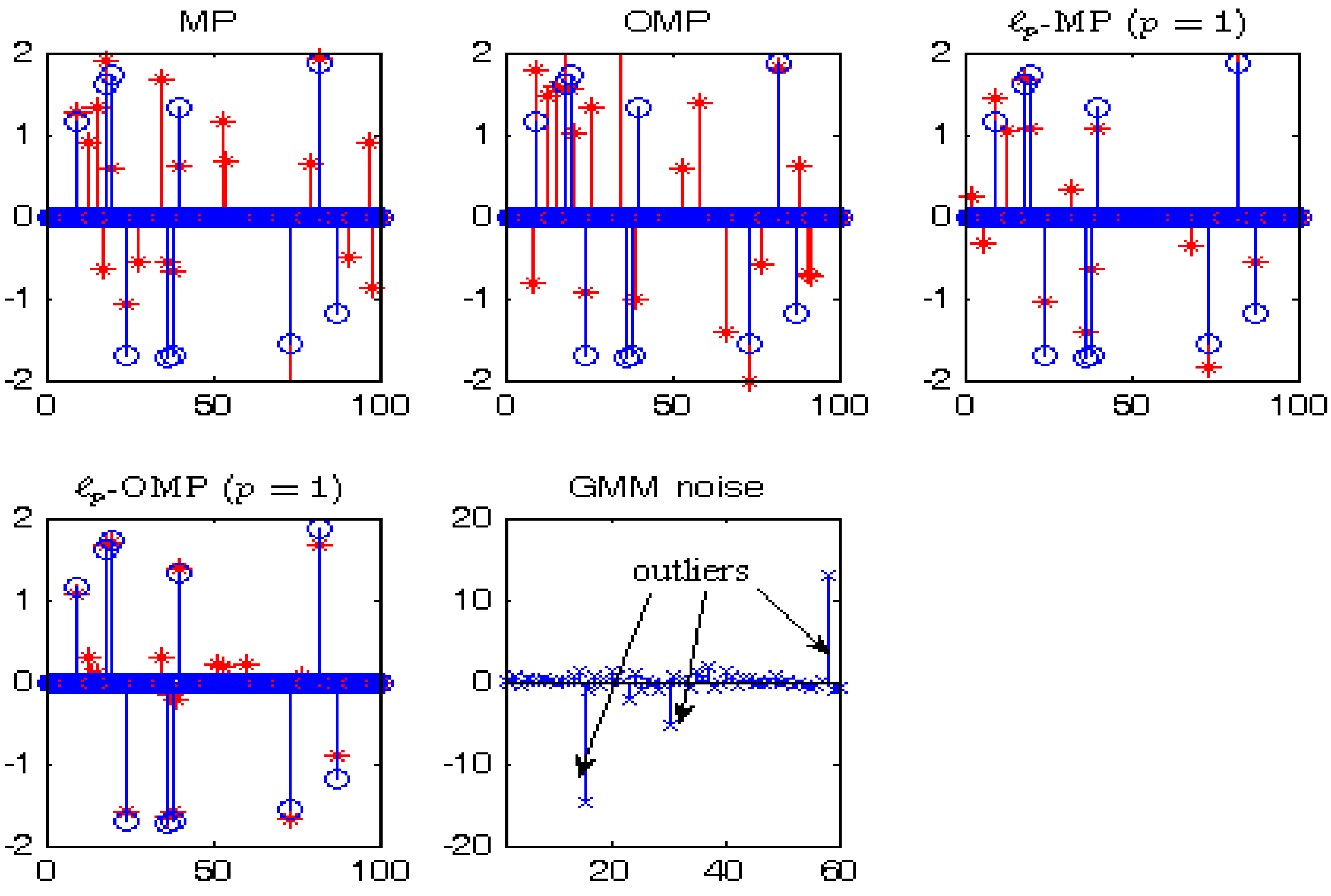
ℓ_p-MP and ℓ_p-OMP with $p = 1$ or $p = 1.2$ are compared with MP and OMP.



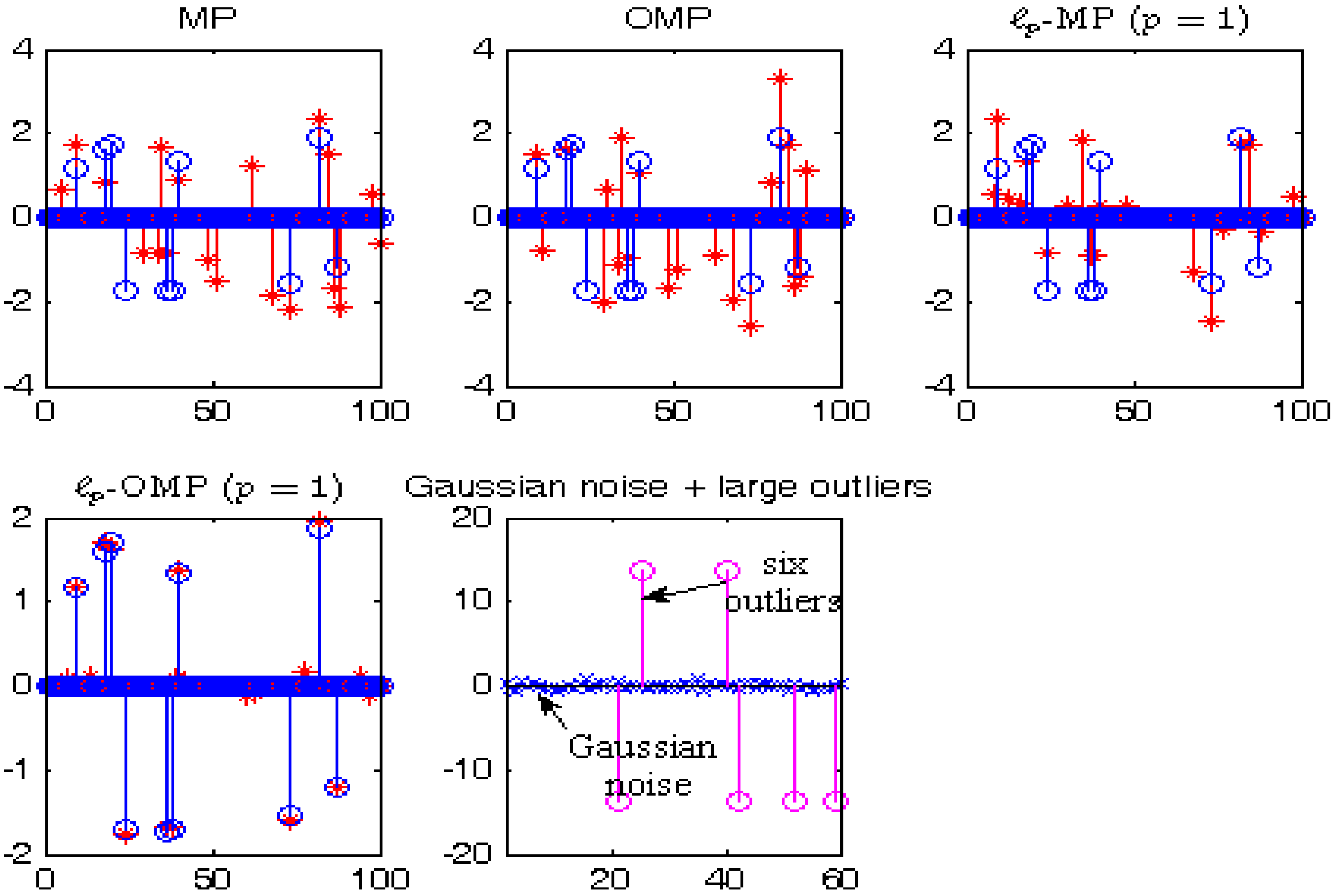
$\|\boldsymbol{r}^k\|_p^p / \|\boldsymbol{b}\|_p^p$ vs Iteration Number in Noiseless and GMM cases



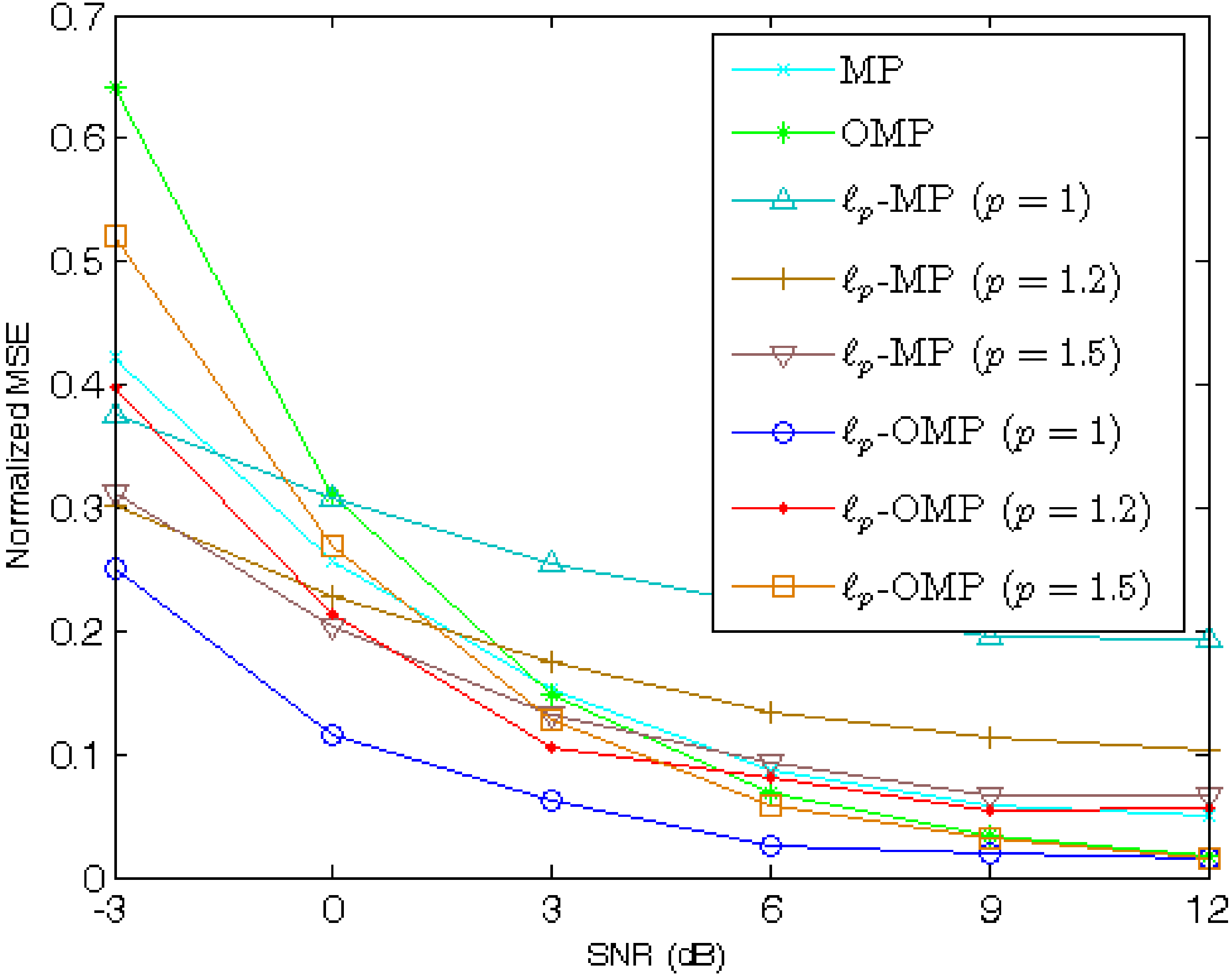
Sparse Recovery in Noiseless case



Sparse Recovery in GMM noise



Sparse Recovery in Salt-and-Pepper noise



MSE for $\boldsymbol{b}$ versus SNR in GMM noise

Harmonic Retrieval

The observation vector $\boldsymbol{b} = [b_1 \ \cdots \ b_N]^T$ is now:

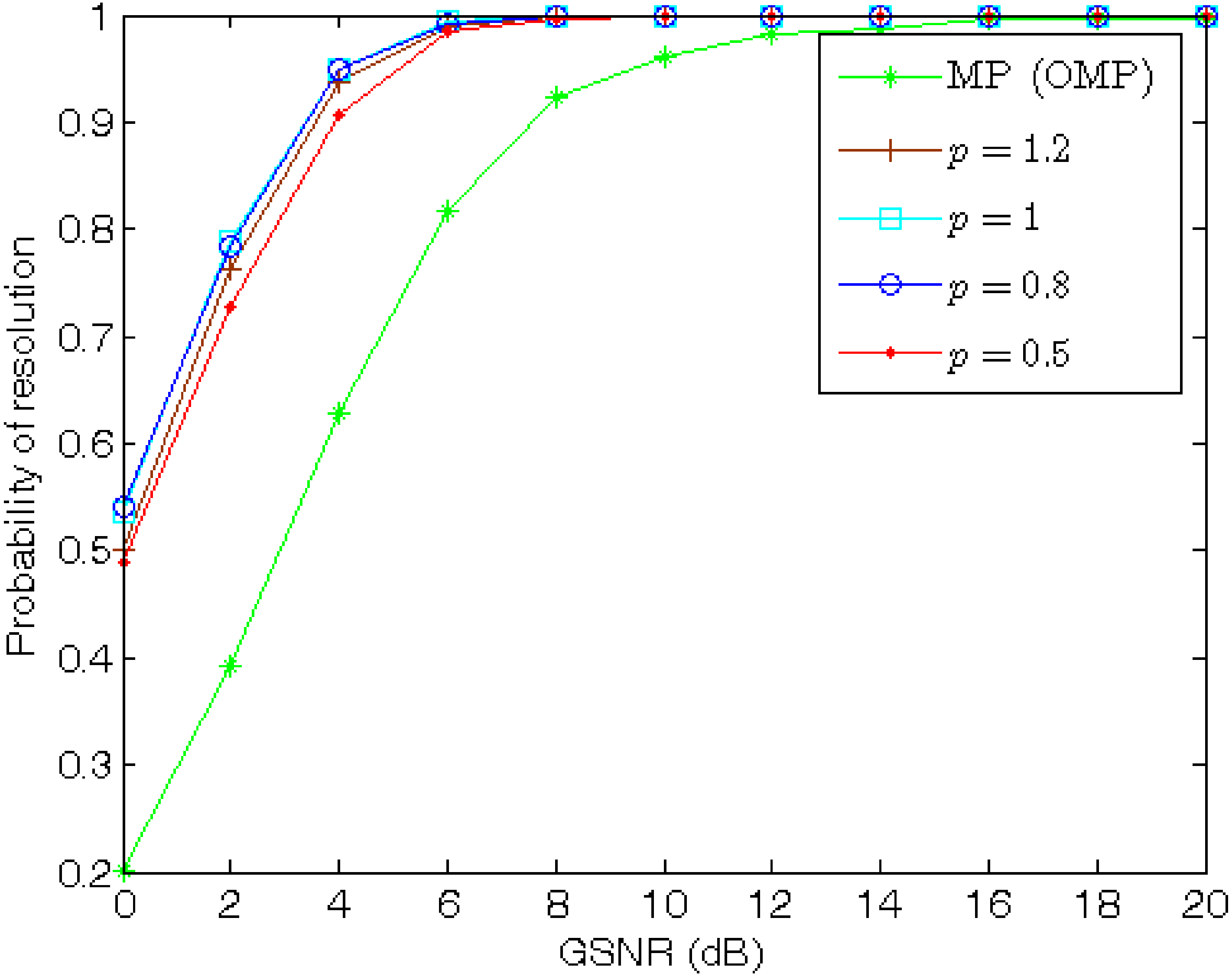
$$b_n = \sum_{k=1}^{K} \beta_k e^{j\omega_k n} + v_n, \quad K = 2, \ N = 50$$

The frequency matrix $\boldsymbol{A}$ is:

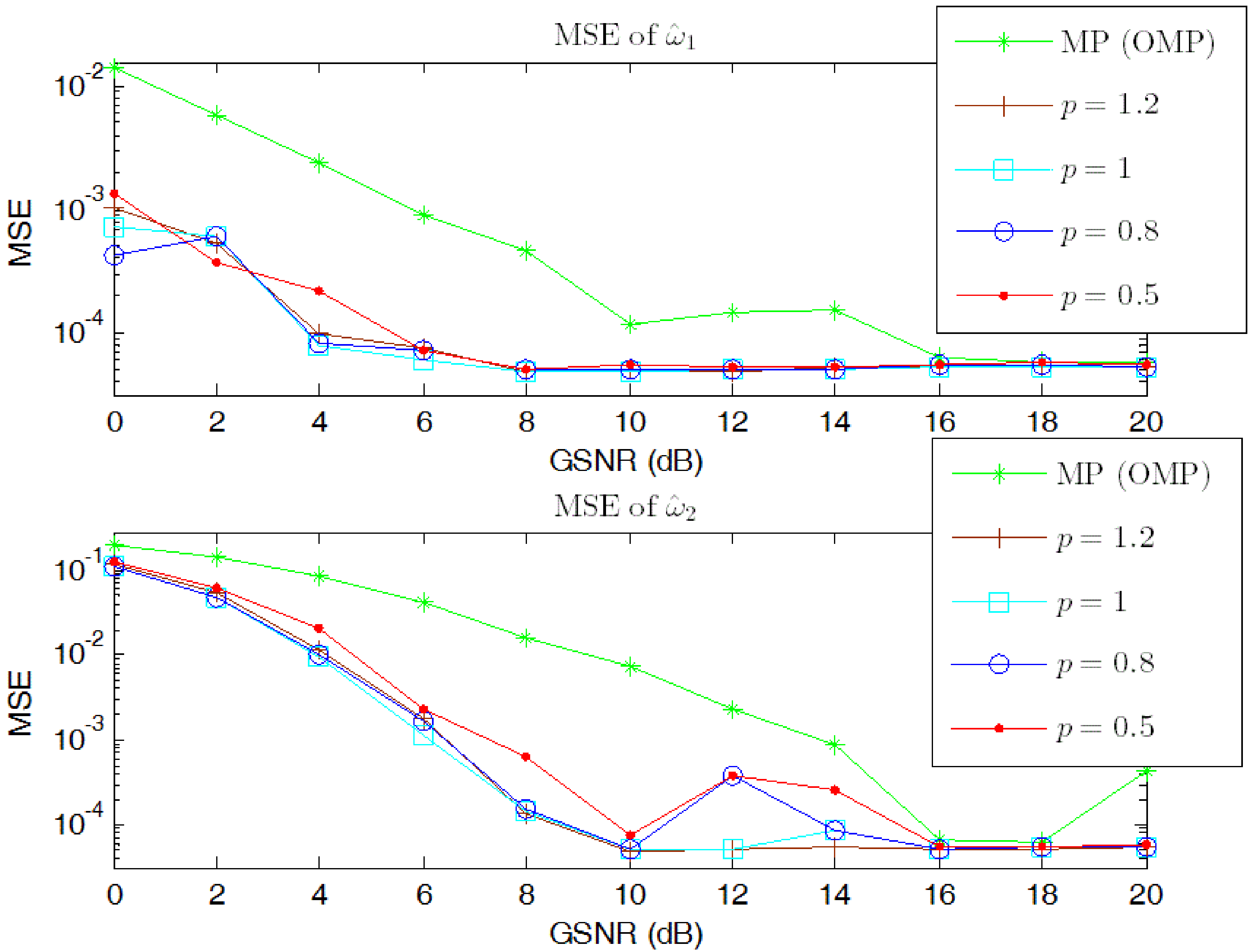
$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{j\omega_1} & e^{j\omega_2} & \cdots & e^{j\omega_M} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j\omega_1(N-1)} & e^{j\omega_2(N-1)} & \cdots & e^{j\omega_M(N-1)} \end{bmatrix} \in \mathbb{C}^{N \times M}, \quad M = 1000$$

$\boldsymbol{x} = [x_1 \ \cdots \ x_M]^T$ contains unknowns associated with M bins:

$$x_m = \begin{cases} \beta_k, & \text{if } \omega_m = \omega_k \\ 0, & \text{if } \omega_m \notin \{\omega_k\}_{k=1}^{K} \end{cases}$$



Probability of Resolution vs Generalized SNR in SαS noise



MSE vs Generalized SNR in SαS noise

Summary

- Novel concepts of ℓ_p-correlation and ℓ_p-orthogonality are devised and they generalize the standard correlation and orthogonality definitions in the inner product space.

- ℓ_p-correlation provides similarity measure of two vectors in ℓ_p-space where $p > 0$, and its computational efficient realizations are developed.

- ℓ_p-space versions of MP, OMP and weak MP are derived and they outperform the ℓ_2-norm based counterparts in the presence of impulsive noise or outliers.

List of References

[1] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, NY: Springer, 2010.

[2] M. Lustig, D. Donoho and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol.58, no.6, pp.1182-1195, 2007.

[3] J.A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol.50, no.10, pp.2231-2242, 2004.

[4] J.A. Tropp and A.C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol.53, no.12, pp.4655-4666, 2007.

[5] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society, Series B*, vol.58, no.1, pp.267-288, 1996.

[6] S.S. Chen, D.L. Donoho and M.A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol.20, no.1, pp.33-61, 1998.

[7] M. Figueiredo, R. Nowak and S. Wright, “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems,” *IEEE Journal of Selected Topics on Signal Processing*, vol.1, no.4, pp.586-597, 2007.

[8] A.M. Zoubir, V. Koivunen, Y. Chakhchoukh and M. Muma, “Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts,” *IEEE Signal Processing Magazine*, vol.29, no.4, pp.61-80, 2012.

[9] W.-J. Zeng, H.C. So and X. Jiang, “Outlier-robust greedy pursuit algorithms in ℓ_p-space for sparse approximation,” *to appear in IEEE Transactions on Signal Processing*, *2016*.