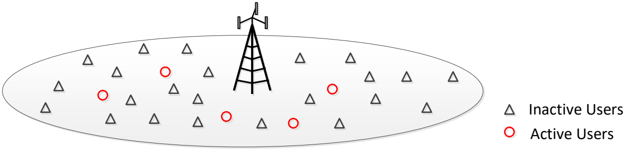
Massive Random Access with Massive MIMO: Covariance Based Detection

Wei Yu

Joint Work with Zhilin Chen, Foad Sohrabi, Ya-Feng Liu

University of Toronto

Massive Random Access for Internet-of-Things (IoT)



- Large number of devices with sporadic activity
- Low latency random access scheme for massive users is required
- Non-orthogonal signature sequences need to be used
- User activity detection (user identification) performed at base station (BS)

System Model

- BS equipped with M antennas
- N single-antenna devices, K of which are active at a time
- Each device is associated with a length-L unique signature sequence $\mathbf{s}_n$
- Channel $\mathbf{h}_n$ of user n includes both (i.i.d.) Rayleigh and large-scale fading
- For single-cell system, received signal $\mathbf{Y} \in \mathbb{C}^{L \times M}$ at the BS is

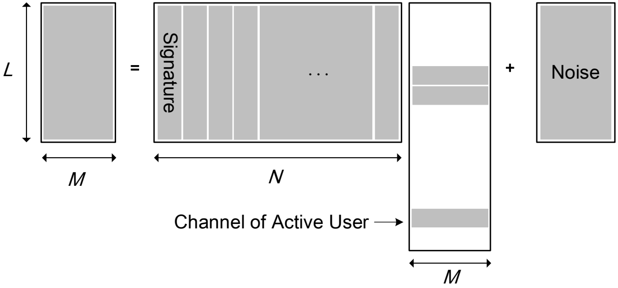
$$\mathbf{Y} = \sum_{n=1}^{N} \alpha_n \mathbf{s}_n \mathbf{h}_n^T + \mathbf{Z} = \mathbf{S}\mathbf{X} + \mathbf{Z}, \tag{1}$$

where

- $\alpha_n \in \{1, 0\}$ activity indicator; $\mathbf{Z} \in \mathbb{C}^{L \times M}$ Gaussian noise with variance σ^2
- $\mathbf{S} \triangleq [\mathbf{s}_1, \ldots, \mathbf{s}_N] \in \mathbb{C}^{L \times N}$; $\mathbf{X} \triangleq [\alpha_1 \mathbf{h}_1, \cdots, \alpha_N \mathbf{h}_N]^T \in \mathbb{C}^{N \times M}$

Joint Sparse Activity Detection and Channel Estimation

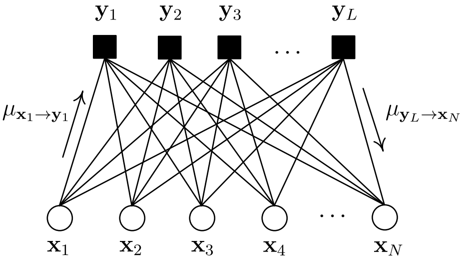
Aim to identify the K non-zero rows of $\mathbf{X}$ from $\mathbf{Y} = \mathbf{S}\mathbf{X} + \mathbf{Z}$.



- Multiple measurement vector (MMV) problem in compressed sensing
  - Columns of $\mathbf{X}$ share the same sparsity pattern, i.e., row sparsity
- Efficiently solved by the approximate message passing (AMP) algorithm

Approximate Message Passing (AMP)

Derived from graphical model of $\mathbf{Y} = \mathbf{S}\mathbf{X} + \mathbf{Z}$ [Donoho-Maleki-Montanari'09]



- Suppose entries of $\mathbf{S}$ are i.i.d. random
- Aim to compute the marginals of the joint distribution $p(\mathbf{X}, \mathbf{Y})$
- Approximate $\mu_{\mathbf{y} \to \mathbf{x}}$ as Gaussian in the large system limit $N, L \to \infty$
- Further simplify the messages such that only $N + L$ messages are tracked

Intuitive Interpretation of AMP

- Matched filtering $\longrightarrow$ Denoising $\longrightarrow$ Computing and correcting the residual

$$\eta(y;\theta) = \begin{cases} y-\theta, & y>\theta \\ 0, & -\theta \le y \le \theta \\ y+\theta, & y<-\theta \end{cases} \tag{2}$$

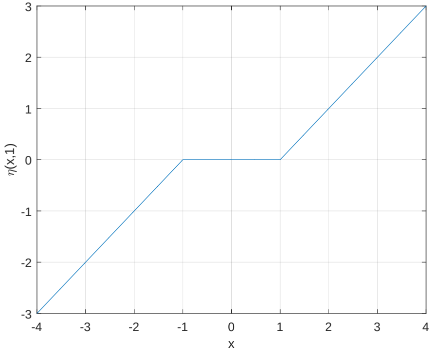


Figure: Soft thresholding function with $\theta = 1$

AMP Algorithm for MIMO

- The AMP algorithm expressed in matrix form:

$$\mathbf{X}^{t+1} = \eta_t(\mathbf{S}^H\mathbf{R}^t + \mathbf{X}^t), \tag{3}$$

$$\mathbf{R}^{t+1} = \mathbf{Y} - \mathbf{S}\mathbf{X}^{t+1} + \frac{N}{L}\mathbf{R}^t\langle\eta_t'(\mathbf{S}^H\mathbf{R}^t + \mathbf{X}^t)\rangle, \tag{4}$$

  where
  - $\mathbf{X}^{t+1}$, estimate at iteration $t+1$;
  - $\mathbf{R}^{t+1}$, residual at iteration $t+1$;
  - $\eta_t(\cdot)$, a non-linear function known as denoiser that performs on each row
  - $\langle\cdot\rangle$, sample averaging operation
- Works well if M is fixed, and L, N, $K \to \infty$.
- Complexity: $\mathcal{O}(NLM)$ + complexity of $\eta_t(\cdot)$ per iteration

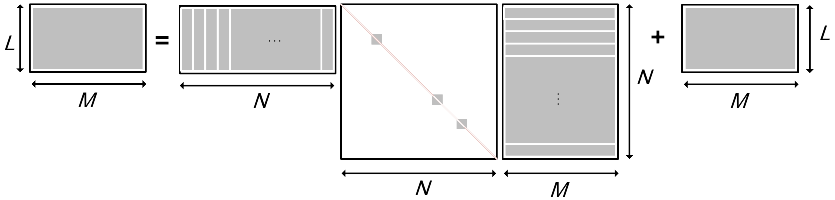
  But what if M is large? AMP becomes increasingly difficult to converge.

Joint Activity Detection and Large-Scale Fading Estimation

Key Assumption: We only need activity α_n and do not need $\mathbf{h}_n$.

Reformulate sparse activity detection as a large-scale-fading estimation problem:

$$\mathbf{Y} = \sum_{n=1}^{N} \alpha_n \mathbf{s}_n \mathbf{h}_n^T + \mathbf{Z} \triangleq \mathbf{S}\boldsymbol{\Gamma}^{\frac{1}{2}}\tilde{\mathbf{H}} + \mathbf{Z} \tag{5}$$

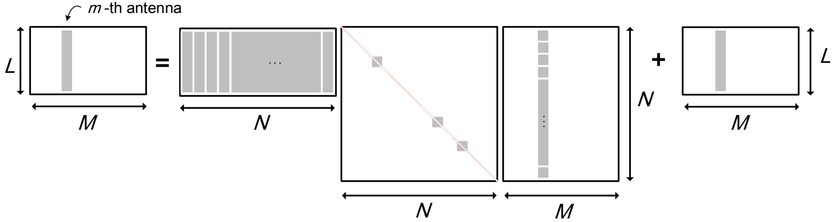


- $\mathbf{S} \triangleq [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N] \in \mathbb{C}^{L \times N}$, signature matrix
- $\boldsymbol{\Gamma} \triangleq \text{diag}\{\alpha_1\beta_1, \alpha_2\beta_2, \cdots, \alpha_N\beta_N\} \in \mathbb{R}^{N \times N}$, where β_n is large-scale fading
- $\tilde{\mathbf{H}} \triangleq \left[\mathbf{h}_1/\sqrt{\beta_1}, \mathbf{h}_2/\sqrt{\beta_2}, \cdots, \mathbf{h}_N/\sqrt{\beta_N}\right]^T \in \mathbb{C}^{N \times M}$, normalized channel matrix

Statistics of the Received Signal

Let $\mathbf{y}_m$ be the received signal at the m-th antenna, and let $\tilde{\mathbf{h}}_m$ be the normalized channel and $\mathbf{z}_m$ be the noise. Then, $\mathbf{y}_m$ can be expressed as

$$\mathbf{y}_m = \mathbf{S}\mathbf{\Gamma}^{\frac{1}{2}}\tilde{\mathbf{h}}_m + \mathbf{z}_m \tag{6}$$



- Model: Small-scale fading is i.i.d. Rayleigh across M received antennas.
- Then, $\tilde{\mathbf{h}}_m$ follows $\mathcal{CN}(\mathbf{0}, \mathbf{I})$. Also, $\mathbf{z}_m$ follows $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$.
- Therefore, given $\mathbf{\Gamma}$, $\mathbf{y}_m$ is i.i.d. across m as $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}$.

Maximum Likelihood Estimation of Γ

The sparse device activity is included in the diagonal matrix $\mathbf{\Gamma}$, which can be estimated using the maximum likelihood estimation (MLE) as:

$$
\begin{aligned}
\min_{\mathbf{\Gamma} \geq \mathbf{0}} \quad f(\mathbf{\Gamma}) &:= -\frac{1}{M} \log p(\mathbf{Y}|\mathbf{\Gamma}) && \longleftarrow \text{minimization of negative log-likelihood} \\
&= -\frac{1}{M} \sum_{m=1}^{M} \log p(\mathbf{y}_m|\mathbf{\Gamma}) && \longleftarrow \text{i.i.d. over antennas} \\
&= -\frac{1}{M} \sum_{m=1}^{M} \log \left(\frac{1}{|\pi \mathbf{\Sigma}|} \exp\left(-\mathbf{y}_m^H \mathbf{\Sigma}^{-1} \mathbf{y}_m\right) \right) && \longleftarrow \text{Gaussian distribution} \\
&= -\frac{1}{M} \sum_{m=1}^{M} \log \left(\frac{1}{|\pi \mathbf{\Sigma}|} \right) - \frac{1}{M} \sum_{m=1}^{M} \log \left(\exp\left(-\mathbf{y}_m^H \mathbf{\Sigma}^{-1} \mathbf{y}_m\right)\right) \\
&= \log|\mathbf{\Sigma}| + \frac{1}{M} \sum_{m=1}^{M} \operatorname{tr}\left(\mathbf{\Sigma}^{-1} \mathbf{y}_m \mathbf{y}_m^H\right) + const. && \longleftarrow \mathbf{x}^H \mathbf{A} \mathbf{x} = \operatorname{tr}\left(\mathbf{A}\mathbf{x}\mathbf{x}^H\right) \\
&= \log|\mathbf{\Sigma}| + \operatorname{tr}\left(\mathbf{\Sigma}^{-1} \frac{1}{M} \sum_{m=1}^{M} \mathbf{y}_m \mathbf{y}_m^H\right) + const. && (7)
\end{aligned}
$$

Sample Covariance as a Sufficient Statistic

Define the sample covariance matrix of the received signal as

$$\hat{\mathbf{\Sigma}} \triangleq \frac{1}{M}\sum_{m=1}^{M}\mathbf{y}_m\mathbf{y}_m^H = \frac{1}{M}\mathbf{Y}\mathbf{Y}^H. \tag{8}$$

With the sample covariance matrix, the MLE of $\mathbf{\Gamma}$ can be expressed as

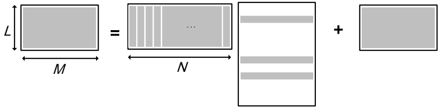
$$\begin{aligned}\min_{\mathbf{\Gamma}\geq\mathbf{0}} \quad f(\mathbf{\Gamma}) :&= \log|\mathbf{\Sigma}| + \operatorname{tr}\left(\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}\right) + const. \\ &= \log|\mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}| + \operatorname{tr}\left((\mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I})^{-1}\hat{\mathbf{\Sigma}}\right) + const.\end{aligned} \tag{9}$$

- $\hat{\mathbf{\Sigma}}$ is computed by averaging over different antennas, not time slots
- $\hat{\mathbf{\Sigma}}$ is a sufficient statistics since $f(\mathbf{\Gamma})$ depends on $\mathbf{Y}$ only through $\hat{\mathbf{\Sigma}}$
- The size of the MLE problem depends on N, L only, not M.

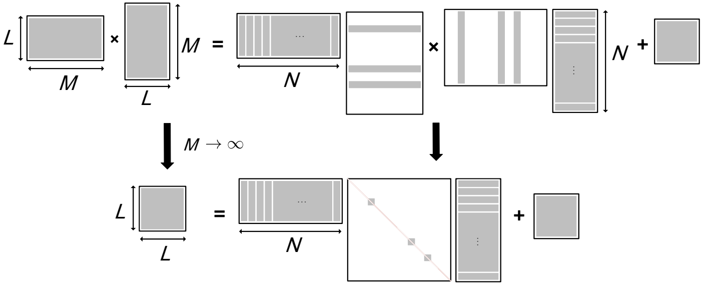
A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire: "Non-Bayesian Activity Detection, Large-Scale Fading Coefficient Estimation, and Unsourced Random Access with a Massive MIMO Receiver", *IEEE Trans. Inf. Theory*, May 2021. http://arxiv.org/abs/1910.11266

Covariance Based Sparse Activity Detection

Instead of jointly estimating the channel, i.e., the non-zero rows in $\mathbf{X}$ based on $\mathbf{Y}$:



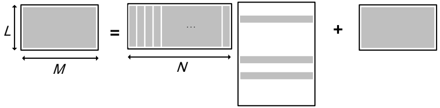
We now estimate large-scale fading $\mathbf{\Gamma}$ based on $\hat{\mathbf{\Sigma}} = \frac{1}{M}\mathbf{Y}\mathbf{Y}^H$:



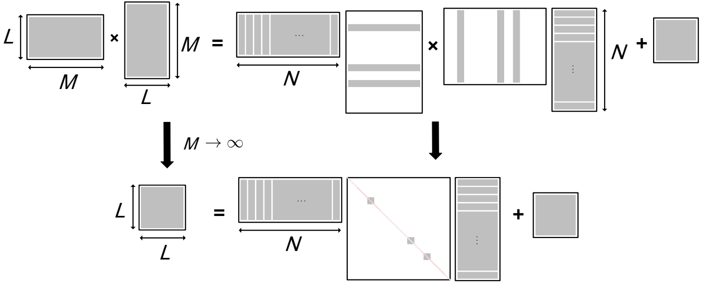
In the massive MIMO regime, i.e., if we let $M \to \infty$, this can be thought of detecting a diagonal sparse matrix from the sample covariance.

Covariance Based Sparse Activity Detection

Instead of jointly estimating the channel, i.e., the non-zero rows in $\mathbf{X}$ based on $\mathbf{Y}$:



We now estimate large-scale fading $\mathbf{\Gamma}$ based on $\hat{\mathbf{\Sigma}} = \frac{1}{M}\mathbf{Y}\mathbf{Y}^H$:



Crucial Advantage: Instead of detecting KM variables based on LM observations, we now detect K variables based on L^2 observations!

Covariance Based Sparse Activity Detection

To estimate $\mathbf{\Gamma}$, need to solve the optimization problem

$$\begin{aligned}\min_{\mathbf{\Gamma}\geq\mathbf{0}} \quad f(\mathbf{\Gamma}) &:= -\frac{1}{M}\log p(\mathbf{Y}|\mathbf{\Gamma}) \\ &= \log\left|\mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}\right| + \mathrm{tr}\left(\left(\mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}\right)^{-1}\hat{\mathbf{\Sigma}}\right) + const. \end{aligned} \tag{10}$$

- $f(\mathbf{\Gamma})$ is non-convex (since it is concave function + convex function)
  - Expectation-maximization [Wipf-Rao '07] (Sparse Bayesian Learning)
  - Coordinate descent [Haghighatshoar-Jung-Caire '18]
- Observe: In the large M limit, $f(\mathbf{\Gamma})$ is minimized by the true value $\mathbf{\Gamma}^0$:

$$\hat{\mathbf{\Sigma}} \triangleq \frac{1}{M}\sum_{m=1}^{M}\mathbf{y}_m\mathbf{y}_m^H \quad \rightarrow \quad \mathbf{\Sigma}^0 \triangleq \mathbf{S}\mathbf{\Gamma}^0\mathbf{S}^H + \sigma^2\mathbf{I}, \qquad \text{as} \quad M\rightarrow\infty. \tag{11}$$

Now consider the optimization (10) with $\hat{\mathbf{\Sigma}} = \mathbf{\Sigma}^0$, optimizing over $\mathbf{\Sigma}$ as in:

$$\min_{\mathbf{\Sigma}} \log|\mathbf{\Sigma}| + \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^0). \tag{12}$$

By taking derivative, we see $\mathbf{\Sigma}^{\mathrm{opt}} = \mathbf{\Sigma}^0$. For finite M, we need to solve (10).

Coordinate Descent for Solving the MLE problem

Let γ_n be the n-th diagonal entry of $\mathbf{\Gamma}$. The MLE can be expressed as

$$\min_{\gamma_1,\ldots,\gamma_N \geq 0} \log \left| \sum_{n=1}^{N} \gamma_n \mathbf{s}_n \mathbf{s}_n^H + \sigma^2 \mathbf{I} \right| + \operatorname{tr}\left(\left(\sum_{n=1}^{N} \gamma_n \mathbf{s}_n \mathbf{s}_n^H + \sigma^2 \mathbf{I} \right)^{-1} \hat{\mathbf{\Sigma}} \right). \tag{13}$$

- **Basic Idea**: Update the coordinates $\gamma_1, \ldots, \gamma_N$ alternatively
- **Coordinate update**: Let $\hat{\gamma}_n, \forall n$ be the current estimates. Update $\hat{\gamma}_k$ with other $\hat{\gamma}_n, n \neq k$ fixed at a time. Let $\hat{\gamma}_k + d$ be the update. Determine d by

$$\min_{d \geq -\hat{\gamma}_k} \quad \log\left(1 + d\mathbf{s}_k^H \tilde{\mathbf{\Sigma}}^{-1} \mathbf{s}_k\right) - \frac{d\mathbf{s}_k^H \tilde{\mathbf{\Sigma}}^{-1} \hat{\mathbf{\Sigma}} \tilde{\mathbf{\Sigma}}^{-1} \mathbf{s}_k}{1 + d\mathbf{s}_k^H \tilde{\mathbf{\Sigma}}^{-1} \mathbf{s}_k}. \tag{14}$$

  - $\tilde{\mathbf{\Sigma}} = \sum_{n=1}^{N} \hat{\gamma}_n \mathbf{s}_n \mathbf{s}_n^H + \sigma^2 \mathbf{I}$ is the current value of the covariance based on $\hat{\gamma}_n$.
  - The constraint $d \geq -\hat{\gamma}_k$ ensures the new $\hat{\gamma}_k + d$ is always non-negative.
  - By taking the derivative of the objective in (14), a closed-form solution is

$$d = \max\left\{ \frac{\mathbf{s}_k^H \tilde{\mathbf{\Sigma}}^{-1} \hat{\mathbf{\Sigma}} \tilde{\mathbf{\Sigma}}^{-1} \mathbf{s}_k - \mathbf{s}_k^H \tilde{\mathbf{\Sigma}}^{-1} \mathbf{s}_k}{(\mathbf{s}_k^H \tilde{\mathbf{\Sigma}}^{-1} \mathbf{s}_k)^2}, -\hat{\gamma}_k \right\}. \tag{15}$$

- **Advantages**: Efficient due to closed-form solution; empirically performs well.

Coordinate Descent for Device Activity Detection

We use two loops in the coordinate decent. The inner loop update all $\gamma_1, \ldots, \gamma_N$ in a random permuted order to ensure that each coordinate will be visited once.

Algorithm 1 Coordinate descent for device activity detection

Initialize $\hat{\boldsymbol{\gamma}} = \mathbf{0}, \tilde{\boldsymbol{\Sigma}} = \sigma^2 \mathbf{I}, \tilde{\boldsymbol{\Sigma}}^{-1} = \sigma^{-2}\mathbf{I}$.

for $i = 1, 2, \ldots, T$ **do**

  Select a random permutation $i_1, i_2, \ldots, i_N$ of indices $\{1, 2, \ldots, N\}$.

  for $n = 1$ to N **do**

    $d = \max\left\{ \frac{\mathbf{s}_{i_n}^H \tilde{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\Sigma}} \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{s}_{i_n} - \mathbf{s}_{i_n}^H \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{s}_{i_n}}{(\mathbf{s}_{i_n}^H \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{s}_{i_n})^2}, -\hat{\gamma}_{i_n} \right\}$

    $\hat{\gamma}_{i_n} = \hat{\gamma}_{i_n} + d$ $\leftarrow$ Coordinate descent update

    $\tilde{\boldsymbol{\Sigma}}^{-1} = \tilde{\boldsymbol{\Sigma}}^{-1} - d \frac{\tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{s}_{i_n} \mathbf{s}_{i_n}^H \tilde{\boldsymbol{\Sigma}}^{-1}}{1 + d \mathbf{s}_{i_n}^H \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{s}_{i_n}}$ $\leftarrow$ Rank-1 update of estimated covariance

  end for

end for

Output $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}_1, \ldots, \hat{\gamma}_N]^T$. Declare the activity with thresholding on $\hat{\boldsymbol{\gamma}}$.

Covariance Matching Approach

Recall that the MLE aims to recover $\mathbf{\Gamma}$ by solving the following problem

$$\min_{\mathbf{\Gamma} \geq \mathbf{0}} \quad f(\mathbf{\Gamma}) := -\frac{1}{M} \log p(\mathbf{Y}|\mathbf{\Gamma}) = \log|\mathbf{\Sigma}| + \operatorname{tr}\left(\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}\right) + const. \tag{16}$$

- The objective can be seen as the distance between $\hat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma} = \mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}$ measured in the log-det Bregman matrix divergence.
- The MLE aims to match the sample covariance $\hat{\mathbf{\Sigma}}$ to the true covariance $\mathbf{\Sigma}$.

We can also use other distance metrics. With Frobenius norm as metric, we get

$$\min_{\mathbf{\Gamma} \geq \mathbf{0}} \quad \|\mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I} - \hat{\mathbf{\Sigma}}\|_F^2 \tag{17}$$

- This method is also known as non-negative least square (NNLS).
- The objective is convex. Coordinate descent can also be used to solve NNLS.
- A scaling law on N, L, K, and M has been established under NNLS.

MLE versus NNLS for Device Activity Detection

We compare the detection performance of MLE and NNLS via simulations.

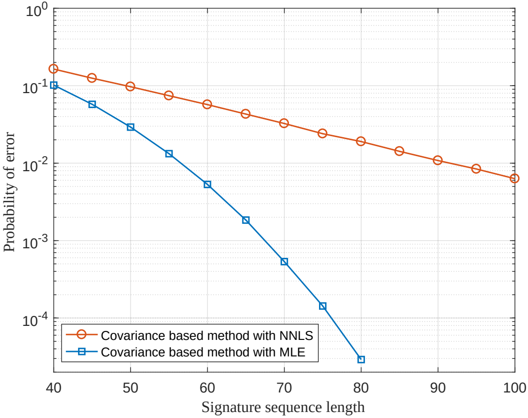


Figure: Performance comparison of MLE and NNLS. $N = 2000$, $K = 100$, and $M = 64$. MLE outperforms NNLS. The performance gap becomes more substantial as L increases.

Activity Detection with Known Large-Scale Fading

The covariance approach detects the device activity by estimating $\gamma_n \triangleq \alpha_n \beta_n$. There are scenarios in which the large-scale fading β_n is known at the BS, only the activities α_n need to be estimated.

- Maximizing the log-likelihood function of $\mathbf{Y}$ given $\alpha_1, \ldots, \alpha_N$ can be cast as

$$\begin{aligned}
\min_{\alpha_1, \ldots, \alpha_N} \quad f(\alpha_1, \ldots, \alpha_N) :&= -\frac{1}{M} \log p(\mathbf{Y} | \alpha_1, \ldots, \alpha_N) \\
&= -\frac{1}{M} \sum_{m=1}^{M} \log p(\mathbf{y}_m | \alpha_1, \ldots, \alpha_N) \\
&= -\frac{1}{M} \sum_{m=1}^{M} \log \left(\frac{1}{|\pi \mathbf{\Sigma}|} \exp\left(-\mathbf{y}_m^H \mathbf{\Sigma}^{-1} \mathbf{y}_m \right) \right) \\
&= \log |\mathbf{\Sigma}| + \operatorname{tr}\left(\mathbf{\Sigma}^{-1} \hat{\mathbf{\Sigma}} \right) + \mathit{const}. \qquad (18)
\end{aligned}$$

Note that $p(\mathbf{y}_m | \alpha_1, \ldots, \alpha_N)$ remains Gaussian with covariance $\mathbf{\Sigma}$.

Activity Detection with Known Large-Scale Fading

- The problem of detecting the binary activity indicator α_n is now:

$$\min_{\{\alpha_n\}} \quad \log\left|\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}\right| + \operatorname{tr}\left((\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I})^{-1}\hat{\boldsymbol{\Sigma}}\right) \tag{19a}$$

$$\text{s. t.} \quad \alpha_n \in \{0,1\},\ n = 1,2,\ldots,N \tag{19b}$$

- Binary α_n is challenging to deal with. We relax the constraint such that

$$\alpha_n \in [0,1], \qquad n = 1,2,\ldots,N \tag{20}$$

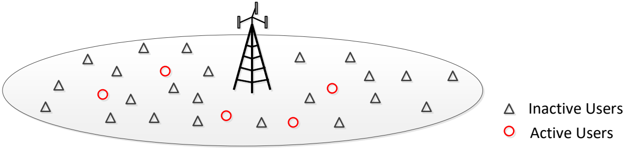
- The relaxed problem can be solved by coordinated descent with minor modifications:

$$d = \min\left\{\max\left\{\frac{\mathbf{s}_k^H\tilde{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{s}_k - \mathbf{s}_k^H\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{s}_k}{\beta_k(\mathbf{s}_k^H\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{s}_k)^2}, -\hat{\alpha}_k\right\}, 1-\hat{\alpha}_k\right\}. \tag{21}$$

- With unknown large-scale fading β_n, we estimate $\gamma_n = \alpha_n\beta_n$ in $[0,\infty]$. With known large-scale fading β_n, we estimate α_n in $[0,1]$.

Recap of Problem Formulations

- Sparse user activity detection with channel $\alpha_n \mathbf{h}_n \sim \alpha_n \sqrt{\beta}_n \mathcal{CN}(\mathbf{0}, \mathbf{I})$:



- If channel estimate is needed for subsequent data transmission:
  - We can use AMP, which gives a rough estimate of the instantaneous $\mathbf{h}_n$.
- If only user activities (α_n) are needed and large-scale fading is not known:
  - We can estimate large-scale fading ($\alpha_n \beta_n$) using the covariance method.
- If the users are not mobile and large-scale fading (β_n) is known:
  - We can modify the covariance method to estimate α_n.

Comparison of AMP vs Covariance Approaches

	Compressed Sensing (AMP)	Covariance Based Estimation
Derived from	Approx. marginals of $p(\mathbf{X}, \mathbf{Y})$	Maximization of $p(\mathbf{Y}\|\mathbf{\Gamma})$
Prior needed	Sparsity level for design of $\eta_t(\cdot)$	None (deterministic $\mathbf{\Gamma}$)
Estimate	Activities α_n and channels $\mathbf{h}_n$	Activities α_n and large-scale fading β_n
Preferred regime	Fix $\epsilon \triangleq \frac{K}{N}$, $\delta \triangleq \frac{L}{N}$, and M Let $N, L, K \to \infty$	Fix N, K, L Let $M \to \infty$
Complexity	Roughly $\mathcal{O}(NLM)$ per iteration	Roughly $\mathcal{O}(NL^2)$ via CD per iteration

Remark on Complexity

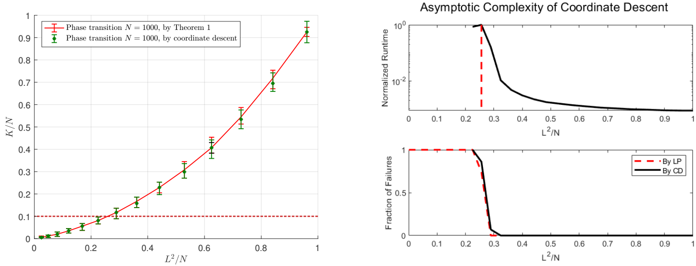
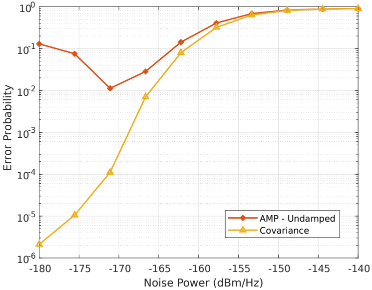


Figure: Numerical runtime for $\hat{\gamma}$ to converge to an ϵ norm ball around γ for fixed $\frac{K}{N} = 0.1$

- Each coordinate descent step requires $O(L^2)$ operations so updating all N coordinates resulting in a complexity of $O(L^2 N)$ per iteration.
- The average number of iterations required to converge to solution increases as the operating point approaches the phase transition boundary.
- The complexity of each iteration grows as $O(L^2)$, but the complexity of the overall algorithm decreases with L.

Instability of AMP at High SNR

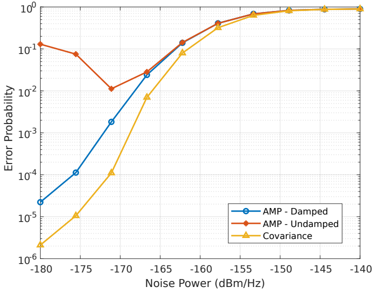
- Total users $N = 1000$, active users $K = 100$, BS antennas $M = 64$, $L = 110$.



- Users are uniformly distributed between 0.8 and 1km from the BS.
- User transmit power is 13dBm; Path-loss model $128.1 + 37.6 \log(d[\text{in km}])$
- Error probability is the probability chosen such that $N \cdot P_{FA} \approx K \cdot P_{MD}$

Damping for AMP

- Total users $N = 1000$, active users $K = 100$, BS antennas $M = 64$, $L = 110$.



- Consider a *damping term* in the AMP update, improving stability and convergence, making AMP more effective with large M.

$$\mathbf{X}^{t+1} = (1-\alpha)\eta_t\left(\mathbf{S}^H\mathbf{R}^t + \mathbf{X}^t\right) + \alpha\mathbf{X}^t,$$

where $\alpha \in [0, 1]$ is a damping factor ($\alpha = 0$ for standard AMP).

Numerical Comparison of AMP vs. Covariance Approach

- Total users $N = 1000$, Active users $K = 50$, Number of antennas $M = 8$

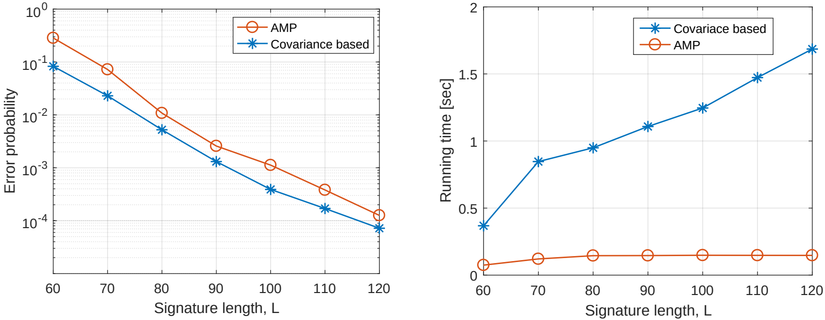


Figure: Performance and complexity of AMP vs covariance based estimation

- All users are located in the cell-edge (1000m) with transmit power 23dBm.
- Path-loss model $128.1 + 37.6\log(d[\text{in km}])$.
- Error probability is defined as the average of $\frac{\#(\text{Incorrectly detected users})}{K}$

Numerical Comparison of AMP vs. Covariance Approach

- Total users $N = 1000$, Active users $K = 50$, Signature length $L = 100$

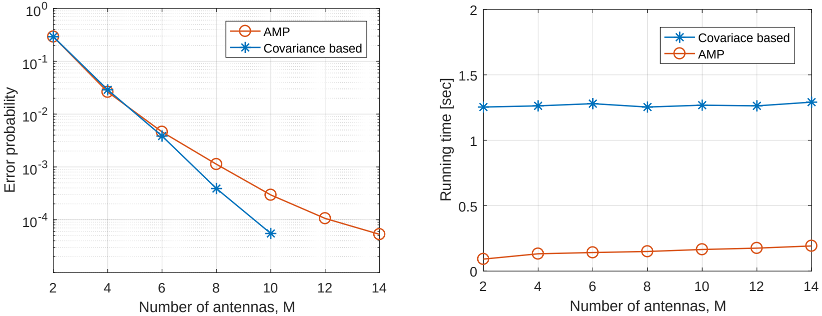
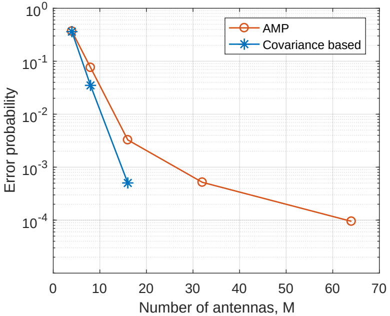


Figure: Performance and complexity of AMP vs covariance based estimation

- All users are located at the cell-edge (1000m) with transmit power 23dBm.
- Path-loss model $128.1 + 37.6\log(d[\text{in km}])$.
- Error probability is defined as the average of $\frac{\#(\text{Incorrectly detected users})}{K}$

Numerical Comparison of AMP vs. Covariance Approach

- Total users $N = 1000$, Active users $K = 90$, Signature length $L = 100$



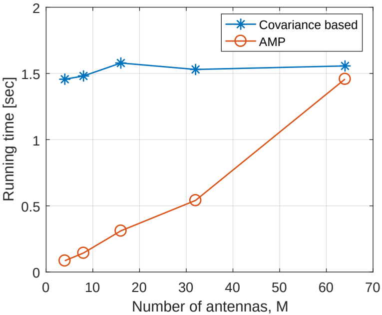


Figure: Performance and complexity of AMP vs covariance based estimation

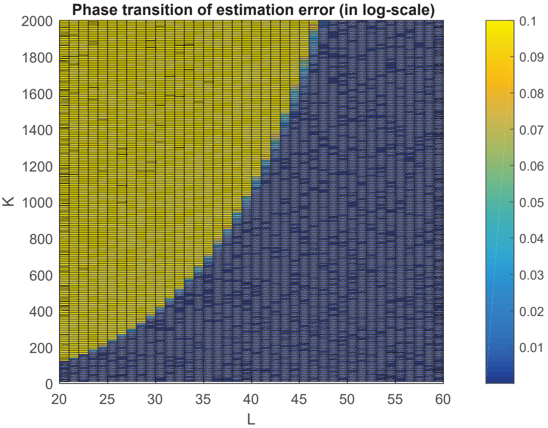
- All users are located at the cell-edge (1000m) with transmit power 23dBm.
- Path-loss model $128.1 + 37.6 \log(d[\text{in km}])$.
- Error probability is defined as the average of $\frac{\#(\text{Incorrectly detected users})}{K}$

AMP vs Covariance Approach

- Objectives:
  - Both algorithms perform sparse activity detection for massive random access.
  - AMP aims to recover the channels as well.
- Performance:
  - AMP and covariance approach have similar performance if $K \ll L$ and M small
  - Covariance approach is more effective in exploiting large M and when $K \gtrapprox L$.
- Complexity:
  - AMP is more computationally efficient when $K \ll L$ and M small.
- Crucial advantage of covariance method:
  - Being able to accommodate $K \gg L$ (!)

Scaling Law of the Covariance Approach

Suppose high SNR, perfect sampled covariance matrix $\hat{\boldsymbol{\Sigma}}$ ($M \to \infty$), we plot the estimation error of $\boldsymbol{\Gamma}$ under different (K, L) with $N = 2000$



Analysis

Analysis of AMP via State Evolution

The performance of AMP at each iteration can be predicted in the asymptotic regime where $L \to \infty, N \to \infty$ with fixed $\frac{L}{N}$

- $\mathbf{S}^H \mathbf{r}^t + \mathbf{x}^t$ can be modeled as signal plus noise, i.e., $\mathbf{x} + \mathbf{v}^t$
- $\mathbf{v}^t$ is i.i.d. Gaussian noise with variance τ_t tracked by state evolution equation

$$\tau_{t+1}^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}|\eta_t(X + \tau_t Z) - X|^2 \tag{22}$$

  for the $M = 1$ case.

- Interpretation of state evolution: Vector estimation $\mathbf{y} = \mathbf{S}\mathbf{x} + \mathbf{z}$ is reduced to uncoupled scalar estimation $(\mathbf{x}^t + \mathbf{S}^H \mathbf{r}^t)_i = x_i + v_i^t$

Analysis of Covariance Approach via Fisher Info Matrix

Recall the MLE formulation, and let γ denote the diagonal entries of $\mathbf{\Gamma}$

$$\begin{aligned}\min_{\gamma \geq \mathbf{0}} \quad f(\gamma) := -\frac{1}{M} \log p(\mathbf{Y}|\gamma) &= -\frac{1}{M} \sum_{m=1}^{M} \log p(\mathbf{y}_m|\gamma) \\ &= \log |\mathbf{S\Gamma S}^H + \sigma^2 \mathbf{I}| + \operatorname{tr}\left((\mathbf{S\Gamma S}^H + \sigma^2 \mathbf{I})^{-1} \hat{\mathbf{\Sigma}} \right) + const. \end{aligned} \tag{23}$$

- Analyzing the solution to (23) under coordinate descent is hard.
- Instead, let's analyze the true optimum of (23), i.e., MLE solution $\hat{\gamma}^{ML}$.
- Investigate asymptotic property of $\hat{\gamma}^{ML}$ in the massive MIMO regime.
- The Fisher information matrix, denoted by $\mathbf{J}(\gamma)$, plays a critical role in the asymptotic analysis. The (i,j)-th entry of $\mathbf{J}(\gamma)$ is defined as

$$[\mathbf{J}(\gamma)]_{ij} = \mathbb{E}\left[\frac{\partial \log p(\mathbf{Y}|\gamma)}{\partial \gamma_i} \frac{\partial \log p(\mathbf{Y}|\gamma)}{\partial \gamma_j} \right]. \tag{24}$$

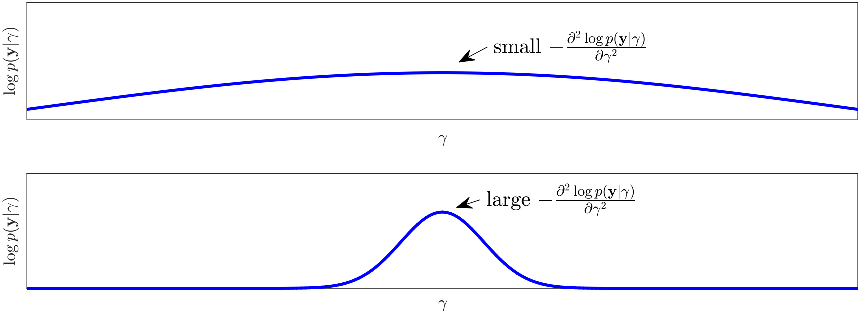
- Key assumption for the analysis: $M \to \infty$.

Fisher Information Matrix

- The Fisher Information matrix can be also written as the negative expected second derivative of the log-likelihood function

$$[\mathbf{J}(\gamma)]_{ij} = \mathbb{E}\left[\frac{\partial \log p(\mathbf{Y}|\gamma)}{\partial \gamma_i}\frac{\partial \log p(\mathbf{Y}|\gamma)}{\partial \gamma_j}\right] = -\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{Y}|\gamma)}{\partial \gamma_i \partial \gamma_j}\right] \quad (25)$$

- **Intuitive interpretation**: Fisher information matrix measures how informative the likelihood function is, and how effective the MLE can be



Cramer-Rao Bound and Asymptotic Property of MLE

Fisher information matrix plays a critical role in classic estimation theory.

- **Cramer-Rao bound**: Let γ be a parameter, and let $\hat{\gamma}$ be an unbiased estimator of γ. Then the covariance of estimation error is lower bounded by

$$\mathbb{E}\left[(\hat{\gamma}-\gamma)(\hat{\gamma}-\gamma)^T\right] \geq \mathbf{J}^{-1}(\gamma) \tag{26}$$

- **Asymptotic properties of the MLE**: Let $\hat{\gamma}^{ML}$ be the maximum likelihood estimator of γ. Then, under certain regularity conditions, as $M \to \infty$

$$\text{Consistency:} \quad \hat{\gamma}^{ML} \xrightarrow{P} \gamma \tag{27}$$

$$\text{Asymptotic normality:} \quad \sqrt{M}(\hat{\gamma}^{ML}-\gamma) \xrightarrow{D} \mathcal{N}(\mathbf{0}, M\mathbf{J}^{-1}(\gamma)) \tag{28}$$

It means that the maximum likelihood estimator $\hat{\gamma}^{ML}$ is asymptotically unbiased and asymptotically attains the Cramer-Rao bound, i.e., asymptotically efficient.

Regularity Conditions

- The regularity conditions for consistency and asymptotic normality include
  - The true parameter γ^0 is identifiable, i.e.,, there exists no other $\gamma' \neq \gamma^0$ with
$$p(\mathbf{Y}|\gamma^0) = p(\mathbf{Y}|\gamma'). \tag{29}$$
  - The true parameter should be in the interior of the feasible region, as otherwise $\hat{\gamma}^{ML} - \gamma^0$ cannot be Gaussian distributed.
- These conditions are usually reasonable and mild.
- But, these conditions are NOT always satisfied for sparse activity detection.
  - The identifiability may not be guaranteed when
$$N \gg L^2, \tag{30}$$
i.e., when the dimension of γ^0 is larger than the dimensions of the sample covariance $\hat{\mathbf{\Sigma}}$, there are too many parameters to estimate.
  - The true parameter γ^0 in fact always lies on the boundary of its parameter space $[0, \infty)^N$, because most of the entries of γ^0 are zero.

Need new analysis!

Fisher Information Matrix in Sparse Activity Detection

We first derive the Fisher information matrix for the activity detection problem:

$$[\mathbf{J}(\gamma)]_{ij} = -\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{Y}|\gamma)}{\partial\gamma_i\partial\gamma_j}\right]. \tag{31}$$

- $p(\mathbf{y}_m|\gamma)$ is Gaussian, the second derivative of $\log p(\mathbf{Y}|\gamma) = \sum_m \log p(\mathbf{y}_m|\gamma)$ is

$$\begin{aligned}\frac{\partial^2 \log p(\mathbf{Y}|\gamma)}{\partial\gamma_i\partial\gamma_j} &= M\,\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{s}_j\mathbf{s}_j^H\mathbf{\Sigma}^{-1}\mathbf{s}_i\mathbf{s}_i^H) - M\,\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{s}_j\mathbf{s}_j^H\mathbf{\Sigma}^{-1}\mathbf{s}_i\mathbf{s}_i^H\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}) \\ &\quad - M\,\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{s}_i\mathbf{s}_i^H\mathbf{\Sigma}^{-1}\mathbf{s}_j\mathbf{s}_j^H\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}).\end{aligned} \tag{32}$$

- Taking the expectation of $\hat{\mathbf{\Sigma}}$ using the fact that $\mathbb{E}[\hat{\mathbf{\Sigma}}] = \mathbf{\Sigma}$ gives

$$-\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{Y}|\gamma)}{\partial\gamma_i\partial\gamma_j}\right] = M(\mathbf{s}_i^H\mathbf{\Sigma}^{-1}\mathbf{s}_j)(\mathbf{s}_j^H\mathbf{\Sigma}^{-1}\mathbf{s}_i). \tag{33}$$

Fisher Information Matrix in Sparse Activity Detection

- The Fisher information matrix $\mathbf{J}(\gamma)$ can be further written in a matrix form as

$$\mathbf{J}(\gamma) = M\left(\mathbf{P} \odot \mathbf{P}^*\right), \tag{34}$$

where $\mathbf{P} \triangleq \mathbf{S}^H \left(\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}\right)^{-1}\mathbf{S}$; $\odot$ element-wise product; $(\cdot)^*$ conjugate

- $\mathbf{J}(\gamma)$ is a real symmetric matrix of dimensions $N \times N$, whose rank satisfies:

$$\text{Rank}(\mathbf{P} \odot \mathbf{P}^*) \overset{(a)}{\leq} \text{Rank}(\mathbf{P})^2 \overset{(b)}{\leq} L^2, \tag{35}$$

where

  - (a) is due to $\text{Rank}(\mathbf{U} \odot \mathbf{V}) \leq \text{Rank}(\mathbf{U})\,\text{Rank}(\mathbf{V})$;
  - (b) is due to $\text{Rank}(\mathbf{P}) \leq \text{Rank}(\mathbf{S}) \leq \min\{N, L\}$.

- Thus $\mathbf{J}(\gamma)$ is rank-deficient if $N > L^2$ since $\mathbf{P} \odot \mathbf{P}^*$ is of size $N \times N$.

Our new analysis takes rank-deficiency of $\mathbf{J}(\gamma)$ into consideration

Performance Analysis of Activity Detection

Since the regularity conditions may not hold in the sparse activity detection problem, we need to ask:

- What are the conditions on the system parameters such that $\hat{\gamma}^{ML}$ can approach the true parameter γ^0 as $M \to \infty$?
- This helps identify the desired operating regime of the system parameters for getting an accurate estimate $\hat{\gamma}^{ML}$ via MLE with massive MIMO

- If M is finite, how is the estimation error $\hat{\gamma} - \gamma^0$ distributed?
- This helps characterize the error probabilities in device activity detection.

We answer these questions by examining the "null space" of the Fisher information matrix.

Necessary and Sufficient Condition for $\hat{\gamma}^{ML} \to \gamma^0$

Theorem

Let $\mathcal{I}$ be an index set corresponding to zero entries of γ^0, i.e., $\mathcal{I} \triangleq \{i \mid \gamma_i^0 = 0\}$. We define two sets $\mathcal{N}, \mathcal{C}$ in the space $\mathbb{R}^N$, respectively, as follows

$$\mathcal{N} \triangleq \{\mathbf{x} \mid \mathbf{x}^T \mathbf{J}(\gamma^0)\mathbf{x} = 0, \mathbf{x} \in \mathbb{R}^N\}, \tag{36}$$

$$\mathcal{C} \triangleq \{\mathbf{x} \mid x_i \geq 0, i \in \mathcal{I}, \mathbf{x} \in \mathbb{R}^N\}, \tag{37}$$

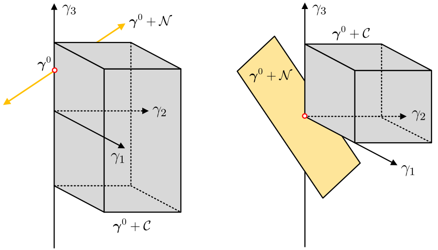
where x_i is the i-th entry of $\mathbf{x}$. Then a necessary and sufficient condition for the consistency of $\hat{\gamma}^{ML}$, i.e., $\hat{\gamma}^{ML} \to \gamma^0$ as $M \to \infty$, is $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$.

$\mathcal{N}$ is the "null space" of $\mathbf{J}(\gamma^0)$; $\mathcal{C}$ is a cone with non-negative entries indexed by $\mathcal{I}$.

This condition leads to a phase analysis for the covariance based method, i.e., set of (N, L, K) outside of which $\hat{\gamma}^{ML}$ cannot approach γ^0 even in the large M limit.

Interpretation of the Condition

- $\mathcal{N}$ corresponds to all directions in $\mathbb{R}^N$ along which likelihood stays constant. In these directions, the true parameter cannot be identified via the likelihood.
- $\mathcal{C}$ is the directions along which parameters remain within constraint $\mathbb{R}_+^N$
- $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ ensures that the true parameter $\boldsymbol{\gamma}^0$ is uniquely identifiable via the likelihood in its feasible neighborhood, also termed as local identifiability



- Local identifiability is clearly necessary.
- Sufficiency due to equivalence of local and global identifiability in this case.
- A necessary condition for $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ is that $dim(\mathcal{N}) < |\mathcal{I}|$.
- Since $dim(\mathcal{N})$ is roughly $N - L^2$ and $|\mathcal{I}| = N - K$, we have $K < L^2$.

Numerically Verify the Condition via $\mathcal{M}^+$ Criterion

Proposition

Let $\mathcal{I} \triangleq \{i \mid \gamma_i^0 = 0\}$ and $\mathcal{I}^c \triangleq \{i \mid \gamma_i^0 > 0\}$ be two index sets with $|\mathcal{I}| = N - K$ and $|\mathcal{I}^c| = K$.We define three submatrices of $\mathbf{J}(\boldsymbol{\gamma}^0) \in \mathbb{R}^{N\times N}$ as follows.

- $\mathbf{A} \in \mathbb{R}^{(N-K)\times(N-K)}$, *row indices and column indices from* $\mathcal{I}$
- $\mathbf{B} \in \mathbb{R}^{(N-K)\times K}$, *row indices from* $\mathcal{I}$ *and column indices from* $\mathcal{I}^c$
- $\mathbf{C} \in \mathbb{R}^{K\times K}$, *row indices and column indices from* $\mathcal{I}^c$

If $\mathbf{C}$ *is invertible, then the condition* $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ *is equivalent to the feasibility of*

$$\text{find} \quad \mathbf{x} \tag{38a}$$

$$\text{subject to} \quad (\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)\mathbf{x} > \mathbf{0}, \tag{38b}$$

where vector $\mathbf{x} \in \mathbb{R}^{N-K}$.

Note that matrix $\mathbf{M}$ satisfying $\mathbf{M}^T\mathbf{x} > \mathbf{0}$ for some $\mathbf{x}$, i.e., row span intersecting the positive orthant, is referred to as $\mathcal{M}^+$ [Bruckstein-Elad-Zibulevsky'08].

Proof based on analyzing the null space of $\mathbf{J}(\boldsymbol{\gamma}^0)$ and that $\forall\mathbf{M}$: (i) $\mathbf{M}\mathbf{x} = \mathbf{0}$ has no solution for $\mathbf{x} \geq \mathbf{0}$ and $\mathbf{x} \neq \mathbf{0}$, is equivalent to (ii) $\mathbf{M}^T\mathbf{x} > \mathbf{0}$ has solutions.

Covariance Matching Perspective

Analyzing the optimization problem:

$$\min_{\gamma \succeq \mathbf{0}} \quad f(\gamma) = \log |\mathbf{S\Gamma S}^H + \sigma^2 \mathbf{I}| + \operatorname{tr}\left((\mathbf{S\Gamma S}^H + \sigma^2 \mathbf{I})^{-1} \hat{\mathbf{\Sigma}}\right) \tag{39}$$

- By taking the derivative, we see that ideally we need: $\mathbf{S\Gamma S}^H + \sigma^2 \mathbf{I} = \hat{\mathbf{\Sigma}}$
- For finite M, usually $\mathbf{S\Gamma S}^H + \sigma^2 \mathbf{I} \neq \hat{\mathbf{\Sigma}}$ since $\hat{\mathbf{\Sigma}}$ is the *sample* covariance
- For $M \to \infty$, $\mathbf{S\Gamma S}^H + \sigma^2 \mathbf{I} = \hat{\mathbf{\Sigma}}$ holds at true γ^0, i.e., γ^0 minimizes $f(\gamma)$.

Intuition

$\hat{\gamma}^{ML} \to \gamma^0$ *as* $M \to \infty$ $\iff$ γ^0 *uniquely minimizes* $f(\gamma)$ *in the limit* $M \to \infty$
$\iff$ γ^0 *is the unique solution to* $\mathbf{S\Gamma S}^H + \sigma^2 \mathbf{I} = \hat{\mathbf{\Sigma}}$ *in the limit* $M \to \infty$.

A necessary and sufficient condition for the consistency of $\hat{\gamma}^{ML}$ can be derived by studying the uniqueness of $\mathbf{S\Gamma S}^H + \sigma^2 \mathbf{I} = \hat{\mathbf{\Sigma}}$ in the limit $M \to \infty$.

Equivalent Necessary and Sufficient Condition

Proposition

Let $\widehat{\mathbf{S}} \in \mathbb{C}^{L^2 \times N}$ *be the column-wise Kronecker product (Khatri-Rao product) of* $\mathbf{S}^*$ *and* $\mathbf{S}$*, i.e.,* $\widehat{\mathbf{S}} \triangleq [\mathbf{s}_1^* \otimes \mathbf{s}_1, \ldots, \mathbf{s}_N^* \otimes \mathbf{s}_N]$. *We define a set* $\widetilde{\mathcal{N}}$ *in the space* $\mathbb{R}^N$ *as*

$$\widetilde{\mathcal{N}} \triangleq \{\mathbf{x} \mid \widehat{\mathbf{S}}\mathbf{x} = \mathbf{0}, \mathbf{x} \in \mathbb{R}^N\}. \tag{40}$$

Then a necessary and sufficient condition for γ^0 *being the unique solution to* $\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I} = \hat{\boldsymbol{\Sigma}}$ *in the limit* $M \to \infty$*, is* $\widetilde{\mathcal{N}} \cap \mathcal{C} = \{\mathbf{0}\}$*, where* $\mathcal{C}$ *is as in* (37).

The proof is obtained by vectorizing $\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I} = \hat{\boldsymbol{\Sigma}}$ in the limit $M \to \infty$, and studying the resulting linear equation.

Proposition

We have that $\widetilde{\mathcal{N}}$ *defined in* (40) *and* $\mathcal{N}$ *defined in* (36) *are identical. Thus, the condition* $\widetilde{\mathcal{N}} \cap \mathcal{C} = \{\mathbf{0}\}$ *is equivalent to* $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$

Key advantage of using $\widetilde{\mathcal{N}}$ is that it depends only on S and is independent of SNR.

Alternative Way to Numerically Verify the Condition

Theorem

Let $\mathbf{r}_i^T = [s_{i1}, s_{i2}, \ldots, s_{iN}]$ *be the i-th row of* $\mathbf{S}$*. Based on* $\mathbf{r}_i^T$*, we construct two sets of row vectors to represent the real and imaginary parts of rows of* $\widehat{\mathbf{S}}$*:*

$$\left\{\mathrm{Re}(\mathbf{r}_i^T) \odot \mathrm{Re}(\mathbf{r}_j^T) + \mathrm{Im}(\mathbf{r}_i^T) \odot \mathrm{Im}(\mathbf{r}_j^T),\ 1 \le i \le j \le L\right\}, \tag{41}$$

$$\left\{\mathrm{Re}(\mathbf{r}_i^T) \odot \mathrm{Im}(\mathbf{r}_j^T) - \mathrm{Im}(\mathbf{r}_i^T) \odot \mathrm{Re}(\mathbf{r}_j^T),\ 1 \le i < j \le L\right\}. \tag{42}$$

Let $\mathbf{D} \in \mathbb{R}^{L^2 \times N}$ *be the matrix formed by all* L^2 *rows from these two sets. The condition* $\widetilde{\mathcal{N}} \cap \mathcal{C} = \{\mathbf{0}\}$ *is equivalent to the infeasibility of the following problem*

$$\begin{aligned} \text{find} \quad & \mathbf{x} && \text{(43a)} \\ \text{subject to} \quad & \mathbf{D}\mathbf{x} = \mathbf{0}, && \text{(43b)} \\ & \|\mathbf{x}\|_1 = 1, && \text{(43c)} \\ & x_i \ge 0, i \in \mathcal{I}, && \text{(43d)} \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^N$ *and the constraint* (43c) *guarantees* $\mathbf{x} \ne \mathbf{0}$*.*

Scaling Law for NNLS Formulation

Theorem (Fengler-Haghighatshoar-Jung-Caire '19)

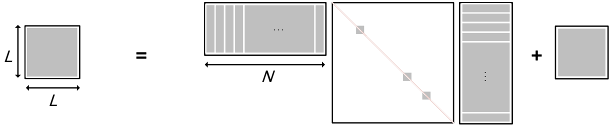
Let $\mathbf{S} \in \mathbb{C}^{L \times N}$ be the signature sequence matrix whose columns are uniformly drawn from the sphere of radius $\sqrt{L}$ in an i.i.d. fashion. There exist some constants c_1, c_2, c_3, and c_4 whose values do not depend on K, L, and N such that if $K \leq c_1 L^2 / \log^2(eN/L^2)$, then with probability at least $1 - \exp(-c_2 L)$, the solution of the NNLS problem, $\hat{\gamma}^{NNLS}$, satisfies

$$\|\gamma^0 - \hat{\gamma}^{NNLS}\|_2 \leq c_3 \left(\sqrt{\frac{L}{K}} + c_4 \right) \frac{\|\mathbf{S}\mathbf{\Gamma}^0\mathbf{S}^H + \sigma^2\mathbf{I} - \hat{\mathbf{\Sigma}}\|_F}{L}. \tag{44}$$

- The derivation is based on restricted isometry property in compressed sensing.
- It implies that the error vanishes as $M \to \infty$, because $\hat{\mathbf{\Sigma}} \to \mathbf{S}\mathbf{\Gamma}^0\mathbf{S}^H + \sigma^2\mathbf{I}$.
- The result is for specific sequence $\mathbf{S}$.
- Since $K < L^2$, we get a simpler form of scaling law: $L^2 \approx K \log^2(N/K)$.

Scaling Law for MLE Formulation

- Scaling laws in compressed sensing:
  - For $\mathbf{Ax} = \mathbf{b}$ with $\mathbf{A}$ satisfying restricted isometry property, the number of measurements needed to recover a K-sparse vector $\mathbf{x}$ of length-N is
  $$L = O(K \log(N/K)).$$
  - For $\hat{\mathbf{\Sigma}} = \mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}$ with $\widehat{\mathbf{S}}$ satisfying robust null space property, the number of measurements needed to receover a K-sparse diagonal matrix $\mathbf{\Gamma}$ of size N^2 is
  $$L^2 = O(K \log^2(N/K)).$$
- Based on the same robust NSP of $\widehat{\mathbf{S}}$, we can derive the scaling law of MLE:



Theorem

Under the same scaling law for K, L, N and for the same randomly chosen $\mathbf{S}$, $\widetilde{\mathcal{N}} \cap \mathcal{C} = \{\mathbf{0}\}$ holds with probability at least $1 - \exp(-c_2 L)$.

Numerical Results – Scaling Law of Covariance Approach

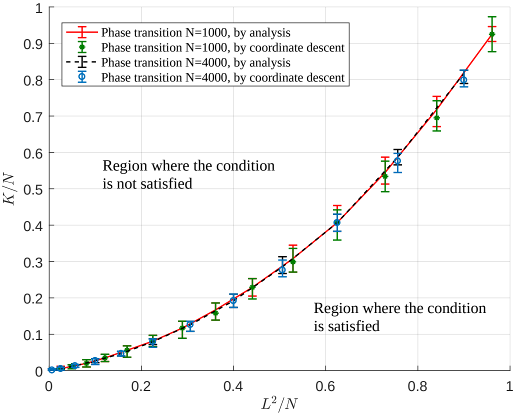
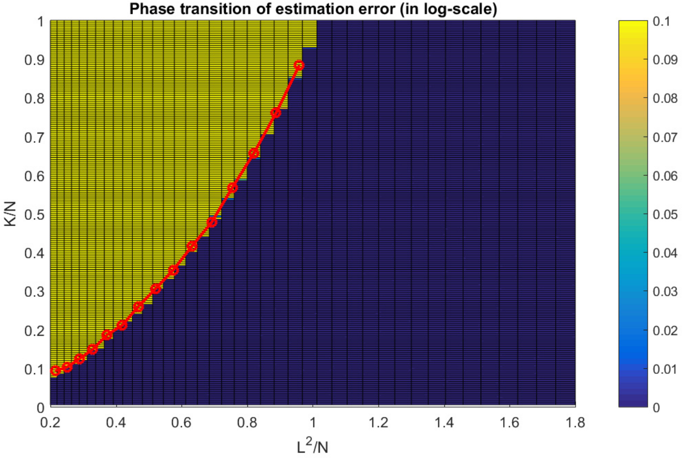


Figure: Phase transition in the space of N, L, K. All users are located at the cell-edge (1000m) with transmit power 23dBm. Path-loss is $128.1 + 37.6\log(d[\text{km}])$. Generated by 100 Monte Carlo simulations. Error bars indicate the range below which all 100 realizations satisfy the condition and above which none satisfies the condition.

Phase Transition of the Covariance Approach

Suppose high SNR, perfect sampled covariance matrix $\hat{\mathbf{\Sigma}}$ ($M \to \infty$), we plot the estimation error of $\mathbf{\Gamma}$ under different $(K/N,\ L^2/N)$ with $N = 2000$



Performance of coordinate descent algorithm is very close to the optimal MLE!

Value of Knowing Large-Scale Fading

The covariance method directly estimates the activity indicator α_n in $[0, 1]$ instead of $\gamma_n = \alpha_n \beta_n$ in $[0, \infty)$. Let $\boldsymbol{\alpha} \triangleq [\alpha_1, \ldots, \alpha_N]^T$ and let the true value of $\boldsymbol{\alpha}$ be $\boldsymbol{\alpha}^0$.

Theorem

Let $\mathcal{I}$ be an index set corresponding to zero entries of $\boldsymbol{\alpha}^0$, i.e., $\mathcal{I} \triangleq \{i \mid \alpha_i^0 = 0\}$. We define two sets $\mathcal{N}, \mathcal{C}$ in the space $\mathbb{R}^N$, respectively, as follows

$$\mathcal{N} \triangleq \{\mathbf{x} \mid \mathbf{x}^T \mathbf{J}(\gamma^0)\mathbf{x} = 0, \mathbf{x} \in \mathbb{R}^N\}, \tag{45}$$

$$\mathcal{C} \triangleq \{\mathbf{x} \mid x_i \geq 0, i \in \mathcal{I}, x_i \leq 0, i \notin \mathcal{I}, \mathbf{x} \in \mathbb{R}^N\}, \tag{46}$$

where x_i is the i-th entry of $\mathbf{x}$. Then a necessary and sufficient condition for the consistency of $\hat{\boldsymbol{\alpha}}^{ML}$, i.e., $\hat{\boldsymbol{\alpha}}^{ML} \to \boldsymbol{\alpha}^0$ as $M \to \infty$, is $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$.

The extra constraint in defining $\mathcal{C}$ is due to the fact that α_n is upper bounded.

Value of Knowing Large-Scale Fading

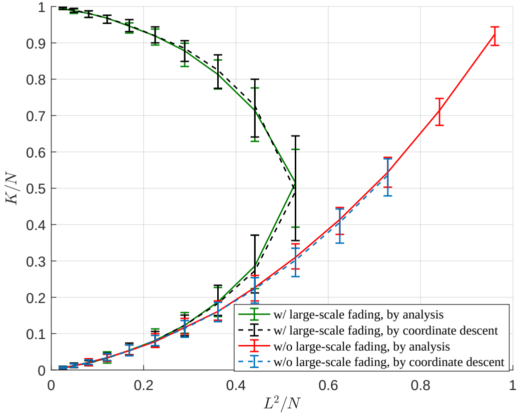


Figure: Phase transition comparison of the cases with and without knowing large-scale fading. $N = 1000$. With known large-scale fading, α_n is both lower and upper bounded.

When $\frac{K}{N} \approx 1$, then inactive users are sparse!

Asymptotic Distribution of the ML Estimation Error

For MLE solutions not on boundary, we have $\sqrt{M}(\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma}) \xrightarrow{D} \mathcal{N}(\mathbf{0}, M\mathbf{J}^{-1}(\boldsymbol{\gamma}))$.
For MLE with boundary constraint: $\mathcal{C} \triangleq \{\mathbf{x} \mid x_i \geq 0, i \in \mathcal{I}, \mathbf{x} \in \mathbb{R}^N\}$:

Theorem

Let $\mathbf{x} \in \mathbb{R}^{N\times 1} \sim \mathcal{N}\left(\mathbf{0}, M\mathbf{J}^{\dagger}(\boldsymbol{\gamma}^0)\right)$, *where* $\dagger$ *denotes Moore-Penrose inverse. Let* $\boldsymbol{\mu} \in \mathbb{R}^{N\times 1}$ *be a solution to the constrained quadratic programming problem*

$$\underset{\boldsymbol{\mu}}{\text{minimize}} \quad \frac{1}{M}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{J}(\boldsymbol{\gamma}^0)(\mathbf{x} - \boldsymbol{\mu}) \tag{47a}$$

$$\text{subject to} \quad \boldsymbol{\mu} \in \mathcal{C}, \tag{47b}$$

where $\mathcal{C}$ *is defined in* (37). *For the case without knowing large-scale fading, assume that* $\hat{\boldsymbol{\gamma}}^{ML} \to \boldsymbol{\gamma}^0$, *then there exists a sequence of* $\boldsymbol{\mu}$ *such that* $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma})$ *has asymptotically the same distribution as* $\boldsymbol{\mu}$ *as* $M \to \infty$.

Note that $\boldsymbol{\mu}$ is random due to the randomness of $\mathbf{x}$.

Detection error can be characterized based on the distribution of $\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma}^0$.

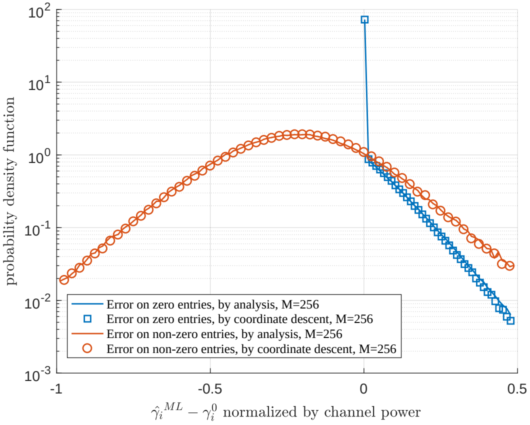


Figure: Probability density functions (PDFs) of the error $\hat{\gamma}_i^{ML} - \gamma_i^0$ (normalized). The parameters are $N = 1000$, $K = 50$, and $L = 20$ ($L^2/N = 0.4$, $K/N = 0.05$). Note that there is a point mass in the distribution of the error for the zero entries. This is the probability that the inactive devices are correctly identified at finite $M = 256$.

Detection Error Probabilities

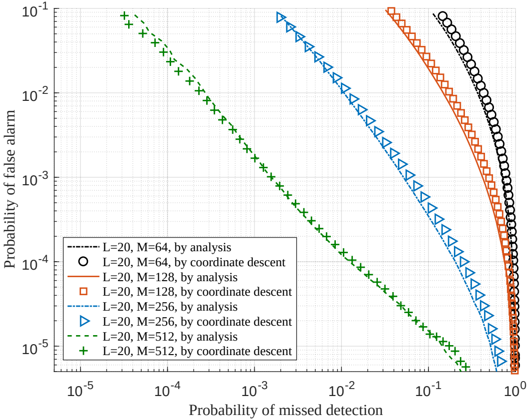
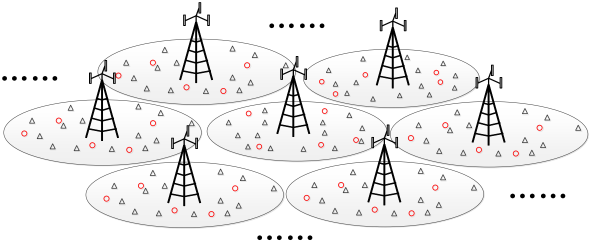


Figure: Probability of missed detection vs. probability of false alarm. The parameters are $N = 1000$, $K = 50$, and $L = 20$ ($L^2/N = 0.4$, $K/N = 0.05$). All users are located at the cell-edge (1000m) with transmit power 23dBm. Path-loss is $128.1 + 37.6 \log(d[\text{km}])$.

User Activity Detection in Multicell Systems

- What is the impact of the inter-cell interference?



- How to overcome the inter-cell interference?

Activity Detection in Multicell Systems

- Multi-cell system with B BSs each equipped with M antennas;
- N single-antenna devices per cell, K of which are active;
- Device n in cell b is assigned a length-L unique signature sequence $\mathbf{s}_{bn}$;
- Received signal $\mathbf{Y}_b \in \mathbb{C}^{L\times M}$ at BS b is

$$\mathbf{Y}_b = \sum_{n=1}^{N} \alpha_{bn}\mathbf{s}_{bn}\mathbf{h}_{bbn}^T + \sum_{j=1,j\neq b}^{j=B} \sum_{n=1}^{N} \alpha_{jn}\mathbf{s}_{jn}\mathbf{h}_{bjn}^T + \mathbf{Z}_b$$

$$= \mathbf{S}_b\mathbf{X}_{bb} + \sum_{j=1,j\neq b}^{j=B} \mathbf{S}_j\mathbf{X}_{bj} + \mathbf{Z}_b, \tag{48}$$

where

- $\alpha_{bn} \in \{1, 0\}$ activity indicator; $\mathbf{Z}_b \in \mathbb{C}^{L\times M}$ Gaussian noise with variance σ^2.
- $\mathbf{h}_{bjn} \in \mathbb{C}^{M\times 1}$ is the channel from user n in cell j to BS b
- $\mathbf{S}_j \triangleq [\mathbf{s}_{j1}, \cdots, \mathbf{s}_{jN}] \in \mathbb{C}^{L\times N}$; $\mathbf{X}_{bj} \triangleq [\alpha_{j1}\mathbf{h}_{bj1}, \cdots, \alpha_{jN}\mathbf{h}_{bjN}]^T \in \mathbb{C}^{N\times M}$

The inter-cell interference brings performance degradation for activity detection.

Covariance Based Activity Detection for Multi-cell

To use the covariance approach, the signal at BS b is re-written as

$$\begin{aligned}
\mathbf{Y}_b &= \sum_{n=1}^{N} \alpha_{bn} \mathbf{s}_{bn} \mathbf{h}_{bbn}^T + \sum_{j=1, j\neq b}^{j=B} \sum_{n=1}^{N} \alpha_{jn} \mathbf{s}_{jn} \mathbf{h}_{bjn}^T + \mathbf{Z}_b \\
&= \mathbf{S}_b \mathbf{A}_b \mathbf{G}_{bb}^{\frac{1}{2}} \tilde{\mathbf{H}}_{bb} + \sum_{j=1, j\neq b}^{j=B} \mathbf{S}_j \mathbf{A}_j \mathbf{G}_{bj}^{\frac{1}{2}} \tilde{\mathbf{H}}_{bj} + \mathbf{Z}_b \\
&= \mathbf{S}_b \boldsymbol{\Gamma}_{bb}^{\frac{1}{2}} \tilde{\mathbf{H}}_{bb} + \sum_{j=1, j\neq b}^{j=B} \mathbf{S}_j \boldsymbol{\Gamma}_{bj}^{\frac{1}{2}} \tilde{\mathbf{H}}_{bj} + \mathbf{Z}_b
\end{aligned} \tag{49}$$

- $\mathbf{S}_j \triangleq [\mathbf{s}_{j1}, \mathbf{s}_{j2}, \cdots, \mathbf{s}_{jN}] \in \mathbb{C}^{L\times N}$; $\mathbf{A}_j \triangleq \text{diag}\{\alpha_{j1}, \alpha_{j2}, \cdots, \alpha_{jN}\} \in \{0,1\}^{N\times N}$
- $\mathbf{G}_{bj} \triangleq \text{diag}\{\beta_{bj1}, \beta_{bj2}, \cdots, \beta_{bjN}\} \in \mathbb{R}^{N\times N}$ large-scale fading matrix
- $\boldsymbol{\Gamma}_{bj} \triangleq \text{diag}\{\alpha_{j1}\beta_{bj1}, \alpha_{j2}\beta_{bj2}, \cdots, \alpha_{jN}\beta_{bjN}\} \in \mathbb{R}^{N\times N}$
- $\tilde{\mathbf{H}}_{bj} \triangleq \left[\mathbf{h}_{bj1}/\sqrt{\beta_{bj1}}, \cdots, \mathbf{h}_{bjN}/\sqrt{\beta_{bjN}}\right]^T \in \mathbb{C}^{N\times M}$, normalized channel

Similar to single-cell case, all $\boldsymbol{\Gamma}_{bj}$ are treated as deterministic unknown parameters and all $\tilde{\mathbf{H}}_{bj}$ are treated as random samples.

Cooperative Activity Detection via Covariance Approach

- Assume that each BS is equipped with a large-scale antenna array.
- **Cooperative detection**: To alleviate the impact of inter-cell interference, we further consider BS cooperation by assuming all BSs are connected to a CU.
- Depending on whether the large-scale fading matrices $\mathbf{G}_{bj}, \forall b, j$ are known, the device activity detection problem can be formulated differently.
- When $\mathbf{G}_{bj}$ are not known, we need to estimate $\mathbf{\Gamma}_{bj} = \mathbf{A}_j \mathbf{G}_{bj}$, $\forall b, j$, which has

$$\boxed{B^2 N \text{ unknown parameters}}$$

- When $\mathbf{G}_{bj}$ are known, we only need to estimate $\mathbf{A}_b, \forall b$, which contains

$$\boxed{BN \text{ unknown parameters}}$$

Device activity detection is much easier if large-scale fading is known!

Cooperative Detection with Unknown Large-scale Fading

We aim to estimate $\mathbf{\Gamma}_{bj} = \mathbf{A}_j \mathbf{G}_{bj}, \forall b, j$ from the received signals $\mathbf{Y}_b, \forall b$.
The likelihood function of $\mathbf{Y}_b$'s given $\mathbf{\Gamma}_{bj}$'s can be expressed as

$$\begin{aligned} p(\mathbf{Y}_1, \ldots, \mathbf{Y}_B | \mathbf{\Gamma}_{11}, \mathbf{\Gamma}_{12}, \ldots, \mathbf{\Gamma}_{BB}) &= \prod_{b=1}^{B} p(\mathbf{Y}_b | \mathbf{\Gamma}_{11}, \mathbf{\Gamma}_{12}, \ldots, \mathbf{\Gamma}_{BB}) \\ &= \prod_{b=1}^{B} \frac{1}{|\pi \mathbf{\Sigma}_b|^M} \exp\left(-\operatorname{tr}\left(M \mathbf{\Sigma}_b^{-1} \hat{\mathbf{\Sigma}}_b\right)\right). \end{aligned} \tag{50}$$

The MLE problem can be cast as minimization of negative log-likelihood:

$$\min_{\{\mathbf{\Gamma}_{bj}\}} \quad \sum_{b=1}^{B} \left(\log|\mathbf{\Sigma}_b| + \operatorname{tr}\left(\mathbf{\Sigma}_b^{-1} \hat{\mathbf{\Sigma}}_b\right)\right) \tag{51a}$$

$$\text{s.t.} \quad \gamma_{bjn} \in [0, \infty), \forall b, j, n \tag{51b}$$

- The problem can be solved using coordinate descent.

Cooperative Detection with Known Large-scale Fading

Assuming that all large-scale fading matrices $\mathbf{G}_{bj}$'s, we directly estimate the device activity $\mathbf{A}_b$'s using the MLE. The likelihood function of $\mathbf{Y}_b$'s can be expressed as

$$\begin{aligned} p(\mathbf{Y}_1, \ldots, \mathbf{Y}_B | \mathbf{A}_1, \ldots, \mathbf{A}_B) &= \prod_{b=1}^{B} p(\mathbf{Y}_b | \mathbf{A}_1, \ldots, \mathbf{A}_B) \\ &= \prod_{b=1}^{B} \frac{1}{|\pi \mathbf{\Sigma}_b|^M} \exp\left(-\operatorname{tr}\left(M \mathbf{\Sigma}_b^{-1} \hat{\mathbf{\Sigma}}_b\right)\right). \end{aligned} \tag{52}$$

Since the activity α_{bn} is binary, the maximization of likelihood can be cast as

$$\min_{\{\mathbf{A}_b\}} \quad \sum_{b=1}^{B} \left(\log|\mathbf{\Sigma}_b| + \operatorname{tr}\left(\mathbf{\Sigma}_b^{-1} \hat{\mathbf{\Sigma}}_b\right)\right) \tag{53a}$$

$$\text{s. t.} \quad \alpha_{bn} \in \{0, 1\}, \forall b, n \tag{53b}$$

Known large-scale fading: Single-cell and multicell have same phase transition

- Multicell problem: Find a BN-dim sparse vector in $(BN - BL^2)$-dim subspace.
- Single-cell problem: Find a N-dim sparse vector in $(N - L^2)$-dim subspace.

Performance of Covariance Based Detection for Multi-cell

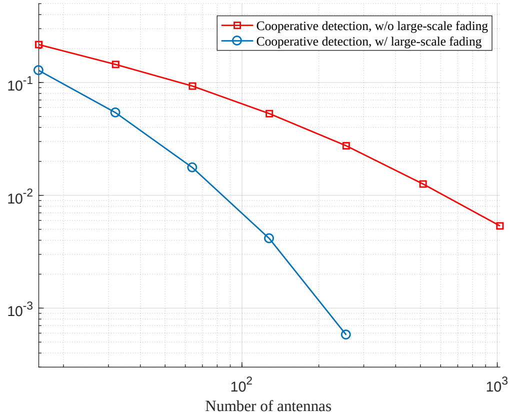


Figure: Performance comparison of the multicell covariance approach with and without knowing large-scale fading. $B = 7$, $N = 200$, $K = 20$, and $L = 20$. We observe that knowing the large-scale fading brings substantial improvement.

Performance of Covariance Based Detection for Multi-cell

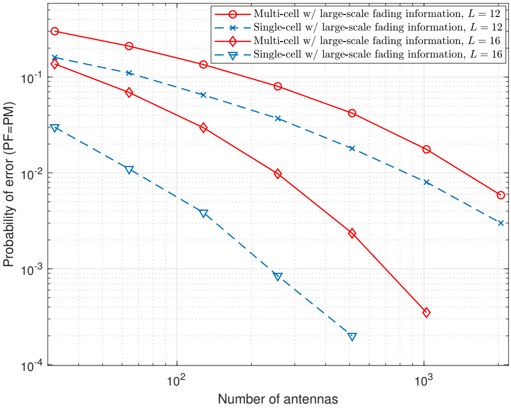


Figure: Performance comparison of the multicell covariance approach with single-cell system, with knowing large-scale fading. $B = 7$, $N = 200$, $K = 20$.

Conclusions

- Device activity detection for massive random access in machine-type and IoT communications is a sparse recovery problem.
- Two detection algorithms for user activity detection:
  - Signal-based AMP for estimating the user activity and the exact channel.
  - Covariance-based MLE for estimating the user activity only.
- Analyses for AMP and the covariance approach:
  - State evolution for AMP: Low complexity, works best for small M.
  - Fisher information matrix for covariance approach: Suited for massive MIMO.
- Advantage of covariance-based approach is that it can handle $K = O(L^2)$, as compared to AMP which can only handle $K = O(L)$.

Further Information

Zhilin Chen, Foad Sohrabi, Ya-Feng Liu, Wei Yu,
"Phase Transition Analysis for Covariance Based Massive Random Access with Massive MIMO",
[Online] available: https://arxiv.org/abs/2003.04175, March 2020.

Zhilin Chen, Foad Sohrabi, and Wei Yu,
"Sparse Activity Detection in Multi-Cell Massive MIMO Exploiting Channel Large-Scale Fading",
To appear in *IEEE Transactions on Signal Processing*, 2021.