Covariance-Based Activity Detection in Cooperative Multi-Cell Massive MIMO: Scaling Law and Efficient Algorithms

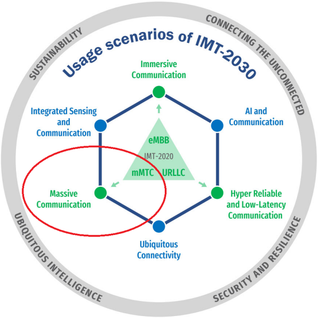
Ya-Feng Liu

State Key Laboratory of Scientific/Engineering Computing
Institute of Computational Mathematics and Scientific/Engineering Computing
Academy of Mathematics and Systems Science
Chinese Academy of Sciences, Beijing, China
Email: yafliu@lsec.cc.ac.cn

Algorithmic Structures for Uncoordinated Communications and Statistical Inference in Exceedingly Large Spaces Workshop

BIRS, Banff, March 10–15, 2024

Massive Connectivity for IoTs

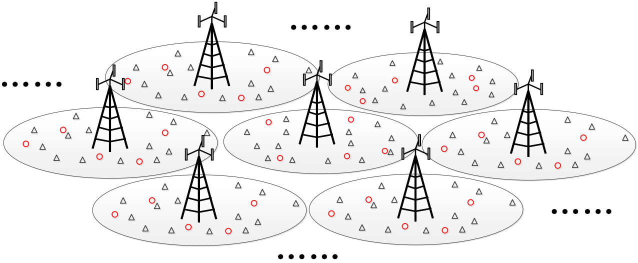


- From massive machine-type communication (mMTC) in 5G to massive communication in 6G
- Please see Prof. Wei Yu's talk for more about the background such as massive random access, massive MIMO, covariance-based approach, scaling law, coordinate descent (CD) algorithms,...
- Mainly focus on covariance-based activity detection in multi-cell massive MIMO systems

Outline

- Problem Formulation
- Scaling Law
- Efficient CD Algorithms
- Simulation Results
- Concluding Remarks

Cooperative Multi-Cell MIMO System



- A multi-cell system consists of B cells, and each of cell contains
  - one base station (BS) equipped with M antennas;
  - N single-antenna devices, K of which are active during any coherence interval.
- The signals received at all BSs are collected and jointly processed at the central unit.

Channel Model

- The device n in cell b is preassigned a unique signature sequence $\mathbf{s}_{bn} \in \mathbb{C}^L$.
- The channel between device n in cell b and BS j is $\sqrt{g_{jbn}}\mathbf{h}_{jbn}$, where
  - $g_{jbn} \geq 0$ is the large-scale fading coefficient;
  - $\mathbf{h}_{jbn} \in \mathbb{C}^M$ is the Rayleigh fading coefficient following $\mathcal{CN}(\mathbf{0}, \mathbf{I})$.
- The activity of device n in cell b is indicated as $a_{bn} \in \{0, 1\}$:
  - $a_{bn} = 1$ indicates that the device is active;
  - $a_{bn} = 0$ indicates that the device is inactive.
- $\mathbf{W}_b \in \mathbb{C}^{L \times M}$ is the additive white Gaussian noise that follows $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$.

Channel Model

Then the received signal $\mathbf{Y}_b \in \mathbb{C}^{L\times M}$ at BS b can be expressed as

$$\begin{aligned}\mathbf{Y}_b &= \sum_{n=1}^{N} a_{bn}\mathbf{s}_{bn} g_{bbn}^{\frac{1}{2}} \mathbf{h}_{bbn}^T + \sum_{j\neq b}\sum_{n=1}^{N} a_{jn}\mathbf{s}_{jn} g_{bjn}^{\frac{1}{2}} \mathbf{h}_{bjn}^T + \mathbf{W}_b,\\ &= \mathbf{S}_b\mathbf{A}_b\mathbf{G}_{bb}^{\frac{1}{2}}\mathbf{H}_{bb} + \sum_{j\neq b}\mathbf{S}_j\mathbf{A}_j\mathbf{G}_{bj}^{\frac{1}{2}}\mathbf{H}_{bj} + \mathbf{W}_b,\end{aligned}$$

where

- the signature sequence matrix $\mathbf{S}_b = [\mathbf{s}_{b1}, \ldots, \mathbf{s}_{bN}] \in \mathbb{C}^{L\times N}$ and the large-scale fading coefficient matrices $\mathbf{G}_{bj} = \mathrm{diag}(g_{bj1}, \ldots, g_{bjN}) \in \mathbb{R}^{N\times N}$ for all j are assumed to be known (e.g., when all devices' locations are fixed);
- the matrices $\mathbf{H}_{bj} = [\mathbf{h}_{bj1}, \ldots, \mathbf{h}_{bjN}]^T \in \mathbb{C}^{N\times M}$ and the noise $\mathbf{W}_b \in \mathbb{C}^{L\times M}$ are unknown, but they are Gaussian distributed;
- the problem is to estimate $\mathbf{A}_b = \mathrm{diag}(a_{b1}, \ldots, a_{bN}) \in \mathbb{R}^{N\times N}$ for all b from the received signals $\{\mathbf{Y}_b\}_{b=1}^{B}$.

Distribution of $\{\mathbf{Y}_b\}_{b=1}^{B}$

- Let $\mathbf{a} = \left[\mathbf{a}_1^T, \ldots, \mathbf{a}_B^T\right]^T \in \mathbb{R}^{BN}$, where $\mathbf{a}_b = \left[a_{b1}, \ldots, a_{BN}\right]^T \in \mathbb{R}^N$ denotes the diagonal entries of $\mathbf{A}_b$.
- For a given (deterministic) $\mathbf{a}$, the columns of the received signal $\mathbf{Y}_b$ denoted by $\mathbf{y}_{bm},\ m = 1, \ldots, M$ are i.i.d. Gaussian vectors:

$$\mathbf{y}_{bm} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}_b\right),$$

where the (true) covariance matrix $\mathbf{\Sigma}_b \in \mathbb{C}^{L \times L}$ is given by

$$\mathbf{\Sigma}_b = \sum_{j=1}^{B} \mathbf{S}_j \mathbf{G}_{bj} \mathbf{A}_j \mathbf{S}_j^H + \sigma^2 \mathbf{I}.$$

Recovery via MLE

- To recover $\mathbf{a}$ by using maximum likelihood estimation (MLE):

$$
\begin{aligned}
\underset{\mathbf{0}\le\mathbf{a}\le\mathbf{1}}{\text{minimize}} \quad F(\mathbf{a}) &\triangleq -\frac{1}{M}\log \mathrm{P}(\mathbf{Y}_1,\ldots,\mathbf{Y}_B \mid \mathbf{a}) \\
&= -\frac{1}{M}\sum_{b=1}^{B}\log \mathrm{P}(\mathbf{Y}_b \mid \mathbf{a}) \\
&= \sum_{b=1}^{B}\left(-\frac{1}{M}\sum_{m=1}^{M}\log \mathrm{P}(\mathbf{y}_{bm} \mid \mathbf{a})\right) \\
&= \sum_{b=1}^{B}\left(\log|\boldsymbol{\Sigma}_b| + \mathrm{tr}\left(\boldsymbol{\Sigma}_b^{-1}\widehat{\boldsymbol{\Sigma}}_b\right)\right),
\end{aligned}
$$

where $\widehat{\boldsymbol{\Sigma}}_b$ is the sample covariance matrix defined as

$$
\widehat{\boldsymbol{\Sigma}}_b \triangleq \frac{1}{M}\mathbf{Y}_b\mathbf{Y}_b^H = \frac{1}{M}\sum_{m=1}^{M}\mathbf{y}_{bm}\mathbf{y}_{bm}^H.
$$

Covariance-Based Detection in Multi-Cell MIMO Systems

- The activity vector $\mathbf{a}$ can be estimated by solving the following MLE problem [Chen-Sohrabi-Yu; 2021]:

$$\begin{aligned} \underset{\mathbf{a}}{\text{minimize}} \quad & \sum_{b=1}^{B} \left(\log |\mathbf{\Sigma}_b| + \operatorname{tr}\left(\mathbf{\Sigma}_b^{-1} \widehat{\mathbf{\Sigma}}_b \right) \right) && \text{(1a)} \\ \text{subject to} \quad & \mathbf{a} \in [0,1]^{BN}, && \text{(1b)} \end{aligned}$$

  where

  - $\mathbf{a} = [\mathbf{a}_1^T, \ldots, \mathbf{a}_B^T]^T \in \mathbb{R}^{BN}$ indicates the activity of all the devices;
  - $\mathbf{\Sigma}_b = \sum_{j=1}^{B} \mathbf{S}_j \mathbf{A}_j \mathbf{G}_{bj} \mathbf{S}_j^H + \sigma^2 \mathbf{I}$ is the true covariance matrix of $\mathbf{Y}_b$;
  - $\widehat{\mathbf{\Sigma}}_b = \frac{1}{M} \mathbf{Y}_b \mathbf{Y}_b^H = \frac{1}{M} \sum_{m=1}^{M} \mathbf{y}_{bm} \mathbf{y}_{bm}^H$ is the sample covariance matrix.

- In the rest of this talk, we focus on the MLE problem formulated in (1).

Two Important Problems about MLE

Consider the following two important problems about the following MLE problem:

$$\begin{aligned} \underset{\mathbf{a}}{\text{minimize}} \quad & \sum_{b=1}^{B} \left(\log |\boldsymbol{\Sigma}_b| + \text{tr}\left(\boldsymbol{\Sigma}_b^{-1} \widehat{\boldsymbol{\Sigma}}_b \right) \right) \\ \text{subject to} \quad & \mathbf{a} \in [0,1]^{BN}. \end{aligned}$$

Q1: Scaling law: What is the detection performance limit of the MLE formulation as the number of antennas M goes to infinity? How the number of cells B (and the inter-cell interference) affects the detection performance?

Q2: Algorithms: How to design efficient algorithms for solving the MLE problem to achieve accurate and fast activity detection? (Caution! Do not overlook the summation due to multiple cells, which results into a highly nonlinear objective function.)

Two Important Problems about MLE

Consider the following two important problems about the following MLE problem:

$$\begin{aligned} \underset{\mathbf{a}}{\text{minimize}} \quad & \sum_{b=1}^{B} \left(\log |\mathbf{\Sigma}_b| + \operatorname{tr}\left(\mathbf{\Sigma}_b^{-1} \widehat{\mathbf{\Sigma}}_b \right) \right) \\ \text{subject to} \quad & \mathbf{a} \in [0,1]^{BN}. \end{aligned}$$

Q1: Scaling law: What is the detection performance limit of the MLE formulation as the number of antennas M goes to infinity? How the number of cells B (and the inter-cell interference) affects the detection performance?

Q2: Algorithms: How to design efficient algorithms for solving the MLE problem to achieve accurate and fast activity detection? (Caution! Do not overlook the summation due to multiple cells, which results into a highly nonlinear objective function.)

- Let us first answer Q1.

Consistency of the MLE

Lemma 1 (Chen-Sohrabi-Yu, 2021)

Consider the MLE problem (1) *with given* $\mathbf{S}$, $\{\mathbf{G}_b\}_{b=1}^{B}$, *and* σ_w^2. *Let matrix* $\widetilde{\mathbf{S}}$ *be defined as*

$$\widetilde{\mathbf{S}} \triangleq [\mathbf{s}_{11}^{*} \otimes \mathbf{s}_{11}, \mathbf{s}_{12}^{*} \otimes \mathbf{s}_{12}, \ldots, \mathbf{s}_{BN}^{*} \otimes \mathbf{s}_{BN}] \in \mathbb{C}^{L^2 \times BN}. \tag{3}$$

Let $\hat{\mathbf{a}}^{(M)}$ *be the solution to* (1) *when the number of antennas* M *is given and let* $\mathbf{a}^{\circ}$ *be the true activity indicator vector whose* $B(N-K)$ *zero entries are indexed by* $\mathcal{I}$, *i.e.,*

$$\mathcal{I} \triangleq \{i \mid a_i^{\circ} = 0\}.$$

Define two sets

$$\mathcal{N} \triangleq \{\mathbf{x} \in \mathbb{R}^{BN} \mid \widetilde{\mathbf{S}}\mathbf{G}_b\mathbf{x} = \mathbf{0}, \forall\, b\}, \tag{4}$$

$$\mathcal{C} \triangleq \{\mathbf{x} \in \mathbb{R}^{BN} \mid x_i \geq 0 \text{ if } i \in \mathcal{I},\, x_i \leq 0 \text{ if } i \notin \mathcal{I}\}. \tag{5}$$

Then a necessary and sufficient condition for $\hat{\mathbf{a}}^{(M)} \to \mathbf{a}^{\circ}$ *as* $M \to \infty$ *is that the intersection of* $\mathcal{N}$ *and* $\mathcal{C}$ *is the zero vector, i.e.,* $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$.

Remarks on Lemma 1

- The condition $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ implies that the likelihood function is uniquely identifiable in the feasible neighborhood of $\mathbf{a}^\circ$:
  - the subspace $\mathcal{N}$ contains all directions from $\mathbf{a}^\circ$ along which the likelihood function $p(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_B \mid \mathbf{a}^\circ)$ remains unchanged;
  - the cone $\mathcal{C}$ contains all directions starting from $\mathbf{a}^\circ$ towards the feasible region.
- There is generally no closed-form characterization of $\mathcal{N} \cap \mathcal{C}$.
- The scaling law analysis is to characterize the feasible set of system parameters such as $N, K, L,$ and B under which $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ holds true.
- Recall $\mathcal{N} = \{\mathbf{x} \in \mathbb{R}^{BN} \mid \widetilde{\mathbf{S}}\mathbf{G}_b\mathbf{x} = \mathbf{0}, \forall\, b\}$. The signature sequences and the large-scale fading coefficients are critical in the scaling law analysis because they are involved in the definition of $\mathcal{N}$.

Assumption on Signature Sequences

Recall $\mathbf{S} = [\mathbf{S}_1, \ldots, \mathbf{S}_B] \in \mathbb{C}^{L \times BN}$, where $\mathbf{S}_b = [\mathbf{s}_{b1}, \ldots, \mathbf{s}_{bN}] \in \mathbb{C}^{L \times N}$.

Assumption 1 (Generation of the signature sequence matrix)

The signature sequence matrix $\mathbf{S}$ is generated from one of the following two ways, and the corresponding signature sequences are called Type I and Type II, respectively:

Type I: draw the components of $\mathbf{S}$ uniformly and independently from the discrete set $\left\{\pm\frac{\sqrt{2}}{2} \pm \imath\frac{\sqrt{2}}{2}\right\}$, i.e., draw the columns of $\mathbf{S}$ randomly and uniformly from the discrete set $\left\{\pm\frac{\sqrt{2}}{2} \pm \imath\frac{\sqrt{2}}{2}\right\}^L$ (where $\imath$ is the imaginary unit);

Type II: draw the columns of $\mathbf{S}$ uniformly and independently from the complex sphere of radius $\sqrt{L}$.

- Both types of sequences are normalized with the length of $\sqrt{L}$;
- Type I is better than Type II in terms of the complexity of generating and storing the signature sequences.

Statistical Properties of Signature Sequences

Theorem 1 (Wang-L.-Wang-Yu, 2023)

For both the Type I and Type II sequences stated in Assumption 1, the following holds. For any given parameter $\bar{\rho} \in (0,1)$, there exist constants c_1 and c_2 depending only on $\bar{\rho}$ such that if

$$s \leq c_1 L^2 / \log^2(eBN/L^2),$$

then with probability at least $1 - \exp(-c_2 L)$, the matrix $\widetilde{\mathbf{S}}$ defined in (3) has the stable null space property of order s with parameters $\rho \in (0, \bar{\rho})$.
More precisely, for any $\mathbf{v} \in \mathbb{R}^{BN}$ that satisfies $\widetilde{\mathbf{S}}\mathbf{v} = \mathbf{0}$, the following inequality holds for any index set $\mathcal{S} \subseteq \{1, 2, \ldots, BN\}$ with $|\mathcal{S}| \leq s$:

$$\|\mathbf{v}_{\mathcal{S}}\|_1 \leq \rho \|\mathbf{v}_{\mathcal{S}^c}\|_1,$$

where $\mathbf{v}_{\mathcal{S}}$ is a sub-vector of $\mathbf{v}$ with entries from $\mathcal{S}$, and $\mathcal{S}^c$ is the complementary set of $\mathcal{S}$ with respect to $\{1, 2, \ldots, BN\}$.

- This conclusion of Type II signature sequences was previously proved in [Fengler-Haghighatshoar-Jung-Caire, 2021].

Assumption on the Path-Loss Model

Recall $\mathcal{N} = \{\mathbf{x} \in \mathbb{R}^{BN} \mid \widetilde{\mathbf{S}}\mathbf{G}_b\mathbf{x} = \mathbf{0},\ \forall\, b\}$.

Assumption 2

The multi-cell system consists of B hexagonal cells with radius R. In this system, the large-scale fading components are inversely proportional to the distance raised to the power γ, i.e.,

$$g_{bjn} = P_0 \left(\frac{D_0}{D_{bjn}} \right)^{\gamma},$$

where P_0 is the received power at the point with distance D_0 from the transmitting antenna, D_{bjn} is the BS-device distance between device n in cell j and BS b, and γ is the path-loss exponent.

Properties of Large-Scale Fading Coefficients

Lemma 2 (Wang-Liu-Wang-Yu, 2023)

Suppose that Assumption 2 holds true with $\gamma > 2$. *Then, there exists a constant* $C > 0$ *depending only on* γ, P_0, D_0, *and* R *defined in Assumption 2, such that for each BS* b, *the large-scale fading coefficients satisfy*

$$\sum_{j=1,\, j\neq b}^{B} \left(\max_{1\leq n\leq N} g_{bjn} \right) \leq C. \tag{6}$$

- The summation in the left-hand side of (6) is upper bounded by a constant C that is independent of B.
- $\gamma > 2$ is a sufficient condition for (6), and it typically holds for most channel models and application scenarios.

Scaling Law Result

Theorem 2 (Wang-L.-Wang-Yu, 2023)

For both Type I and Type II sequences in Assumption 1 and under Assumption 2 with $\gamma > 2$, *there exist constants* c_1 *and* $c_2 > 0$, *independent of system parameters* K, L, N, *and* B, *such that if*

$$K \leq c_1 L^2 / \log^2(eBN/L^2),$$

then the condition $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ *in Lemma 1 holds with probability at least* $1 - \exp(-c_2 L)$.

- The maximum number of active devices K that can be detected correctly in each cell increases quadratically with L and decreases logarithmically with B.

Scaling Law Result

Theorem 2 (Wang-L.-Wang-Yu, 2023)

For both Type I and Type II sequences in Assumption 1 and under Assumption 2 with $\gamma > 2$*, there exist constants* c_1 *and* $c_2 > 0$*, independent of system parameters* K, L, N, *and* B*, such that if*

$$K \leq c_1 L^2 / \log^2(eBN/L^2),$$

then the condition $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ *in Lemma 1 holds with probability at least* $1 - \exp(-c_2 L)$.

- The maximum number of active devices K that can be detected correctly in each cell increases quadratically with L and decreases logarithmically with B.
- Implication: the maximum number of active devices that can be correctly detected in each cell in the multi-cell scenario is almost identical to that in the single-cell scenario [Fengler-Haghighatshoar-Jung-Caire, 2021] [Chen-Sohrabi-L.-Yu, 2022].

Recap

- Statistical properties of the two types of signature sequences.
- Good detection performance of MLE (under mild assumptions):
  - scaling law: $K = \mathcal{O}(L^2 / \log^2(BN/L^2))$;
  - estimation error: $\hat{\mathbf{a}}^{(M)} - \mathbf{a}^{\circ} = \mathcal{O}\left(\frac{1}{\sqrt{M}}\right)$ (skipped).

Recap

- Statistical properties of the two types of signature sequences.
- Good detection performance of MLE (under mild assumptions):
  - scaling law: $K = \mathcal{O}(L^2/\log^2(BN/L^2))$;
  - estimation error: $\hat{\mathbf{a}}^{(M)} - \mathbf{a}^\circ = \mathcal{O}\left(\frac{1}{\sqrt{M}}\right)$ (skipped).
- Q2: how to achieve fast and accurate activity detection based on the MLE formulation?
- We shall design efficient CD algorithms for solving the MLE problem.

MLE Problem

- Need to solve the following nonconvex MLE problem:

$$\underset{\mathbf{a}\in[0,1]^{BN}}{\text{minimize}} \quad F(\mathbf{a}) \triangleq \sum_{b=1}^{B} f_b(\mathbf{a}),$$

where $f_b(\mathbf{a}) \triangleq \log|\mathbf{\Sigma}_b| + \text{tr}\left(\mathbf{\Sigma}_b^{-1}\widehat{\mathbf{\Sigma}}_b\right)$.

- Define the following nonnegative KKT violation vector

$$\mathbb{V}(\mathbf{a}) \triangleq |\text{Proj}(\mathbf{a} - \nabla F(\mathbf{a})) - \mathbf{a}| \in \mathbb{R}_+^{BN}.$$

- Then solving the above problem is equivalent to finding a point that satisfies its KKT condition, i.e., $\mathbb{V}(\mathbf{a}) = \mathbf{0}$.
- Goal: for a given tolerance $\epsilon > 0$, find a feasible point $\mathbf{a}$ with $\left\|\mathbb{V}(\mathbf{a})\right\|_\infty \leq \epsilon$.

Review of A SOTA CD Algorithm

- Random permuted CD [Chen-Sohrabi-Yu, 2021]: at each iteration the algorithm randomly permutes the indices of all coordinates and then updates all coordinates one by one according to the order in the permutation.

- For any given coordinate (b, n) of device n in cell b, the algorithm solves the following one-dimensional optimization problem:

$$\underset{d\in[-a_{bn},\,1-a_{bn}]}{\text{minimize}} \quad \sum_{j=1}^{B}\left(\log\left(1+d\,g_{jbn}\,\mathbf{s}_{bn}^{H}\boldsymbol{\Sigma}_j^{-1}\mathbf{s}_{bn}\right)-\frac{d\,g_{jbn}\,\mathbf{s}_{bn}^{H}\boldsymbol{\Sigma}_j^{-1}\widehat{\boldsymbol{\Sigma}}_j\boldsymbol{\Sigma}_j^{-1}\mathbf{s}_{bn}}{1+d\,g_{jbn}\,\mathbf{s}_{bn}^{H}\boldsymbol{\Sigma}_j^{-1}\mathbf{s}_{bn}}\right) \quad (7)$$

  to possibly update a_{bn}.

- Unlike the single-cell case, problem (7) has no closed-form solution.

- Problem (7) can be solved by finding the roots of a polynomial of degree $2B-1$ with a complexity of $\mathcal{O}(B^3)$.

Remarks on the SOTA CD Algorithm

(i) The total complexity of updating one coordinate is $\mathcal{O}(BL^2 + B^3)$, including solving the suproblem for d and updating $\boldsymbol{\Sigma}_b^{-1}$ for all $b = 1, \ldots, B$.

(ii) The current CD algorithm might become inefficient and involves many unnecessary coordinate updates when B and N are large, i.e., there are many subproblems with the solution being $d \approx 0$.

Remarks on the SOTA CD Algorithm

(i) The total complexity of updating one coordinate is $\mathcal{O}(BL^2 + B^3)$, including solving the suproblem for d and updating $\mathbf{\Sigma}_b^{-1}$ for all $b = 1, \ldots, B$.

(ii) The current CD algorithm might become inefficient and involves many unnecessary coordinate updates when B and N are large, i.e., there are many subproblems with the solution being $d \approx 0$.

- Propose two "simple" acceleration techniques to overcome the above two problems, which lead to two accelerated CD algorithms:
  - inexact CD, where the complexity of solving the subproblem is $\mathcal{O}(BL^2)$;
  - active set CD, where only the coordinates in a carefully selected set (called active set) are updated.

Proposed Inexact CD Algorithm

- Rewrite the one-dimensional subproblem (7) with given (b, n) as follows:

$$\underset{d \in [-a_{bn},\, 1-a_{bn}]}{\text{minimize}} \quad \sum_{j=1}^{B} f_j(\mathbf{a} + d\,\mathbf{e}_{bn}), \tag{8}$$

  where

  - $\mathbf{e}_{bn} \in \mathbb{R}^{BN}$ is an all-zero vector except its (b, n)-th component being 1;
  - $f_j(\mathbf{a} + d\,\mathbf{e}_{bn}) = f_j(\mathbf{a}) + \log\left(1 + d\, g_{jbn}\, \mathbf{s}_{bn}^H \boldsymbol{\Sigma}_j^{-1} \mathbf{s}_{bn}\right) - \frac{d\, g_{jbn}\, \mathbf{s}_{bn}^H \boldsymbol{\Sigma}_j^{-1} \widehat{\boldsymbol{\Sigma}}_j \boldsymbol{\Sigma}_j^{-1} \mathbf{s}_{bn}}{1 + d\, g_{jbn}\, \mathbf{s}_{bn}^H \boldsymbol{\Sigma}_j^{-1} \mathbf{s}_{bn}}.$

- Motivation: update coordinate a_{bn} with a lower complexity by solving problem (8) inexactly in a controllable fashion.

- Observations: (i) the large-scale fading coefficient g_{jbn} appears in the j-th term multiplied with d in (8); and (ii) $g_{bbn} \gg g_{jbn}$ for all $j \neq b$ (due to the path-loss model).

Proposed Inexact CD Algorithm

- Idea: construct a simple yet tight approximation of problem (8) by retaining the b-th dominant term and approximating the other terms.
- The approximate problem is given by

$$d^{(\mu)} = \underset{d \in [-a_{bn},\, 1-a_{bn}]}{\operatorname{argmin}} \quad f_b(\mathbf{a} + d\,\mathbf{e}_{bn}) + \sum_{j=1,\, j\neq b}^{B} \left(f_j(\mathbf{a}) + [\nabla f_j(\mathbf{a})]_{bn}\, d\right) + \frac{\mu}{2} d^2. \quad (9)$$

- The parameter μ is chosen (e.g., by the line search) such that $d^{(\mu)}$ satisfies the following sufficient decrease condition:

$$\sum_{j=1,\, j\neq b}^{B} f_j\left(\mathbf{a} + d^{(\mu)}\mathbf{e}_{bn}\right) \leq \sum_{j=1,\, j\neq b}^{B} \left(f_j(\mathbf{a}) + [\nabla f_j(\mathbf{a})]_{bn}\, d^{(\mu)}\right) + \frac{\mu}{2}\left(d^{(\mu)}\right)^2.$$

- The approximate problem in (9) can be solved by the cubic formula with complexity $\mathcal{O}(1)$.

CD and Inexact CD Algorithms

Algorithm 1 CD algorithm for solving the MLE problem

Initialize $\mathbf{a} = \mathbf{0}$, $\boldsymbol{\Sigma}_b^{-1} = \sigma_w^{-2}\mathbf{I}$, $1 \leq b \leq B$, and $\epsilon > 0$;
repeat
    Randomly select a permutation $\{(b,n)_1, (b,n)_2, \ldots, (b,n)_{BN}\}$ of the coordinate indices $\{(1,1), (1,2), \ldots, (1,N), \ldots, (B,1), (B,2), \ldots, (B,N)\}$ of $\mathbf{a}$;
    for $(b,n) = (b,n)_1$ to $(b,n)_{BN}$ **do**
        If CD: Apply the root-finding algorithm [McNamee, 2007] to solve subproblem (7) *exactly* to obtain $\hat{d}$, and set $d = \hat{d}$;
        If inexact CD: Use the cubic formula to solve the approximate subproblem (9) to obtain $\bar{d}$, and set $d = \bar{d}$;
        $a_{bn} \leftarrow a_{bn} + d$;
        $\boldsymbol{\Sigma}_j^{-1} \leftarrow \boldsymbol{\Sigma}_j^{-1} - \frac{d\, g_{jbn}\, \boldsymbol{\Sigma}_j^{-1}\mathbf{s}_{bn}\mathbf{s}_{bn}^H\boldsymbol{\Sigma}_j^{-1}}{1 + d\, g_{jbn}\, \mathbf{s}_{bn}^H\boldsymbol{\Sigma}_j^{-1}\mathbf{s}_{bn}}$, $j = 1, \ldots, B$;
    end for
until $\left\|\mathbb{V}(\mathbf{a})\right\|_\infty \leq \epsilon$;
Output $\mathbf{a}$.

Active Set CD Algorithm

- Motivation of the active set CD algorithm:
  - both CD and inexact CD algorithms might do many unnecessary coordinate updates where subproblems (7) and (9) are solved but the corresponding a_{bn} almost does not change and the objective does not decrease "sufficiently";
  - in this case, the solution d of subproblems (7) and (9) has a very small magnitude, i.e.,

$$d \approx 0.$$

- The active set idea: select a subset of coordinates that have the most potential of decreasing the objective to reduce the number of unnecessary coordinate updates.

Proposed Active Set CD Algorithm

- At each iteration, the active set CD algorithm
  - first judiciously selects an "active" set of coordinates;
  - then updates the coordinates in the active set once.
- The proposed selection strategy for the active set $\mathcal{A}^{(k)}$ is expressed as

$$\mathcal{A}^{(k)} = \left\{(b,n) \mid \left[\mathbb{V}(\mathbf{a}^{(k)})\right]_{bn} \geq \omega^{(k)}\right\},$$

  where $\omega^{(k)} \geq 0$ is a properly selected threshold parameter at each iteration.
- Intuition: updating the coordinate a_{bn} with a larger $\left[\mathbb{V}(\mathbf{a})\right]_{bn}$ is expected to yield a larger decrease in the objective function. [Do not treat all coordinates equally as in CD!]
- The choice of $\omega^{(k)}$ is crucial in achieving the balance between decreasing the objective function and reducing the cardinality of the active set (and hence the computational cost) at the k-th iteration.

Proposed Active Set CD Algorithm

Algorithm 2 Active set CD algorithm for solving the MLE problem

```
Initialize $\mathbf{a}^{(0)} = \mathbf{0}$, $k = 0$, and $\epsilon > 0$;
**repeat**
    Update $\omega^{(k)}$;
    Select the active set $\mathcal{A}^{(k)} = \left\{(b, n) \mid \left[\mathbb{V}(\mathbf{a}^{(k)})\right]_{bn} \geq \omega^{(k)}\right\}$;
    Apply lines 5–8 in Algorithm 1 to update all coordinates in $\mathcal{A}^{(k)}$ *only once*
    in the order of a random permutation;
**until** $\left\|\mathbb{V}(\mathbf{a}^{(k)})\right\|_\infty \leq \epsilon$;
Output $\mathbf{a}^{(k)}$.
```

- Note that $\mathbb{V}(\mathbf{a}^{(k)})$ will become smaller and smaller as k increases.
- The cardinality of the selected active set, $|\mathcal{A}^{(k)}|$, is expected to be significantly less than BN and gradually decreases as the iteration goes on.
- The active set strategy can accelerate both CD and inexact CD algorithms.

Convergence Guarantee of Active Set CD Algorithm

Convergence and iteration complexity properties of the proposed active set CD algorithm:

Theorem 3 (Wang-L.-Wang-Yu, 2023)

For any given error tolerance $\epsilon > 0$, *let* $\omega^{(k)}$ *satisfy the condition*

$$\epsilon \leq \omega^{(k)} \leq \max\left\{\left\|\mathbb{V}(\mathbf{a}^{(k)})\right\|_\infty, \epsilon\right\}.$$

Then, the proposed active set CD Algorithm 2 will terminate (i.e., finding a feasible point $\mathbf{a}$ *which satisfies* $\left\|\mathbb{V}(\mathbf{a})\right\|_\infty \leq \epsilon$*) within* $\mathcal{O}(1/\epsilon^2)$ *iterations.*

Recap

Table 1: A summary of per-iteration complexity comparison.

	Vanilla CD	Inexact CD	Active set CD	Active set inexact CD
Total number of updated coordinates	BN		$\lvert\mathcal{A}^{(k)}\rvert$	
Complexity of updating one coordinate	$\mathcal{O}(BL^2+B^3)$	$\mathcal{O}(BL^2)$	$\mathcal{O}(BL^2+B^3)$	$\mathcal{O}(BL^2)$

- Inexact CD:
  - the total complexity of updating one coordinate is $\mathcal{O}(BL^2)$;
  - the per-iteration total complexity (of updating all coordinates) is $\mathcal{O}(BN \times BL^2)$.
- Active set CD:
  - the total number of updated coordinates is $\lvert\mathcal{A}^{(k)}\rvert$, which is generally significantly less than BN;
  - terminate within $\mathcal{O}(\varepsilon^{-2})$ iterations to return a point satisfying $\lVert\mathbb{V}(\mathbf{a})\rVert_\infty \leq \epsilon$.

Simulation Settings and Our Goals

- Simulation settings:
  - consider a multi-cell system consisting of hexagonal cells, and all potential devices within each cell are uniformly distributed.
  - the radius of each cell is $500\,\mathrm{m}$.
  - the channel path-loss is $128.1 + 37.6\log 10(d)$ (satisfying Assumption 2 with $\gamma = 3.76$), where d is the corresponding BS-device distance in km.
  - the transmit power of each device is $23\,\mathrm{dBm}$ and the background noise power is $-169\,\mathrm{dBm/Hz}$ over $10\,\mathrm{MHz}$.
- Two goals in this part:
  - detection performance comparison of different types of signature sequences
  - computational efficiency of proposed inexact and active set CD algorithms

Detection Performance Comparison

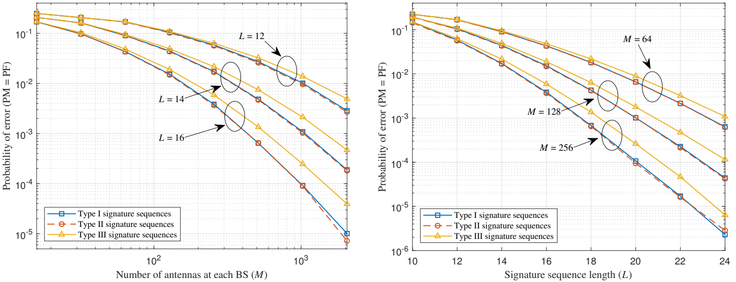


Figure 1: Detection performance comparison of three different types of signature sequences for different M and L ($B = 7$, $N = 200$, and $K = 20$).

- Almost the same detection performance can be obtained by using Type I and Type II signature sequences.
- The detection performance of Type III signature sequences (generated from an i.i.d. complex Gaussian distribution) is worse than the other two.
- Key message: normalization of the signature sequences is crucial to the detection performance, especially to the FA error performance. Please check our paper on more simulation results and explanations on this.

Benchmarks

- Consider the following two algorithms as benchmarks:
  - vanilla CD [Chen-Sohrabi-Yu, 2021];
  - clustering-based CD [Ganesan-Björnson-Larsson, 2021], which utilizes the signals from T dominant BSs for a given user (the number of clusters T is chosen to be $1, 2, 3$ in our implementation).
- Our proposed three algorithms:
  - inexact CD;
  - active set CD;
  - active set inexact CD.
- Parameter settings:
  - the signature sequences used are Type I.
  - $B = 7,\ K = 20,\ L = 20$, and $M = 128$, $\epsilon = 10^{-3}$, $\beta = 2$, and

$$\omega^{(k)} = \max\left\{5^{-k-1}\left\|\mathbb{V}(\mathbf{a}^{(k)})\right\|_\infty,\ \epsilon\right\}.$$

Comparison of Probability of Error versus Running Time

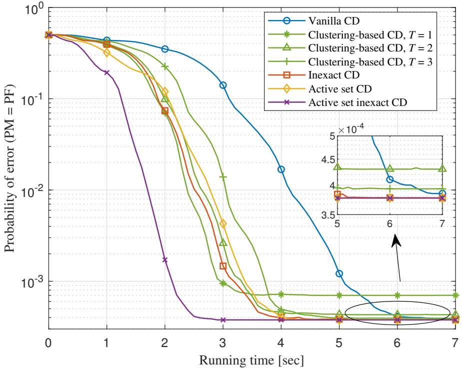


Figure 2: Comparison of the probability of error of the proposed algorithms and the two benchmarks versus the running time.

- Active set inexact CD is the most efficient algorithm.
- Clustering-based CD is efficient but its probability of error is worse than the other algorithms (due to the coarse approximation).

Average Running Time Comparison

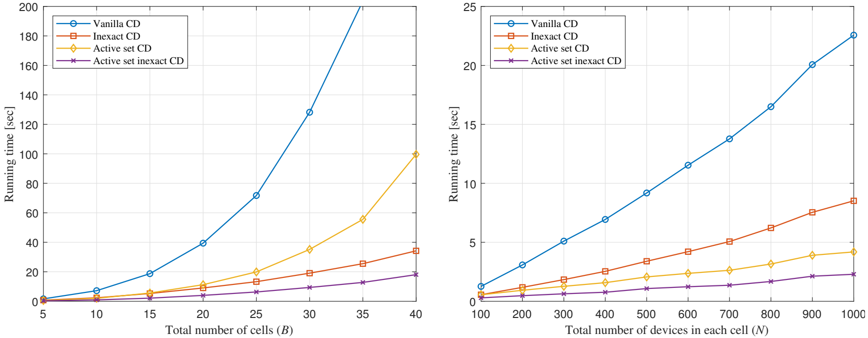


Figure 3: Average running time comparison of the proposed algorithms and the vanilla CD algorithm for different B and N.

- As B increases, the algorithms that solve subproblem (7) inexactly, inexact CD and active set inexact CD, are much more efficient.
- As N increases, the algorithms that use the active set selection strategy, active set CD and active set inexact CD, are significantly more efficient.

Comparison of Number of Updated Coordinates

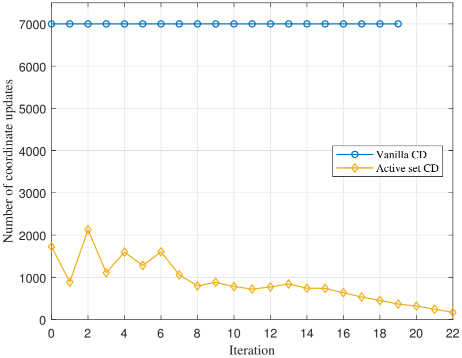


Figure 4: Comparison of the number of updated coordinates of vanilla CD and active set CD over the iteration.

- The number of updated coordinates of active set CD at each iteration is significantly less than that of vanilla CD at each iteration.
- The overall number of iterations of active set CD is slightly more than that of vanilla CD.

Practical Issues and Extensions

I: Joint device data and activity detection when each device has only a few bits of data to transmit [Chen-Sohrabi-L.-Yu, 2019]

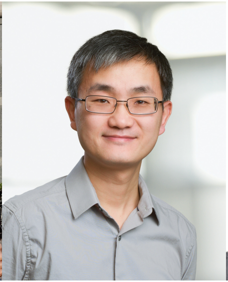
II: Joint device activity and delay detection when active devices asynchronously transmit their preassigned signature sequences [Li-Lin-L.-Ai-Wu, 2023]

III: Device activity detection when the BS is equipped with low-resolution ADCs [Wang-L.-Wang-Liu-Pan-Cui, 2023]

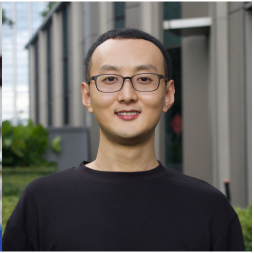
Concluding Remarks

- Covariance-based activity detection in cooperative multi-cell massive MIMO systems
- Statistical properties of two types of signature sequences and scaling law result
- Two efficient accelerated CD algorithms:
  - inexact CD: reducing the complexity of updating one coordinate by exploiting the special structure of the objective function;
  - active set CD: reducing the total number of unnecessary coordinate updates by using the optimality condition;
  - convergence and iteration complexity guarantees of the proposed CD algorithms.
- Practical issues and extensions

My Collaborators on Massive Random Access





Reference

- Z. Wang, Y.-F. Liu, Z. Wang, and W. Yu,"Covariance-based activity detection in cooperative multi-cell massive MIMO: Scaling law and efficient algorithms," 2023. [Online]. Available: https://arxiv.org/abs/2311.15299
- Y.-F. Liu, W. Yu, Z. Wang, Z. Chen, and F. Sohrabi, "Grant-free random access via covariance-based approach," in *Next Generation Multiple Access.*, (Editors: Yuanwei Liu, Liang Liu, Zhiguo Ding, and Xuemin (Sherman) Shen), Wiley, 2023.
- Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection in multi-cell massive MIMO exploiting channel large-scale fading," *IEEE Trans. Signal Process.*, vol. 69, pp. 3768–3781, Jun. 2021.
- Z. Chen, F. Sohrabi, Y.-F. Liu, and W. Yu, "Phase transition analysis for covariance-based massive random access with massive MIMO," *IEEE Trans. Inf. Theory*, vol. 68, no. 3, pp. 1696–1715, Mar. 2022.
- 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*, vol. 67, no. 5, pp. 2925–2951, May 2021.
- U. K. Ganesan, E. Björnson, and E. G. Larsson, "Clustering-based activity detection algorithms for grant-free random access in cell-free massive MIMO," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7520–7530, Nov. 2021.

Contact Information

Many Thanks for Your Attention!

- Homepage: http://lsec.cc.ac.cn/~yafliu/
- E-mail: yafliu@lsec.cc.ac.cn