An Efficient Active Set Algorithm for Covariance Based Joint Data and Activity Detection for Massive Random Access with Massive MIMO

Ziyue Wang, Zhilin Chen, Ya-Feng Liu, Foad Sohrabi, and Wei Yu

University of Chinese Academy of Sciences
IEEE ICASSP, Toronto, Canada, June 6–11, 2021
Let us focus on a fundamental problem in mMTC:

- A large number of devices with sporadic activity
- Low latency random access scheme for massive users is required
- Non-orthogonal signature sequences are preferred to orthogonal sequences
- User activity detection (user identification) performed at base station (BS)
System Model

- Single cell with one BS equipped with $M$ antennas
- $N$ single-antenna devices, $K$ of which are active at a time
- Each active device wishes to transmit $J$ bits of data to the BS
- Each device $n$ has a unique signature sequence set $S_n = \{s_{n,1}, s_{n,2}, \ldots, s_{n,Q}\}$, where $s_{n,q} \in \mathbb{C}^{L \times 1}$, $1 \leq q \leq Q \triangleq 2^J$, and $L$ is the signature sequence length
- Channel $\sqrt{g_n}h_n \in \mathbb{C}^{M \times 1}$ of user $n$ includes both
  - Large-scale fading component $g_n \geq 0$
  - Rayleigh fading component $h_n \in \mathbb{C}^{M \times 1}$ following the i.i.d. complex Gaussian distribution
System Model

- Whether or not $s_{n,q}$ is transmitted is indicated as $\chi_{n,q} \in \{0, 1\}$, which satisfies $\sum_{q=1}^{Q} \chi_{n,q} \in \{0, 1\}$
  - $\sum_{q=1}^{Q} \chi_{n,q} = 1$ indicates that device $n$ is active
  - $\sum_{q=1}^{Q} \chi_{n,q} = 0$ indicates that device $n$ is inactive

- Define
  - $S_n = [s_{n,1}, \ldots, s_{n,Q}] \in \mathbb{C}^{L \times Q}$, and $S = [S_1, \ldots, S_N] \in \mathbb{C}^{L \times NQ}$;
  - $D_n = \sqrt{g_n} \text{diag}\{\chi_{n,1}, \ldots, \chi_{n,Q}\} \in \mathbb{C}^{Q \times Q}$, and $\Gamma^{1/2} = \text{diag}\{D_1, \ldots, D_N\} \in \mathbb{C}^{NQ \times NQ}$;
  - $H_n = [h_n, \ldots, h_n]^T \in \mathbb{C}^{Q \times M}$ for all $n$, and $H = [H_1^T, \ldots, H_N^T]^T \in \mathbb{C}^{NQ \times M}$. 
The received signal $Y \in \mathbb{C}^{L \times M}$ at the BS can be expressed as

$$Y = \sum_{n=1}^{N} \sum_{q=1}^{Q} \chi_{n,q} s_{n,q} \sqrt{g_{n}} h_{n}^{T} + W$$

$$= S \Gamma^{1/2} H + W. \quad (1)$$

where $W \in \mathbb{C}^{L \times M}$ is the effective i.i.d. Gaussian noise whose variance $\sigma_{w}^{2}$ is the background noise power normalized by the device transmit power.
The joint activity and data detection problem can be formulated as the maximum likelihood estimation (MLE) problem [Haghighatshoar-Jung-Caire '18]

\[
\begin{align*}
\min_{\gamma} & \quad \log |\mathbf{S}\Gamma\mathbf{S}^H + \sigma_w^2 \mathbf{I}| + \text{Tr} \left( (\mathbf{S}\Gamma\mathbf{S}^H + \sigma_w^2 \mathbf{I})^{-1} \hat{\Sigma} \right) \\
\text{s. t.} & \quad \gamma \geq 0.
\end{align*}
\]

The sample covariance matrix \( \hat{\Sigma} = \mathbf{Y}\mathbf{Y}^H / M \) is computed by averaging over different antennas.

Let \( \gamma \in \mathbb{C}^{NQ \times 1} \) denote the diagonal entries of \( \Gamma \), i.e., \( \gamma = [\gamma_1^T, \ldots, \gamma_N^T]^T \), where \( \gamma_n = [\gamma_{n,1}, \ldots, \gamma_{n,Q}]^T \in \mathbb{C}^{Q \times 1} \) with \( \gamma_{n,q} = g_n \chi_{n,q} \geq 0 \).
Problem Formulation and Analysis

- Let $f(\gamma)$ denote the objective function of problem (2). The gradient of $f(\gamma)$ with respect to $\gamma_{n,q}$ is

$$[\nabla f(\gamma)]_{n,q} = s_{n,q}^H \Sigma^{-1} s_{n,q} - s_{n,q}^H \hat{\Sigma} \Sigma^{-1} s_{n,q}.$$

- The first-order (necessary) optimality condition of problem (2) is

$$[\nabla f(\gamma)]_{n,q} \begin{cases} = 0, & \text{if } \gamma_{n,q} > 0; \\ \geq 0, & \text{if } \gamma_{n,q} = 0, \end{cases} \forall q, n, \tag{3}$$

- Let $\cdot_+$ denote the projection operator onto the nonnegative orthant. Then (3) is equivalent to

$$[\gamma - \nabla f(\gamma)]_+ - \gamma = 0.$$
Motivation

- Existing algorithms for solving problem (2):
  - Coordinate descent (CD) [Haghighatshoar-Jung-Caire ’18]: Iteratively update every single coordinate (it admits a closed-form solution for each update)
  - Expectation-maximization (EM) [Wipf-Rao ’07]
  - Sparse iterative covariance-based estimation (SPICE) [Yang-Li-Stoica-Xie ’18]
- None of the above algorithms take advantage of the sparsity of true solution
- This paper proposes the active set algorithm
  - It fully exploits the sparsity of its true solution
  - At each iteration, it first judiciously selects an active set
  - It solves the small dimensional subproblem defined over the active set
  - It is more effective than the CD algorithm (the state-of-the-art algorithm)
Active Set Algorithm

- To fully exploit the sparsity of the true solution of (2), the active set should
  - contain the indices of active sequences
  - have the smallest possible cardinality

- At the \( k \)-th iteration, the proposed selection strategy of the active set \( \mathcal{A}^k \) is
  \[
  \mathcal{A}^k = \left\{ (i, q) \mid \gamma_{i,q}^k > \omega_k \text{ or } [\nabla f(\gamma^k)]_{i,q} < -\nu_k \right\},
  \]

  where \( \omega_k, \nu_k > 0 \) and \( \omega_k \downarrow 0 \) and \( \nu_k \downarrow 0 \) (monotonically decrease and converge to zero).
Once the active set $A^k$ is selected, solve the following subproblem

\[
\begin{align*}
\min & \quad \hat{f}(\gamma_{A^k}) \tag{5a} \\
\text{s. t.} & \quad \gamma_{A^k} \geq 0, \tag{5b}
\end{align*}
\]

where $\gamma_{A^k}$ is the subvector of $\gamma$ indexed by $A^k$ and $\hat{f}(\gamma_{A^k})$ is $f(\gamma)$ defined over $\gamma_{A^k}$ with all the other variables fixed being zero.

If the set $A^k$ in (5) is properly chosen, the dimension of problem (5) is potentially much smaller than that of problem (2).

Apply the spectral PG algorithm [Birgin-Martínez-Raydan '00] to solve the subproblem in (5) until $\gamma_{A^k}^{k+1}$ satisfying

\[
\left\| \left[ \gamma_{A^k}^{k+1} - \nabla \hat{f}(\gamma_{A^k}^{k+1}) \right]_+ - \gamma_{A^k}^{k+1} \right\| < \varepsilon_k, \tag{6}
\]

where $\varepsilon_k > 0$ is the solution tolerance at the $k$-th iteration.
The pseudocodes of the proposed algorithm are given in Algorithm 1.

**Algorithm 1** Proposed active set PG algorithm for solving problem (2)

1. **Initialize:** $\gamma^0 = 0$, $k = 0$, $\{\omega_k, \nu_k, \varepsilon_k\}_{k \geq 0}$, and $\varepsilon > 0$;
2. **repeat**
3. Select the active set $A^k$ according to (4);
4. Apply the spectral PG algorithm to solve the subproblem (5) until (6) is satisfied;
5. Set $k \leftarrow k + 1$;
6. **until** $\|\gamma^k - \nabla f(\gamma^k) + \gamma^k\| < \varepsilon$
7. **Output:** $\gamma^k$
Theorem

For any given tolerance $\varepsilon > 0$, suppose that the parameters $\omega_k$ and $\nu_k$ in (4) satisfy $\omega_k \downarrow 0$ and $\nu_k \downarrow 0$ and the parameter $\epsilon_k$ in (6) satisfy $\lim_{k \to \infty} \epsilon_k < \varepsilon$, then the active set PG Algorithm 1 will terminate within a finite number of iterations.

- The convergence property is mainly because of
  - the activity set selection strategy in (4) (choices of parameters $\omega_k$ and $\nu_k$)
  - the convergence property of the spectral PG algorithm

- A not careful selection of the active set might lead to oscillation or divergence
Simulation Results

- The power spectrum density of the background noise is $-169$dBm/Hz over 10 MHz and the transmit power of each device is 25dBm.

- A single cell of radius 1000m, all devices are located in the cell edge, $g_n$’s are the same for all devices.

- All signature sequences from i.i.d. complex Gaussian distribution with zero mean and unit variance.

- Parameters setting: $M = 256$, $L = 150$, and $J = 1$ (and thus $Q = 2$), $K/N = 0.1$ (10% of the total devices are active).
Simulation Results

- Compare the proposed Algorithm 1 with
  - random CD algorithm: apply the random CD algorithm to solve problem (2)
  - Ideal CD algorithm: apply the CD algorithm to solve problem (2) defined over the indices of active sequences
  - Ideal PG algorithm: apply the PG algorithm to solve problem (2) defined over the indices of active sequences

- Parameters setting:

  \[
  \omega_k = 10^{-6-k}, \quad \varepsilon_k = \max \left\{ 10^{-k}, 0.8 \times 10^{-3} \right\},
  \]

  \[
  \nu_k = \min \left\{ 10^{4-k}, 0.5 \left\| \min_{n,q} \left\{ [\nabla f(\gamma^k)]_{n,q} \right\} \right\| \right\}.
  \]

- Average over 500 Monte-Carlo runs.
Simulation Results

Figure: Performance of the proposed active set PG algorithm.

- The average ratio of $|\mathcal{A}^k|/K$ is in the interval $[1.5, 2.5]$.
- Algorithm 1 will generally terminate within 4–7 iterations.
- The proposed active set selection strategy (4) is very efficient.
Simulation Results

Algorithm 1 significantly outperforms the random CD algorithm.

Algorithm 1 even achieves slightly better efficiency than the ideal CD algorithm.

Figure: Average CPU time comparison.
The main contribution of this paper is

- Propose a computationally efficient active set algorithm for activity detection
- Provide the convergence guarantee for the proposed algorithm
- Show the efficiency of the proposed algorithm by simulations


