Wideband Massive MIMO Channel Estimation via Sequential Atomic Norm Minimization

Stelios Stefanatos[1] Mahdi Barzegar Khalilsarai[2] Gerhard Wunder[1]

[1] Department of Mathematics and Computer Science
Freie Universität Berlin, Germany

[2] Department of Telecommunication Systems
Technische Universität Berlin, Germany

GlobalSIP 2018

Outline

Outline

Introduction

Freie Universität Berlin

Challenges of "Massive" Communications

- Extremely large number of UEs
- Short-length transmissions
- Extremely large signal space

Critical System Design Goal

Employ channel estimation procedures that

- provide reliable estimates
- are of low complexity
- require small training overhead

In this Work:

1. A low-complexity, ANM-based channel estimator for uplink wideband mMIMO is proposed
2. MSE performance characterized by tight lower bounds
3. Close to optimal for low-to-moderate number of propagation paths

Introduction

Freie Universität Berlin

Challenges of "Massive" Communications

- Extremely large number of UEs
- Short-length transmissions
- Extremely large signal space

Critical System Design Goal

Employ channel estimation procedures that

- provide reliable estimates
- are of low complexity
- require small training overhead

In this Work:

1. A low-complexity, ANM-based channel estimator for uplink wideband mMIMO is proposed
2. MSE performance characterized by tight lower bounds
3. Close to optimal for low-to-moderate number of propagation paths

System Model (1/2)

- Single cell, uplink
- ULA of $M \gg 1$ antennas at BS
- Single-antenna UE
- OFDM signaling with $N \gg 1$ subcarriers
- Link characterized by the unknown space-frequency transfer matrix $\mathbf{H} \in \mathbb{C}^{M \times N}$
- UE transmits pilot symbols over a set $\mathcal{N}_p \subseteq \{0, 1, \ldots, N-1\}$ of N_p subcarriers
- BS utilizes the observations from a set $\mathcal{M}_p \subseteq \{0, 1, \ldots, M-1\}$ of M_p antennas

Assumption

Sets $\mathcal{N}_p$, $\mathcal{M}_p$ are selected randomly and uniformly from $\{0, 1, \ldots, N-1\}$, $\{0, 1, \ldots, M-1\}$, respectively

- motivated by compressive sensing theory
- results in an robust and multiuser-fair design
- allows for tractable analysis

- N_p, M_p are design parameters to be specified

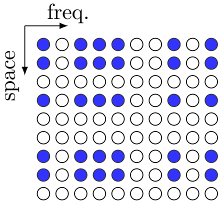
System Model (2/2)

Freie Universität Berlin

- Observed $M_p \times N_p$ signal at the BS (all-ones pilot symbols):

$$\mathbf{Y} = \mathbf{S}_{\mathcal{M}_p} \mathbf{H} \mathbf{S}_{\mathcal{N}_p}^T + \mathbf{Z}$$

$\mathbf{S}_{\mathcal{M}_p} \in \mathbb{R}^{M_p \times M}$, $\mathbf{S}_{\mathcal{N}_p} \in \mathbb{R}^{N_p \times M}$: downsampling matrices
$\mathbf{Z} \in \mathbb{C}^{M_p \times N_p}$: : AWGN of variance σ^2



Receiver Task

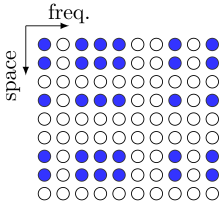
Obtain a low-complexity and accurate estimate of the MN elements of $\mathbf{H}$ given the $M_p N_p < MN$ observations in $\mathbf{Y}$

- Underdetermined linear system

Key concept

Exploit structural properties of the physical channel

System Model (2/2)



- Observed $M_p \times N_p$ signal at the BS (all-ones pilot symbols):

$$\mathbf{Y} = \mathbf{S}_{\mathcal{M}_p} \mathbf{H} \mathbf{S}_{\mathcal{N}_p}^T + \mathbf{Z}$$

$\mathbf{S}_{\mathcal{M}_p} \in \mathbb{R}^{M_p \times M}$, $\mathbf{S}_{\mathcal{N}_p} \in \mathbb{R}^{N_p \times M}$: downsampling matrices

$\mathbf{Z} \in \mathbb{C}^{M_p \times N_p}$: : AWGN of variance σ^2

Receiver Task

Obtain a low-complexity and accurate estimate of the MN elements of $\mathbf{H}$ given the $M_p N_p < MN$ observations in $\mathbf{Y}$

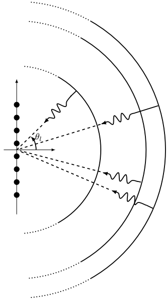
- Underdetermined linear system

Key concept

Exploit structural properties of the physical channel

$$H[n;m] = \sum_{l=0}^{L-1} c_l e^{-i2\pi m\theta_l} e^{-i2\pi n\tau_l}, n \in [N], m \in [M]$$

L	: number of paths
$c_l \in \mathbb{C}$	: gain of lth path
$\theta_l \in [0,1]$	: angle of arrival (AoA) of lth path (normalized)
$\tau_l \in [0,1]$	: delay of lth path (normalized)



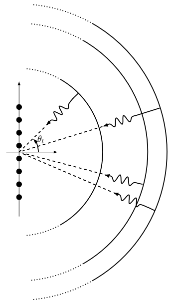
- Channel described by $3L \ll MN$ path parameters $\{(\rho_l, \theta_l, \tau_l)\}_{l=0}^{L-1}$ in the angle-delay domain
- Maximum Likelihood (ML) detection of path parameters:

$$\{(\hat{c}_l, \hat{\tau}_l, \hat{\theta}_l)\}_{l=0}^{L-1} = \underset{\{(c_l,\tau_l,\theta_l)\}_{l=0}^{L-1}}{\arg\min} \left\| \mathbf{Y} - \mathbf{S}_{\mathcal{M}_p} \mathbf{H} \mathbf{S}_{\mathcal{N}_p}^T \right\|^2$$

NP-hard problem $\Longrightarrow$ suboptimal solutions necessary

Parametric Wideband Massive MIMO Channel Model

$$H[n;m] = \sum_{l=0}^{L-1} c_l e^{-i2\pi m\theta_l} e^{-i2\pi n\tau_l}, n \in [N], m \in [M]$$



L	: number of paths
$c_l \in \mathbb{C}$	: gain of lth path
$\theta_l \in [0,1]$	: angle of arrival (AoA) of lth path (normalized)
$\tau_l \in [0,1]$	: delay of lth path (normalized)

- Channel described by $3L \ll MN$ path parameters $\{(\rho_l, \theta_l, \tau_l)\}_{l=0}^{L-1}$ in the angle-delay domain
- Maximum Likelihood (ML) detection of path parameters:

$$\{(\hat{c}_l, \hat{\tau}_l, \hat{\theta}_l)\}_{l=0}^{L-1} = \underset{\{(c_l,\tau_l,\theta_l)\}_{l=0}^{L-1}}{\arg\min} \left\| \mathbf{Y} - \mathbf{S}_{\mathcal{M}_p} \mathbf{H} \mathbf{S}_{\mathcal{N}_p}^T \right\|^2$$

NP-hard problem $\Longrightarrow$ suboptimal solutions necessary

Outline

Freie Universität Berlin

Channel Estimation via Atomic Norm Minimization

Freie Universität Berlin

- Define the *atom set* (manifold)

$$\mathcal{A} \triangleq \left\{ \mathbf{f}_M(\theta)\mathbf{f}_N^H(\tau) : (\theta,\tau) \in [0,1] \times [0,1] \right\}$$

  - $\mathbf{f}_M(\theta) \triangleq [1, e^{-i2\pi\theta}, \ldots, e^{-i2\pi\theta(M-1)}]^T$ and similarly for $\mathbf{f}_N(\tau)$

- Rationale for this set: $\mathbf{H} = \sum_{l=0}^{L-1} c_l \mathbf{f}_M(\theta_l)\mathbf{f}_N^H(\tau_l)$, i.e., $\mathbf{H} \in \text{span}(\mathcal{A})$

Definition (Atomic Norm)

The atomic norm of an arbitrary matrix $\mathbf{X} \in \mathbb{C}^{M \times N}$ w.r.t. $\mathcal{A}$ is

$$\|\mathbf{X}\|_{\mathcal{A}} \triangleq \inf_{\substack{c_l \in \mathbb{C}, \\ \theta_l, \tau_l \in [0,1]}} \left\{ \sum_l |c_l| \,\middle|\, \mathbf{X} = \sum_l c_l \mathbf{f}_M(\theta)\mathbf{f}_N^H(\tau) \right\}$$

- Extension of the standard ℓ_1-norm

Channel Estimation via Atomic Norm Minimization

$$\hat{\mathbf{H}} = \underset{\mathbf{X} \in \mathbb{C}^{M \times N}}{\text{argmin}} \left\{ \|\mathbf{X}\|_{\mathcal{A}} \,\middle|\, \|\mathbf{Y} - \mathbf{S}_{\mathcal{M}_p}\mathbf{X}\mathbf{S}_{\mathcal{N}_p}^T\| \leq \|\hat{\mathbf{Z}}\| \right\}$$

- Define the *atom set* (manifold)

$$\mathcal{A} \triangleq \left\{ \mathbf{f}_M(\theta)\mathbf{f}_N^H(\tau) : (\theta, \tau) \in [0,1] \times [0,1] \right\}$$

  - $\mathbf{f}_M(\theta) \triangleq [1, e^{-i2\pi\theta}, \ldots, e^{-i2\pi\theta(M-1)}]^T$ and similarly for $\mathbf{f}_N(\tau)$
- Rationale for this set: $\mathbf{H} = \sum_{l=0}^{L-1} c_l \mathbf{f}_M(\theta_l)\mathbf{f}_N^H(\tau_l)$, i.e., $\mathbf{H} \in \text{span}(\mathcal{A})$

Definition (Atomic Norm)

The atomic norm of an arbitrary matrix $\mathbf{X} \in \mathbb{C}^{M \times N}$ w.r.t. $\mathcal{A}$ is

$$\|\mathbf{X}\|_{\mathcal{A}} \triangleq \inf_{\substack{c_l \in \mathbb{C}, \\ \theta_l, \tau_l \in [0,1]}} \left\{ \sum_l |c_l| \,\middle|\, \mathbf{X} = \sum_l c_l \mathbf{f}_M(\theta)\mathbf{f}_N^H(\tau) \right\}$$

- Extension of the standard ℓ_1-norm

Channel Estimation via Atomic Norm Minimization

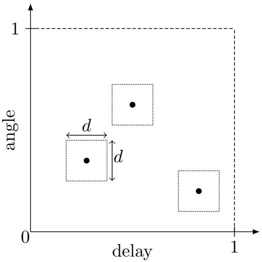
$$\hat{\mathbf{H}} = \underset{\mathbf{X} \in \mathbb{C}^{M \times N}}{\text{argmin}} \left\{ \|\mathbf{X}\|_{\mathcal{A}} \,\middle|\, \|\mathbf{Y} - \mathbf{S}_{\mathcal{M}_p}\mathbf{X}\mathbf{S}_{\mathcal{N}_p}^T\| \le \|\hat{\mathbf{Z}}\| \right\}$$

Freie Universität Berlin

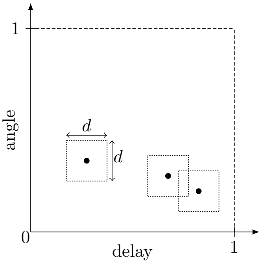
Theorem (Informal Statement)

Under (a) noiseless conditions and (b) sufficiently large N_p, M_p, perfect recovery of $\mathbf{H}$ can be achieved with high probability as long as channel paths are sufficiently separated in the delay-angle domain, i.e.,

$$\min_{l \neq l'} \max \{|\theta_l - \theta_{l'}|, |\tau_l - \tau_{l'}|\} > d \approx 1/\min\{M, N\}$$



(a) separable paths



(b) non-separable paths

Freie Universität Berlin

- Computation of $\|\cdot\|_{\mathcal{A}}$ can be formulated as an SDP problem, resulting in a convex program for obtaining $\hat{\mathbf{H}}$:

$$\left\{ \begin{array}{cc} \underset{\hat{\mathbf{H}}\in\mathbb{C}^{M\times N},\mathbf{u}\in\mathbb{C}^{MN},t>0}{\text{minimize}} & \frac{1}{2}\left(\text{tr}\left\{\mathbf{T}_{\text{2D}}(\mathbf{u})\right\}+t\right) \\ \text{subject to} & \begin{pmatrix} \mathbf{T}_{\text{2D}}(\mathbf{u}) & \text{vec}(\hat{\mathbf{H}}) \\ \text{vec}(\hat{\mathbf{H}})^H & t \end{pmatrix} \succeq \mathbf{0}, \\ & \|\mathbf{Y}-\mathbf{S}_{\mathcal{M}_p}\hat{\mathbf{H}}\mathbf{S}_{\mathcal{N}_p}^T\| \leq \|\hat{\mathbf{Z}}\| \end{array} \right\}$$

- $\mathbf{T}_{\text{2D}}(\mathbf{u}) \in \mathbb{C}^{MN\times MN}$: block Toeplitz matrix
- Angle-delay pairs of paths can be estimated from the Vandermonde Decomposition of $\mathbf{T}_{\text{2D}}(\mathbf{u})$
  - denoising gains when L is known

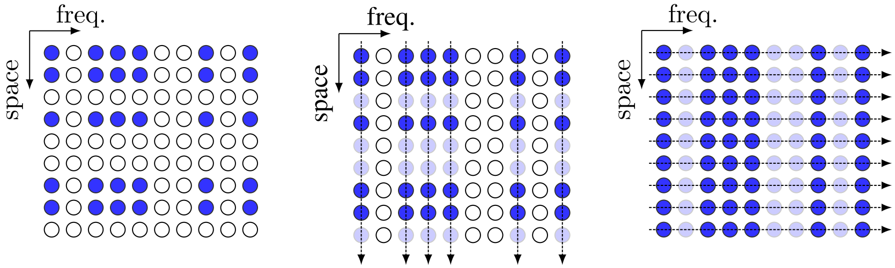
- Complexity of solution: $\mathcal{O}(MN)$
- Impractical when $M \gg 1$ and/or $N \gg 1 \Longrightarrow$ Low-complexity alternatives needed

Proposed Approach (1/2)

Freie Universität Berlin

Basic Idea

Decouple the spatial and frequency dimensions, treating them sequentially as Multiple Measurement Vectors (MMV) estimation problems and apply ANM-based estimation to each



(a) observation (b) interpolate over space (c) interpolate over frequency

Proposed Approach (2/2)

Freie Universität Berlin

1. Spatial dimension interpolation:
   - Rewrite the observation matrix as $\mathbf{Y} = \mathbf{S}_{\mathcal{M}_p} \underbrace{\mathbf{H}\mathbf{S}_{\mathcal{N}_p}^T}_{\triangleq \mathbf{H}_1} + \mathbf{Z}$
   - Note that $\mathbf{H}_1$ can be written as $\mathbf{H}_1 = \sum_{l=0}^{L-1} c_l \mathbf{f}_M(\theta_l)\mathbf{b}_{1,l}^H$, $\mathbf{b}_{1,l} \triangleq \mathbf{S}_{\mathcal{N}_p}\mathbf{f}_N(\tau_l)$
   - By ignoring the structure of $\{\mathbf{b}_{1,l}\}$ *and* noise, an estimate of $\mathbf{H}_1$ can be obtained as

$$\hat{\mathbf{H}}_1 = \underset{\mathbf{X}\in\mathbb{C}^{M\times N_p}}{\text{argmin}} \left\{ \|\mathbf{X}\|_{\mathcal{A}_{\text{MMV}_1}} \,\middle|\, \mathbf{Y} = \mathbf{S}_{\mathcal{M}_p}\mathbf{X} \right\},$$

   where $\mathcal{A}_{\text{MMV}_1} \triangleq \left\{ \mathbf{f}_M(\theta)\mathbf{b}_1^H, \theta \in [0,1], \mathbf{b}_1 \in \mathbb{C}^{M_p}, \|\mathbf{b}_1\|^2 = 1 \right\}$
   - Denoise the estimate exploiting that there are L paths
2. Frequency dimension interpolation:
   - Repeat the same approach treating now $\hat{\mathbf{H}}_1 \in \mathbb{C}^{M\times N_p}$ as the partial observations of the complete channel matrix with structure $\mathbf{H} = \sum_{l=0}^{L-1} c_l \mathbf{b}_{2,l}\mathbf{f}_N^H(\tau_l)$

SDP implementation with complexity order $\mathcal{O}(M+N) \ll \mathcal{O}(MN)$

- Exact characterization of ANM-based estimation performance extremely difficult
  - Resort to bounds

Theorem (Universal Bound)

The per-element MSE of any unbiased estimator of $\mathbf{H}$ is lower bounded as

$$\frac{1}{MN}\mathbb{E}(\|\hat{\mathbf{H}} - \mathbf{H}\|^2) \geq \frac{2L\sigma^2}{M_p N_p},$$

where the expectation is over the statistics of noise, $\mathcal{N}_p$, $\mathcal{M}_p$.

- bound is looser than the CRLB, i.e., non-achievable, in general
- Trade off N_p for $M_p \Longrightarrow N_p \geq L$ is not required in massive MIMO
- Scales as $\mathcal{O}(L)$
- Bound holds with no assumptions on path separability

MSE Bound for Proposed Algorithm

- The following result can serve as an approximation of the MSE performance

Theorem

Under the assumption that the error $\hat{\mathbf{H}}_1 - \mathbf{H}_1$ *consists of i.i.d., zero mean, Gaussian elements, the per-element MSE of any unbiased estimator of* $\mathbf{H}$ *from* $\hat{\mathbf{H}}_1$ *that treats the rows of* $\mathbf{H}$ *as MMV, is lower bounded as*

$$\frac{1}{MN}\mathbb{E}(\|\hat{\mathbf{H}} - \mathbf{H}\|^2) \geq \frac{L^2\sigma^2(1+2N_p)(1+2M)}{4MM_pN_p^2} \approx \frac{L^2\sigma^2}{M_pN_p} \text{ (for } N_p, M \gg 1\text{).}$$

- obtained under assumptions for the spatial-interpolation estimate that do not hold
- $L/2$ times greater than the universal bound
- Scales as $\mathcal{O}(L^2)$ instead of $\mathcal{O}(L)$

Outline

System Setup

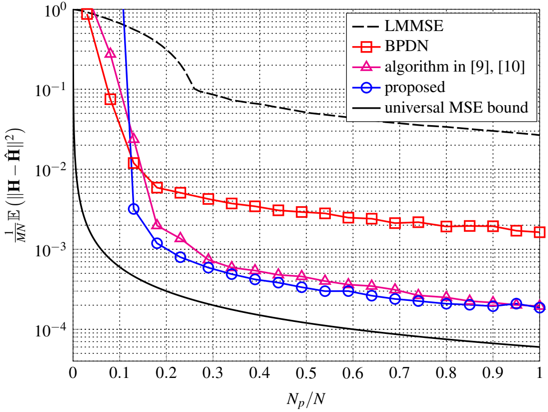
Freie Universität Berlin

- $M = 100$ ULA elements with full antenna observations ($M_p = M$)
- $N = 100$ OFDM subcarriers
  - 2D ANM-based estimation practically infeasible
- i.i.d. paths with $\theta_l \sim U[0,1], \tau_l \sim U[0,1/4], c_l \sim \mathcal{CN}(0,1/L)$
  - No restrictions on the path separability
- average SNR $= 1/\sigma^2 = 10$ dB
- Compare MSE of proposed algorithm with:
  1. naive LS with $N_p = N$ (MSE $= \sigma^2 = 10^{-1}$)
  2. LMMSE interpolator
  3. conventional (oversampled) ℓ_1-norm minimization (BPDN)
  4. $\mathcal{O}(N+M)$-complexity ANM-based approach (with path separability)[1,2]
  5. universal bound

[1] Z. Tian, Z. Zhang, and Y. Wang, "Low-complexity optimization for two-dimensional DoA estimation via decoupled ANM," *ICASSP 2017*

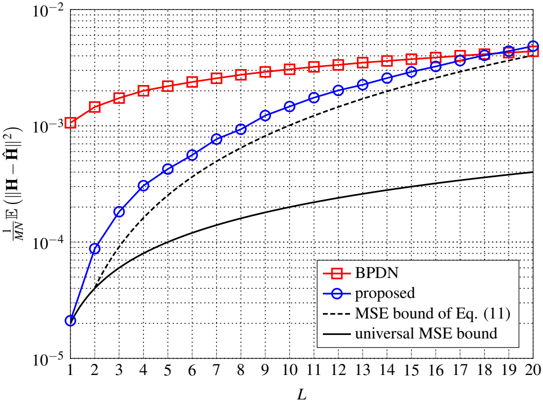
[2] J.-F. Cai, W. Xu, and Y. Yang, "Large scale 2D spectral compressed sensing in continuous domain," *ICASSP 2017*

MSE dependence on number of pilot subcarriers



- Results averaging over $\mathbf{H}$, $\mathcal{N}_p$, $\mathbf{Z}$
- $L = 3$ (very sparse channel)
- proposed performs very close to optimal and outperforms other approaches
- massive MIMO offers potential for (huge) denoising and/or training overhead gains

MSE dependence on number of paths



- $N_p = N$ (full observations)
- analysis closely follows MSE of proposed algorithm
- MSE scaling as $\mathcal{O}(L^2)$ eventually results in worse performance than BPDN
- For low-to-moderate L, proposed approach provides significant better performance

Outline

Freie Universität Berlin

Conclusion and Future Work

Freie Universität Berlin

- An ANM-based algorithm for wideband massive MIMO channel estimator was proposed
- Performs close to optimal for low-to-moderate number of paths w/o any assumptions on path separability
- Possible extensions:
  - time-varying channels, mutli-antenna UEs
  - multi-user, multi-cell setting