Active Uplink Sensing Beamformer Design via Bayesian Cramér-Rao Bound Dual Optimization

Nadim Ghaddar and Wei Yu
Department of Electrical and Computer Engineering
University of Toronto, Toronto, Canada
nadim.ghaddar@utoronto.ca, weiyu@ece.utoronto.ca

***Abstract*—This paper presents a novel optimization framework for solving active sensing problems in wireless communications, in which a base station equipped with massive multiple-input multiple-output (MIMO) and a limited number of radio-frequency chains aims to estimate the channel parameters of a sensing target. Specifically, the receive beamforming matrix at the BS is designed sequentially through optimizing the Bayesian Cramér-Rao bound (B-CRB) metric at each sensing stage, while satisfying a rank constraint and that the receive beamformers must be implementable by analog phase shifters. The proposed approach tackles this B-CRB minimization problem in the Lagrangian dual domain. This dual optimization approach has the advantage of reducing the dimension of the search space from the number of antenna elements to the number of channel parameters, which is typically much smaller for sparse mmWave channels. We propose efficient numerical methods for obtaining the primal solution from the dual and subsequentially setting the phase shifts in each active sensing stage based on this approach. Finally, we demonstrate the benefits of the proposed approach as compared to existing beamforming strategies.**

I. Introduction

The advent of millimeter-wave (mmWave) technology has enabled the use of large-scale antenna arrays in wireless communication systems, thereby addressing the ever-increasing demand for high data rates through highly directional beamforming. Nonetheless, constructing such directional beamformers requires accurate estimates of the high-dimensional wireless channels, which is particularly challenging in practical systems, especially when the number of radio-frequency (RF) chains is smaller than the number of antenna elements. In this paper, we consider the uplink directional beam alignment problem for multiple-input multiple-output (MIMO) systems that are RF-chain-limited. In such systems, it is desirable that the beamformers can be realized using simple analog components such as analog phase shifters [1], [2], which typically adds even more complexity to the design problem.

In particular, we consider the setting in which a base station (BS) equipped with massive MIMO and a limited number of RF chains aims to acquire the channel in an uplink mmWave environment, specifically the angles-of-arrival (AoAs) of (potentially) multiple paths. Due to the constraint on the number of RF chains, the BS can only make a low-dimensional observation of the received pilot signals through adjusting the phase shifts of a limited number of analog beamformers (which is equal to the number of RF chains), but can do so over multiple sensing stages in an adaptive manner.

Previous works have shown that adaptive sensing strategies for AoA estimation can achieve higher estimation accuracy with a much smaller number of pilot symbols as compared to their non-adaptive counterparts [3]–[7]. The majority of these adaptive beam alignment methods select the sensing vectors from a pre-designed beamforming codebook [3], [4], e.g., based on bisection search over the desired range of AoAs, but the performances of such methods are limited by the quality of the codebook, which is not easy to design. On the other hand, deep-learning-based solutions for active sensing have also been proposed in the literature, e.g. [5]–[7], but these methods require training a new model for each specific channel model, and their generalizability to varying channel conditions or varying number of sensing stages is challenging.

In this paper, we provide an analytical method for beamformer design in active sensing, which does not utilize a pre-designed beamforming codebook. Specifically, we explicitly track the posterior distribution of the AoAs and design the sensing beamforming vectors of the next stage by optimizing a Bayesian Cramér–Rao bound (B-CRB) metric of the AoAs, which is a lower bound on the mean-squared error (MSE) of any unbiased estimator that utilizes the received symbols up to that stage [8]–[12]. The B-CRB minimization problem is not easy to solve numerically when there are multiple paths and when the BS is equipped with a large number of antennas. In this paper, we solve a relaxed version of this problem in the Lagrangian dual domain. We show that the dual problem simplifies to that of maximizing the beamforming gains in some particular directions of interest, which can be found analytically. The phase shifters can then be tuned based on the phases of the resulting solution. This reduces the dimension of the search space from the number of antenna elements to one that depends only on the number of parameters to be estimated, which is typically much smaller in sparse mmWave channels. The technique used is the uplink counterpart of [12], which uses Lagrangian duality to design transmit waveforms in a downlink integrated sensing and communication (ISAC) system.

II. System Model

A. Active Sensing Model and Performance Metrics

Consider the uplink of a MIMO system, in which a BS equipped with N antennas and M RF chains ($M < N$) serves a single-antenna user, as illustrated in Fig. 1. In

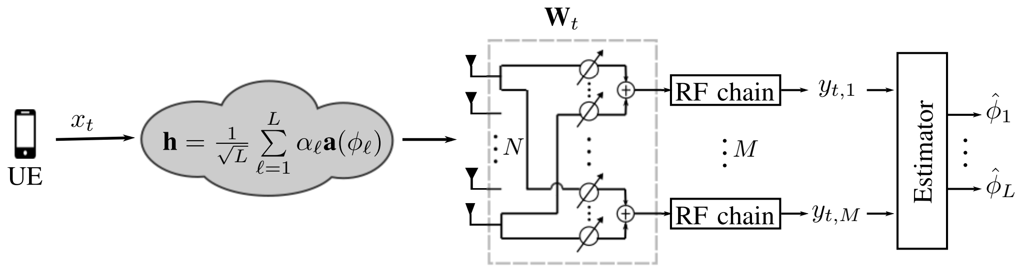


Fig. 1: Block diagram of a MIMO system with analog beamforming at the BS.

order to estimate the channel from the user to the BS, the user sends T pilot symbols $\{x_t\}, t = 1, \ldots, T$, with power P, i.e., $x_t = \sqrt{P}, \forall t$. Due to the limited number of RF chains, the BS can only observe the pilot symbols through a lower-dimensional analog combiner (i.e., sensing matrix) $\mathbf{W}_t \in \mathbb{C}^{N \times M}$. In this case, the received signal at time t, $1 \le t \le T$, can be expressed as

$$\mathbf{y}_t = \sqrt{P}\mathbf{W}_t^H \mathbf{h} + \mathbf{W}_t^H \mathbf{z}_t, \tag{1}$$

where:

- $\mathbf{h} = \frac{1}{\sqrt{L}} \sum_{\ell=1}^{L} \alpha_\ell \mathbf{a}(\phi_\ell)$ is a vector of channel gains between the user and the BS, with
  - L being the number of reflected paths,
  - $\alpha_\ell \sim \mathcal{CN}(0, 1)$ being the complex fading coefficient corresponding to ℓth path,
  - ϕ_ℓ being the angle-of-arrival (AoA) of the received signal from the ℓth path, and is assumed to be uniformly distributed over a range of interest $[\phi_{\min}, \phi_{\max}]$,
  - $\mathbf{a}(\phi_\ell)$ being the array response vector of a uniform linear array with half-wavelength antenna spacing, i.e.,

$$\mathbf{a}(\phi_\ell) = \begin{bmatrix} 1 & e^{j\pi \sin\phi_\ell} & \cdots & e^{j(N-1)\pi \sin\phi_\ell} \end{bmatrix}^T, \tag{2}$$

- $\mathbf{z}_t \in \mathbb{C}^N$ is a vector of i.i.d. zero-mean circularly-symmetric complex Gaussians with unit variance.

Since the sensing vectors in an RF-chain-limited system are typically implemented using a network of phase shifters, the elements of $\mathbf{W}_t$ should satisfy a constant modulus constraint, i.e., we should have $|[\mathbf{W}_t]_{i,j}| = \frac{1}{\sqrt{N}}$ for each i, j, t.

Here, we consider an active sensing strategy in which $\mathbf{W}_t$ can be designed in a sequential adaptive manner as a function of previously received symbols, i.e.,

$$\mathbf{W}_t = g_t(\mathbf{y}_{1:t-1}, \mathbf{W}_{1:t-1}) \tag{3}$$

for some sensing strategy $g_t : \mathbb{C}^{M(t-1)} \times \mathbb{C}^{N \times M(t-1)} \to \mathbb{C}^{N \times M}$ that needs to satisfy a constant-modulus constraint.

After observing T symbols $\mathbf{y}_{1:T} = (\mathbf{y}_1, \ldots, \mathbf{y}_T)$, the BS computes $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_L)$ as a function of all received measurements and sensing matrices,

$$\hat{\boldsymbol{\phi}} = f(\mathbf{y}_{1:T}, \mathbf{W}_{1:T}), \tag{4}$$

for some estimation scheme $f : \mathbb{C}^{MT} \times \mathbb{C}^{N \times MT} \to \mathbb{R}^L$.

This active beamforming design problem is motivated by a wide range of applications, including localization [13] and downlink beamforming for time-division duplex (TDD) systems [14]. In this paper, we consider the setting in which the channel fading coefficients $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_L)$ are regarded as nuisance parameters, i.e., the goal is only to estimate the AoAs.

The performance of the estimator is measured by the MSE, defined as

$$\text{MSE} = \frac{1}{L}\,\mathbb{E}\left[\|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}\|^2\right], \tag{5}$$

where the expectation is over the distribution of all stochastic parameters of the model, i.e., $\boldsymbol{\phi}$, $\boldsymbol{\alpha}$, and $\mathbf{z}_{1:T}$.

B. Problem Formulation

Based on the above performance metric, the active sensing problem for AoA estimation can be formulated as:

$$\underset{\{g_t(\cdot,\cdot)\}_{t=1}^T, f(\cdot,\cdot)}{\text{minimize}} \quad \mathbb{E}\left[\|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}\|^2\right] \tag{6a}$$

$$\text{subject to} \quad \mathbf{W}_t = g_t(\mathbf{y}_{1:t-1}, \mathbf{W}_{1:t-1}),\ \forall\, t = 1, \ldots, T, \tag{6b}$$

$$\hat{\boldsymbol{\phi}} = f(\mathbf{y}_{1:T}, \mathbf{W}_{1:T}). \tag{6c}$$

Solving the optimization problem (6) jointly over the functions $\{g_t(\cdot,\cdot)\}_{t=1}^T$ and $f(\cdot,\cdot)$ is challenging.

To make the problem more tractable, this paper considers a Bayesian formulation in which a prior distribution of $\boldsymbol{\phi}$ is assumed at the beginning, and then the prior is updated in each subsequent sensing stage. We adopt sensing strategies $\{g_t^{\text{BCRB}}(\cdot,\cdot)\}_{t=1}^T$ that minimize the Bayesian Cramér–Rao bound (B-CRB) as a function of the prior at each sensing stage t, then finally a minimum mean-squared error (MMSE) estimator $f^{\text{MMSE}}(\cdot,\cdot)$ to estimate the AoAs.

Unlike the classical CRB, which depends on the unknown parameters $\boldsymbol{\phi}$ [15], [16], the B-CRB provides a lower bound on the MSE averaged over the prior distribution of $\boldsymbol{\phi}$ [17]. Specifically, when applied to the active sensing problem, we have that after observing the first $t-1$ measurements, any unbiased estimator $\hat{\boldsymbol{\phi}}$ must satisfy

$$\mathbb{E}\left[(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})^T \,\middle|\, \mathbf{Y}_{1:t-1}\right] \succeq \mathbf{J}_t^{-1}(\mathbf{W}_t), \tag{7}$$

where $\succeq$ denotes inequality with respect to the positive semidefinite (PSD) cone, $\mathbf{J}_t(\mathbf{W}_t) \in \mathbb{C}^{L\times L}$ is the Bayesian Fisher information matrix (B-FIM) computed over the prior distribution of $\boldsymbol{\phi}$, and the expectation is taken over the joint distribution of $(\mathbf{Y}_t, \boldsymbol{\phi})$ given the measurements $\mathbf{Y}_{1:t-1}$ observed so far. This gives a lower bound on the MSE of estimating $\boldsymbol{\phi}$ in the t-th stage as

$$\text{MSE}(t) \geq \frac{1}{L}\,\mathbb{E}\left[\text{tr}\left(\mathbf{J}_t^{-1}(\mathbf{W}_t)\right)\right], \tag{8}$$

where the expectation on the right-hand side is over the distribution of $\mathbf{Y}_{1:t-1}$.

This paper proposes to minimize a lower bound on the MSE in each sensing stage given $\mathbf{Y}_{1:t-1}$. Specifically, we design the sensing matrix in the t-th stage by solving the following optimization problem:

$$\underset{\mathbf{W}_t}{\text{minimize}} \qquad \text{tr}\left(\mathbf{J}_t^{-1}(\mathbf{W}_t)\right). \tag{9}$$

In this way, the sensing matrix $\mathbf{W}_t$ is adapted based on the previous received symbols. At the end of the t-th stage, we compute the posterior distribution of $\boldsymbol{\phi}$ and use it as the prior distribution for the design of $\mathbf{W}_{t+1}$ in the next stage.

For ease of implementation, the analog beamformers $\mathbf{W}_t$ typically need to have constant modulus, so it can be implemented using analog phase shifters. Because of this, we need to impose an additional constraint

$$|[\mathbf{W}_t]_{i,j}| = \frac{1}{\sqrt{N}} \qquad \forall\, i,j. \tag{10}$$

This constraint is not easy to deal with analytically. We begin the exposition without this constraint, then subsequently obtain a constant-modulus solution afterwards.

III. Proposed Active Sensing Approach

A. Derivation of B-FIM and Problem Reformulation

The B-CRB has been used in prior work for downlink sensing beamforming optimization, e.g., [8], [12]. Following similar derivations as in [8, Appendices A and B], the entries of the B-FIM for the uplink channel model (1) can be expressed as

$$\begin{aligned}[\mathbf{J}_t(\mathbf{W}_t)]_{i,j} &= -\,\mathbb{E}\left[\frac{\partial^2 \log f(\mathbf{y}_t, \boldsymbol{\phi}\,|\,\boldsymbol{y}_{1:t-1})}{\partial\phi_i\partial\phi_j}\Big|\mathbf{Y}_{1:t-1}\right] \\ &\triangleq \left[\mathbf{J}_t^{(D)}(\mathbf{W}_t)\right]_{i,j} + \left[\mathbf{J}_{t-1}^{(P)}\right]_{i,j},\end{aligned} \tag{11}$$

where $\mathbf{J}_t^{(D)}(\mathbf{W}_t)$ is the part of the B-FIM corresponding to the data measurement model,

$$\begin{aligned}\left[\mathbf{J}_t^{(D)}(\mathbf{W}_t)\right]_{i,j} &= -\,\mathbb{E}\left[\frac{\partial^2 \log f(\mathbf{y}_t\,|\,\boldsymbol{\phi}, \mathbf{y}_{1:t-1})}{\partial\phi_i\partial\phi_j}\Big|\mathbf{Y}_{1:t-1}\right] \\ &= 2P\Re\left\{\text{tr}\left(\mathbb{E}\left[\dot{\mathbf{h}}_i(\boldsymbol{\phi})\dot{\mathbf{h}}_j^H(\boldsymbol{\phi})\big|\mathbf{Y}_{1:t-1}\right]\mathbf{R}_t\right)\right\},\end{aligned} \tag{12}$$

and $\mathbf{J}_{t-1}^{(P)}$ is the part of the B-FIM corresponding to the posterior distribution,

$$\begin{aligned}\left[\mathbf{J}_{t-1}^{(P)}\right]_{i,j} &= -\,\mathbb{E}\left[\frac{\partial^2 \log f(\boldsymbol{\phi}\,|\,\mathbf{y}_{1:t-1})}{\partial\phi_i\partial\phi_j}\Big|\mathbf{Y}_{1:t-1}\right] \\ &= 2P\textstyle\sum_{\tau=1}^{t-1}\Re\left\{\text{tr}\left(\mathbb{E}\left[\dot{\mathbf{h}}_i(\boldsymbol{\phi})\dot{\mathbf{h}}_j^H(\boldsymbol{\phi})\big|\mathbf{Y}_{1:t-1}\right]\mathbf{R}_\tau\right)\right\} \\ &\quad + 2P\textstyle\sum_{\tau=1}^{t-1}\Re\left\{\text{tr}\left(\mathbb{E}\left[\mathbf{h}(\boldsymbol{\phi})\ddot{\mathbf{h}}_{ij}^H(\boldsymbol{\phi})\big|\mathbf{Y}_{1:t-1}\right]\mathbf{R}_\tau\right)\right\} \\ &\quad - 2\sqrt{P}\textstyle\sum_{\tau=1}^{t-1}\Re\left\{\mathbf{y}_\tau^H(\mathbf{W}_\tau^H\mathbf{W}_\tau)^{-1}\mathbf{W}_\tau^H\,\mathbb{E}\left[\ddot{\mathbf{h}}_{ij}(\boldsymbol{\phi})\big|\mathbf{Y}_{1:t-1}\right]\right\} \\ &\quad - \mathbb{E}\left[\frac{\partial^2\log f(\boldsymbol{\phi})}{\partial\phi_i\partial\phi_j}\Big|\mathbf{Y}_{1:t-1}\right],\end{aligned} \tag{13}$$

where $\mathbf{R}_\tau = \mathbf{W}_\tau(\mathbf{W}_\tau^H\mathbf{W}_\tau)^{-1}\mathbf{W}_\tau^H$, for $\tau = 1,\ldots,t$, is the orthogonal projection matrix on the range space of $\mathbf{W}_\tau$, $\dot{\mathbf{h}}_i(\boldsymbol{\phi}) = \frac{\partial\mathbf{h}(\boldsymbol{\phi})}{\partial\phi_i}$ and $\ddot{\mathbf{h}}_{ij}(\boldsymbol{\phi}) = \frac{\partial^2\mathbf{h}(\boldsymbol{\phi})}{\partial\phi_i\partial\phi_j}$. Note that $\mathbf{J}_{t-1}^{(P)}$ does not depend on $\mathbf{W}_t$, and hence, can be viewed as a constant matrix in the optimization problem (9).

We now make some observations about the optimization problem (9). First, the B-FIM depends on $\mathbf{W}_t$ only through $\mathbf{R}_t$; in fact, when expressed as $\mathbf{J}_t(\mathbf{R}_t)$, the B-FIM is an affine function of $\mathbf{R}_t$. So, without the additional constraint (10), the optimization problem (9) can be rewritten as

$$\begin{aligned}\underset{\mathbf{R}_t}{\text{minimize}} \quad & \text{tr}\left(\mathbf{J}_t^{-1}(\mathbf{R}_t)\right) && \text{(14a)}\\ \text{subject to} \quad & \mathbf{R}_t \text{ is an orthogonal projection matrix,} && \text{(14b)}\\ & \text{rank}(\mathbf{R}_t) = M. && \text{(14c)}\end{aligned}$$

Second, the minimization of the trace of $\mathbf{J}_t^{-1}(\mathbf{R}_t)$ in (14) has a convenient reformulation using Schur complement [8], [16]:

$$\begin{aligned}\underset{\mathbf{R}_t, d_1,\ldots,d_L}{\text{minimize}} \quad & \sum_{\ell=1}^{L} d_\ell && \text{(15a)}\\ \text{subject to} \quad & \begin{bmatrix}\mathbf{J}_t(\mathbf{R}_t) & \mathbf{e}_\ell\\ \mathbf{e}_\ell^T & d_\ell\end{bmatrix} \succeq 0, \quad \forall\ell, && \text{(15b)}\\ & \mathbf{R}_t \text{ is an orthogonal projection matrix,} && \text{(15c)}\\ & \text{rank}(\mathbf{R}_t) = M, && \text{(15d)}\end{aligned}$$

where $\mathbf{e}_\ell$ denotes the ℓ-th column of the $L\times L$ identity matrix. This reformulation of the problem is crucial for obtaining an efficient numerical algorithm for solving (9). Note that, compared to the downlink problem considered in [12], the uplink B-CRB optimization problem (15) has an additional (non-convex) orthogonal projection constraint (i.e., (15c)), whereas the corresponding downlink problem has a power constraint (which is linear in the optimization variable $\mathbf{R}_t$). The orthogonal projection constraint follows from the fact that the sensing matrix $\mathbf{W}_t$ in the uplink model (1) affects the noise covariance matrix, which is not the case for downlink.

B. B-CRB Optimization via Duality

Our method for solving (15) mirrors the solution to the downlink counterpart in [12], while accounting for the additional orthogonal projection constraint. In particular, we show that (15) can be solved efficiently in the Lagrangian dual domain, despite the non-convex constraints. The optimal solution in the dual domain also gives rise to an optimal primal solution for (9). Toward this end, we proceed as in [12] and dualize (15) with respect to the constraint (15b). Let $\tilde{\boldsymbol{\Lambda}}_1,\ldots,\tilde{\boldsymbol{\Lambda}}_L$ denote the dual variables defined as

$$\tilde{\boldsymbol{\Lambda}}_\ell = \begin{bmatrix}\boldsymbol{\Gamma}_\ell & -\boldsymbol{\lambda}_\ell\\ -\boldsymbol{\lambda}_\ell^T & \nu_\ell\end{bmatrix}. \tag{16}$$

Then, the dual problem can be written as

$$\max_{\tilde{\mathbf{\Lambda}}_\ell \succeq 0\,\forall \ell} \min_{\substack{\mathbf{R}_t \in \mathcal{R},\\ d_1,\ldots,d_L}} \sum_{\ell=1}^{L} \Big(d_\ell(1-\nu_\ell) + 2\mathbf{e}_\ell^T \boldsymbol{\lambda}_\ell - \mathrm{tr}(\mathbf{\Gamma}_\ell \mathbf{J}_t(\mathbf{R}_t)) \Big), \tag{17}$$

where $\mathcal{R}$ denotes the constraints (15c) and (15d). Optimizing over $d_1,\ldots,d_L$, we conclude that $\nu_\ell^* = 1$ for all ℓ. Then, the dual problem becomes

$$\max_{\mathbf{\Gamma}_\ell \succeq \boldsymbol{\lambda}_\ell \boldsymbol{\lambda}_\ell^H \,\forall \ell} \min_{\mathbf{R}_t \in \mathcal{R}} \sum_{\ell=1}^{L} \Big(2\mathbf{e}_\ell^T \boldsymbol{\lambda}_\ell - \mathrm{tr}(\mathbf{\Gamma}_\ell \mathbf{J}_t(\mathbf{R}_t)) \Big), \tag{18}$$

where the condition $\mathbf{\Gamma}_\ell \succeq \boldsymbol{\lambda}_\ell \boldsymbol{\lambda}_\ell^H$ follows from the Schur complement of the constraint $\tilde{\mathbf{\Lambda}}_\ell \succeq 0$. Recall that

$$\mathbf{J}_t(\mathbf{R}_t) = \mathbf{J}_t^{(D)}(\mathbf{R}_t) + \mathbf{J}_{t-1}^{(P)}, \tag{19}$$

hence, solving the inner minimization in (18) is equivalent to solving the following problem

$$\max_{\mathbf{R}_t \in \mathcal{R}} \quad \mathrm{tr}\left(\mathbf{A} \mathbf{J}_t^{(D)}(\mathbf{R}_t) \right), \tag{20}$$

where $\mathbf{A} \triangleq \sum_{\ell=1}^{L} \mathbf{\Gamma}_\ell \succeq 0$. Note that, for any $\mathbf{A} \succeq 0$, we have

$$\begin{aligned}
&\mathrm{tr}\left(\mathbf{A} \mathbf{J}_t^{(D)}(\mathbf{R}_t) \right) = \sum_{i,j=1}^{L} [\mathbf{A}]_{i,j} \left[\mathbf{J}_t^{(D)}(\mathbf{W}_t) \right]_{i,j} \\
&= 2P \Re \left\{ \mathrm{tr}\left(\sum_{i,j=1}^{L} [\mathbf{A}]_{i,j}\, \mathbb{E}\left[\dot{\mathbf{h}}_i(\boldsymbol{\phi}) \dot{\mathbf{h}}_j^H(\boldsymbol{\phi}) \middle| \mathbf{Y}_{1:t-1} \right] \mathbf{R}_t \right) \right\} \\
&= 2P \mathrm{tr}\left(\mathbf{G}_{\mathbf{A}} \mathbf{R}_t \right),
\end{aligned} \tag{21}$$

where

$$\mathbf{G}_{\mathbf{A}} \triangleq \mathbb{E}\left[\dot{\mathbf{H}}(\boldsymbol{\phi}) \mathbf{A} \dot{\mathbf{H}}^H(\boldsymbol{\phi}) \,\middle|\, \mathbf{Y}_{1:t-1} \right], \tag{22}$$

and $\dot{\mathbf{H}}(\boldsymbol{\phi}) = \begin{bmatrix} \dot{\mathbf{h}}_1(\boldsymbol{\phi}) & \cdots & \dot{\mathbf{h}}_L(\boldsymbol{\phi}) \end{bmatrix}$. Since $\mathbf{A} \succeq 0$, we have $\mathbf{G}_{\mathbf{A}} \succeq 0$. Thus, the inner minimization in (18) is equivalent to

$$\begin{aligned}
&\underset{\mathbf{R}_t}{\text{maximize}} && \mathrm{tr}\left(\mathbf{G}_{\mathbf{A}} \mathbf{R}_t \right) && \text{(23a)} \\
&\text{subject to} && \mathbf{R}_t \text{ is an orthogonal projection matrix,} && \text{(23b)} \\
& && \mathrm{rank}(\mathbf{R}_t) = M. && \text{(23c)}
\end{aligned}$$

The key observation is that (23) has the following analytic solution based on the eigenvalue decomposition of $\mathbf{G}_{\mathbf{A}}$:

$$\mathbf{R}_t^* = \tilde{\mathbf{W}}_t^* (\tilde{\mathbf{W}}_t^*)^H, \tag{24}$$

where $\tilde{\mathbf{W}}_t^* = \begin{bmatrix} \tilde{\mathbf{w}}_1^* & \ldots & \tilde{\mathbf{w}}_M^* \end{bmatrix}$, and $\tilde{\mathbf{w}}_j^*$ is the eigenvector of $\mathbf{G}_{\mathbf{A}}$ corresponding to the jth largest eigenvalue. Note that since $\mathbf{G}_{\mathbf{A}}$ is a PSD matrix, the eigenvectors can be chosen to be mutually orthogonal, in which case $\mathbf{R}_t^*$ is an orthogonal projection matrix. The dual problem (18) can then be written as

$$\underset{\mathbf{\Gamma}_\ell \succeq \boldsymbol{\lambda}_\ell \boldsymbol{\lambda}_\ell^H \,\forall \ell}{\text{maximize}} \sum_{\ell=1}^{L} \left(2\mathbf{e}_\ell^T \boldsymbol{\lambda}_\ell - \mathrm{tr}(\mathbf{\Gamma}_\ell \mathbf{J}_{t-1}^{(P)}) \right) - 2P \sum_{m=1}^{M} \mu_m\left(\mathbf{G}_{\mathbf{A}} \right), \tag{25}$$

where $\mu_m(\cdot)$ denotes the mth largest eigenvalue of a matrix.

Observe that, for any $\mathbf{\Gamma}_\ell \succeq \boldsymbol{\lambda}_\ell \boldsymbol{\lambda}_\ell^H$, we have that $\mathrm{tr}(\mathbf{\Gamma}_\ell \mathbf{J}_{t-1}^{(P)}) \geq \mathrm{tr}(\boldsymbol{\lambda}_\ell \boldsymbol{\lambda}_\ell^H \mathbf{J}_{t-1}^{(P)})$ (since $\mathbf{J}_{t-1}^{(P)}$ is a PSD matrix). Also, we have $\mathbf{A} = \sum_{\ell=1}^{L} \mathbf{\Gamma}_\ell \succeq \sum_{\ell=1}^{L} \boldsymbol{\lambda}_\ell \boldsymbol{\lambda}_\ell^H \triangleq \mathbf{B}$, and hence,

$$\textstyle\sum_{m=1}^{M} \mu_m\left(\mathbf{G}_{\mathbf{A}} \right) \geq \sum_{m=1}^{M} \mu_m(\mathbf{G}_{\mathbf{B}}). \tag{26}$$

It follows that the optimal solution of (25) should satisfy $\mathbf{A} = \mathbf{B}$, so that $\mathbf{\Gamma}_\ell^* = \boldsymbol{\lambda}_\ell^* (\boldsymbol{\lambda}_\ell^*)^H$ for each ℓ. Therefore, by denoting $\mathbf{\Lambda} \triangleq \begin{bmatrix} \boldsymbol{\lambda}_1 & \cdots & \boldsymbol{\lambda}_L \end{bmatrix} \in \mathbb{C}^{L \times L}$, we can write $\mathbf{B} = \mathbf{\Lambda}\mathbf{\Lambda}^H$, and the optimization problem (25) becomes

$$\underset{\mathbf{\Lambda}}{\text{maximize}} \quad 2\mathrm{tr}(\mathbf{\Lambda}) - \mathrm{tr}\left(\mathbf{\Lambda}^H \mathbf{J}_{t-1}^{(P)} \mathbf{\Lambda} \right) - 2P \sum_{m=1}^{M} \mu_m\left(\mathbf{G}_{\mathbf{\Lambda}\mathbf{\Lambda}^H} \right). \tag{27}$$

The objective function in (27) is concave. The optimal $\mathbf{\Lambda}^*$ can be found efficiently using convex optimization methods.

The main conclusion here is that despite the non-convex orthogonal projection and rank constraints in the primal problem (15), under mild conditions, a global optimum solution of (15) can be obtained from the optimal dual solution $\mathbf{\Lambda}^*$. This is due to the fact that the optimization of the Lagrangian, which boils down to the optimization problem (23), can be solved to global optimality, despite the non-convex constraints. The primal solution corresponding to the optimal $\mathbf{\Lambda}^*$ can be obtained by taking the eigenvectors of $\mathbf{G}_{\mathbf{\Lambda}^*(\mathbf{\Lambda}^*)^H}$ corresponding to the M largest eigenvalues, calling them $\{\tilde{\mathbf{w}}_j^*\}_{j=1}^{M}$, and stacking them as

$$\tilde{\mathbf{W}}_t^* = \begin{bmatrix} \tilde{\mathbf{w}}_1^* & \ldots & \tilde{\mathbf{w}}_M^* \end{bmatrix}. \tag{28}$$

The global optimality of the above solution can be established whenever the dual variable $\mathbf{\Lambda}^*$ that maximizes (27) is finite (which is indeed the case as long as the prior distribution is not degenerate), and whenever the primal solution corresponding to the optimal $\mathbf{\Lambda}^*$ is feasible and unique (which holds when the eigenvalues of $\mathbf{G}_{\mathbf{\Lambda}^*(\mathbf{\Lambda}^*)^H}$ are distinct).

To understand the solution given by (28), recall from (23) that solving the inner minimization in (18) is equivalent to maximizing $\mathrm{tr}\left(\mathbf{G}_{\mathbf{A}} \mathbf{R}_t \right)$. This can be interpreted as maximizing the sum of beamforming powers along particular directions given by the top-M eigenvectors of $\mathbf{G}_{\mathbf{A}}$. In other words, solving the dual problem allows to identify particular spatial regions of interest, which, when probed, minimize a lower bound on the error of AoA estimation.

In terms of computational complexity, note that the dual problem is of dimension $L \times L$, where L is the number of channel parameters, while the primal problem has dimension $N \times M$. Thus, the dual problem has much lower dimension. Furthermore, obtaining the primal solution depends on finding the top-M eigenvectors of an $N \times N$ matrix, which can be done efficiently using the power iteration algorithm, especially when the matrix has sparsity structure.

The development thus far does not yet take into account the constant modulus constraint (10). When the constraint (10) is included, we can no longer guarantee global optimality. Instead, we propose the following heuristic of setting

$$[\mathbf{W}_t]_{i,j} = \frac{1}{\sqrt{N}} \exp\left\{ j \arg\left(\left[\tilde{\mathbf{W}}_t^* \right]_{i,j} \right) \right\}, \tag{29}$$

where $\arg(\cdot)$ denotes the phase of a complex number, and $\tilde{\mathbf{W}}_t^*$ is the solution given by (28) for the problem without the constant modulus constraint.

C. Kalman Filter Tracking of Channel Gains

Solving the dual problem (27) requires the computation of the matrix $\mathbf{G}_{\mathbf{\Lambda}\mathbf{\Lambda}^H}$ given by

$$\mathbf{G}_{\mathbf{\Lambda}\mathbf{\Lambda}^H} = \mathbb{E}\left[\dot{\mathbf{H}}(\boldsymbol{\phi})\mathbf{\Lambda}\mathbf{\Lambda}^H\dot{\mathbf{H}}^H(\boldsymbol{\phi}) \,\middle|\, \mathbf{Y}_{1:t-1}\right], \tag{30}$$

where the expectation is over the posterior distribution of $\boldsymbol{\phi}$ given $\mathbf{Y}_{1:t-1}$. Computing this posterior distribution requires averaging over the conditional distribution of the fading coefficients $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_L)$ given the AoAs $\boldsymbol{\phi}$ and the observed sequence of measurements so far $\mathbf{y}_{1:t-1}$. The key observation is that this conditional distribution is complex Gaussian, which follows from the linearity of the measurement model (1) with respect to $\boldsymbol{\alpha}$. The mean and covariance matrix of this conditional distribution can be tracked using Kalman filtering. In particular, let $\mu_{\boldsymbol{\alpha}|\boldsymbol{\phi}}^{(t)}$ and $\Sigma_{\boldsymbol{\alpha}|\boldsymbol{\phi}}^{(t)}$ denote respectively the mean and covariance matrix of $\boldsymbol{\alpha}$ given the AoAs $\boldsymbol{\phi}$ and the first t observed measurements $\mathbf{y}_{1:t}$. Then, we can write

$$\begin{aligned} \mu_{\boldsymbol{\alpha}|\boldsymbol{\phi}}^{(t)} &= \mu_{\boldsymbol{\alpha}|\boldsymbol{\phi}}^{(t-1)} + \Sigma_{\boldsymbol{\alpha}\mathbf{y}|\boldsymbol{\phi}}^{(t-1)}\big(\Sigma_{\mathbf{y}|\boldsymbol{\phi}}^{(t-1)}\big)^{-1}\left(\mathbf{y}_t - \mu_{\mathbf{y}|\boldsymbol{\phi}}^{(t-1)}\right), \\ \Sigma_{\boldsymbol{\alpha}|\boldsymbol{\phi}}^{(t)} &= \Sigma_{\boldsymbol{\alpha}|\boldsymbol{\phi}}^{(t-1)} - \Sigma_{\boldsymbol{\alpha}\mathbf{y}|\boldsymbol{\phi}}^{(t-1)}\big(\Sigma_{\mathbf{y}|\boldsymbol{\phi}}^{(t-1)}\big)^{-1}\big(\Sigma_{\boldsymbol{\alpha}\mathbf{y}|\boldsymbol{\phi}}^{(t-1)}\big)^{H}, \end{aligned} \tag{31}$$

where we have used

$$\begin{aligned} \mu_{\mathbf{y}|\boldsymbol{\phi}}^{(t-1)} &= \mathbf{C}_{\boldsymbol{\phi},t}^H \mu_{\boldsymbol{\alpha}|\boldsymbol{\phi}}^{(t-1)}, \\ \Sigma_{\boldsymbol{\alpha}\mathbf{y}|\boldsymbol{\phi}}^{(t-1)} &= \Sigma_{\boldsymbol{\alpha}|\boldsymbol{\phi}}^{(t-1)}\mathbf{C}_{\boldsymbol{\phi},t}, \\ \Sigma_{\mathbf{y}|\boldsymbol{\phi}}^{(t-1)} &= \mathbf{C}_{\boldsymbol{\phi},t}^H\Sigma_{\boldsymbol{\alpha}|\boldsymbol{\phi}}^{(t-1)}\mathbf{C}_{\boldsymbol{\phi},t} + \mathbf{W}_t^H\mathbf{W}_t, \end{aligned} \tag{32}$$

and

$$\mathbf{C}_{\boldsymbol{\phi},t} = \sqrt{\frac{P}{L}}\begin{bmatrix}\mathbf{a}^H(\phi_1)\\ \vdots \\ \mathbf{a}^H(\phi_L)\end{bmatrix}\mathbf{W}_t. \tag{33}$$

Since the channel gains are initially assumed to have a zero-mean complex Gaussian distribution with unit variance, the tracking of the mean and covariance matrix is initialized with $\mu_{\boldsymbol{\alpha}|\boldsymbol{\phi}}^{(0)} = \mathbf{0}_{L\times 1}$ and $\Sigma_{\boldsymbol{\alpha}|\boldsymbol{\phi}}^{(0)} = \mathbf{I}_L$ for each parameter vector $\boldsymbol{\phi}$. Hence, if $\pi_{\boldsymbol{\phi}}^{(t)}$ denotes the posterior probability density function of $\boldsymbol{\phi}$ given the first t measurements $\mathbf{y}_{1:t}$, then $\pi_{\boldsymbol{\phi}}^{(t)}$ can be computed recursively as follows:

$$\pi_{\boldsymbol{\phi}}^{(t)} = \frac{\pi_{\boldsymbol{\phi}}^{(t-1)} f(\mathbf{y}_t \,|\, \boldsymbol{\phi}, \mathbf{y}_{1:t-1})}{\int_{\boldsymbol{\phi}'} \pi_{\boldsymbol{\phi}'}^{(t-1)} f(\mathbf{y}_t \,|\, \boldsymbol{\phi}', \mathbf{y}_{1:t-1})\mathrm{d}\boldsymbol{\phi}'}, \tag{34}$$

where the denominator in (34) is a normalization term that does not depend on $\boldsymbol{\phi}$. Note that $\mathbf{Y}_t \,|\, \boldsymbol{\phi}, \mathbf{Y}_{1:t-1} \sim \mathcal{CN}\left(\mu_{\mathbf{y}|\boldsymbol{\phi}}^{(t-1)}, \Sigma_{\mathbf{y}|\boldsymbol{\phi}}^{(t-1)}\right)$, where $\mu_{\mathbf{y}|\boldsymbol{\phi}}^{(t-1)}$ and $\Sigma_{\mathbf{y}|\boldsymbol{\phi}}^{(t-1)}$ are as computed in (32). Finally, we point out that a similar Kalman filter tracking procedure is developed in [5] and [18] for the special case of a single-path channel (i.e., $L = 1$) and a BS equipped with a single RF chain (i.e., $M = 1$).

IV. Numerical Results

In this section, we evaluate the performance of the proposed active sensing strategy for AoA estimation as compared to several existing analog beamforming strategies in the literature.

1) Compressed sensing with random beamforming: In this baseline, the AoA estimation problem is formulated as an L-sparse recovery problem. Specifically, the sensing matrices for all T measurements are generated randomly satisfying the constant-modulus constraint. Let $\bar{\mathbf{W}} = [\mathbf{W}_1, \dots, \mathbf{W}_T]$ be the collection of all T sensing matrices. If the AoAs are chosen from a grid set of size K that is large enough, and denoting $\mathbf{A} = [\mathbf{a}(\phi_1), \dots, \mathbf{a}(\phi_K)]$ as the collection of all K array response vectors, then the received baseband symbols in the T time frames can be expressed as $\bar{\mathbf{y}} = \sqrt{\frac{P}{L}}\bar{\mathbf{W}}^H\mathbf{A}\mathbf{x} + \mathbf{n}$, where $\bar{\mathbf{y}} = \left[\mathbf{y}_1^H, \dots, \mathbf{y}_T^H\right]^H$, $\mathbf{x}$ is an unknown L-sparse vector, and $\mathbf{n}$ is the effective Gaussian noise. Hence, the problem is equivalent to recovering the support of $\mathbf{x}$, which can be done using compressed sensing methods, e.g., the orthogonal matching pursuit (OMP) algorithm.

2) Coordinate descent: The B-CRB optimization problem (9) can be solved using the iterative coordinate descent (CD) algorithm over the phase shifts of the beamformers. In particular, by denoting $[\mathbf{W}_t]_{n,m} = \frac{1}{\sqrt{N}}e^{j[\boldsymbol{\theta}_t]_{n,m}}$ for each entry of the sensing matrix $\mathbf{W}_t$, we can view (9) as an optimization over the phase matrix $\boldsymbol{\theta}_t$. Using the bisection method, we can minimize the objective function in (9) with respect to each entry $[\boldsymbol{\theta}_t]_{n,m}$, while holding the rest as fixed. Iterating this approach over the coordinates of $\boldsymbol{\theta}_t$ gives the CD solution.

3) Deep learning solutions: Several deep-learning-based solutions have been proposed in the literature in order to adaptively design sensing beamformers for AoA estimation. In particular, the approach taken in [6] employs a long short-term memory (LSTM) to model the temporal correlation between the received symbols. The model is trained for a given number of pilot symbols. In the following, we consider two cases of this model: 1) when the model is trained for the exact number of pilot symbols that are used in testing, and 2) when there is a mismatch between the number of pilot symbols used in training, and the number of pilot symbols used in testing.

The performance of the proposed active sensing strategy is compared with the baseline strategies for a channel model with $L = 2$ paths. The AoAs are assumed to be uniformly distributed over $[\phi_{\min}, \phi_{\max}] = [-\frac{\pi}{3}, \frac{\pi}{3}]$, and the BS is equipped with $N = 32$ antennas and $M = 2$ RF chains. The number of measurements made is $T = 7$. For the compressed sensing method and the computation of the posterior distribution of the AoA's, we use a grid set of size $K = 1024$.

Fig. 2 shows the plot of the average MSE for the different beamforming strategies versus the SNR. The proposed active sensing strategy achieves better AoA estimation performance compared to the compressed sensing channel estimation approach (i.e., OMP) and the CD algorithm for optimizing the B-CRB metric. However, it falls short compared to the LSTM approach of [6] when the number of sensing stages used in training and testing are perfectly matched. However, when

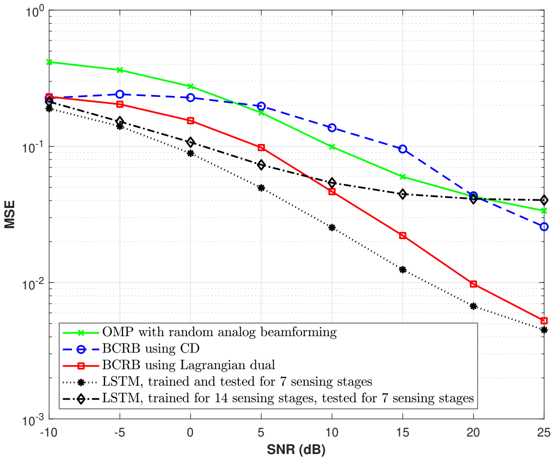


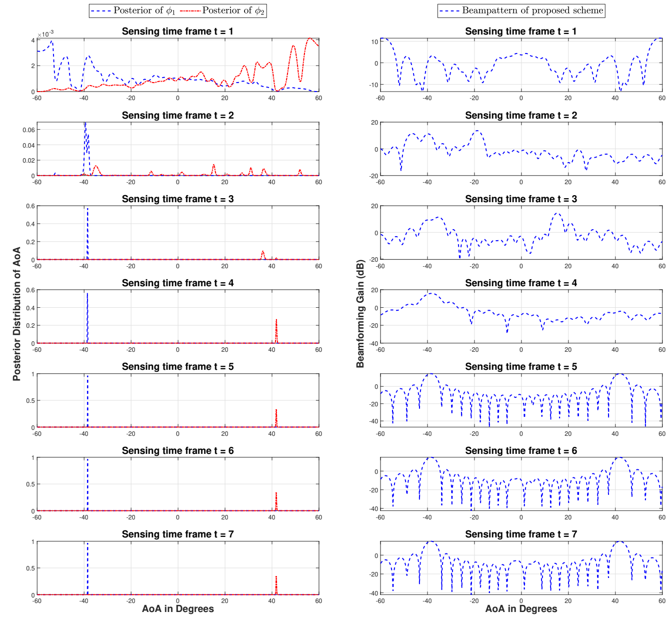
Fig. 2: Average MSE versus SNR for a system model with $N = 32$ antennas, $L = 2$ paths, $T = 7$ sensing stages, and $M = 2$ RF chains.

there is a mismatch between the number of sensing stages used in training the LSTM model and the number of sensing stages used in testing, the performance of the LSTM approach degrades significantly, which highlights the drawback of the learning-based solution for active sensing. In this case, the analytic approach of this paper has a significant advantage.

Fig. 3 plots the beamforming pattern as well as the AoA posterior distribution at the end of the t-th sensing stage when the sensing vectors are designed using the proposed active sensing strategy for an $\text{SNR} = 25$ dB. The true AoAs are $\phi_1 = -38.53^\circ$ and $\phi_2 = 41.85^\circ$. We can see that as the number of sensing stages increases, the marginal posterior distributions converge to highly concentrated distributions with peaks at the true AoAs. Moreover, the generated sensing vectors gradually focus the energy in the direction of the true AoAs. Thus, the designed beamformers have intuitive interpretation.

V. Conclusion

This paper develops an analytic approach for designing an adaptive sequence of beamforming vectors for the angle-of-arrival estimation problem in uplink massive MIMO systems with limited number of RF chains based on optimizing the B-CRB metric in each stage. The proposed adaptive beamforming design is superior to conventional approaches and more robust than the learning-based approach.



(a) (b)

Fig. 3: (a) Posterior distributions of the AoAs, and (b) beamforming patterns of the sensing vectors when using the proposed scheme for a system model with $N = 32$ antennas, $L = 2$ paths, $T = 7$ sensing stages, $M = 2$ RF chains, and $\text{SNR} = 25$ dB. The true AoAs are $\phi_1 = -38.53^\circ$ and $\phi_2 = 41.85^\circ$, and the fading coefficients are $|\alpha_1|^2 = 0.44$ and $|\alpha_2|^2 = 0.71$.

References

[1] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.

[2] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, 2016.

[3] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, 2014.

[4] S.-E. Chiu, N. Ronquillo, and T. Javidi, "Active learning and CSI acquisition for mmWave initial alignment," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2474–2489, 2019.

[5] F. Sohrabi, Z. Chen, and W. Yu, "Deep active learning approach to adaptive beamforming for mmWave initial alignment," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2347–2360, 2021.

[6] F. Sohrabi, T. Jiang, W. Cui, and W. Yu, "Active sensing for communications by learning," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1780–1794, 2022.

[7] T. Jiang and W. Yu, "Active sensing for reciprocal MIMO channels," *IEEE Trans. Signal Process.*, vol. 72, pp. 2905–2920, 2024.

[8] W. Huleihel, J. Tabrikian, and R. Shavit, "Optimal adaptive waveform design for cognitive MIMO radar," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 5075–5089, 2013.

[9] F. Liu, Y.-F. Liu, A. Li, C. Masouros, and Y. C. Eldar, "Cramér-Rao bound optimization for joint radar-communication beamforming," *IEEE Trans. Signal Process.*, vol. 70, pp. 240–253, 2022.

[10] X. Song, J. Xu, F. Liu, T. X. Han, and Y. C. Eldar, "Intelligent reflecting surface enabled sensing: Cramér-Rao bound optimization," *IEEE Trans. Signal Process.*, vol. 71, pp. 2011–2026, 2023.

[11] M. Zhu, L. Li, S. Xia, and T.-H. Chang, "Information and sensing beamforming optimization for multi-user multi-target MIMO ISAC systems," in *IEEE Int. Conf. Acous., Speech, Signal Process. (ICASSP)*, 2023, pp. 1–5.

[12] K. M. Attiah and W. Yu, "Beamforming design for integrated sensing and communications using uplink-downlink duality," in *Proc. IEEE Int. Symp. Inf. Theory*, 2024, pp. 2808–2813.

[13] M. Li and Y. Lu, "Angle-of-arrival estimation for localization and communication in wireless networks," in *Eur. Signal Process. Conf. (EUSIPCO)*, 2008, pp. 1–5.

[14] L. Zhao, D. W. K. Ng, and J. Yuan, "Multi-user precoding and channel estimation for hybrid millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1576–1590, 2017.

[15] I. Bekkerman and J. Tabrikian, "Target detection and localization using MIMO radars and sonars," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3873–3883, 2006.

[16] J. Li, L. Xu, P. Stoica, K. W. Forsythe, and D. W. Bliss, "Range compression and waveform optimization for MIMO radar: A Cramér–Rao bound based study," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 218–232, 2008.

[17] H. Van Trees, *Detection, Estimation and Modulation Theory, Part 1*. New York: Wiley, 1968.

[18] N. Ronquillo, S.-E. Chiu, and T. Javidi, "Sequential learning of CSI for mmWave initial alignment," in *Proc. 53rd Asilomar Conf. Sig. Syst. Comput.*, 2019, pp. 1278–1283.