# Successive Localization and Beamforming in 5G mmWave MIMO Communication Systems

Bingpeng Zhou, *Member, IEEE*, An Liu, *Senior Member, IEEE*, and Vincent Lau, *Fellow, IEEE*

***Abstract*—Beamforming is an attractive technique to improve the system performance for multi-input multi-output (MIMO) communications. Previous works mainly focus on improving the data transmission quality. However, the potential of beamforming for improving the localization quality is not yet fully studied. In this paper, we focus on active beamforming to reduce the user equipment (UE) localization error for millimeter-wave MIMO systems. Such beamforming for localization is of challenge because its optimization cost function (e.g., the localization error bound) also depends on the actual UE location and instantaneous channel states, which are unknown in advance. To address this challenge, a novel successive localization and beamforming (SLAB) scheme is proposed, where the long-term UE location and the instantaneous channel state will be jointly estimated and then the beamforming vector will be successively optimized as per the obtained estimation results. The proposed SLAB scheme will yield a sequence of beamforming weights and UE location estimates, which will converge to the stationary point of the associated optimization problem. Simulation results show that the proposed SLAB scheme achievesa huge performance gain for UE localization compared with state-of-the-art baselines.**

***Index Terms*—5G localization, mmWave MIMO, beamforming, Cramer-Rao lower bound.**

## I. Introduction

WIRELESS localization is of increasing importance for 5G communications, particularly for the millimeter-wave (mmWave) multiple-input-multiple-output (MIMO) systems, due to the expected rising demands of localization-based services in the future [1]–[3].

A number of papers regarding the localization of mmWave systems [4]– [8] have been published. For instance, active beamforming (BF) is proposed in [9] to enhance the localization accuracy of distributed antenna systems. However, active BF for localization of a mmWave MIMO system has not yet been studied. Compared with radio resource optimizations for improving communication qualities (such as maximizing the sum capacity [10] and the error bit probability [11], and minimizing the outage probability [12], the mean squared error [13] and the power subject to minimal rate constraint [14]), radio resource optimization targeting for localization is quite challenging for the following reasons.

- *Parameter Uncertainties*: The cost function (i.e., Cramer-Rao lower bound (CRLB) [8] on localization errors) of the BF optimization depends on the BF vector, channel states and the actual user location. However, the actual user location and channel states are unknown in advance.
- *Non-convexity Of the BF Optimization*: The localization-oriented BF optimization is a non-convex problem. Traditional brute-force solutions may result in a poor solution or lead to a high computational cost.

In [15]–[18], power and bandwidth optimization for localization (by optimizing the CRLB) was studied, where the user location is assumed to be known. To overcome the first challenge, a solution from a robust optimization method is used to optimize the radio resource for localization [19]–[21]. Specifically, a worst case CRLB with respect to (w.r.t.) an uncertainty set of UE location parameters is used as the optimization objective. However, such a robust optimization method is usually over-conservative, especially when the uncertainty set is large. This uncertainty jeopardizes the associated performance gain. To overcome the con-convexity challenge, the successive convex approximation (SCA) or majorization minimization (MM) methods are commonly used. Please refer to [22]–[25] for the details. However, the brute-force application of these algorithms will result in poor performance.

In this paper, we focus on the active BF optimization for 5G mmWave localization. To overcome the above challenges, we propose a novel successive localization and beamforming (SLAB) scheme which does not require a known UE location or the known uncertainty set of the user location. The proposed SLAB performs an alternating optimization of base station BF vectors (namely the *beamforming refinement* (BFR)), long-term user equipment (UE) location parameters and instantaneous channel state (namely the *localization and channel estimation* (LCE)). The proposed SLAB solution generates a sequence of location estimates and BF updates, which is shown to approach the performance of the *genie-aided BF strategy* (the optimal BF with a known UE location). As a result, the proposed solution achieves significant performance gains in localization accuracy

Manuscript received May 14, 2018; revised October 24, 2018 and January 7, 2019; accepted January 15, 2019. Date of publication January 23, 2019; date of current version February 7, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mats Bengtsson. This work was supported by the Science and Technology Plan of Shenzhen under Grant JCYJ20170818114014753. *(Corresponding authors: An Liu and Bingpeng Zhou.)*

B. Zhou is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, and also with the Shenzhen Research Institute, Hong Kong University of Science and Technology, Shenzhen 518057, China (e-mail: eebzhou@ust.hk).

A. Liu is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310000, China (e-mail: anliu@zju.edu.cn).

V. Lau is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: eeknlau@ust.hk).

Digital Object Identifier 10.1109/TSP.2019.2894789

1053-587X © 2019 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

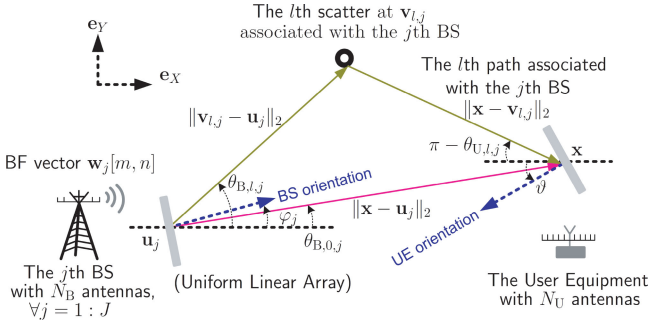



Fig. 1. Illustration of the mmWave MIMO system for localization and BF.

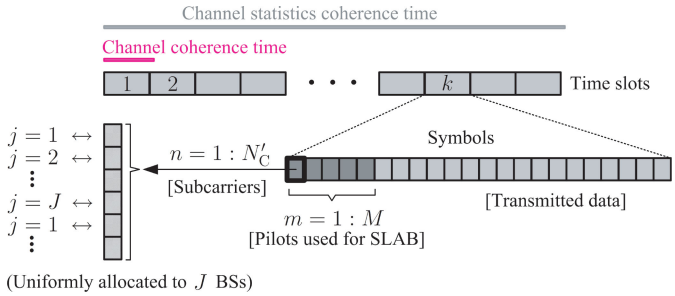



Fig. 2. The frame structure and subcarrier allocation of SLAB.

over various state-of-the-art baselines. The following summarizes the main contributions.

- *Successive Localization and BF Optimization*: The proposed SLAB algorithm based on alternating optimization of two subproblems, namely BFR and LCE, does not require any prior assumption on the UE location. Moreover, it can achieve a significant performance gain over existing robust optimization approaches. Unlike the conventional UE localization methods, such as [26] and [27], which first perform the intermediate step of time-of-arrival (ToA) or angle-of-arrival (AoA) estimation combined with trilateration or triangulation, the proposed LCE algorithm directly estimates the UE and scatterer locations, with problem-specific update rule designs.
- *Successive Non-Convex Approximation*: To overcome the challenge of non-convexity, we propose a novel *successive non-convex approximation* method to derive an efficient BFR algorithm for the SLAB problem. In traditional SCA algorithm designs, a convex surrogate function is used to derive low complexity iterations. In contrast, in our BFR algorithm, we exploit a specific problem structure and propose a non-convex surrogate function with closed-form iterations, which simultaneously preserves the important inherent structure of the BFR subproblem. Thus, the proposed BFR algorithm has fast convergence and good performance.
- *Convergence Analysis of SLAB*: The convergence analysis of SLAB is non-trivial, since it involves a coupled dynamic issue between the fast-time-scale channel states and the varying BF strategies. We have addressed this issue in associated convergence proof.

The remainder of this paper is organized as follows. Section II presents system model. Problem formulation and outline of the proposed SLAB scheme are presented in Section III. The novel LCE and BRF algorithms of SLAB are elaborated in Section IV and V, respectively. In Section VI, the associated convergence is analyzed. Simulations results are presented in Section VII. Finally, Section VIII concludes the paper.

## II. System Model

We consider a mmWave system with $J$ base stations (BSs), one UE and $N'_{\mathrm{C}}$ subcarriers. Each BS has $N_{\mathrm{B}}$ antennas, whereas the UE has $N_{\mathrm{U}}$ antennas, as shown in Fig. 1. In addition, we consider the uniform linear antenna array for UE and BS's. However, the SLAB scheme proposed in this paper can be applied to an arbitrary-shaped antenna array. At each time slot (indexed by $k$), BSs will transmit downlink pilot signals for the UE to jointly estimate the UE location (including orientation) and channel gains based on the received signals. In the following, we will elaborate the frame structure of the system, the channel model, the geometric model for UE localization, and the received signal model.

### A. Frame Structure

The frame structure of the SLAB scheme is illustrated in Fig. 2. We focus on a coherence time interval of channel statistics within which the ToA, angle-of-departure (AoD) and AoA of each channel path and the distribution of small-scale fading coefficients are invariant. The coherence time of channel statistics consists of several time slots, and the small-scale fading coefficients are constant within each time slot. Moreover, each time slot consists of a number of symbols, where the first $M$ symbols are used to transmit training BF vectors for joint localization and channel estimation, and the rest are used to transmit data, as illustrated in Fig. 2. Let $\boldsymbol{\omega}_j[n,m]\in\mathbb{C}^{N_{\mathrm{B}}}$ be the $m$th training BF vector transmitted from the $N_{\mathrm{B}}$ antennas of the $j$th BS on the $n$th subcarrier.[1]

In order to ensure the effectiveness of SLAB, we assume the received signals from different BSs can be identified by UE, via a frequency-division-based coordinated multiple point transmission technique [29]. To be specific, the $N'_{\mathrm{C}}$ subcarriers of each pilot are fairly allocated to those $J$ BSs via some predefined scheduling procedure, and thus each BS has $N_{\mathrm{C}}=N'_{\mathrm{C}}/J$ subcarriers to transmit training BF pilots (we assume $N_C$ is an integer). For instance, the subcarriers of the $j$th BS are given by $\{j,j+J,\ldots,j+(N_{\mathrm{C}}-1)J\}$. For brevity, we use $\Theta_j=\{j,j+J,\ldots,j+(N_{\mathrm{C}}-1)J\}$ to denote the index set of subcarriers associated with the $j$th BS. Let $\boldsymbol{\omega}\in\mathbb{C}^{N_{\mathrm{B}}JN_{\mathrm{C}}M}=\mathrm{vec}\,[\boldsymbol{\omega}_j[n,m]|\forall n\in\Theta_j,\forall m=1:M,\forall j=1:J]$ be the collection of the training BF vectors.

It should be noted that the realization of BF vector $\boldsymbol{\omega}$ will be determined at each BFR stage, and thus it will be viewed as an unknown parameter to be optimized in the BFR stage. Once it is optimized, it will keep invariant (a known signal) at the LCE

[1]We slightly abuse the use of "BF vector" without any ambiguity, which incorporates both the BF matrix and the training pilot [28].

stage. In addition, during the BFR phase at each time slot $k$, the channel $\mathbf{h}[k]$ is still unknown. Therefore, we can only adapt the training BF vector $\boldsymbol{\omega}$ according to the channel statistics, instead of the instantaneous channel $\mathbf{h}[k]$. Moreover, we have no assumption on the structure of training sequence,[2] and we will use a general-form expression of beamformer $\boldsymbol{\omega}$.

### B. Channel Model With Limited Scattering

We consider the limited scattering in the mmWave channel. Specifically, let $\mathbf{u}_j \in \mathbb{R}^2$ and $\varphi_j \in [-\pi, \pi)$ be the known coordinate and angular position, respectively, of the $j$th BS. Let $\mathbf{x} \in \mathbb{R}^2$ and $\vartheta \in [-\pi, \pi)$ be the unknown position and orientation, respectively, of the UE. We assume there are $L+1$ paths in the scattering channel ($l = 0$ for the line-of-sight (LOS) path, and $l > 0$ for the non-line-of-sight (NLOS) path), where $L$ is the maximum number of paths between the UE and any BS. Let $\tau_{l,j}$, $\theta_{\mathrm{B},l,j}$ and $\theta_{\mathrm{U},l,j}$ denote the ToA, AoD and AoA, respectively, of the $l$th path associated with the $j$th BS, which are unknown scalars. For each NLOS path, there is a scatterer with an unknown location, as shown in Fig. 1. Let $\mathbf{v}_{l,j} \in \mathbb{R}^2$ be the unknown location of the scatter associated with the $l$th path and the $j$th BS.

For given large-scale multipath parameters $\tau_{l,j}$, $\theta_{\mathrm{B},l,j}$, $\theta_{\mathrm{U},l,j}$, $\forall l$, the channel matrix $\widetilde{\mathbf{H}}_j[k,n] \in \mathbb{C}^{N_\mathrm{B} \times N_\mathrm{U}}$ between the $j$th BS and the UE on the $n$th subcarrier is given by [5]

$$\widetilde{\mathbf{H}}_j[k,n] = \mathbf{A}_{\mathrm{U},j}[n]\mathbf{H}_j[k,n]\mathbf{A}_{\mathrm{B},j}^{\mathrm{H}}[n],$$

where $\mathbf{A}_{\mathrm{B},j}[n] \in \mathbb{C}^{N_\mathrm{B} \times (L+1)}$ is the steering matrix of the $j$th BS antenna array on the $n$th subcarrier, and $\mathbf{A}_{\mathrm{U},j}[n] \in \mathbb{C}^{N_\mathrm{U} \times (L+1)}$ is the response matrix of UE antenna array on the $n$th subcarrier, which depend on the unknown angular parameters such as $\theta_{\mathrm{B},l,j}$ and $\theta_{\mathrm{U},l,j}$, given by

$$\mathbf{A}_{\mathrm{B},j}[n] = [\mathbf{a}_{\mathrm{B},n}(\theta_{\mathrm{B},0,j}), \ldots, \mathbf{a}_{\mathrm{B},n}(\theta_{\mathrm{B},L,j})], \tag{1}$$

$$\mathbf{a}_{\mathrm{B},n}(\theta_{\mathrm{B},l,j}) = \mathrm{vec}[e^{-\mathrm{j}\frac{d_\mathrm{B}\pi}{\lambda_n}(t-1)\sin\theta_{\mathrm{B},l,j}} | \forall t = 1 : N_\mathrm{B}], \tag{2}$$

$$\mathbf{A}_{\mathrm{U},j}[n] = [\mathbf{a}_{\mathrm{U},n}(\theta_{\mathrm{U},0,j}), \ldots, \mathbf{a}_{\mathrm{U},n}(\theta_{\mathrm{U},L,j})], \tag{3}$$

$$\mathbf{a}_{\mathrm{U},n}(\theta_{\mathrm{U},l,j}) = \mathrm{vec}[e^{-\mathrm{j}\frac{d_\mathrm{A}\pi}{\lambda_n}(r-1)\sin\theta_{\mathrm{U},l,j}} | \forall r = 1 : N_\mathrm{U}], \tag{4}$$

in which $\mathrm{j} = \sqrt{-1}$, $\lambda_n$ denotes the wavelength associated with subcarrier $n$, and $d_\mathrm{A}$ is the distance between the antenna elements of each BS, which are known scalars.

In addition, $\mathbf{H}_j[k,n] \in \mathbb{C}^{(L+1)\times(L+1)}$ is the frequency-domain channel matrix on the $n$th subcarrier:

$$\mathbf{H}_j[k,n] = \sqrt{N_\mathrm{B}N_\mathrm{U}}\mathrm{diag}\big\{h_{l,j}[k]e^{-\mathrm{j}2\pi\frac{n}{N_\mathrm{C}'T_s}\tau_{l,j}} | \forall l = 0 : L\big\},$$

where $h_{l,j}[k]$ denotes the small-scale fading coefficient of the $l$th path associated with the $j$th BS at time slot $k$ (note that we absorb the path loss into the small-scale fading coefficient), and $T_s$ is the sampling period. For convenience, let $\mathbf{h}[k] \in \mathbb{C}^{J(L+1)} = \mathrm{vec}[h_{l,j}[k] | \forall l = 0 : L, \forall j = 1 : J]$ denote the small-scale fading channel vector, which is an unknown variable. We assume $\mathbf{h}[k] \sim \mathcal{CN}(\mathbf{h}[k] | \mathbf{0}, \boldsymbol{\Sigma})$ is independent and identically distributed over different time slots, where $\boldsymbol{\Sigma}$ is the variance matrix of $\mathbf{h}[k]$ which is assumed to be known.

[2] In practice, the training sequence structure will affect the receiver-end SNR and hence the UE localization error, as will be implied in (20).

### C. Geometric Model for UE Localization

Let $\boldsymbol{\alpha} \in \mathbb{R}^{2JL+3} = [\mathbf{x}^\top, \vartheta, \mathbf{v}^\top]^\top$ be the set of location parameters, where $\mathbf{v} \in \mathbb{R}^{2JL} = \mathrm{vec}[\mathbf{v}_{l,j} | \forall l = 1 : L, \forall j = 1 : J]$ is the collection of unknown scatter locations. The relationship between $\{\mathbf{x}, \mathbf{v}_{l,j}, \vartheta\}$ and $\{\tau_{l,j}, \theta_{\mathrm{B},l,j}, \theta_{\mathrm{U},l,j}\}$ is given by

$$\tau_{0,j} = \frac{\|\mathbf{x} - \mathbf{u}_j\|_2}{c}, \tag{5}$$

$$\tau_{l,j} = \frac{\|\mathbf{x} - \mathbf{v}_{l,j}\|_2 + \|\mathbf{u}_j - \mathbf{v}_{l,j}\|_2}{c},\ l > 0, \tag{6}$$

$$\theta_{\mathrm{B},0,j} = \arccos\left(\frac{(\mathbf{x} - \mathbf{u}_j)^\top \mathbf{e}_X}{\|\mathbf{x} - \mathbf{u}_j\|_2}\right) - \varphi_j, \tag{7}$$

$$\theta_{\mathrm{B},l,j} = \arccos\left(\frac{(\mathbf{v}_{l,j} - \mathbf{u}_j)^\top \mathbf{e}_X}{\|\mathbf{v}_{l,j} - \mathbf{u}_j\|_2}\right) - \varphi_j,\ l > 0, \tag{8}$$

$$\theta_{\mathrm{U},0,j} = \pi + \arccos\left(\frac{(\mathbf{x} - \mathbf{u}_j)^\top \mathbf{e}_X}{\|\mathbf{x} - \mathbf{u}_j\|_2}\right) - \vartheta, \tag{9}$$

$$\theta_{\mathrm{U},l,j} = \pi + \arccos\left(\frac{(\mathbf{x} - \mathbf{v}_{l,j})^\top \mathbf{e}_X}{\|\mathbf{x} - \mathbf{v}_{l,j}\|_2}\right) - \vartheta,\ l > 0, \tag{10}$$

where $c$ is the light speed, and $\mathbf{e}_X = [1, 0]^T$.

### D. Received Signal Model

Let $\mathbf{z}_j[k,n,m] \in \mathbb{C}^{N_\mathrm{U}}$ be the observation signal, i.e., the $m$th received pilot signal vector on subcarrier $n$ from the $j$th BS at the $k$th time slot, which is given by [5]

$$\mathbf{z}_j[k,n,m] = \widetilde{\mathbf{H}}_j[k,n]\,\boldsymbol{\omega}_j[n,m] + \boldsymbol{\epsilon}_j[k,n,m], \tag{11}$$

where $\boldsymbol{\epsilon}_j[k,n,m] \in \mathbb{C}^{N_\mathrm{U}}$ denotes the measurement noise vector at the UE side, and we generally assume $\epsilon_j[k,n,m] \sim \mathcal{CN}(\epsilon_j[k,n,m] | \mathbf{0}_{N_\mathrm{U}}, \sigma^2\mathbf{I}_{N_\mathrm{U}})$ with the variance $\sigma^2$. Let $\mathbf{z}[k] \in \mathbb{C}^{N_\mathrm{U}N_\mathrm{C}MJ} = \mathrm{vec}[\mathbf{z}_j[k,n,m] | \forall n \in \Theta_j, \forall m = 1 : M, \forall j = 1 : J]$ and $\boldsymbol{\epsilon}[k] \in \mathbb{C}^{N_\mathrm{U}N_\mathrm{C}MJ} = \mathrm{vec}[\boldsymbol{\epsilon}_j[k,n,m] | \forall n \in \Theta_j, \forall m = 1 : M, \forall j = 1 : J]$ be the collection of received pilot signals and noise vectors, respectively. Then, given a training BF vector $\boldsymbol{\omega}$, the measurement signal $\mathbf{z}[k]$ can be expressed as a function of the slow-timescale location parameter $\boldsymbol{\alpha}$, the fast-timescale channel $\mathbf{h}[k]$ and the noise vector $\boldsymbol{\epsilon}[k]$ as

$$\mathbf{z}[k] = \mathbf{g}(\boldsymbol{\alpha}, \mathbf{h}[k]; \boldsymbol{\omega}) + \boldsymbol{\epsilon}[k], \tag{12}$$

where $\mathbf{g}(\boldsymbol{\alpha}, \mathbf{h}[k]; \boldsymbol{\omega})$ is the measurement function, given by

$$\mathbf{g}(\boldsymbol{\alpha}, \mathbf{h}[k]; \boldsymbol{\omega}) = \mathbf{G}(\boldsymbol{\alpha}; \boldsymbol{\omega})\mathbf{h}[k], \tag{13}$$

in which $\mathbf{G}(\boldsymbol{\alpha}; \boldsymbol{\omega}) \in \mathbb{C}^{N_\mathrm{U}N_\mathrm{C}MJ \times J(L+1)}$ is called the coefficient matrix of channel vector $\mathbf{h}[k]$, which is dependent on the

unknown location parameter $\boldsymbol{\alpha}$ and given by [5]

$$\mathbf{G}(\boldsymbol{\alpha};\boldsymbol{\omega}) = \mathrm{diag}[\mathbf{G}_j|\forall j = 1:J], \quad (14)$$

$$\mathbf{G}_j \in \mathbb{C}^{N_{\mathrm{U}}N_{\mathrm{C}}M\times(L+1)} = \mathrm{vec}[\mathbf{g}_j^{(r)\mathrm{H}}[n,m]|\forall r,\forall n,\forall m], \quad (15)$$

$$\mathbf{g}_j^{(r)}[n,m] \in \mathbb{C}^{L+1} = \mathrm{vec}[\mathrm{g}_{l,j}^{(r)}[n,m]|\forall l = 0:L], \quad (16)$$

$$\mathrm{g}_{l,j}^{(r)}[n,m] = \boldsymbol{\omega}_j^{\mathrm{H}}[n,m]\boldsymbol{\mu}_{l,j,n}^{(r)}, \quad (17)$$

$$\boldsymbol{\mu}_{l,j,n}^{(r)} \in \mathbb{C}^{N_{\mathrm{B}}} = \mathrm{vec}[\mu_{l,j,n}^{(r,t)}|\forall t = 1:N_{\mathrm{B}}], \quad (18)$$

$$\mu_{l,j,n}^{(r,t)} = \mathrm{a}_{\mathrm{U},n}^{(r)}(\theta_{\mathrm{U},l,j})e^{-\mathrm{j}2\pi\frac{n}{N_{\mathrm{C}}Ts}\tau_{l,j}}(\mathrm{a}_{\mathrm{B},n}^{(t)}(\theta_{\mathrm{B},l,j}))^*, \quad (19)$$

where $\mathrm{a}_{\mathrm{U},n}^{(r)}(\theta_{\mathrm{U},l,j})$ and $\mathrm{a}_{\mathrm{B},n}^{(t)}(\theta_{\mathrm{B},l,j})$ denote the $r$th and the $t$th elements of $\mathbf{a}_{\mathrm{U},n}(\theta_{\mathrm{U},l,j})$ and $\mathbf{a}_{\mathrm{B},n}(\theta_{\mathrm{B},l,j})$ in (4) and (2), respectively. Note that, $\mathrm{a}_{\mathrm{U},n}^{(r)}(\theta_{\mathrm{U},l,j})$ and $\mathrm{a}_{\mathrm{B},n}^{(t)}(\theta_{\mathrm{B},l,j})$ are functions of location parameters $\mathbf{x}$, $\vartheta$ and $\mathbf{v}_{l,j}$ via (5)–(10).

## III. The Proposed SLAB Scheme

We first formulate the localization-oriented beamforming (LOB) problem and then point out its challenges. After that, we outline the proposed SLAB scheme and explain how it addresses the challenges. In the next two sections, we will elaborate the proposed SLAB scheme.

### A. Problem Formulation Of LOB

Let $\boldsymbol{\beta}[k] \in \mathbb{C}^{3JL+J+3} = [\boldsymbol{\alpha};\mathbf{h}[k]]$ be the joint variable of $\boldsymbol{\alpha}$ and $\mathbf{h}[k]$. At each time slot $k$, an LCE algorithm is used to obtain a new estimate of $\boldsymbol{\beta}[k]$. The minimum mean squared error (MSE) of a specific LCE algorithm for $\boldsymbol{\beta}[k]$ is given by

$$\mathrm{MSE}_{\boldsymbol{\beta}[k]}(\boldsymbol{\omega}) = \mathbb{E}_{\mathbf{h}[k],\boldsymbol{\epsilon}[k]}\{\|\boldsymbol{\beta}[k] - \hat{\boldsymbol{\beta}}[k]\|_2^2\},$$

where the expectation is taken w.r.t. $\boldsymbol{\epsilon}[k]$ and $\mathbf{h}[k]$. It should be noted that $\mathrm{MSE}_{\boldsymbol{\beta}[k]}(\boldsymbol{\omega})$ depends on the BF vector $\boldsymbol{\omega}$. Yet, there is no closed-form expression of $\mathrm{MSE}_{\boldsymbol{\beta}[k]}(\boldsymbol{\omega})$ due to the nonlinear model $\mathbf{G}(\boldsymbol{\alpha};\boldsymbol{\omega})$ w.r.t. $\boldsymbol{\alpha}$. To address this challenge, we first obtain a closed-form lower bound, i.e., CRLB [30], for the MSE of an arbitrary LCE algorithm, and then we use the obtained CRLB as a performance metric to optimize the BF vector, as elaborated later. This is effective since the reduction of the lower bound usually means the reduction of the actual MSE, which has been widely adopted, as in [15]–[21].

*Lemma 1 (Lower Bound of MSE):* For a given training BF vector $\boldsymbol{\omega}$, the MSE performance of an arbitrary LCE algorithm, denoted by $\mathrm{MSE}_{\boldsymbol{\beta}[k]}(\boldsymbol{\omega})$, is bounded from below as

$$\mathrm{MSE}_{\boldsymbol{\beta}[k]}(\boldsymbol{\omega}) \geq \mathrm{trace}\big(\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega})\big),$$

where $\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega}) \in \mathbb{S}^{3JL+J+3}$ is called the long-term CRLB of $\boldsymbol{\beta}[k]$ (in terms of the average over small-scale fading $\mathbf{h}[k]$), given by (20) shown at the bottom of this page, where $\mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega}) \in \mathbb{C}^{(2JL+3)\times J(L+1)}$ and $\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega}) \in \mathbb{C}^{J\times J(L+1)}$ will be given by (49) and (54), respectively, in Appendix A.

*Proof:* The proof is given in Appendix A. ■

Given this long-term CRLB, we propose to design the BF vector $\boldsymbol{\omega}$ as the following minimization problem:

$$\mathscr{P}_{\mathrm{LOB}}: \ \boldsymbol{\omega}_\star = \arg\min_{\boldsymbol{\omega}} \mathrm{trace}\big(\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega})\big), \quad (21)$$

$$\mathrm{s.t.}\ \|\boldsymbol{\omega}_j[n,m]\|_2 \leq 1,\ \forall n,m,j, \quad (22)$$

where the BF vector $\boldsymbol{\omega}$ is viewed as an unknown parameter to be optimized. It should be noted that we consider to adopt the "long-term" CRLB $\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega})$ adaptive to the statistics of $\mathbf{h}[k]$ in the cost function of BF optimization, since the "instantaneous" CRLB adaptive to the instantaneous $\mathbf{h}[k]$ is practically unavailable because the instantaneous $\mathbf{h}[k]$ is still unknown before BF training at time slot $k$.[3]

*Challenge*. However, there are two technical challenges in the above BF optimization problem $\mathscr{P}_{\mathrm{LOB}}$:

- $\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega})$ used for LOB is dependent on the true value of $\boldsymbol{\alpha}$ that is, however, unknown beforehand.
- $\mathscr{P}_{\mathrm{LOB}}$ is a non-convex problem w.r.t. $\boldsymbol{\omega}$.

For the unknown location parameters in the cost function, conventional methods [19]–[21] in a power allocation regime resort to optimizing the worst case CRLB within an uncertain set of UE location parameters. These methods are usually over-conservative, especially when the uncertainty set of the UE location is large, thus jeopardizing the associated performance gain. Regarding the non-convex problem, conventional algorithms such as [31] resort to optimizing the BF vector $\boldsymbol{\omega}$ via maximizing the Fisher information matrix (FIM, i.e., inverse CRLB), where the FIM constraint is formulated as a linear matrix inequality (LMI). However, due to the semidefinite positive approximation and rank relaxation, there will be a non-ignorable performance loss in these methods.

[3]The CRLB $\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega})$ in (20) contains two types of CRLB sub-matrices, i.e., the location-related CRLB (the $(2JL+3)\times(2JL+3)$ left-top submatrix, denoted by $\bar{\mathfrak{B}}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega})$) and the channel-related CRLB (the $(JL+J)\times(JL+J)$ right-bottom submatrix, denoted by $\bar{\mathfrak{B}}_{\mathbf{h}[k]}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega})$), which might be in different scales. In practice, we can adjust them into the same scale via matrices normalization over the corresponding sub-matrix traces, as illustrated in (23),

$$\tilde{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega}) = \begin{bmatrix} \frac{\bar{\mathfrak{B}}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega})}{\mathrm{trace}\left(\bar{\mathfrak{B}}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\mathbf{w}[0])\right)} & \mathbf{0}_{(2JL+3)\times J(L+1)} \\ \mathbf{0}^{\mathrm{H}}_{(2JL+3)\times J(L+1)} & \frac{\bar{\mathfrak{B}}_{\mathbf{h}[k]}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega})}{\mathrm{trace}\left(\bar{\mathfrak{B}}_{\mathbf{h}[k]}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\mathbf{w}[0])\right)} \end{bmatrix}. \quad (23)$$

where $\mathbf{w}[0]$ is the realization of BF vector $\boldsymbol{\omega}$ at the initial time slot $k=0$. Then, the normalized CRLB $\tilde{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega})$ can be adopted as the new cost function in (21) to take care of the different scaling issue, which will not affect the structure of the proposed BFR method.

$$\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\boldsymbol{\Sigma};\boldsymbol{\omega}) = \sigma^2\left(\sum_{r,n,m}\begin{bmatrix} \mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega})\boldsymbol{\Sigma}\big(\mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega})\big)^{\mathrm{H}} & \mathbf{0}_{(2JL+3)\times J(L+1)} \\ \mathbf{0}^{\mathrm{H}}_{(2JL+3)\times J(L+1)} & \big(\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega})\big)^{\mathrm{H}}\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega}) \end{bmatrix}\right)^{-1}. \quad (20)$$

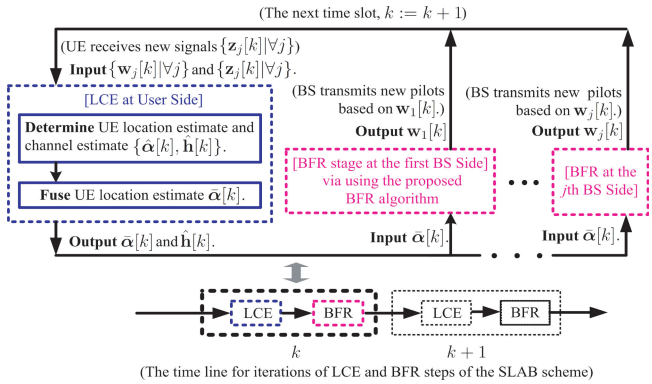



Fig. 3. Illustration of the main SLAB diagram and the associated time line.

### B. Outline of the SLAB Scheme

To address the above challenges, we propose a novel SLAB scheme to find a stationary point of $\mathscr{P}_{\rm LOB}$ (the optimized training BF vector), and simultaneously obtain the estimate of location parameters and channel state at each time slot.

Specifically, starting from an initial BF vector $\mathbf{w}[1]$ at time slot $k=1$,[4] we alternately update the estimate of joint parameter $\big[\boldsymbol{\alpha}; \mathbf{h}[k]\big]$ and optimize the BF choice $\mathbf{w}[k]$, until the obtained BF sequence $\{\mathbf{w}[k]\}$ and the location parameter estimate sequence $\{\bar{\boldsymbol{\alpha}}[k]\}$ converge. Thus, at each time slot $k$, the SLAB scheme is comprised of two stages, i.e., the joint localization and channel estimation (LCE) stage and the BF refinement (BFR) stage. At the LCE stage, the estimate of location parameter $\boldsymbol{\alpha}$ is updated and the channel parameter $\mathbf{h}[k]$ is estimated based on the fresh measurements $\mathbf{z}[k]$. Then, at the BFR stage, the BF vector $\boldsymbol{\omega}$ will be refined based on the updated location parameter estimate $\bar{\boldsymbol{\alpha}}[k]$ from the LCE stage (via minimizing the long-term CRLB $\mathrm{trace}\big(\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\bar{\boldsymbol{\alpha}}[k], \boldsymbol{\Sigma}; \boldsymbol{\omega})\big)$ over $\boldsymbol{\omega}$). Afterwards, BSs will transmit new pilot signals by using the newly optimized BF vector $\mathbf{w}[k+1]$, and the new measurement $\mathbf{z}[k+1]$ will be received at the UE. Based on $\mathbf{z}[k+1]$, the UE will further update its location parameter estimate $\bar{\boldsymbol{\alpha}}[k+1]$ using the LCE algorithm (i.e., the new LCE stage again). This alternating optimization process between LCE and BFR will repeat until $\{\mathbf{w}[k]\}$ and $\{\bar{\boldsymbol{\alpha}}[k]\}$ converge. The main diagram of SLAB is illustrated in Fig. 3. The LCE and BFR stages are outlined as follows.

*Outline of the LCE Stage:* At the $k$th time slot, the LCE stage is to determine the location estimate $\bar{\boldsymbol{\alpha}}[k]$ (based on a fusion criterion) and the channel estimate $\hat{\mathbf{h}}[k]$, as follows.

Firstly, a new estimate $\hat{\boldsymbol{\beta}}[k] = \big[\hat{\boldsymbol{\alpha}}[k]; \hat{\mathbf{h}}[k]\big]$ is determined by solving the following optimization problem:

$$\mathscr{A}_{\rm SLAB}^{(\rm LCE)}: \hat{\boldsymbol{\beta}}[k] = \arg\min_{\boldsymbol{\beta}} \|\mathbf{z}[k] - \mathbf{g}(\boldsymbol{\beta}; \mathbf{w}[k])\|_2^2, \tag{24}$$

where it should be noted that the measurement $\mathbf{z}[k]$ used for LCE depends on the training BF vector $\mathbf{w}[k]$ obtained in the previous BFR stage. The details of the LCE algorithm to estimate $\hat{\boldsymbol{\alpha}}[k]$ and $\hat{\mathbf{h}}[k]$ will be elaborated in Section IV.

[4]We use $\mathbf{w}[k]$ to denote the realization of $\boldsymbol{\omega}$ at the $k$th time slot.

Secondly, once a new estimate $\hat{\boldsymbol{\alpha}}[k]$ is obtained as above, the UE location parameter $\bar{\boldsymbol{\alpha}}[k]$ will be updated by fusing the new estimate $\hat{\boldsymbol{\alpha}}[k]$ from fresh measurement $\mathbf{z}[k]$ and the previous result $\bar{\boldsymbol{\alpha}}[k-1]$ in the following manner:

$$\bar{\boldsymbol{\alpha}}[k] = (\boldsymbol{\Lambda}[k])^{-1}\big(\boldsymbol{\Lambda}[k-1]\bar{\boldsymbol{\alpha}}[k-1] + \boldsymbol{\Lambda}^{\sharp}[k]\hat{\boldsymbol{\alpha}}[k]\big), \tag{25}$$

where $\boldsymbol{\Lambda}[k-1] \in \mathbb{S}^{2JL+3}$ is the overall estimation precision matrix at the previous time slot, $\boldsymbol{\Lambda}^{\sharp}[k] \in \mathbb{S}^{2JL+3}$ is the estimation precision matrix associated with the fresh measurement $\mathbf{z}[k]$ at the current time slot, and $\boldsymbol{\Lambda}[k] \in \mathbb{S}^{2JL+3}$ is the overall estimation precision matrix obtained so far, given by[5]

$$\boldsymbol{\Lambda}[k] = \boldsymbol{\Lambda}[k-1] + \boldsymbol{\Lambda}^{\sharp}[k], \tag{26}$$

$$\boldsymbol{\Lambda}^{\sharp}[k] = \big(\mathfrak{B}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}[k], \hat{\mathbf{h}}[k]; \mathbf{w}[k])\big)^{-1}, \tag{27}$$

where $\mathfrak{B}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}[k], \hat{\mathbf{h}}[k]; \mathbf{w}[k]) \in \mathbb{S}^{2JL+3}$ is the CRLB of $\boldsymbol{\alpha}$, depending on the BF vector $\mathbf{w}[k]$ obtained at the previous BFR stage. The CRLB of $\boldsymbol{\alpha}$ is actually the $(2JL+3)\times(2JL+3)$ top-left sub-matrix of CRLB matrix $\mathfrak{B}_{\boldsymbol{\beta}[k]}(\hat{\boldsymbol{\alpha}}[k], \hat{\mathbf{h}}[k]; \mathbf{w}[k])$ of joint variable $\boldsymbol{\beta}[k]$; that is to say, $\mathfrak{B}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}[k], \hat{\mathbf{h}}[k]; \mathbf{w}[k]) = [\mathfrak{B}_{\boldsymbol{\beta}[k]}(\hat{\boldsymbol{\alpha}}[k], \hat{\mathbf{h}}[k]; \mathbf{w}[k])]_{[1:(2JL+3)]\times[1:(2JL+3)]}$, and the joint-variable CRLB $\mathfrak{B}_{\boldsymbol{\beta}[k]}(\hat{\boldsymbol{\alpha}}[k], \hat{\mathbf{h}}[k]; \mathbf{w}[k]) \in \mathbb{S}^{3JL+J+3}$ is given by (47). We set the initial fusion precision $\boldsymbol{\Lambda}[0] = \mathbf{0}$.

*Outline of the BFR Stage:* The BFR stage is to derive a refined BF vector $\mathbf{w}[k+1]$ for the next LCE stage at time slot $(k+1)$ by minimizing an approximate long-term CRLB, i.e.,

$$\mathscr{A}_{\rm SLAB}^{(\rm BFR)}: \mathbf{w}[k+1] = \arg\min_{\boldsymbol{\omega}} \mathrm{trace}\big(\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\bar{\boldsymbol{\alpha}}[k], \boldsymbol{\Sigma}; \boldsymbol{\omega})\big) \tag{28}$$

$$\text{s.t. } \|\boldsymbol{\omega}_j[n,m]\|_2 \le 1, \forall j, n, m. \tag{29}$$

$\mathscr{A}_{\rm SLAB}^{(\rm BFR)}$ is non-convex w.r.t. $\boldsymbol{\omega}$, and the BFR algorithm to solve $\mathscr{A}_{\rm SLAB}^{(\rm BFR)}$ will be given later in Section V. As such, the overall SLAB scheme is summarized in Algorithm 1. By alternately repeating the LCE and BFR stages, the SLAB scheme will yield a gradually-refined BF vector $\mathbf{w}[k]$ and location parameter estimate $\bar{\boldsymbol{\alpha}}[k]$, until $\mathbf{w}[k]$ and $\bar{\boldsymbol{\alpha}}[k]$ converge.

## IV. The Proposed LCE Algorithm

In this section we will elaborate the LCE algorithm to solve $\mathscr{A}_{\rm SLAB}^{(\rm LCE)}$ and then we will elaborate the BFR algorithm to solve $\mathscr{A}_{\rm SLAB}^{(\rm BFR)}$ in the next section.

### A. Nature of the LCE Problem

It should be noted that $\mathscr{A}_{\rm SLAB}^{(\rm LCE)}$ is a non-convex optimization problem w.r.t. $(\boldsymbol{\alpha}, \mathbf{h}[k])$. A number of optimization methods can be used to find the globally optimal solution to a non-convex problem, e.g., the branch-and-band algorithm [32]. However, such a global search algorithm usually has a high computational

[5]Note that we have assumed that $\mathbf{h}[k]$ is independently and identically distributed, and thus there is no channel estimate gain from the previous time slots. Hence, the filtering of $\hat{\boldsymbol{\beta}}[k]$ will equivalently reduce to the filtering of $\bar{\boldsymbol{\alpha}}[k]$ (given by (25)) in conjunction with the independent estimate of $\mathbf{h}[k]$ (will be given by (30)), without loss of filtering performance.

**Algorithm 1:** The Proposed SLAB Scheme.

**Input**: The initial measurement $\mathbf{z}[k]$, $k=1$, and channel variance matrix $\mathbf{\Sigma}$.
1: Determine the initial BF choice $\mathbf{w}[1]$.
2: **While** not satisfied **do** (for $k=1:K$)
3: Determine the geometric parameter estimate $\bar{\boldsymbol{\alpha}}[k]$ and channel estimate $\hat{\mathbf{h}}[k]$ from $\mathbf{z}[k]$ associated with BF choice $\mathbf{w}[k]$, by using **Algorithm 2** (LCE).
4: Determine the BF vector $\mathbf{w}[k+1]$ based on $\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\bar{\boldsymbol{\alpha}}[k],\mathbf{\Sigma};\boldsymbol{\omega})$, by using **Algorithm 3** (BFR).
5: Receive new $\mathbf{z}[k+1]$ dependent on $\mathbf{w}[k+1]$.
6: **End**
**Output**: $\{\bar{\boldsymbol{\alpha}}[k],\hat{\mathbf{h}}[k]\}$, and $\mathbf{w}[k]$.

complexity. Hence, we propose a successive linear least square (SLLS)-based low-cost algorithm to find a stationary solution to $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{LCE})}$ via leveraging the hybrid convexity/non-convexity structure.

Since the measurement function $\mathbf{g}(\boldsymbol{\alpha},\mathbf{h}[k];\mathbf{w}[k])$ in (13) is linear w.r.t. the channel $\mathbf{h}[k]$, there is a hidden convex structure w.r.t. $\mathbf{h}[k]$. As a result, the LCE problem $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{LCE})}$ in (24) is convex for $\mathbf{h}[k]$ (conditioned on $\bar{\boldsymbol{\alpha}}[k]$) but non-convex for $\boldsymbol{\alpha}$.

### *B. The SLLS-Based LCE Algorithm*

By exploiting such a hidden convex structure, we decompose the LCE problem $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{LCE})}$ into two components, (convex) channel estimate and (non-convex) UE localization. Then, we solve $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{LCE})}$ by alternately optimizing the convex component $\mathbf{h}[k]$ and the non-convex component $\boldsymbol{\alpha}$.

*1) Channel Estimate:* For the convex component $\mathbf{h}[k]$, we derive its least square solution (which is the optimal estimate due to the linear Gaussian model (12)), depending on a fresh guess $\boldsymbol{\alpha}_{[i]}^{\natural}[k]$ of $\boldsymbol{\alpha}$ (elaborated by (31)). At the $i$th iteration, given $\boldsymbol{\alpha}_{[i]}^{\natural}[k]$, $\mathbf{h}_{[i]}^{\natural}[k]$ is given by

$$\mathbf{h}_{[i]}^{\natural}[k]=\big(\mathbf{G}^{\mathrm{H}}(\boldsymbol{\alpha}_{[i]}^{\natural}[k])\big)^{\dagger}\mathbf{z}[k], \tag{30}$$

where $^{\dagger}$ is the pseudo-inverse, and $\mathbf{G}(\boldsymbol{\alpha}_{[i]}^{\natural}[k])$ dependent on $\boldsymbol{\alpha}_{[i]}^{\natural}[k]$ is given by (14). Finally, when the iteration converges, we have $\hat{\mathbf{h}}[k]=\mathbf{h}_{[i]}^{\natural}[k]$. The fresh guess $\boldsymbol{\alpha}_{[i]}^{\natural}[k]$ at the $i$th iteration within the $k$th time slot is determined as follows.

*2) UE Localization:* For the non-convex problem associated with $\boldsymbol{\alpha}$, we employ the SLLS method to find a stationary solution to $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{LCE})}$ for $\boldsymbol{\alpha}$, by exploiting a convex approximation to the cost function in (24). Specifically, we iteratively solve the convex subproblems $\mathscr{A}_{\mathrm{SLAB},[i]}^{(\mathrm{LCE})}$ to first find a candidate update $\boldsymbol{\alpha}_{[i+1]}^{\circ}[k]$, for $i=1:M_{\mathrm{I}}$,

$$\mathscr{A}_{\mathrm{SLAB},[i]}^{(\mathrm{LCE})}:\boldsymbol{\alpha}_{[i+1]}^{\circ}[k]=\arg\min_{\boldsymbol{\alpha}} f_{\mathrm{S}}\big(\boldsymbol{\alpha};\boldsymbol{\alpha}_{[i]}^{\natural}[k],\mathbf{h}_{[i]}^{\natural}[k]\big), \tag{31}$$

where $f_{\mathrm{S}}\big(\boldsymbol{\alpha};\boldsymbol{\alpha}_{[i]}^{\natural}[k],\mathbf{h}_{[i]}^{\natural}[k]\big)$ denotes the surrogate function of the original cost function in (24), given by (32) shown at the bottom of this page, where $\nabla_{\boldsymbol{\alpha}}\mathbf{g}\big(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k]\big)\in\mathbb{C}^{(2JL+3)\times JMN_{\mathrm{U}}N_{\mathrm{C}}}$ is given by (56) in Appendix B.

Note that $f_{\mathrm{S}}(\boldsymbol{\alpha};\boldsymbol{\alpha}_{[i]}^{\natural}[k],\mathbf{h}_{[i]}^{\natural}[k])$ is different from the standard gradient-based surrogate function $f_{\mathrm{G}}\big(\boldsymbol{\alpha};\boldsymbol{\alpha}_{[i]}^{\natural}[k],\mathbf{h}_{[i]}^{\natural}[k]\big)$ given by (33) shown at the bottom of this page, that is usually adopted in conventional localization methods, for instance [33] and [34]. In addition, $f_{\mathrm{S}}(\boldsymbol{\alpha};\boldsymbol{\alpha}_{[i]}^{\natural}[k],\mathbf{h}_{[i]}^{\natural}[k])$ preserves the second-order structure of the original cost function in (24). Therefore, it can lead to a faster convergence speed than conventional gradient-based methods, which will be elaborated and confirmed later by theorem 2 and also verified by simulations.

At each iteration, since $\mathscr{A}_{\mathrm{SLAB},[i]}^{(\mathrm{LCE})}$ is strictly convex, we can give the closed-form expression of $\boldsymbol{\alpha}_{[i+1]}^{\circ}[k]$ as in (34) shown at the bottom of this page, where $\mathbf{p}_{[i]}[k]$ is its update direction that is different from the gradient.

Then, given $\mathbf{p}_{[i]}[k]$, the new update $\boldsymbol{\alpha}_{[i+1]}^{\natural}[k]$ is given by

$$\boldsymbol{\alpha}_{[i+1]}^{\natural}[k]=\boldsymbol{\alpha}_{[i]}^{\natural}[k]+\gamma_{[i]}\mathbf{p}_{[i]}[k], \tag{35}$$

where $\gamma_{[i]}$ is the step size subject to the Armijo rule (36) shown at the bottom of the next page, in which $f(\boldsymbol{\alpha};\mathbf{h}_{[i]}^{\natural}[k])=\|\mathbf{z}[k]-\mathbf{g}(\boldsymbol{\alpha};\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k])\|_2^2$ is the cost function depending on $\mathbf{h}_{[i]}^{\natural}[k]$, and $\nabla_{\boldsymbol{\alpha}}f(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k])=\nabla_{\boldsymbol{\alpha}}^{\mathrm{H}}\mathbf{g}(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k])(\mathbf{g}(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k])-\mathbf{z}[k])\in\mathbb{R}^{2JL+3}$ is the gradient vector of

$$f_{\mathrm{S}}\big(\boldsymbol{\alpha};\boldsymbol{\alpha}_{[i]}^{\natural}[k],\mathbf{h}_{[i]}^{\natural}[k]\big)=\big\|\mathbf{z}[k]-\mathbf{g}\big(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k]\big)-\nabla_{\boldsymbol{\alpha}}^{\mathrm{H}}\mathbf{g}\big(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k]\big)\big(\boldsymbol{\alpha}-\boldsymbol{\alpha}_{[i]}^{\natural}[k]\big)\big\|_2^2 \tag{32}$$

$$\begin{aligned}f_{\mathrm{G}}\big(\boldsymbol{\alpha};\boldsymbol{\alpha}_{[i]}^{\natural}[k],\mathbf{h}_{[i]}^{\natural}[k]\big)=\;&\|\mathbf{z}[k]-\mathbf{g}\big(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k]\big)\|_2^2+\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_{[i]}^{\natural}[k]\|_2^2\\&-2\big(\mathbf{z}[k]-\mathbf{g}\big(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k]\big)\big)^{\mathrm{H}}\nabla_{\boldsymbol{\alpha}}^{\mathrm{H}}\mathbf{g}\big(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k]\big)\big(\boldsymbol{\alpha}-\boldsymbol{\alpha}_{[i]}^{\natural}[k]\big).\end{aligned} \tag{33}$$

$$\boldsymbol{\alpha}_{[i+1]}^{\circ}[k]=\boldsymbol{\alpha}_{[i]}^{\natural}[k]+\underbrace{\big(\nabla_{\boldsymbol{\alpha}}^{\mathrm{H}}\mathbf{g}\big(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k]\big)\big)^{\dagger}\big(\mathbf{z}[k]-\mathbf{g}\big(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k]\big)\big)}_{\mathbf{p}_{[i]}[k]}. \tag{34}$$

the original cost function $f(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k])$ w.r.t. $\boldsymbol{\alpha}$ around $\boldsymbol{\alpha}=\boldsymbol{\alpha}_{[i]}^{\natural}[k]$. A legal $\gamma_{[i]}$ can be obtained by starting from a certain $\gamma_{[i]}>0$ and repeatedly trying $\gamma_{[i]}=\nu\gamma_{[i]}$ with $\nu\in(0,1)$ until (36) is satisfied.

The obtained solution in (34) combined with (35) and (25) will finally result in a closed-form undate of $\bar{\boldsymbol{\alpha}}[k]$ for the non-convex problem $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{LCE})}$, as summarized in Algorithm 2.

### C. Analysis of SLLS-Based Location Update

We shall analyze the convergence of SLLS-based localization and quantify the associated convergence rate.

*Theorem 1 (Convergence of SLLS-based Localization):* If the gradient matrix $\nabla_{\boldsymbol{\alpha}}\mathbf{g}(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k])$ is full-column-rank, then $\boldsymbol{\alpha}_{[i]}^{\natural}[k]$ will converge to a stationary point of the problem $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{LCE})}$ in (24), as the iteration number $i\to\infty$.

*Proof:* See Appendix C. ■

It should be noted that the update direction of the LCE-based localization, i.e., $\mathbf{p}_{[i]}[k]$, is different from that of the traditional gradient descent (GD)-based methods [33], [34], and this can achieve a second-order convergence rate as given in the following theorem. Intuitively, this is because the proposed surrogate function retains part of the second-order structure of the original objective function.

*Theorem 2 (Second-Order Convergence Rate of SLLS):* If the gradient matrix $\nabla_{\boldsymbol{\alpha}}\mathbf{g}(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k],\mathbf{w}[k])$ is full-column-rank and the initial point $\boldsymbol{\alpha}_{[0]}^{\natural}[k]$ is sufficiently close to an arbitrary locally optimal solution $\boldsymbol{\alpha}^{\bullet}[k]$ to $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{LCE})}$, then the convergence of the SLLS-based location parameter estimate error is quadratic, i.e.,

$$\left\|\boldsymbol{\alpha}^{\bullet}[k]-\boldsymbol{\alpha}_{[i+1]}^{\natural}[k]\right\|_2=\mathcal{O}\left(\left\|\boldsymbol{\alpha}^{\bullet}[k]-\boldsymbol{\alpha}_{[i]}^{\natural}[k]\right\|_2^2\right). \tag{37}$$

*Proof:* See Appendix D. ■

### D. Summary of LCE Algorithm

Given an initial point $\boldsymbol{\alpha}_{[0]}^{\sharp}[k]$, the inner iterations of LCE stage can find a stationary solution of $(\boldsymbol{\alpha}_{[i]}^{\sharp}[k],\mathbf{h}_{[i]}^{\sharp}[k])$ to $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{LCE})}$ in (24), as $i\to\infty$. Once inner iterations converge, the fresh estimate of $\boldsymbol{\alpha}$ and $\mathbf{h}[k]$ is determined by $\hat{\boldsymbol{\alpha}}[k]=\boldsymbol{\alpha}_{[i]}^{\natural}[k]$ and $\hat{\mathbf{h}}[k]=\mathbf{h}_{[i]}^{\natural}[k]$, respectively. Then, based on $\hat{\boldsymbol{\alpha}}[k]$ and the previous result $\bar{\boldsymbol{\alpha}}[k-1]$, the geometric parameter estimate will be updated to be $\bar{\boldsymbol{\alpha}}[k]$ as per (25). The pseudo-code of the LCE approach for parameter fusion is summarized in Algorithm 2.

In addition, at the initial time slot, we use multiple parameter samples to generate a good initial point $\boldsymbol{\alpha}_{[0]}^{\sharp}[1]$. Namely, we randomly generate multiple samples over the parameter space of $\boldsymbol{\alpha}$, try all samples and then choose the best sample with the minimum cost function value (see (24)) as the initial point. In the following time slot $k\geq 2$, the initial point is chosen to be result of the last time slot, i.e., $\boldsymbol{\alpha}_{[0]}^{\sharp}[k]=\bar{\boldsymbol{\alpha}}[k-1]$.

**Algorithm 2:** The Proposed SLLS-based LCE Algorithm.

**Input**: The measurements $\mathbf{z}[k]$ and $\bar{\boldsymbol{\alpha}}[k-1]$.
1: Initialize $\boldsymbol{\alpha}_{[i]}^{\natural}[k]$ (for $i=0$).
2: **While** not converge **do** (for $i=1:M_{\mathrm{I}}$)
3: Determine the channel state $\mathbf{h}_{[i]}^{\natural}[k]$ based on (30).
4: Find the best solution $\boldsymbol{\alpha}_{[i]}^{\natural}[k]$ to $\mathscr{A}_{\mathrm{SLAB},[i]}^{(\mathrm{LCE})}$ based on the closed-from update equation (35), given the previous-iteration result $\boldsymbol{\alpha}_{[i-1]}^{\natural}[k]$ and $\mathbf{h}_{[i-1]}^{\natural}[k]$.
5: **End**
6: Determine the fresh estimate $\hat{\boldsymbol{\alpha}}[k]=\Re\{\boldsymbol{\alpha}_{[i]}^{\natural}[k]\}$.
7: Update the location parameter $\bar{\boldsymbol{\alpha}}[k]$ by fusing $\hat{\boldsymbol{\alpha}}[k]$ and $\bar{\boldsymbol{\alpha}}[k-1]$ based on (25).
8: Determine the channel estimate $\hat{\mathbf{h}}[k]=\mathbf{h}_{[i]}^{\natural}[k]$.
**Output**: The geometric parameter estimate $\bar{\boldsymbol{\alpha}}[k]$ and channel estimate $\hat{\mathbf{h}}[k]$.

## V. The Proposed BFR Algorithm

In this section, we will elaborate the BFR algorithm to solve subproblem $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{BFR})}$ in the proposed SLAB scheme. Let's start to explicate the motivation and then we give an outline of the proposed BFR algorithm to explain how it addresses the challenge in subproblem $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{BFR})}$.

### A. Outline of the Proposed BFR Method

For the non-convex optimization in power allocation, conventional algorithms, e.g., [31], resort to optimizing the BF vector $\boldsymbol{\omega}$ to maximize the localization FIM, where the FIM constraint is formulated as an LMI form. However, due to the semidefinite positive approximation and rank relaxation, there will be a non-ignorable performance loss in these methods.

To address the con-convex problem $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{BFR})}$ in (28), in this section, we propose a novel successive concave optimization algorithm (named BFR) to iteratively optimize $\boldsymbol{\omega}$, which will achieve a stationary solution to $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{BFR})}$ (as elaborated later). We will show that at each time slot, the best BF vector for UE localization is the principal eigenvector (one with the largest eigenvalue) of the certain BFR feature matrix that is dependent on the LCE result $\bar{\boldsymbol{\alpha}}[k]$.

Specifically, let $\mathbf{w}_{[i]}[k]$ be the BF solution at the $i$th iteration of the BFR algorithm, and let $\kappa(\boldsymbol{\omega})$ be the cost function of $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{BFR})}$, i.e., $\kappa(\boldsymbol{\omega})=\mathrm{trace}((\bar{\boldsymbol{\mathcal{J}}}_{\boldsymbol{\beta}[k]}(\bar{\boldsymbol{\alpha}}[k],\boldsymbol{\Sigma};\boldsymbol{\omega}))^{-1})$, where $\bar{\boldsymbol{\mathcal{J}}}_{\boldsymbol{\beta}[k]}(\bar{\boldsymbol{\alpha}}[k],\boldsymbol{\Sigma};\boldsymbol{\omega})\in\mathbb{S}^{3JL+J+3}$ is the FIM of $\boldsymbol{\beta}[k]$, given by (38) shown at the bottom of the next page, where $\bar{\boldsymbol{\mathfrak{B}}}_{\mathbf{h}[k]}(\bar{\boldsymbol{\alpha}}[k],\boldsymbol{\Sigma};\mathbf{w}[0])$ and $\bar{\boldsymbol{\mathfrak{B}}}_{\mathbf{h}[k]}(\bar{\boldsymbol{\alpha}}[k],\boldsymbol{\Sigma};\mathbf{w}[0])$ are given by (23). The surrogate function at the $(i+1)$th iteration, denoted as $\kappa_{\mathrm{S}}(\boldsymbol{\omega};\mathbf{w}_{[i]}[k])$, is chosen to be a concave approximation of cost function $\kappa(\boldsymbol{\omega})$ around $\mathbf{w}_{[i]}[k]$, given by (39) shown at the bottom of the next page, where the informative

$$f(\boldsymbol{\alpha}_{[i]}^{\natural}[k]+\gamma_{[i]}\mathbf{p}_{[i]}[k];\mathbf{h}_{[i]}^{\natural}[k])\leq f(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k])+a\gamma_{[i]}\Re\{\nabla_{\boldsymbol{\alpha}}^{\mathrm{H}}f(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k])\mathbf{p}_{[i]}[k]\},\ \text{for some } a>0. \tag{36}$$

term $\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k]) \in \mathbb{S}^{N_{\rm B}}$, called the feature matrix of LCE, is given by (40) shown at the bottom of this page, where $C(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k])$ is a constant independent to $\boldsymbol{\omega}$, while $\boldsymbol{\mathcal{G}}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k]) \in \mathbb{S}^{N_{\rm B}}$ and $\boldsymbol{\mathcal{H}}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k]) \in \mathbb{S}^{N_{\rm B}}$ is given by (60) and (61), respectively, in Appendix E.

This surrogate $\kappa_{\rm S}(\boldsymbol{\omega};\mathbf{w}_{[i]}[k])$ in (39) is based on the first-order Taylor expansion of $\kappa(\boldsymbol{\omega})$ around $\bar{\boldsymbol{\mathcal{J}}}_{\boldsymbol{\beta}[k]}(\bar{\boldsymbol{\alpha}}[k],\boldsymbol{\Sigma};\boldsymbol{\omega}) = \bar{\boldsymbol{\mathcal{J}}}_{\boldsymbol{\beta}[k]}(\bar{\boldsymbol{\alpha}}[k],\boldsymbol{\Sigma};\mathbf{w}_{[i]}[k])$. Its properties, which are useful for understanding the structure of the proposed BFR algorithm and its convergence behavior, are summarized as follows.

*Lemma 2:* [Properties of $\kappa_{\rm S}(\boldsymbol{\omega};\mathbf{w}_{[i]}[k])$] The surrogate function $\kappa_{\rm S}(\boldsymbol{\omega};\mathbf{w}_{[i]}[k])$ is locally tight around $\boldsymbol{\omega} = \mathbf{w}_{[i]}[k]$, i.e., (i) $\kappa_{\rm S}(\mathbf{w}_{[i]}[k];\mathbf{w}_{[i]}[k]) = \kappa(\mathbf{w}_{[i]}[k])$, and (ii) its gradient vector satisfies $\nabla_{\boldsymbol{\omega}}\kappa_{\rm S}(\mathbf{w}_{[i]}[k];\mathbf{w}_{[i]}[k]) = \nabla_{\boldsymbol{\omega}}\kappa(\mathbf{w}_{[i]}[k])$, where $\nabla_{\boldsymbol{\omega}_j[n,m]}\kappa(\mathbf{w}_{[i]}[k]) \in \mathbb{C}^{N_{\rm B}}$ is given by (46), $\forall j,n,m$.

*Proof:* These properties can be directly verified based on (28), (39) and (46), by letting $\boldsymbol{\omega} = \mathbf{w}_{[i]}[k]$. ■

Note that we have used a concave surrogate function, instead of a convex one as in conventional SCA methods. This is because $\kappa_{\rm S}(\boldsymbol{\omega};\mathbf{w}_{[i]}[k])$ preserves some structure of the original problem $\mathscr{A}^{(\rm BFR)}_{\rm SLAB}$. Hence, it is a better approximation of $\kappa(\boldsymbol{\omega})$ with faster convergence. Moreover, we can obtain a closed-form BF update in (44) even if $\kappa_{\rm S}(\boldsymbol{\omega};\mathbf{w}_{[i]}[k])$ is non-convex. Thus, the resulting algorithm has a low computational cost.

### B. Determination of the BF Update Direction

At the $(i+1)$th iteration of the BFR algorithm, given the previous BF vector $\mathbf{w}_{[i]}[k]$, the following minimization problem is first solved to obtain the candidate update $\mathbf{w}^{\sharp}_{[i+1]}[k]$:

$$\mathscr{A}^{(\rm BFR)}_{{\rm SLAB},[i+1]}:\ \mathbf{w}^{\sharp}_{[i+1]}[k] = \arg\min_{\boldsymbol{\omega}} \kappa_{\rm S}(\boldsymbol{\omega};\mathbf{w}_{[i]}[k]) \quad \text{s.t. } \|\boldsymbol{\omega}_j[n,m]\|_2 \le 1, \forall j,n,m, \tag{41}$$

and $\mathbf{w}^{\sharp}_{[i+1]}[k]$ is used to determine a feasible update direction vector $\mathbf{d}_{[i+1]}[k] \in \mathbb{C}^{N_{\rm B}JN_{\rm C}M}$ for BF as follows,

$$\mathbf{d}_{[i+1]}[k] = \mathbf{w}^{\sharp}_{[i+1]}[k] - \mathbf{w}_{[i]}[k]. \tag{42}$$

Since the channel coefficients $\{h_{l,j}[k]|\forall l,\forall j\}$ are identically and independently distributed, $\mathscr{A}^{(\rm BFR)}_{{\rm SLAB},[i+1]}$ in (41) can be decomposed into a series of subproblems $\{\mathscr{A}^{(\rm BFR)}_{j,[i+1]}|j=1:J\}$ associated with all BSs, as given in (43) shown at the bottom of this page. As a result, the BFR problem $\mathscr{A}^{(\rm BFR)}_{{\rm SLAB},[i+1]}$ associated with various BSs can be solved in a parallel manner, which will significantly reduce the calculation time. In addition, although the subproblem $\mathscr{A}^{(\rm BFR)}_{j,[i+1]}$ is non-convex w.r.t. $\boldsymbol{\omega}$, we find that it is identical to Rayleigh quotient maximization [35], and thus each component $\mathbf{w}^{\sharp}_{j,[i+1]}[k,n,m]$ of $\mathbf{w}^{\sharp}_{[i+1]}[k]$ in (41) is exactly the principal eigenvector (the eigenvector associated with the largest eigenvalue) of $\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k])$.

### C. Update of the BF Vector

Once obtaining $\mathbf{d}_{[i+1]}[k]$ via (42), the BF vector will be updated based on the Armijo rule as follows,

$$\mathbf{w}_{j,[i+1]}[k,n,m] = \mathbf{w}_{j,[i]}[k,n,m] + \gamma'_{[i]}\mathbf{d}_{j,[i+1]}[k,n,m], \tag{44}$$

where $\gamma'_{[i]} > 0$ is the step length determined by the Armijo rule [40] in (45) shown at the bottom of the next page, where $\nabla_{\boldsymbol{\omega}_j[n,m]}\kappa(\mathbf{w}_{j,[i]}[k,n,m])$ is the derivative of $\kappa(\boldsymbol{\omega}_j[n,m])$ w.r.t. $\boldsymbol{\omega}_j[n,m]$ at $\boldsymbol{\omega}_j[n,m] = \mathbf{w}_{j,[i]}[k,n,m]$, which is given by (46) shown at the bottom of the next page, and $\mathbf{I}_{N_{\rm B}}$ is the $N_{\rm B}\times N_{\rm B}$ identity matrix. Specifically, starting with a certain step size $\gamma'_{[i]} > 0$, the Armijo rule repeatedly decreases $\gamma'_{[i]}$ as $\gamma'_{[i]} = \nu\gamma'_{[i]}$ for some $\nu \in (0,1)$ until the condition in (45) is satisfied.

$$\bar{\boldsymbol{\mathcal{J}}}_{\boldsymbol{\beta}[k]}(\bar{\boldsymbol{\alpha}}[k],\boldsymbol{\Sigma};\boldsymbol{\omega}) = \sigma^{-2}\sum_{r,n,m}\begin{bmatrix}\mathbf{K}^{(r)}_{n,m}\left(\bar{\boldsymbol{\alpha}}[k];\boldsymbol{\omega}\right)\boldsymbol{\Sigma}\left(\mathbf{K}^{(r)}_{n,m}\left(\bar{\boldsymbol{\alpha}}[k];\boldsymbol{\omega}\right)\right)^{\rm H} & \mathbf{0}_{(2JL+3)\times J(L+1)} \\ \mathbf{0}^{\rm H}_{(2JL+3)\times J(L+1)} & \left(\mathbf{G}^{(r)}_{n,m}(\bar{\boldsymbol{\alpha}}[k];\boldsymbol{\omega})\right)^{\rm H}\mathbf{G}^{(r)}_{n,m}(\bar{\boldsymbol{\alpha}}[k];\boldsymbol{\omega})\end{bmatrix}. \tag{38}$$

$$\kappa_{\rm S}(\boldsymbol{\omega};\mathbf{w}_{[i]}[k]) = C(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k]) - \sum_{\substack{n\in\Theta_j,\\ j=1:J,\\ m=1:M}}\boldsymbol{\omega}^{\top}_j[n,m]\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k])\,\boldsymbol{\omega}^{*}_j[n,m]. \tag{39}$$

$$\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k]) = \underbrace{\boldsymbol{\mathcal{G}}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k])}_{\text{UE localization feature matrix}} + \underbrace{\boldsymbol{\mathcal{H}}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k])}_{\text{Channel estimate feature matrix}}. \tag{40}$$

$$\mathscr{A}^{(\rm BFR)}_{j,[i+1]}:\ \mathbf{w}^{\sharp}_{j,[i+1]}[k,n,m] = \arg\max_{\boldsymbol{\omega}_j[n,m]}\frac{\boldsymbol{\omega}^{\top}_j[n,m]\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k])\,\boldsymbol{\omega}^{*}_j[n,m]}{\|\boldsymbol{\omega}_j[n,m]\|^2_2}. \tag{43}$$

**Algorithm 3:** The Proposed BFR Algorithm.

**Input**: LCE result $\bar{\boldsymbol{\alpha}}[k]$ and channel variance matrix $\boldsymbol{\Sigma}$.
1: Determine an initial point $\hat{\mathbf{w}}_{[0]}[k]$,
2: **While** not converge **do** (for $i = 1 : M_I$)
3: Determine the feature matrix $\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k]; \mathbf{w}_{[i]}[k])$ based on $\mathbf{w}_{[i]}[k]$ and $\bar{\boldsymbol{\alpha}}[k]$, $\forall n, j$.
4: Determine $\mathbf{w}^{\sharp}_{j,[i]}[k,n,m]$ based on the spectral decomposition of $\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k]; \mathbf{w}_{[i]}[k])$, $\forall n, m, j$.
5: Determine the BF direction $\mathbf{d}_{j,[i+1]}[k,n,m]$ based on (42), $\forall n, m, j$.
6: Determine the step length $\gamma'_{[i]}$ as per (45).
7: Update the BF solution $\mathbf{w}_{[i]}[k]$ based on (44).
8: **End**
9: Return $\mathbf{w}[k+1] = \mathbf{w}_{[i]}[k]$ for the next round of LCE.
**Output**: The best BF choice $\mathbf{w}[k+1]$.

This will finally result in a closed-form update of $\mathbf{w}[k]$ for the non-convex BFR problem $\mathscr{A}^{(\mathrm{BFR})}_{\mathrm{SLAB}}$ in (28), via combining with (44), as summarized in Algorithm 3.

If $\mathbf{d}_{[i+1]}[k]$ is a descent direction, (i.e., the inner product $\nabla^{\mathrm{H}}_{\boldsymbol{\omega}} \kappa\big(\mathbf{w}_{[i]}[k]\big)\mathbf{d}_{[i+1]}[k] < 0$ for any non-stationary $\mathbf{w}_{[i]}[k]$), this Armijo rule will ensure a sufficient reduction of cost function value such that the obtained BF solution $\mathbf{w}_{[i]}[k]$ converges to a stationary solution to the overall BFR problem $\mathscr{A}^{(\mathrm{BFR})}_{\mathrm{SLAB}}$. However, it is highly non-trivial to prove that $\mathbf{d}_{[i+1]}[k]$ is a descent direction when the surrogate function is non-convex. This challenge will be addressed in Section VI.

### D. Summary of the Overall BFR Algorithm

The pseudo-code of the novel BFR algorithm is summarized in Algorithm 3. Start from an initial point $\mathbf{w}_{j,[0]}[k,n,m]$, $\forall j, n, m$. The proposed BFR algorithm will first determine $\mathbf{w}^{\sharp}_{j,[i+1]}[k,n,m]$ based on the spectral decomposition of feature matrix $\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k]; \mathbf{w}_{[i]}[k])$ dependent on the previous BF solution $\mathbf{w}_{[i]}[k]$, and then determine the feasible BF update direction $\mathbf{d}_{j,[i+1]}[k,n,m]$ as per (42), and finally update the BF vector as per (44). Once obtaining the new BF choice $\mathbf{w}[k+1]$, each BS will transmit the new symbols based on $\mathbf{w}[k+1]$, and thus the new measurement $\mathbf{z}_{(k+1)}$ used for the next LCE stage is dependent on $\mathbf{w}[k+1]$.

*The Characteristics of our BFR Solution*: The BFR-related feature matrix $\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k]; \mathbf{w}_{[i]}[k])$ is the combination of two individual features $\boldsymbol{\mathcal{G}}_{j,n}(\bar{\boldsymbol{\alpha}}[k]; \mathbf{w}_{[i]}[k])$ and $\boldsymbol{\mathcal{H}}_{j,n}(\bar{\boldsymbol{\alpha}}[k]; \mathbf{w}_{[i]}[k])$ w.r.t. the UE localization and channel estimate, respectively, as given in (40). This structure explicitly shows how the physical nature of UE localization and channel estimate (characterized by their respective feature matrices) affect the BFR result. In addition, we can also see from (40) that the localization-based BF optimization and the channel-estimate-based BF optimization in the BFR problem can be decoupled from each other due to the problem-specific surrogation function design in (39). Furthermore, the obtained BF solution is adaptive to the pilot symbols. In addition, unlike the traditional gradient-descent algorithm where the Taylor expansion is directly applied w.r.t. the variable (say $\boldsymbol{\omega}$), we apply the first-order expansion w.r.t. the information matrix $\bar{\boldsymbol{\mathcal{J}}}_{\boldsymbol{\beta}[k]}(\bar{\boldsymbol{\alpha}}[k], \boldsymbol{\Sigma}; \boldsymbol{\omega})$, rather than $\boldsymbol{\omega}$. Hence, our surrogate function $\kappa_{\mathrm{S}}(\bullet)$ preserves the inherent structure of $\bar{\boldsymbol{\mathcal{J}}}_{\boldsymbol{\beta}[k]}(\bar{\boldsymbol{\alpha}}[k], \boldsymbol{\Sigma}; \boldsymbol{\omega})$. This will lead to a more accurate approximation and thus result in a fast convergence of the BF optimization.

## VI. Convergence Analysis

By using the proposed SLAB scheme above, we can obtain two sequences $\{\bar{\boldsymbol{\alpha}}[k] | \forall k = 1 : K\}$ and $\{\mathbf{w}[k] | \forall k = 1 : K\}$ of the UE location parameter and the BF vector, respectively. In this section, we shall establish their convergence behaviors, which justifies the proposed SLAB scheme.

### A. Challenges and Assumptions

**Challenges.** It is not easy to establish the convergence of the SLAB scheme, due to the following challenges.

- *Uniqueness of the Globally Optimal Solution*: The uniqueness of the globally optimal solution of $\boldsymbol{\beta}[k]$ to the LCE problem $\mathscr{A}^{(\mathrm{LCE})}_{\mathrm{SLAB}}$ is not clear, due to the nonlinear system model. This issue concerns the solvability of the mmWave MIMO-based localization problem and the stability of the obtained BF sequence.
- *Coupling Dynamics*: In the proposed SLAB scheme, the channel estimate and BF refinement are coupled with each other. Thus, the coupled dynamics of $\mathbf{h}[k]$ and $\mathbf{w}[k]$ will affect the convergence of the SLAB scheme.

**Assumptions.** To facilitate the SLAB convergence analysis, we have the following assumptions on the SLAB system.

(A1) $N_{\mathrm{C}} N_{\mathrm{U}} JM \geq 3JL + J + 3$ is satisfied.
(A2) Gradient matrix $\nabla_{\boldsymbol{\beta}[k]} \mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k]; \mathbf{w}[k])$ is full-column-rank, where $\boldsymbol{\beta}^{\mathrm{true}}[k]$ denotes the true value of $\boldsymbol{\beta}[k]$.
(A3) The mean of measurement error $\boldsymbol{\epsilon}[k]$ is zero.

The first assumption means the number of pilots should be not less than the number of unknown parameters in the SLAB scheme, which is satisfied by usual mmWave MIMO systems. The second assumption means the rank of the gradient matrix should be equal to the number of unknown parameters, which

$$\kappa\big(\mathbf{w}_{j,[i]}[k,n,m] + \gamma'_{[i]}\mathbf{d}_{j,[i+1]}[k,n,m]\big) \leq \kappa\big(\mathbf{w}_{j,[i]}[k,n,m]\big) + a\gamma'_{[i]}\Re\{\nabla^{\mathrm{H}}_{\boldsymbol{\omega}_j[n,m]} \kappa\big(\mathbf{w}_{j,[i]}[k,n,m]\big)\mathbf{d}_{j,[i+1]}[k,n,m]\}. \tag{45}$$

$$\nabla_{\boldsymbol{\omega}_j[n,m]} \kappa\big(\mathbf{w}_{j,[i]}[k,n,m]\big) = -2\frac{\big(\|\mathbf{w}_{j,[i]}[k,n,m]\|_2^2 \mathbf{I}_{N_{\mathrm{B}}} - \mathbf{w}^*_{j,[i]}[k,n,m]\mathbf{w}^{\top}_{j,[i]}[k,n,m]\big)\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k]; \mathbf{w}_{[i]}[k])\, \mathbf{w}^*_{j,[i]}[k,n,m]}{\|\mathbf{w}_{j,[i]}[k,n,m]\|_2^4}. \tag{46}$$

is usually true for the mmWave system (due to A1) and can be verified by numerical results.

We also assume a global search method, e.g., a branch-and-bound algorithm [32], is used for the LCE problem such that a globally optimal solution to $\mathscr{A}_{\rm SLAB}^{(\rm LCE)}$ is obtained at each LCE step. It should be noted that although the convergence is established for the case of a high-complexity global LCE algorithm, simulation results show that the proposed SLAB scheme with the low-cost SLLS-based LCE algorithm in Section IV still achieves a large performance gain over baselines.

### B. Convergence Behavior

*1) Convergence of LCE:* We give the following Lemma 3 to address the first challenge above.

*Lemma 3 (Uniqueness of the Globally Optimal Solution):* If A1 is satisfied, there is a unique globally optimal solution $(\boldsymbol{\alpha}^\star[k], \mathbf{h}^\star[k])$ to $\mathscr{A}_{\rm SLAB}^{(\rm LCE)}$ in (24) at each LCE stage.

*Proof:* The proof is given in Appendix F. ■

Lemma 3 ensures the mmWave MIMO-based localization problem is solvable. We further have the following theorem to quantify the order of $\|\boldsymbol{\beta}^\star[k] - \boldsymbol{\beta}^{\rm true}[k]\|_2$ w.r.t. $\|\boldsymbol{\epsilon}[k]\|_2$.

*Theorem 3 (Bounded LCE Error):* Assume A1–A2 are satisfied. At each time slot $k$, the LCE error is bounded from above, i.e., $\|\boldsymbol{\beta}^\star[k] - \boldsymbol{\beta}^{\rm true}[k]\|_2 \sim \mathcal{O}(\frac{\|\boldsymbol{\epsilon}[k]\|_2}{\|\nabla_{\boldsymbol{\beta}[k]}^{\rm H}\mathbf{g}(\boldsymbol{\beta}[k],\mathbf{w}[k])\|_2})$.

*Proof:* The proof is presented in Appendix G. ■

This means, if those conditions are satisfied, the LCE error is finite and proportional to the noise power.

*2) Convergence of BFR:* Since the first-order optimality condition [36] presents the necessary condition of the optimal solution, we use the mismatch of this condition w.r.t. the obtained BF solution to characterize its optimality and to quantify the order of this mismatch w.r.t. measurement errors. Prior to this, we give a theorem to show that the obtained BF solution at each time slot $k$ is a stationary solution to the BFR problem $\mathscr{A}_{\rm SLAB}^{(\rm BFR)}$, for a given $\bar{\boldsymbol{\alpha}}[k]$.

*Theorem 4 (Stationary Solution to BFR Problem):* Assume A1–A3 are satisfied. At each time slot $k$, the BFR algorithm converges to a stationary solution $\mathbf{w}[k]$ to problem $\mathscr{A}_{\rm SLAB}^{(\rm BFR)}$, as the iteration number $i \to \infty$.

*Proof:* The proof is presented in Appendix H. ■

Due to the inevitable LCE error in $\bar{\boldsymbol{\alpha}}[k]$, the optimality of the BF solution $\mathbf{w}_{[i]}[k]$ will be affected, which is quantified by the first-order optimality condition mismatch as follows.

*Theorem 5 (Bounded Optimality Condition Mismatch):* If A1–A3 are satisfied, at time slot $k$, the obtained stationary BF solution $\mathbf{w}[k]$ approximately satisfies the first-order optimality condition associated with $\boldsymbol{\alpha}^{\rm true}$, with a deviation $W_{\rm mis}$:

$$\Re\{\nabla_{\boldsymbol{\omega}}^{\rm H}\mathrm{trace}\big(\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}\big(\boldsymbol{\alpha}^{\rm true},\boldsymbol{\Sigma};\mathbf{w}[k]\big)\big)(\boldsymbol{\omega}-\mathbf{w}[k])\} \geq W_{\rm mis}$$

holds $\forall\boldsymbol{\omega}$ such that $\|\boldsymbol{\omega}_j[n,m]\|_2 \leq 1, \forall j,n,m$; and meanwhile the optimality deviation follows $W_{\rm mis} \sim \mathcal{O}(\|\boldsymbol{\epsilon}[k]\|_2)$, where $\nabla_{\boldsymbol{\omega}}\mathrm{trace}\big(\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}\big(\bar{\boldsymbol{\alpha}}[k],\boldsymbol{\Sigma};\mathbf{w}[k]\big)\big) \in \mathbb{C}^{N_{\rm B}JN_{\rm C}M}$ is the derivative of $\mathrm{trace}\big(\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}\big(\bar{\boldsymbol{\alpha}}[k],\boldsymbol{\Sigma};\boldsymbol{\omega}\big)\big)$ w.r.t. $\boldsymbol{\omega}$ at $\boldsymbol{\omega}=\mathbf{w}[k]$.

*Proof:* The proof is presented in Appendix I. ■

This implies that the UE-location-error-caused beam misalignment will be gradually mitigated, due to the gradually decreased UE location estimate error via parameter filtering. For a finite UE location error (or equivalently a finite $\|\boldsymbol{\epsilon}[k]\|_2$), the beam misalignment is bounded. Thus, the proposed BFR algorithm will achieve a robust BF solution against the measurement noise. Therefore, the overall convergence of the SLAB scheme is established.

## VII. Numerical Results

In this section we shall provide numerical results to demonstrate the performance of the proposed SLLS algorithm for UE localization and also to verify the performance gain of the proposed BFR algorithm for BF optimization.

### A. Simulation Settings

We choose the number of BSs to be $J=3$ and the number of subcarriers to be $N_{\rm C}'=30$. We set $L=1$, $M=20$, SNR $=20$ dB, $N_{\rm B}=N_{\rm U}=4$, carrier frequency $f_{\rm C}=6$ GHz, sampling period $T_s=10$ ns and light speed $c=3*10^8$ m/s. Thus, $\lambda_n$ and $d_{\rm A}$ can be determined via $\lambda_n=\frac{c}{\frac{n}{N_{\rm C}'T_s}+f_{\rm C}}$ and $d_{\rm A}=c/f_{\rm C}/2$, respectively. We assume the locations of UE and BSs are at random within a squared area of $10^3\times10^3$ m$^2$, and their orientation angles are also at random. We assume a simple path loss model for each channel, i.e., $h_{l,j}=\frac{h_{l,j}'}{\wp_{l,j}^3}$, where $h_{l,j}'\sim\mathcal{CN}(0,1)$ is the small-scale fading and $\wp_{l,j}$ is the path length, namely, $\wp_{0,j}=\|\mathbf{x}-\mathbf{u}_j\|_2$ for $l=0$ and $\wp_{l,j}=\|\mathbf{u}_j-\mathbf{v}_{l,j}\|_2+\|\mathbf{x}-\mathbf{v}_{l,j}\|_2$ for $l=1:L$. Based on this model, we have that $\boldsymbol{\Sigma}=\mathbb{E}\{\mathbf{h}[k]\mathbf{h}^\top[k]\}=\mathbf{I}_J\otimes\mathrm{diag}\left[\eta_l^2\,|\forall l=0:L\right]$, where $\eta_l=\wp_{l,j}^{-3}$, for each realization of BS locations and scatterer locations in simulations.[6] For LCE, we consider the following algorithms as baselines.

- *(GD-based LCE [33]):* This directly updates the unknown parameters by using the gradient of the LCE problem.
- *(LS-based LCE [34]):* This uses the gradient with an Armijo-type line search (LS) to update the estimates of unknown parameters.

For the BFR, we consider the following baseline methods.

- *(GD-based BFR [33]):* This directly uses the gradient of the BFR problem to update BF vectors.
- *(LMI-based BFR [31]):* This optimizes BF vectors to maximize the localization FIM, (i.e., inverse CRLB) by formulating the FIM constraint as an LMI form.

### B. Simulation Results

*1) The Achieved RMSE of LCE:* The root mean squared errors (RMSEs), at the first time slot with random BF vector, achieved by various localization algorithms are presented in Fig. 4. We can see that the proposed SLLS algorithm can achieve a localization error of 0.028 [m]. In addition, the SLLS algorithm has a faster convergence than baseline algorithms, due to our problem-specific parameter estimate rule design.

[6] We have assumed in this paper that the channel variance matrix $\boldsymbol{\Sigma}$ is known and we adopt the value of $\boldsymbol{\Sigma}$ conditioned on the path loss $\wp_{l,j}^{-3}$, while the instantaneous channel vector $\mathbf{h}[k]$ and scatterer locations are unknown and need to be estimated. In practice, the state of $\boldsymbol{\Sigma}$ can be estiamted by using the variance matrix estimation method [37], [38] or alternatively using the stochastic geometry modeling-based analystical analysis [39].

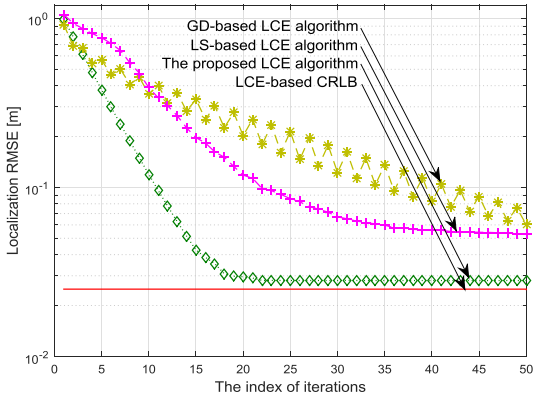



Fig. 4. The achieved localization RMSEs of various algorithms.

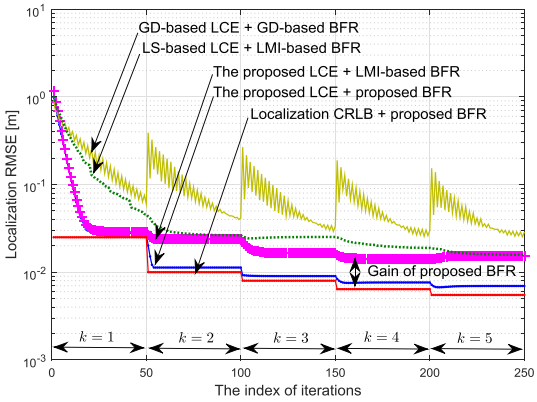



Fig. 5. The achieved localization error with various BFR strategies.

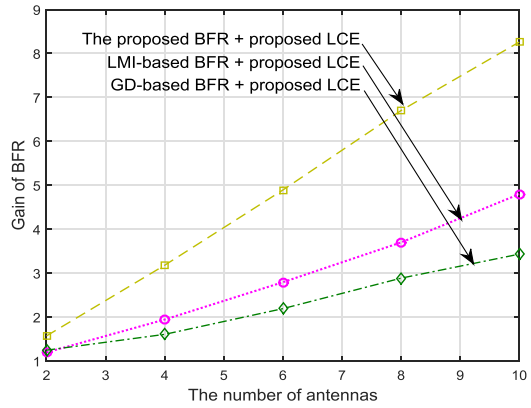



Fig. 6. The achieved BFR gain v.s. the number of antennas (i.e. $N_{\mathrm{B}}$).

*2) Gain of BFR:* As shown in Fig. 5, there will be a localization error reduction owing to BFR. In addition, the proposed BFR algorithm can achieve a satisfactory performance gain in localization accuracy. The achieved BFR gains v.s. the number of transceiver antennas (assuming $N_{\mathrm{U}} = N_{\mathrm{B}}$) are shown in Fig. 6, where the BFR gain is defined as the ratio between the initial CRLB (without the BFR process) and the minimized CRLB (via using a certain BFR algorithm). As expected, the more antennas lead to the larger BFR gain. In addition, due to the non-convex rank-one constraint in the LMI-based BFR baseline method, there will be some performance loss in their BF results. In addition, due to our problem-specific surrogate function design, in which the inherent structure of localization information matrix is preserved, our obtained BF update is more informative than the gradient descent-based baseline, and hence achieves a larger localization performance gain. The proposed BFR algorithm thus outperforms the LMI- and GD-based methods, as shown in Fig. 6.

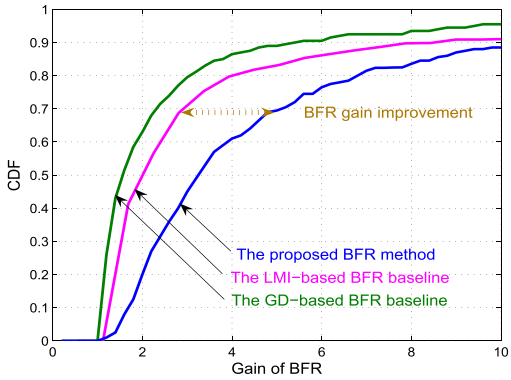



Fig. 7. The CDF of the achieved localization performance gain from BFR.

In addition, we evaluate the achieved BFR gain of our proposed BFR algorithm and the BFR baseline methods in diverse scenarios. Specifically, the SNR is set as a random value ranging from $-20$ dB to 60 dB in simulations. In addition, the BS locations are uniformly distributed within the localization area of $10^3 \times 10^3$ m$^2$, and the BS orientation angles are also uniformly setup within $[0, 2\pi)$, while $N_{\mathrm{B}}$ and $N_{\mathrm{U}}$ are fixed at 6. The settings of other parameters are the same as those in Section VII-A. The cumulative distribution function (CDF) of the achieved BFR gains by different BFR methods is presented in Fig. 7. It is shown that the overall BFR gain of the proposed BFR method is larger than that of BFR baselines under the diverse parameter settings given above, which implies to some extent that the proposed BFR method can also achieve a better performance in some worst-case scenarios with extremely poor SNR conditions or BS deployment.

## VIII. Conclusions

In this paper, localization-oriented BF for mmWave MIMO systems is studied. A novel SLAB scheme is proposed to determine the best BF vector for UE localization and to simultaneously achieve the joint UE localization and channel estimate. A closed-form CRLB is obtained to provide a performance benchmark for UE localization and is also used for BF optimization. The convergence of the proposed SLAB scheme has been established. It is shown that the proposed SLAB scheme can achieve a huge performance gain over existing localization/BFR approaches.

## Appendix A
## Proof of Lemma 1

As per the estimate theory [30], the LCE-based CRLB $\mathfrak{B}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha}, \mathbf{h}[k]; \boldsymbol{\omega})$ will be finally given by

$$\mathfrak{B}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha}, \mathbf{h}[k]; \boldsymbol{\omega}) = \left(\boldsymbol{\mathcal{J}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha}, \mathbf{h}[k]; \boldsymbol{\omega})\right)^{-1}, \quad (47)$$

where $\boldsymbol{\mathcal{J}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\mathbf{h}[k];\boldsymbol{\omega})$ is the FIM given by (48) shown at the bottom of the page, and the derivative $\mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega}) \in \mathbb{C}^{(2JL+3)\times J(L+1)}$ is given by

$$\mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega}) = [\boldsymbol{\varsigma}_{l,j}^{(r)}[n,m]|\forall l = 0:L, \forall j = 1:J]. \quad (49)$$

In addition, vector $\boldsymbol{\varsigma}_{l,j}^{(r)}[n,m] \in \mathbb{C}^{2JL+3}$ is structured as

$$\boldsymbol{\varsigma}_{l,j}^{(r)}[n,m] = \begin{bmatrix} \sum\limits_{t=1:N_{\mathrm{B}}} \omega_j^{(t)}[n,m]\mu_{l,j,n}^{(r,t)*}\boldsymbol{\varrho}_{l,j,n}^{(r,t)} \\ \sum\limits_{t=1:N_{\mathrm{B}}} \omega_j^{(t)}[n,m]\mu_{l,j,n}^{(r,t)*}\rho_{l,j,n}^{(r,t)} \\ \sum\limits_{t=1:N_{\mathrm{B}}} \omega_j^{(t)}[n,m]\mu_{l,j,n}^{(r,t)*}\breve{\boldsymbol{\upsilon}}_{l,j,n}^{(r,t)} \end{bmatrix},$$

where $\mu_{l,j,n}^{(r,t)}$ is given by (19), and $\omega_j^{(t)}[n,m]$ is the $t$th element of $\boldsymbol{\omega}_j[n,m]$. Furthermore, $\boldsymbol{\varrho}_{l,j,n}^{(r,t)} \in \mathbb{R}^2$ and $\rho_{l,j,n}^{(r,t)} \in \mathbb{R}$ are given by (50) and (51) shown at the bottom of the page, respectively, and $\breve{\boldsymbol{\upsilon}}_{l,j,n}^{(r,t)} \in \mathbb{R}^{2JL}$ is given by

$$\left[\breve{\boldsymbol{\upsilon}}_{l,j,n}^{(r,t)}\right]_{2l'-1:2l'} = \begin{cases} \boldsymbol{\upsilon}_{l,j,n}^{(r,t)}, & \text{if } (l' \bmod J) \doteq l, \\ \mathbf{0}_{2\times 1}, & \text{otherwise,} \end{cases} \quad (52)$$

$\forall l' = 1:JL$, where $\boldsymbol{\upsilon}_{l,j,n}^{(r,t)} \in \mathbb{R}^2$ is given by (53) shown at the bottom of this page, and $\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega}) \in \mathbb{C}^{J\times J(L+1)}$ is given by

$$\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega}) = \mathrm{diag}[\mathbf{g}_j^{(r)\mathrm{H}}[n,m]|\forall j = 1:J], \quad (54)$$

where $\mathbf{g}_j^{(r)}[n,m]$ is given by (16).

Based on (48), $\mathbb{E}_{\mathbf{h}[k]}\{\mathfrak{B}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\mathbf{h}[k];\boldsymbol{\omega})\}$ follows (55) shown at the bottom of the next page, as per Jensen's inequality for the inverse function, where we have considered that $\mathbf{h}[k]$ is independent to $\boldsymbol{\alpha}$ and has zero-mean with variance matrix $\boldsymbol{\Sigma}$. Thus, $\mathbb{E}_{\boldsymbol{\epsilon}[k]}\{\|\boldsymbol{\beta}[k]-\hat{\boldsymbol{\beta}}[k]\|_2^2\} \geq \mathrm{trace}\big(\mathfrak{B}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\mathbf{h}[k];\boldsymbol{\omega})\big)$, and Lemma 1 is proved.

## APPENDIX B
## GRADIENT MATRIX

The gradient matrix in (32) is given by

$$\nabla_{\boldsymbol{\alpha}}\mathbf{g}\big(\boldsymbol{\alpha}_{[i]}^{\natural}[k];\mathbf{h}_{[i]}^{\natural}[k]\big) = \begin{bmatrix} \sqrt{N_{\mathrm{B}}N_{\mathrm{U}}}\mathbf{D}_{\mathbf{x}}^{\mathrm{H}}(\boldsymbol{\alpha})\mathbf{B}^{\mathrm{H}}(\mathbf{h}_{[i]}^{\natural}[k]) \\ \sqrt{N_{\mathrm{B}}N_{\mathrm{U}}}\mathbf{d}_{\vartheta}^{\mathrm{H}}(\boldsymbol{\alpha})\mathbf{B}^{\mathrm{H}}(\mathbf{h}_{[i]}^{\natural}[k]) \\ \sqrt{N_{\mathrm{B}}N_{\mathrm{U}}}\mathbf{D}_{\mathbf{v}}^{\mathrm{H}}(\boldsymbol{\alpha})\mathbf{B}^{\mathrm{H}}\big(\mathbf{h}_{[i]}^{\natural}[k]\big) \end{bmatrix} \quad (56)$$

and each term is elaborated as follows.

Firstly, $\mathbf{B}(\mathbf{h}_{[i]}^{\natural}[k]) \in \mathbb{C}^{JMN_{\mathrm{U}}N_{\mathrm{C}}\times JMN_{\mathrm{U}}N_{\mathrm{C}}N_{\mathrm{B}}(L+1)}$ is given by $\mathbf{B}(\mathbf{h}_{[i]}^{\natural}[k]) = \mathrm{diag}[\mathbf{b}_j^{\top}[k,n,m]|\forall r,n,j,m]$, where $\mathbf{b}_j[k,n,m] = (\mathbf{I}_{N_{\mathrm{B}}} \otimes \boldsymbol{\mathcal{H}}_{j,[i]}^{\natural}[k])(\mathbf{w}_j[k,n,m] \otimes \mathbf{1}_{(L+1)})$ and $\boldsymbol{\mathcal{H}}_{j,[i]}^{\natural}[k] = \mathrm{diag}[h_{l,j,[i]}^{\natural}[k]|\forall l = 0:L]$, in which $h_{l,j,[i]}^{\natural}[k]$ is the $(l,j)$th element of $\mathbf{h}_{[i]}^{\natural}[k]$, and $\mathbf{1}_{(L+1)}$ is the $(L+1)$-dimensional full-one vector.

$$\boldsymbol{\mathcal{J}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha},\mathbf{h}[k];\boldsymbol{\omega}) = \sigma^{-2}\sum_{r,n,m}\begin{bmatrix} \mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega})\mathbf{h}^*[k]\mathbf{h}^{\top}[k]\big(\mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega})\big)^{\mathrm{H}} & \mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega})\mathbf{h}^*[k]\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega}) \\ \big(\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega})\big)^{\mathrm{H}}\mathbf{h}^{\top}[k]\big(\mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega})\big)^{\mathrm{H}} & \big(\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega})\big)^{\mathrm{H}}\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha};\boldsymbol{\omega}) \end{bmatrix}. \quad (48)$$

$$\boldsymbol{\varrho}_{l,j,n}^{(r,t)} = \begin{cases} \mathrm{j}2\pi\frac{n}{cN_{\mathrm{C}}'T_s}\frac{\mathbf{x}-\mathbf{u}_j}{\|\mathbf{x}-\mathbf{u}_j\|_2} - \mathrm{j}\frac{d_{\mathrm{B}}\pi}{\lambda_n}(t-1)\frac{\|\mathbf{x}-\mathbf{u}_j\|_2^2\mathbf{e}_Y - (\mathbf{x}-\mathbf{u}_j)(\mathbf{x}-\mathbf{u}_j)^{\mathrm{H}}\mathbf{e}_Y}{\|\mathbf{x}-\mathbf{u}_j\|_2^3} \\ \quad + \mathrm{j}\frac{d_{\mathrm{U}}\pi}{\lambda_n}(r-1)\frac{\cos\left(\arccos\left(\frac{(\mathbf{x}-\mathbf{u}_j)^{\mathrm{H}}\mathbf{e}_X}{\|\mathbf{x}-\mathbf{u}_j\|_2}\right)-\vartheta\right)}{\sqrt{1-\left(\frac{(\mathbf{x}-\mathbf{u}_j)^{\mathrm{H}}\mathbf{e}_X}{\|\mathbf{x}-\mathbf{u}_j\|_2}\right)^2}}\frac{\|\mathbf{x}-\mathbf{u}_j\|_2^2\mathbf{e}_X - (\mathbf{x}-\mathbf{u}_j)(\mathbf{x}-\mathbf{u}_j)^{\mathrm{H}}\mathbf{e}_X}{\|\mathbf{x}-\mathbf{u}_j\|_2^3}, & l=0, \\ \mathrm{j}2\pi\frac{n}{cN_{\mathrm{C}}'T_s}\frac{\mathbf{x}-\mathbf{v}_{l,j}}{\|\mathbf{x}-\mathbf{v}_{l,j}\|_2} \\ \quad + \mathrm{j}\frac{d_{\mathrm{U}}\pi}{\lambda_n}(r-1)\frac{\cos\left(\arccos\left(\frac{(\mathbf{x}-\mathbf{v}_{l,j})^{\mathrm{H}}\mathbf{e}_X}{\|\mathbf{x}-\mathbf{v}_{l,j}\|_2}\right)-\vartheta\right)}{\sqrt{1-\left(\frac{(\mathbf{x}-\mathbf{v}_{l,j})^{\mathrm{H}}\mathbf{e}_X}{\|\mathbf{x}-\mathbf{v}_{l,j}\|_2}\right)^2}}\frac{\|\mathbf{x}-\mathbf{v}_{l,j}\|_2^2\mathbf{e}_X - (\mathbf{x}-\mathbf{v}_{l,j})(\mathbf{x}-\mathbf{v}_{l,j})^{\mathrm{H}}\mathbf{e}_X}{\|\mathbf{x}-\mathbf{v}_{l,j}\|_2^3}, & l>0 \end{cases} \quad (50)$$

$$\rho_{l,j,n}^{(r,t)} = \begin{cases} \mathrm{j}\frac{d_{\mathrm{U}}\pi}{\lambda_n}(r-1)\cos\left(\arccos\left(\frac{(\mathbf{x}-\mathbf{u}_j)^{\mathrm{H}}\mathbf{e}_X}{\|\mathbf{x}-\mathbf{u}_j\|_2}\right)-\vartheta\right), & l=0 \\ \mathrm{j}\frac{d_{\mathrm{U}}\pi}{\lambda_n}(r-1)\cos\left(\arccos\left(\frac{(\mathbf{x}-\mathbf{v}_{l,j})^{\mathrm{H}}\mathbf{e}_X}{\|\mathbf{x}-\mathbf{v}_{l,j}\|_2}\right)-\vartheta\right), & l>0 \end{cases}. \quad (51)$$

$$\begin{aligned} \boldsymbol{\upsilon}_{l,j,n}^{(r,t)} = & -\mathrm{j}2\pi\frac{n}{cN_{\mathrm{C}}'T_s}\frac{\mathbf{x}-\mathbf{v}_{l,j}}{\|\mathbf{x}-\mathbf{v}_{l,j}\|_2} - \mathrm{j}2\pi\frac{n}{cN_{\mathrm{C}}'T_s}\frac{\mathbf{u}_j-\mathbf{v}_{l,j}}{\|\mathbf{u}_j-\mathbf{v}_{l,j}\|_2} + \mathrm{j}\frac{d_{\mathrm{B}}\pi}{\lambda_n}(t-1)\frac{\|\mathbf{u}_j-\mathbf{v}_{l,j}\|_2^2\mathbf{e}_Y - (\mathbf{u}_j-\mathbf{v}_{l,j})(\mathbf{u}_j-\mathbf{v}_{l,j})^{\mathrm{H}}\mathbf{e}_Y}{\|\mathbf{u}_j-\mathbf{v}_{l,j}\|_2^3} \\ & -\mathrm{j}\frac{d_{\mathrm{U}}\pi}{\lambda_n}(r-1)\frac{\cos\left(\arccos\left(\frac{(\mathbf{x}-\mathbf{v}_{l,j})^{\mathrm{H}}\mathbf{e}_X}{\|\mathbf{x}-\mathbf{v}_{l,j}\|_2}\right)-\vartheta\right)}{\sqrt{1-\left(\frac{(\mathbf{x}-\mathbf{v}_{l,j})^{\mathrm{H}}\mathbf{e}_X}{\|\mathbf{x}-\mathbf{v}_{l,j}\|_2}\right)^2}}\frac{\|\mathbf{x}-\mathbf{v}_{l,j}\|_2^2\mathbf{e}_X - (\mathbf{x}-\mathbf{v}_{l,j})(\mathbf{x}-\mathbf{v}_{l,j})^{\mathrm{H}}\mathbf{e}_X}{\|\mathbf{x}-\mathbf{v}_{l,j}\|_2^3}. \end{aligned} \quad (53)$$

Secondly, $\mathbf{D}_{\mathbf{x}}(\boldsymbol{\alpha}) \in \mathbb{C}^{JMN_{\mathrm{C}}N_{\mathrm{U}}N_B(L+1)\times 2} = \mathbf{1}_{(M)} \otimes \mathrm{vec}[(\tilde{\boldsymbol{\varrho}}_{l,j,n}^{(r,t)})^{\mathrm{H}}|\forall l,t,r,n,j]$, where $\tilde{\boldsymbol{\varrho}}_{l,j,n}^{(r,t)} = \mu_{l,j,n}^{(r,t)*}\boldsymbol{\varrho}_{l,j,n}^{(r,t)}$, while $\mu_{l,j,n}^{(r,t)}$ and $\boldsymbol{\varrho}_{l,j,n}^{(r,t)}$ are given by (19) and (50), respectively. Thirdly, $\mathbf{d}_{\vartheta}(\boldsymbol{\alpha}) = \mathrm{vec}[\mu_{l,j,n}^{(r,t)}\rho_{l,j,n}^{(r,t)*}|\forall l,t,r,n,j]$, where $\rho_{l,j,n}^{(r,t)}$ is given by (51). Fourthly, $\mathbf{D}_{\mathbf{v}}(\boldsymbol{\alpha}) \in \mathbb{C}^{JMN_{\mathrm{C}}N_{\mathrm{U}}N_B(L+1)\times 2L} = \mathbf{1}_{(M)} \otimes \mathrm{vec}[(\tilde{\boldsymbol{\upsilon}}_{l,j,n}^{(r,t)})^{\mathrm{H}}|\forall l,t,r,n,j]$, where $\tilde{\boldsymbol{\upsilon}}_{l,j,n}^{(r,t)} \in \mathbb{C}^{2L} = \mu_{l,j,n}^{(r,t)*}\breve{\boldsymbol{\upsilon}}_{l,j,n}^{(r,t)}$, and $\breve{\boldsymbol{\upsilon}}_{l,j,n}^{(r,t)}$ is given by (52), $\forall l = 0:L, \forall j$.

## Appendix C
## Proof of Theorem 1

By inner product calculation, it can be verified that $\mathbf{p}_{[i]}[k]$ is a descent direction, i.e., $\nabla_{\boldsymbol{\alpha}}^{\mathrm{H}} f(\boldsymbol{\alpha}_{[i]}^{\natural}[k]; \mathbf{h}_{[i]}^{\natural}[k])\mathbf{p}_{[i]}[k] < 0$ for any non-stationary $\boldsymbol{\alpha}_{[i]}^{\natural}[k]$, Thus, the SLLS-based location update (35) subject to the Armijo rule (36) will converge to a stationary point to $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{LCE})}$, as per the line search method [40]–[42].

## Appendix D
## Proof of Theorem 2

We first elaborate the convergence rate of $\boldsymbol{\alpha}_{[i+1]}^{\circ}[k]$. Applying the second-order approximation to $\mathbf{g}(\boldsymbol{\alpha}, \mathbf{h}_{[i]}^{\sharp}[k])$, we have

$$\mathbf{z}[k] = \mathbf{g}(\boldsymbol{\alpha}_{[i]}^{\natural}[k], \mathbf{h}_{[i]}^{\sharp}[k]) + \nabla_{\boldsymbol{\alpha}}^{\mathrm{H}}\mathbf{g}(\boldsymbol{\alpha}_{[i]}^{\natural}[k], \mathbf{h}_{[i]}^{\sharp}[k])\big(\boldsymbol{\alpha}^{\bullet}[k] - \boldsymbol{\alpha}_{[i]}^{\natural}[k]\big) + \mathbf{s}(\boldsymbol{\alpha}^{\bullet}[k]; \boldsymbol{\alpha}_{[i]}^{\natural}[k], \mathbf{h}_{[i]}^{\sharp}[k]) + \boldsymbol{\varsigma}, \tag{57}$$

where the second-order term $\mathbf{s}(\boldsymbol{\alpha}^{\bullet}[k]; \boldsymbol{\alpha}_{[i]}^{\natural}[k], \mathbf{h}_{[i]}^{\sharp}[k])$ is given by (58) shown at the bottom of this page, and $\boldsymbol{\varsigma}$ is the higher-order residual error. In addition, $\boldsymbol{\Psi}_{n,m,j}^{(r)} = \nabla_{\boldsymbol{\alpha}}\psi_{n,m,j}^{(r)}(\boldsymbol{\alpha}_{[i]}^{\natural}[k])$ $\nabla_{\boldsymbol{\alpha}}^{\mathrm{H}}\psi_{n,m,j}^{(r)}(\boldsymbol{\alpha}_{[i]}^{\natural}[k])$, where $\psi_{n,m,j}^{(r)}(\boldsymbol{\alpha}_{[i]}^{\natural}[k]) = (\mathbf{g}_{j,[i]}^{(r)}[k,n,m])^{\mathrm{H}}$ $\mathbf{h}_{[i]}^{\sharp}[k]$, and $\mathbf{g}_{j,[i]}^{(r)}[k,n,m]$ is the value of $\mathbf{g}_j^{(r)}[n,m]$ (given by (16)) conditioned on $\boldsymbol{\alpha}_{[i]}^{\natural}[k]$. In (57), we use $\nabla_{\boldsymbol{\alpha}}^{\mathrm{H}}\mathbf{g}(\boldsymbol{\alpha}_{[i]}^{\natural}[k], \mathbf{h}_{[i]}^{\sharp}[k])\nabla_{\boldsymbol{\alpha}}\mathbf{g}(\boldsymbol{\alpha}_{[i]}^{\natural}[k], \mathbf{h}_{[i]}^{\sharp}[k])$ to approximate the Hessian matrix for computational ease (only gradient is needed). Ignoring the residual error, we have (59) shown at the bottom of the next page. If $\nabla_{\boldsymbol{\alpha}}\mathbf{g}(\boldsymbol{\alpha}_{[i]}^{\natural}[k], \mathbf{h}_{[i]}^{\sharp}[k])$ is full-column-rank, then $\|\boldsymbol{\alpha}^{\bullet}[k] - \boldsymbol{\alpha}_{[i+1]}^{\circ}[k]\|_2 = \mathcal{O}(\|\boldsymbol{\alpha}^{\bullet}[k] - \boldsymbol{\alpha}_{[i]}^{\natural}[k]\|_2^2)$.

Based on (36), $\boldsymbol{\alpha}_{[i+1]}^{\natural}[k]$ is a more efficient update than $\boldsymbol{\alpha}_{[i+1]}^{\circ}[k]$, since it can lead to a sufficient decrease in cost function value $f(\boldsymbol{\alpha}_{[i+1]}^{\natural}[k]; \mathbf{h}_{[i]}^{\natural}[k])$. Thus, convergence of $\boldsymbol{\alpha}_{[i+1]}^{\natural}[k]$ is at least quadratic. Theorem 2 is proved.

## Appendix E
## Expression of Feature Matrices

$\boldsymbol{\mathcal{G}}_{j,n}(\bar{\boldsymbol{\alpha}}[k]; \mathbf{w}_{[i]}[k])$ and $\boldsymbol{\mathcal{H}}_{j,n}(\bar{\boldsymbol{\alpha}}[k]; \mathbf{w}_{[i]}[k])$ are given by (60) and (61) shown at the bottom of next page, where $\boldsymbol{\Sigma} = \mathbf{I}_J \otimes \mathrm{diag}\left[\eta_l^2\,|\forall l = 0:L\right]$ with $\eta_l^2$ being its elementary variance, while $\tilde{\boldsymbol{\mathfrak{B}}}_{\boldsymbol{\alpha}}(\mathbf{w}_{[i]}[k])$ and $\tilde{\boldsymbol{\mathfrak{B}}}_{\mathbf{h}_j[k]}(\mathbf{w}_{[i]}[k])$ are CRLBs (dependent on $\mathbf{w}_{[i]}[k]$) associated with $\boldsymbol{\alpha}$ and $\mathbf{h}_j[k]$, respectively, given by (62) and (63) shown at the bottom of the next page. In addition, $\mathbf{R}_{l,j,n}^{(r)}(\bar{\boldsymbol{\alpha}}[k]) \in \mathbb{C}^{(2JL+3)\times N_{\mathrm{B}}} = \mathrm{vec}[\boldsymbol{u}_{l,j,n}^{(r,t)}|\forall t = 1:N_{\mathrm{B}}]$ and $\mathbf{U}_{j,n}^{(r)}(\bar{\boldsymbol{\alpha}}[k]) \in \mathbb{C}^{N_{\mathrm{B}}\times(L+1)} = [\boldsymbol{\mu}_{l,j,n}^{(r)}|\forall l = 0:L]$, where

$$\boldsymbol{u}_{l,j,n}^{(r,t)} = \begin{bmatrix} \mu_{l,j,n}^{(r,t)*}\boldsymbol{\varrho}_{l,j,n}^{(r,t)} \\ \mu_{l,j,n}^{(r,t)}\rho_{l,j,n}^{(r,t)*} \\ \mu_{l,j,n}^{(r,t)*}\breve{\boldsymbol{\upsilon}}_{l,j,n}^{(r,t)} \end{bmatrix},$$

and $\boldsymbol{\varrho}_{l,j,n}^{(r,t)}$, $\rho_{l,j,n}^{(r,t)}$, $\breve{\boldsymbol{\upsilon}}_{l,j,n}^{(r,t)}$ and $\mu_{l,j,n}^{(r,t)}$ is given by (50), (51), (52) and (19), respectively.

## Appendix F
## Proof of Lemma 3

*1) Uniqueness of* $\mathbf{h}^{\star}[k]$*:* Since the system model is linear w.r.t. $\mathbf{h}[k]$ as shown in (13), $\mathbf{h}^{\star}[k]$ is unique at each LCE stage.

*2) Uniqueness of* $\mathbf{x}^{\star}[k]$*:* For easy notation, we use $\mathbf{g}(\mathbf{x})$ to denote $\mathbf{g}(\mathbf{x}, \vartheta, \mathbf{v}, \mathbf{h}[k]; \boldsymbol{\omega})$ (given by (13)) with any point $\vartheta$, $\mathbf{v}$ and $\mathbf{h}[k]$, use $f(\mathbf{x}) = \|\mathbf{z}[k] - \mathbf{g}(\mathbf{x})\|_2^2$ to denote the LCE cost function of $\mathbf{x}$, and we use $\mathbf{x} = \mathbf{x}^{\star}[k] + \psi\boldsymbol{\chi}$ to denote any point of $\mathbf{x}$, where $\boldsymbol{\chi}$ is a unit direction vector and $\psi$ is the length. Then, $f(\mathbf{x})$ is cast as $f(\mathbf{x}^{\star}[k] + \psi\boldsymbol{\chi}) = \|\mathbf{g}(\mathbf{x}^{\star}[k] + \psi\boldsymbol{\chi}) - \mathbf{g}(\mathbf{x}^{\star}[k]) - \boldsymbol{\upsilon}^{\star}[k]\|_2^2$, where $\boldsymbol{\upsilon}^{\star}[k] = \mathbf{z}[k] - \mathbf{g}(\mathbf{x}^{\star}[k])$ (independent of $\mathbf{x}$ and $\psi$). Let $\tilde{f}(\mathbf{x}^{\star}[k] + \psi\boldsymbol{\chi}) = f(\mathbf{x}^{\star}[k] + \psi\boldsymbol{\chi})|_{\tau_{l,j}=0,\forall l,j}$, which is a tight lower bound of $f(\mathbf{x}^{\star}[k] + \psi\boldsymbol{\chi})$, as verified later.

**Example.** Consider a case of $J = M = N_{\mathrm{U}} = N_{\mathrm{B}} = N_{\mathrm{C}} = 1$ and $L = 0$. Then, $f(\mathbf{x}^{\star}[k] + \psi\boldsymbol{\chi})$ and $\tilde{f}(\mathbf{x}^{\star}[k] + \psi\boldsymbol{\chi})$

$$\mathbb{E}_{\mathbf{h}[k]}\{\boldsymbol{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha}, \mathbf{h}[k]; \boldsymbol{\omega})\} = \sigma^2\mathbb{E}_{\mathbf{h}[k]}\left\{\left(\sum_{r,n,m}\begin{bmatrix} \mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha}; \boldsymbol{\omega})\mathbf{h}[k]\mathbf{h}^{\mathrm{H}}[k]\big(\mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha}; \boldsymbol{\omega})\big)^{\mathrm{H}} & \mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha}; \boldsymbol{\omega})\mathbf{h}[k]\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha}; \boldsymbol{\omega}) \\ \big(\mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha}; \boldsymbol{\omega})\mathbf{h}[k]\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha}; \boldsymbol{\omega})\big)^{\mathrm{H}} & \big(\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha}; \boldsymbol{\omega})\big)^{\mathrm{H}}\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha}; \boldsymbol{\omega}) \end{bmatrix}\right)^{-1}\right\}$$

$$\succeq \underbrace{\sigma^2\left(\sum_{r,n,m}\begin{bmatrix} \mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha}; \boldsymbol{\omega})\boldsymbol{\Sigma}\big(\mathbf{K}_{n,m}^{(r)}(\boldsymbol{\alpha}; \boldsymbol{\omega})\big)^{\mathrm{H}} & \mathbf{0}_{(2JL+3)\times J(L+1)} \\ \mathbf{0}_{(2JL+3)\times J(L+1)}^{\mathrm{H}} & \big(\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha}; \boldsymbol{\omega})\big)^{\mathrm{H}}\mathbf{G}_{n,m}^{(r)}(\boldsymbol{\alpha}; \boldsymbol{\omega}) \end{bmatrix}\right)^{-1}}_{\tilde{\boldsymbol{\mathfrak{B}}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha}, \boldsymbol{\Sigma}; \boldsymbol{\omega})}. \tag{55}$$

$$\mathbf{s}(\boldsymbol{\alpha}^{\bullet}[k]; \boldsymbol{\alpha}_{[i]}^{\natural}[k], \mathbf{h}_{[i]}^{\sharp}[k]) = \mathrm{vec}[(\boldsymbol{\alpha}^{\bullet}[k] - \boldsymbol{\alpha}_{[i]}^{\natural}[k])^{\top}\boldsymbol{\Psi}_{n,m,j}^{(r)}(\boldsymbol{\alpha}^{\bullet}[k] - \boldsymbol{\alpha}_{[i]}^{\natural}[k])|\forall n,r,m,j]. \tag{58}$$

reduce to that in (64) and (65) shown at the bottom of this page, respectively. ■

It can be verified by the above simple case that $\tilde{f}(\mathbf{x}^\star[k]+\psi\boldsymbol{\chi})$ satisfies (i) $\tilde{f}(\mathbf{x}^\star[k]) = f(\mathbf{x}^\star[k])$ and (ii) $\tilde{f}(\mathbf{x}^\star[k]+\psi\boldsymbol{\chi}) < f(\mathbf{x}^\star[k]+\psi\boldsymbol{\chi})$, $\forall\boldsymbol{\chi}:\|\boldsymbol{\chi}\|_2=1$ and $\psi\neq 0$. That is to say, $\tilde{f}(\mathbf{x}^\star[k]+\psi\boldsymbol{\chi})$ is a tight lower bound of $f(\mathbf{x}^\star[k]+\psi\boldsymbol{\chi})$.

Moreover, $\tilde{f}(\mathbf{x}^\star[k]+\psi\boldsymbol{\chi})$ also satisfies (i) $\tilde{f}(\mathbf{x}^\star[k]+\psi\boldsymbol{\chi}) = \tilde{f}(\mathbf{x}^\star[k])$ for $\psi=0$; and (ii) $\tilde{f}(\mathbf{x}^\star[k]+\psi\boldsymbol{\chi}) > \tilde{f}(\mathbf{x}^\star[k])$ for $\psi\neq 0$, since the modulus (or $h_{l,j}[k]$) is reduced with transmission distance due to path loss. That is, except $\mathbf{x}^\star[k]$, there is no other point such that $\tilde{f}(\mathbf{x})$ is lower than $\tilde{f}(\mathbf{x}^\star[k])$, i.e., $\mathbf{x}^\star[k]$ is the unique optimal solution to $\mathscr{A}_{\mathrm{SLAB}}^{(\mathrm{LCE})}$.

*3) Uniqueness of* $\mathbf{v}^\star[k]$*:* Its proof is similar to that of $\mathbf{x}$.

*4) Uniqueness of* $\vartheta^\star[k]$*:* We use $f(\vartheta^\star[k]+\nu)$ with $\nu\in(-\pi,\pi]$ to denote the LCE cost function $\|\mathbf{z}[k]-\mathbf{g}(\vartheta)\|_2^2$ (detailed in (66) shown at the bottom of this page) with any point $\mathbf{x}$, $\mathbf{v}$, $\mathbf{h}[k]$, where $\vartheta^\star[k]$ is one optimal solution of $\vartheta$ at each LCE stage. Consider again that simple case. Then, $f\big(\vartheta^\star[k]+\nu\big)$ will reduce to (66), where $C_{\mathrm{LOS},n}^{(r,t)}\big(h_{0,j}[k],\mathbf{x}\big)$ is the constant independent of $\vartheta^\star[k]$ and $\nu$. Thus, we have $f\big(\vartheta^\star[k]+\nu\big) > f\big(\vartheta^\star[k]\big)$, for any $\nu\neq 0$ and $\nu\in(-\pi,\pi]$. Hence, $\vartheta^\star[k]$ is unique, and lemma 3 is proved.

## APPENDIX G
## PROOF OF THEOREM 3

To prove theorem 3, we need to give a lemma regrading the continuity of the LCE system.

*Lemma 4:* If A1–A2 are satisfied, $\|\boldsymbol{\beta}[k]-\boldsymbol{\beta}^{\mathrm{true}}[k]\|_2 \le C_1\frac{\|\mathbf{g}(\boldsymbol{\beta}[k])-\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2}{\|\nabla^{\mathrm{H}}_{\boldsymbol{\beta}[k]}\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2}$ holds, $\forall\boldsymbol{\beta}[k]$, for $C_1\ge 0$.[7]

*Proof:* We have that $\|\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k]+\Delta\boldsymbol{\beta})-\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2 = \|\nabla^{\mathrm{H}}_{\boldsymbol{\beta}[k]}\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\Delta\boldsymbol{\beta}+\mathcal{O}(\|\Delta\boldsymbol{\beta}\|_2^2)\|_2$, and hence we have $\|\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k]+\Delta\boldsymbol{\beta})-\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2 \ge \|\nabla^{\mathrm{H}}_{\boldsymbol{\beta}[k]}\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\Delta\boldsymbol{\beta}\|_2 - \mathcal{O}(\|\Delta\boldsymbol{\beta}\|_2^2)$. As a result, we have the following inequality: $\|\nabla^{\mathrm{H}}_{\boldsymbol{\beta}[k]}\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\Delta\boldsymbol{\beta}\|_2 \le \|\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k]+\Delta\boldsymbol{\beta})-\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2 + \mathcal{O}(\|\Delta\boldsymbol{\beta}\|_2^2)$. In addition, there must be $C_1\ge 0$ such that $\|\nabla^{\mathrm{H}}_{\boldsymbol{\beta}[k]}\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\Delta\boldsymbol{\beta}\|_2 \ge C_1\|\nabla^{\mathrm{H}}_{\boldsymbol{\beta}[k]}\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2\|\Delta\boldsymbol{\beta}\|_2$. Thus, $C_1\|\nabla^{\mathrm{H}}_{\boldsymbol{\beta}[k]}\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2\|\Delta\boldsymbol{\beta}[k]\|_2 \le \|\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k]+\Delta\boldsymbol{\beta}) - \mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2 + \mathcal{O}(\|\Delta\boldsymbol{\beta}\|_2^2)$. Then, we arrive at $\|\Delta\boldsymbol{\beta}\|_2 \ge \frac{\|\Delta\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2+\mathcal{O}(\|\Delta\boldsymbol{\beta}\|_2^2)}{C_1\|\nabla^{\mathrm{H}}_{\boldsymbol{\beta}[k]}\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2}$, in which $\Delta\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k]) = \mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k]+\Delta\boldsymbol{\beta})-\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])$ denotes the range variation.

Based on lemma 3, when there is no measurement error, $(\boldsymbol{\alpha}^\star[k],\mathbf{h}^\star[k])$ is equal to $(\boldsymbol{\alpha}^{\mathrm{true}},\mathbf{h}^{\mathrm{true}}[k])$. This means that for any $\xi>0$, there exists $\zeta>0$ such that $\|\boldsymbol{\beta}^\star[k]-\boldsymbol{\beta}^{\mathrm{true}}[k]\|_2<\xi$ must hold if $\|\boldsymbol{\epsilon}[k]\|_2<\zeta$. Thus, $\mathcal{O}(\|\boldsymbol{\beta}^\star[k]-\boldsymbol{\beta}^{\mathrm{true}}[k]\|_2^2)$ can be safely ignored, and lemma 4 is proved.

Thus, based on lemma 4 we have $\|\boldsymbol{\beta}^\star[k]-\boldsymbol{\beta}^{\mathrm{true}}[k]\|_2 \le \frac{\|\mathbf{g}(\boldsymbol{\beta}^\star[k])-\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2+\mathcal{O}(\|\boldsymbol{\beta}^\star[k]-\boldsymbol{\beta}^{\mathrm{true}}[k]\|_2^2)}{C_1\|\nabla^{\mathrm{H}}_{\boldsymbol{\beta}[k]}\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2}$. In addition, since $\boldsymbol{\beta}^\star[k]$ is the optimal solution and $\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])$ is noiseless, $\|\mathbf{g}(\boldsymbol{\beta}^\star[k])-\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2 \le \|\mathbf{z}[k]-\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2 = \|\boldsymbol{\epsilon}[k]\|_2$.

[7]We drop $\mathbf{w}[k]$ in $\mathbf{g}(\bullet)$ for brevity.

$$\Big\|\underbrace{\big(\nabla^{\mathrm{H}}_{\boldsymbol{\alpha}}\mathbf{g}(\boldsymbol{\alpha}^\natural_{[i]}[k],\mathbf{h}^\sharp_{[i]}[k])\big)^\dagger\big(\mathbf{z}[k]-\mathbf{g}(\boldsymbol{\alpha}^\natural_{[i]}[k],\mathbf{h}^\sharp_{[i]}[k])\big)+\boldsymbol{\alpha}^\natural_{[i]}[k]}_{\boldsymbol{\alpha}^\circ_{[i+1]}[k]}-\boldsymbol{\alpha}^\bullet[k]\Big\|_2 = \big\|\big(\nabla^{\mathrm{H}}_{\boldsymbol{\alpha}}\mathbf{g}(\boldsymbol{\alpha}^\natural_{[i]}[k],\mathbf{h}^\sharp_{[i]}[k])\big)^\dagger\mathbf{s}(\boldsymbol{\alpha}^\bullet[k];\boldsymbol{\alpha}^\natural_{[i]}[k],\mathbf{h}^\sharp_{[i]}[k])\big\|_2$$

$$= \big\|\nabla^{\mathrm{H}}_{\boldsymbol{\alpha}}\mathbf{g}(\boldsymbol{\alpha}^\natural_{[i]}[k],\mathbf{h}^\sharp_{[i]}[k])\big(\boldsymbol{\alpha}^\bullet[k]-\boldsymbol{\alpha}^\natural_{[i]}[k]\big)\big\|_2^2. \tag{59}$$

$$\boldsymbol{\mathcal{G}}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k]) = \sum_{\substack{l=0:L\\ r=1:N_{\mathrm{U}}}}\frac{\eta_l^2\mathbf{R}^{(r)\mathrm{H}}_{l,j,n}\big(\bar{\boldsymbol{\alpha}}[k]\big)\bar{\mathfrak{B}}^{\mathrm{H}}_{\boldsymbol{\alpha}}\big(\mathbf{w}_{[i]}[k]\big)\bar{\mathfrak{B}}_{\boldsymbol{\alpha}}\big(\mathbf{w}_{[i]}[k]\big)\mathbf{R}^{(r)}_{l,j,n}\big(\bar{\boldsymbol{\alpha}}[k]\big)}{\sigma^2}, \tag{60}$$

$$\boldsymbol{\mathcal{H}}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k]) = \sum_{r=1:N_{\mathrm{U}}}\frac{\mathbf{U}^{(r)}_{j,n}\big(\bar{\boldsymbol{\alpha}}[k]\big)\bar{\mathfrak{B}}^{\mathrm{H}}_{\mathbf{h}_j[k]}\big(\mathbf{w}_{[i]}[k]\big)\bar{\mathfrak{B}}_{\mathbf{h}_j[k]}\big(\mathbf{w}_{[i]}[k]\big)\mathbf{U}^{(r)\mathrm{H}}_{j,n}\big(\bar{\boldsymbol{\alpha}}[k]\big)}{\sigma^2}. \tag{61}$$

$$\bar{\mathfrak{B}}_{\boldsymbol{\alpha}}\big(\mathbf{w}_{[i]}[k]\big) = \left(\sum_{l,j,r,n,m}\eta_l^2\mathbf{R}^{(r)}_{l,j,n}(\bar{\boldsymbol{\alpha}}[k])\mathbf{w}^*_{j,[i]}[k,n,m]\mathbf{w}^\top_{j,[i]}[k,n,m]\mathbf{R}^{(r)\mathrm{H}}_{l,j,n}(\bar{\boldsymbol{\alpha}}[k])\right)^{-1}, \tag{62}$$

$$\bar{\mathfrak{B}}_{\mathbf{h}_j[k]}\big(\mathbf{w}_{[i]}[k]\big) = \left(\sum_{r,n,m}\mathbf{U}^{(r)\mathrm{H}}_{j,n}(\bar{\boldsymbol{\alpha}}[k])\mathbf{w}^*_{j,[i]}[k,n,m]\mathbf{w}^\top_{j,[i]}[k,n,m]\mathbf{U}^{(r)}_{j,n}(\bar{\boldsymbol{\alpha}}[k])\right)^{-1}. \tag{63}$$

$$f\big(\mathbf{x}^\star[k]+\psi\boldsymbol{\chi}\big) = \Big\|\underbrace{h_{l,j}[k]\boldsymbol{\omega}_j[n,m]}_{\text{Equivalent modulus}}\underbrace{e^{-\mathrm{j}\frac{2\pi n}{N_{\mathrm{C}}^\prime T_s}\frac{\|\mathbf{x}^\star[k]+\psi\boldsymbol{\chi}-\mathbf{u}_j\|_2}{c}}}_{\text{Equivalent phase}}-h_{l,j}[k]\boldsymbol{\omega}_j[n,m]e^{-\mathrm{j}\frac{2\pi n}{N_{\mathrm{C}}^\prime T_s}\frac{\|\mathbf{x}^\star[k]-\mathbf{u}_j\|_2}{c}}-\upsilon^{(r)\star}_{j,n}[k]\Big\|_2^2, \tag{64}$$

$$\tilde{f}\big(\mathbf{x}^\star[k]+\psi\boldsymbol{\chi}\big) = \big\|h_{l,j}[k]\boldsymbol{\omega}^{(t)}_j[n,m]-h_{l,j}[k]\boldsymbol{\omega}^{(t)}_j[n,m]-\upsilon^{(r)\star}_{j,n}[k]\big\|_2^2. \tag{65}$$

$$f\big(\vartheta^\star[k]+\nu\big) = \Big\|C^{(r,t)}_{\mathrm{LOS},n}\big(h_{0,j}[k],\mathbf{x}\big)\Big(e^{\mathrm{j}\frac{d_{\mathrm{U}}\pi}{\lambda_n}(r-1)\sin(\theta_{\mathrm{U},l,j}-\vartheta^\star[k]-\nu)}-e^{\mathrm{j}\frac{d_{\mathrm{U}}\pi}{\lambda_n}(r-1)\sin(\theta_{\mathrm{U},l,j}-\vartheta^\star[k])}\Big)-\upsilon^{(r)\star}_{j,n}[k]\Big\|_2^2. \tag{66}$$

$$\mathbf{d}^{\mathrm{H}}_{j,[i+1]}[k,n,m]\nabla_{\boldsymbol{\omega}_j[n,m]}\kappa\big(\mathbf{w}_{j,[i]}[k,n,m]\big)=\big(\mathbf{w}^{\sharp}_{j,[i+1]}[k,n,m]\big)^{\mathrm{H}}\nabla_{\boldsymbol{\omega}_j[n,m]}\kappa\big(\mathbf{w}_{j,[i]}[k,n,m]\big) \tag{67}$$

$$=-2\big(\mathbf{w}^{\sharp}_{j,[i+1]}[k,n,m]\big)^{\mathrm{H}}\frac{\big(\|\mathbf{w}_{j,[i]}[k,n,m]\|_2^2\mathbf{I}_{N_{\mathrm{B}}}-\mathbf{w}^{*}_{j,[i]}[k,n,m]\mathbf{w}^{\top}_{j,[i]}[k,n,m]\big)\mathfrak{G}_{j,n}\big(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k]\big)\,\mathbf{w}^{*}_{j,[i]}[k,n,m]}{\|\mathbf{w}_{j,[i]}[k,n,m]\|_2^4} \tag{68}$$

$$=\frac{\big(\mathbf{w}^{\sharp}_{j,[i+1]}[k,n,m]\big)^{\mathrm{H}}\mathbf{w}^{*}_{j,[i]}[k,n,m]}{\|\mathbf{w}_{j,[i]}[k,n,m]\|_2^2}\cdot\Bigg(\underbrace{\frac{\big(\mathbf{w}^{\top}_{j,[i]}[k,n,m]\mathfrak{G}_{j,n}\big(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k]\big)\,\mathbf{w}^{*}_{j,[i]}[k,n,m]}{\|\mathbf{w}_{j,[i]}[k,n,m]\|_2^2}}_{\leq\lambda'_{\max}}-\lambda'_{\max}\Bigg)\leq 0. \tag{69}$$

Thus, $\|\boldsymbol{\beta}^{\star}[k]-\boldsymbol{\beta}^{\mathrm{true}}[k]\|_2\leq\frac{\|\boldsymbol{\epsilon}[k]\|_2+\mathcal{O}(\|\boldsymbol{\beta}^{\star}[k]-\boldsymbol{\beta}^{\mathrm{true}}[k]\|_2^2)}{C_1\|\nabla^{\mathrm{H}}_{\boldsymbol{\beta}[k]}\mathbf{g}(\boldsymbol{\beta}^{\mathrm{true}}[k])\|_2}$, and theorem 3 is proved.

## Appendix H
## Proof of Theorem 4

We will prove $\nabla^{\mathrm{H}}_{\boldsymbol{\omega}_j[n,m]}\kappa\big(\mathbf{w}_{j,[i]}[k,n,m]\big)\mathbf{d}_{j,[i+1]}[k,n,m]\leq 0$, $\forall j,n,m$, and the equality holds only if $\{\mathbf{w}_{j,[i]}[k,n,m]\}$ is an optimal solution of $\mathscr{A}^{(\mathrm{BFR})}_{\mathrm{SLAB},[i+1]}$. Once this is proved, theorem 4 will be proved by combining lemma 2, thus it directly follows from the convergence of the feasible direction method [42] subject to the Armijo rule [40].

Let $\lambda'_{\max}$ be the largest eigenvalue of $\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k])$. Since $\mathbf{w}^{\sharp}_{j,[i+1]}[k,n,m]$ is its principal eigenvector, we have $\big(\mathbf{w}^{\sharp}_{j,[i+1]}[k,n,m]\big)^{\mathrm{H}}\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k])\mathbf{w}_{j,[i]}[k,n,m]=\lambda'_{\max}\big(\mathbf{w}^{\sharp}_{j,[i+1]}[k,n,m]\big)^{\mathrm{H}}\mathbf{w}_{j,[i]}[k,n,m]$. In addition, combining with (46), we know $\mathbf{w}^{\mathrm{H}}_{[i]}[k]\nabla_{\boldsymbol{\omega}}\kappa\big(\mathbf{w}_{[i]}[k]\big)=0$. Thus, we have (67)–(69) shown at the top of this page, and the inequality of the condition is verified.

From (69) we know the equality holds only if $\mathbf{w}_{j,[i]}[k,n,m]$ is the principal eigenvector of $\mathfrak{G}_{j,n}(\bar{\boldsymbol{\alpha}}[k];\mathbf{w}_{[i]}[k])$ which is the optimal solution to the problem in (43) and hence $\mathscr{A}^{(\mathrm{BFR})}_{j,[i+1]}$ in (41) equivalently. Hence, theorem 4 is proved.

## Appendix I
## Proof of Theorem 5

Since $\mathbf{g}(\boldsymbol{\alpha},\mathbf{h}[k];\boldsymbol{\omega})$ is twice differential w.r.t. $\boldsymbol{\omega}$ for any point $\boldsymbol{\omega}$ such that $\nabla_{\boldsymbol{\omega}}\mathbf{g}(\boldsymbol{\alpha},\mathbf{h}[k];\boldsymbol{\omega})$ is element-wise Lipschitz continuous w.r.t. $\boldsymbol{\omega}$, as per theorem 3, $\bar{\boldsymbol{\alpha}}[k]-\boldsymbol{\alpha}^{\mathrm{true}}=\mathcal{O}(\|\boldsymbol{\epsilon}[k]\|_2)$. Thus, we have $\nabla_{\boldsymbol{\omega}}\operatorname{trace}(\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\bar{\boldsymbol{\alpha}}[k],\boldsymbol{\Sigma};\mathbf{w}[k]))=\nabla_{\boldsymbol{\omega}}\operatorname{trace}(\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha}^{\mathrm{true}},\boldsymbol{\Sigma};\mathbf{w}[k]))+\mathcal{O}(\|\boldsymbol{\epsilon}[k]\|_2)$. Note that $\Re\{\nabla^{\mathrm{H}}_{\boldsymbol{\omega}}\operatorname{trace}(\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\bar{\boldsymbol{\alpha}}[k],\boldsymbol{\Sigma};\mathbf{w}[k]))(\boldsymbol{\omega}-\mathbf{w}[k])\}\geq 0$, $\forall\boldsymbol{\omega}$, since $\mathbf{w}[k]$ is a stationary point of the BFR problem given $\bar{\boldsymbol{\alpha}}[k]$. Thus, $\Re\{\nabla^{\mathrm{H}}_{\boldsymbol{\omega}}\operatorname{trace}(\bar{\mathfrak{B}}_{\boldsymbol{\beta}[k]}(\boldsymbol{\alpha}^{\mathrm{true}},\boldsymbol{\Sigma};\mathbf{w}[k]))(\boldsymbol{\omega}-\mathbf{w}[k])\}\geq W_{\mathrm{mis}}$, where $W_{\mathrm{mis}}\sim\mathcal{O}(\|\boldsymbol{\epsilon}[k]\|_2)$, and thus theorem 5 is proved.

## References


[1] R. D. Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch, "Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 102–112, Nov. 2014.

[2] K. Witrisal *et al.*, "High-accuracy localization for assisted living: 5G systems will turn multipath channels from foe to friend," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 59–70, Mar. 2016.

[3] J. A. del Peral-Rosado, J. A Lopez-Salcedo, S. Kim, and G. Seco-Granados, "Feasibility study of 5G-based localization for assisted driving," in *Proc. Int. Conf. Localization GNSS*, 2016, pp. 1–6.

[4] A. Guerra, F. Guidi, and D. Dardari, "On the impact of beamforming strategy on mm-wave localization performance limits," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2017, pp. 809–814.

[5] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter wave MIMO in 5G systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1822–1835, Mar. 2018.

[6] Y. Wang, Y. Wu, and Y. Shen, "Joint spatiotemporal multipath mitigation in large-scale array localization," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 783–797, Feb. 2019.

[7] A. Guerra, F. Guidi, and D. Dardari, "Single-anchor localization and orientation performance limits using massive arrays: MIMO vs. beamforming," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5241–5255, Aug. 2018.

[8] Z. Abu-Shaban, X. Zhou, T. Abhayapala, G. Seco-Granados, and H. Wymeersch, "Error bounds for uplink and downlink 3D localization in 5G mmWave systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 4939–4954, Aug. 2018.

[9] S. Jeong, O. Simeone, A. Haimovich, and J. Kang, "Beamforming design for joint localization and data transmission in distributed antenna system," *IEEE Trans. Veh. Technol.*, vol. 64, no. 1, pp. 62–76, Jan. 2015.

[10] A. J. Tenenbaum and R. S. Adve, "Improved sum-rate optimization in the multiuser MIMO downlink," in *Proc. 42nd Annu. Conf. Inf. Sci. Syst.*, 2008, pp. 984–989.

[11] A. Chelli, K. Kansanen, M. S. Alouini, and I. Balasingham, "On bit error probability and power optimization in multihop millimeter wave relay systems," *IEEE Access*, vol. 6, pp. 3794–3808, 2018.

[12] K. G. Seddik, A. K. Sadek, W. Su, and K. R. Liu, "Outage analysis and optimal power allocation for multinode relay networks," *IEEE Signal Process. Lett.*, vol. 14, no. 6, pp. 377–380, Jun. 2007.

[13] V. Venkateswaran and A.-J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.

[14] Z. Han, Z. Ji, and K. J. Liu, "Power minimization for multi-cell OFDM networks using distributed non-cooperative game approach," in *Proc. Global Telecommun. Conf.*, 2004, vol. 6, pp. 3742–3747.

[15] T. Zhang, A. F. Molisch, Y. Shen, Q. Zhang, H. Feng, and M. Z. Win, "Joint power and bandwidth allocation in wireless cooperative localization networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6527–6540, Oct. 2016.

[16] H. Godrich, A. P. Petropulu, and H. V. Poor, "Power allocation strategies for target localization in distributed multiple-radar architectures," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3226–3240, Jul. 2011.

[17] Y. Shen and M. Z. Win, "Energy efficient location-aware networks," in *Proc. IEEE Int. Conf. Commun.*, 2008, pp. 2995–3001.

[18] W. H. Dai, Y. Shen, and M. Z. Win, "Energy efficient cooperative network localization," in *Proc. IEEE Int. Conf. Commun.*, 2014, pp. pp. 4969–4974.

[19] Y. Shen, W. Dai, and M. Z. Win, "Power optimization for network localization," *IEEE/ACM Trans. Netw.*, vol. 22, no. 4, pp. 1337–1350, Aug. 2014.

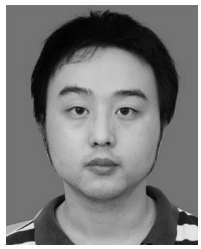
[20] W. L. Li, Y. Shen, Y. J. Zhang, and M. Z. Win, "Robust power allocation for energy-efficient location-aware networks," *IEEE/ACM Trans. Netw.*, vol. 21, no. 6, pp. 1918–1930, Dec. 2013.

[21] W. Dai, Y. Shen, and M. Z. Win, "Energy-efficient network navigation algorithms," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1418–1430, Jul. 2015.

[22] M. Razaviyayn, M. Hong, Z. Q. Luo, and J. S. Pang, "Parallel successive convex approximation for nonsmooth nonconvex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1440–1448.

[23] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, 2017.
[24] A. Liu, V. K. N. Lau, and M.-J. Zhao, "Stochastic successive convex optimization for two-timescale hybrid precoding in massive MIMO," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 432–444, Jun. 2018.
[25] J. Mairal, "Stochastic majorization-minimization algorithms for large-scale optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2283–2291.
[26] B. Wang, J. Liu, and X. Sun, "Mixed sources localization based on sparse signal reconstruction," *IEEE Signal Process. Lett.*, vol. 19, no. 8, pp. 487–490, Aug. 2012.
[27] J. Werner *et al.*, "Sectorized antenna-based DoA estimation and localization: Advanced algorithms and measurements," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2272–2286, Nov. 2015.
[28] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.
[29] D. Lee *et al.*, "Coordinated multipoint transmission and reception in LTE-advanced: Deployment scenarios and operational challenges," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 148–155, Feb. 2012.
[30] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. 2: Detection Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998.
[31] T. Wang, G. Leus, and L. Huang, "Ranging energy optimization for robust sensor positioning based on semidefinite programming," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4777–4787, Dec. 2009.
[32] E. L. Lawler and D. E. Wood, "Branch-and-bound methods: A survey," *Oper. Res.*, vol. 14, no. 4, pp. 699–719, 1966.
[33] A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal recovery," in *Convex Optimization in Signal Processing and Communications*, 2009, pp. 42–88.
[34] Y.-H. Dai, "Conjugate gradient methods with Armijo-type line searches," *Acta Mathematicae Applicatae Sinica*, vol. 18, no. 1, pp. 123–130, 2002.
[35] L.-H. Zhang, "On optimizing the sum of the Rayleigh quotient and the generalized Rayleigh quotient on the unit sphere," *Comput. Optim. Appl.*, vol. 54, no. 1, pp. 111–139, 2013.
[36] J.-J. Ruckmann and A. Shapiro, "First-order optimality conditions in generalized semi-infinite programming," *J. Optim. Theory Appl.*, vol. 101, no. 3, pp. 677–691, 1999.
[37] V. Savaux, F. Bader, and Y. Louet, "A joint MMSE channel and noise variance estimation for OFDM/OQAM modulation," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4254–4266, Nov. 2015.
[38] Y. Sun, P. Babu, and D. P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3576–3590, Jul. 2016.
[39] M. D. Renzo, "Stochastic geometry modeling and analysis of multi-tier millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5038–5057, Sep. 2015.
[40] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 1999.
[41] Z.-J. Shi, "Convergence of line search methods for unconstrained optimization," *Appl. Math. Comput.*, vol. 157, no. 2, pp. 393–405, 2004.
[42] L. Zhang, W. Zhou, and D. Li, "Global convergence of a modified Fletcher–Reeves conjugate gradient method with Armijo-type line search," *Numerische Mathematik*, vol. 104, no. 4, pp. 561–572, 2006.

**Bingpeng Zhou** (S'16–M'17) received the B.Eng. degree from the Zhongyuan University of Technology, Zhengzhou, China, in 2010 and the Ph.D. degree from Southwest Jiaotong University, Chengdu, China, in 2016. He is currently a Postdoctoral Fellow with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong. Prior to this, he was a visiting scholar with the HKUST for six months. He was also a visiting Ph.D. student with the 5G Innovation Centre, University of Surrey, Guildford, U.K., for three months in 2015. His current research interests include wireless localization and tracking for smart city, Bayesian signal processing, and autonomous vehicular networks.



**An Liu** (S'07–M'09–SM'17) received the Ph.D. and B.S. degree in electrical engineering from Peking University, Beijing, China, in 2011 and 2004, respectively. From 2008 to 2010, he was a visiting scholar with the Department of Electrical, Computer, and Energy Engineering, University of Colorado Boulder. He was a Postdoctoral Research Fellow from 2011to 2013, a Visiting Assistant Professor in 2014, and a Research Assistant Professor from 2015 to 2017, with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. He is currently a Distinguished Research Fellow with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. His research interests include wireless communications, stochastic optimization, and compressive sensing.

**Vincent Lau** (SM'04–F'12) received the B.Eng. (distinction first Hons.) degree from The University of Hong Kong, Hong Kong, in 1992 and the Ph.D. degree from the Cambridge University, Cambridge, U.K., in 1997. He was with the Bell Labs from 1997 to 2004 and with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2004. He is currently a Chair Professor and the Founding Director of the Huawei-HKUST Joint Innovation Laboratory, HKUST. He is also elected as an HKIE Fellow, a Croucher Senior Research Fellow, and the Changjiang Chair Professor. He has authored/coauthored more than 300 IEEE journal and conference papers and has contributed to 50 US patents on various wireless systems. His current research interests include stochastic optimization and analysis for wireless systems, massive MIMO, compressed sensing, networked control systems as well as PHY-caching for wireless networks.