Abstract—Fast and accurate analog beam tracking is an important and yet challenging issue in 5G wireless networks, due to the inherent non-convexity of the problem. In this paper, we develop a low-complexity recursive beam tracking algorithm. In static beam tracking scenarios, this algorithm converges to the Cramér-Rao lower bound (CRLB) with very high probability. In dynamic beam tracking scenarios, if combined with a simple TDMA pilot pattern, this algorithm has the potential to track hundreds of independent beams, generated by highly-mobile transmitters/reflectors, with low pilot overhead. Simulations are provided to illustrate the performance gain of this algorithm.

I. INTRODUCTION

The explosively growing data traffic in future wireless systems can be leveraged by using large antenna arrays and higher frequency bands [1]–[6]. However, as the array size grows and the carrier frequency increases, the large number of A/D (or D/A) converters in the fully digital array makes the design infeasible due to high energy consumption and huge hardware cost [6]. A promising alternative is analog beamforming [6]–[11], in which the signals of all antennas are beamformed in the analog domain by using phase shifters, and a single A/D (or D/A) is used for digital processing. This analog beamforming solution has been standardized by IEEE 802.11ad [12] and IEEE 802.15.3c [13], and is actively discussed by several 5G industrial organizations [14], [15].

One fundamental challenge in analog beamforming is how to track the beam directions using limited pilot resources. This challenge is especially difficult when a huge number of beams are generated from many mobile terminals and reflectors with high moving speeds. This challenge has been recognized in the industry as one important research task for 5G massive MIMO and millimeter wave systems, e.g., [16]–[20].

The goal of this paper is to develop an efficient beam tracking algorithm that can achieve high accuracy for tracking a large number of beams with low pilot overhead. To that end, the contributions of this paper are summarized as follows:

- We develop a recursive beam tracking algorithm. In static beam tracking scenarios, its convergence and asymptotic optimality are established in three steps: First, we prove that it converges to a set of stable beam directions, including the real beam direction and some sub-optimal stable beam directions, with probability one (Theorem 1). Second, we prove that under certain conditions, it converges to the real beam direction, instead of other sub-optimal stable directions, with high probability (Theorem 2). Finally, if the step-size parameters are chosen appropriately, then with high probability, the mean square tracking error of this algorithm converges to the minimum Cramér-Rao lower bound (CRLB), and hence the highest convergence speed is achieved. To the extent of our knowledge, this paper presents the first theoretical analysis on the convergence and asymptotic optimality of analog beam tracking in antenna array systems.

- Our simulation results in both static and dynamic beam tracking scenarios suggest that this algorithm can achieve much lower beam tracking error and higher data rate than several state-of-the-art algorithms, with the same amount of pilot overhead. The performance of several beam tracking algorithms is summarized in Table I. In each algorithm, 1000 narrow beams are tracked by using a TDMA round-robin pilot pattern, i.e., each time-slot has one pilot, the pilots are assigned periodically to these 1000 beams, and the duration of each time-slot is 0.2 ms (i.e., one transmission time interval (TTI) in [23], [24]). One can observe that the maximum trackable angular velocity of the proposed algorithm is much faster than those of other algorithms. Furthermore, the proposed algorithm is more robust with respect to the SNR $\rho$ than the other algorithms. In particular, if the receive SNR $\rho$ of each antenna is 10 dB (or 0 dB), the proposed algorithm can track 1000 narrow beams each rotating at an angular velocity of 0.32 rad/s (or 0.23 rad/s), which is 72 mph.

1. The notation “−” denotes that the corresponding algorithm cannot achieve 95% of the channel capacity even at zero angular velocity.

### Table I

| SNR without | Number of | Recursive | IEEE 802.11ad | Least square | Compressed sensing |
|-----------------|--------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| array gain      | antennas          | beam tracking               | 802.11ad [12]               | 802.15.3c [13]              | 802.15.3c [13]              |
| $\rho = 10$dB    | $M = 8$            | 0.32                        | −                           | 0.072                       | 0.04                        |
| $M = 16$        | 0.15               | −                           | 0.018                       | 0.02                        |
| $M = 32$        | 0.074              | 0.001                       | 0.005                       | 0.01                        |
| $\rho = 0$dB    | $M = 8$            | 0.23                        | −                           | −                           | −                           |
| $M = 16$        | 0.13               | −                           | −                           | −                           |
| $M = 32$        | 0.067              | −                           | −                           | −                           |

1. Corresponding author.

This work was supported in part by ONR grant N00014-17-1-2417.

†Department of ECE, the Ohio State University, Columbus OH, 43210, U.S.A.

‡Department of Electronic Engineering, Research Institute of Information Technology, Tsinghua University, Beijing, 100084, China
direction in each time-slot is optimized to maximize the Fisher good tracking performance are: (i) the probing beamforming chosen to ensure a fast convergence speed to the global optimal direction, instead of other local optimal beam directions.

The rest of this paper is organized as follows. In Sections II we describe the system model. In Sections V and VI we propose a recursive beam tracking algorithm that is proven to converge to the minimum CRLB in static beam tracking scenarios. In Section VI, we show that this algorithm converges very fast to the minimum CRLB in the static beam tracking scenarios and achieves a better tradeoff curve between MSE (or data rate) vs. angular velocity in dynamic beam tracking scenarios.

II. RELATED WORK

There has been a large number of recent studies on beam direction estimation/tracking in systems with analog beamforming arrays. Beam sweeping/measurement, scanning many spatial beam directions in a codebook, is widely used to obtain enough information for beam direction estimation/tracking. One possible way is to sweep the channel with narrow beams exhaustively, e.g., [27]–[33], where different methods are used to estimate (e.g., [27]–[31]) or track (e.g., [32], [33]) the beam directions. To save pilot resource, some other sweeping methods are proposed. In [34]–[36], the algorithms sweep the channel with beams in the hierarchical multi-resolution codebooks, where the beam directions can be estimated based on the codebooks. And in [22]–[24], the algorithms sweep the channel with random analog beamforming vectors, where the compressed sensing method is used to estimate the beam directions. What’s more, some most recent works propose to perform pilot training based on the prior knowledge of beam directions, e.g. [37]–[39], where the probing directions are chosen in a smaller angular range depending on the current estimated/tracked beam directions.

In these works, most of the beam direction estimation algorithms are designed under static or quasi-static scenarios without mobility, e.g., [22]–[24], [27]–[31], [34]–[36]. Such assumption will deteriorate the system performance a lot under dynamic scenarios, where beam tracking is a better choice, e.g., [32], [33], [37]–[39]. In [32], a tracking algorithm based on Kalman filtering is proposed. In [33], a probabilistic optimization problem is formulated to model the temporal evolution of the channel under mobility, and the proposed algorithm tracks the beam directions by maximizing the a posteriori probability. However, in [32], [33], an exhaustive beam sweeping stage is required before each tracking iteration, which is a waste of pilot resource. In [37], the authors excavate a temporal variation law of the physical direction and propose a tracking algorithm based on this law. In [38], the authors propose a method to optimize the precoders for pilot training within certain angular ranges. However, in [37], [38], the optimal training scheme with the lowest pilot overhead is not considered. In [39], a tracking algorithm is proposed to maximize the likelihood criterion, where the training scheme requires very small number of pilots. However, its calculation of new estimation in each iteration is of high complexity. Finally, it is worth highlighting that none of these beam-tracking works [32], [33], [37]–[39] studies the convergence property of the beam tracking algorithms.

This paper differs from the existing works in two aspects: First, we design a recursive beam tracking algorithm, which jointly optimizes the analog beamforming vectors for pilot training and the recursive estimator for beam directions. Second, we prove that this recursive beam tracking algorithm converges to the minimum CRLB with high probability, despite of the inherent non-convexity of analog beamforming problems, which ensures the fastest convergence to the real beam directions.

III. NOTATIONS AND MODEL

A. Notations

Lower case letters such as a and A will be used to represent scalars and column vectors, respectively, where |a| denotes the modulus of a and |a|2 denotes the 2-norm of a. Upper case letters such as A will be utilized to denote matrices. For a vector a or a matrix A, its transpose is denoted by aT or AT, and its Hermitian transpose is denoted by aH or AH. Let CN(u, σ2) stand for the circular symmetric complex Gaussian distribution with mean u and variance σ2, and N(u, σ2) stand for the real Gaussian distribution with mean u and variance σ2. The sets of integers and real numbers are written as ℤ and ℝ, respectively. Expectation is denoted by E[·] and the imaginary part of a variable x is denoted by Im{r}. The natural logarithm of x is denoted by log(x). The phase of a complex number z is obtained by $\angle z$.

B. System Model

Consider the linear antenna array receiver in Fig. 1, where $M$ antennas are placed along a line, with a distance $d$ between neighboring antennas. The antennas are connected by phase shifters to a single radio frequency (RF) chain, and the phase shifters are controlled digitally to steer the beam. In time-slot $n$, a narrow-beam pilot signal arrives at the antenna array...
from an angle-of-arrival (AoA) \( \theta_n \in [-\pi/2, \pi/2] \). Hence, the steering vector of this arriving beam is
\[
a(x_n) = \left[ 1, e^{j2\pi n x_n}, \ldots, e^{j2\pi (M-1)n x_n} \right]^T,
\]
(1)
where \( x_n = \sin(\theta_n) \) is the sine of the AoA \( \theta_n \) and \( \lambda \) is the wavelength. The channel response is \( h(x_n) = \beta a(x_n) \), where \( \beta \) is the complex channel response at the first antenna.

Let \( w_{mn} \in [-\pi, \pi] \) be the phase shift in radians provided by the \( m \)-th path of the first antenna. Then, the analog beamforming vector steered by the phase shifters is
\[
w_n = \frac{1}{\sqrt{M}} \left[ e^{jw_{1n}}, e^{jw_{2n}}, \ldots, e^{jw_{Mn}} \right]^T.
\]
(2)
Combining the output signals of the phase shifters and dividing the obtained summation signal by \( \beta \), yields
\[
y_n = w_n^H a(x_n) + \frac{z_n}{\sqrt{\rho}},
\]
(3)
where \( \rho = |\beta|^2/\sigma^2 \) is the SNR at each antenna, \( \sigma^2 \) is the noise power, and the \( z_n \)'s are i.i.d. circularly symmetric complex Gaussian random variables with zero mean and unity variance. Given \( x_n \) and \( w_n \), the conditional probability density function of \( y_n \) is
\[
p(y_n | x_n, w_n) = \frac{\rho}{\pi} e^{-\rho|y_n-w_n^H a(x_n)|^2}.
\]
(4)

A beam tracker determines the analog beamforming vector \( w_n \) and provides an estimate \( \hat{x}_n \) of the sine \( x_n \) of the AoA over time.\(^2\) From a control system perspective, \( x_n \) is the system state, \( \hat{x}_n \) is the estimate of the system state, the beamforming vector \( w_n \) is the control action, and \( y_n \) is a noisy observation that is determined by a non-linear (non-convex) function of the system state \( x_n \) and control action \( w_n \). Let \( \xi = (w_1, w_2, \ldots, \hat{x}_1, \hat{x}_2, \ldots) \) represent a beam tracking policy. We consider the set \( \Xi \) of causal beam tracking policies: At the end of time-slot \( n \), the estimate \( \hat{x}_n \) of time-slot \( n \) and the control action \( w_{n+1} \) of time-slot \( n+1 \) are determined by using the history of the control actions \( (w_1, \ldots, w_n) \), the estimates \( (\hat{x}_1, \ldots, \hat{x}_n) \) and the observations \( (y_1, \ldots, y_n) \).

In Sections IV and V we consider static beam tracking scenarios, i.e., \( x_n = x \) for all time-slot \( n \) and develop a recursive beam tracking algorithm that is asymptotically optimal with very high probability. In Section VI we evaluate this algorithm in dynamic beam tracking scenarios: By employing a simple TDMA round-robin pilot structure, this algorithm can potentially track hundreds of beams from high mobility transmitters/reflectors.

IV. PROBLEM FORMULATION AND PERFORMANCE BOUND

Consider static beam tracking scenarios, where \( x_n = x \) for all time-slot \( n \). Given any \( n \), the beam tracking problem can be formulated as
\[
\text{MSE}_{\text{opt}} \triangleq \min_{\xi \in \Xi} \mathbb{E}\left[ (\hat{x}_n - x)^2 \right]
\]
(5)
subject to
\[
\mathbb{E}[\hat{x}_n] = x,
\]
(6)
where the constraint (6) ensures that \( \hat{x}_n \) is an un-biased estimate of \( x \). Problem (5) is a constrained sequential control and estimation problem that is difficult, if not impossible, to solve optimally. First, the system is partially observed through the observation \( y_n \). Second, both the control action \( w_n \) and the estimate \( \hat{x}_n \) need to be optimized in Problem (5). On the one hand, because only the phase shifts \( (w_{1n}, \ldots, w_{Mn}) \) in (3) are controllable, the optimal control of \( w_n \) is a non-convex optimization problem. On the other hand, as shown in Fig. 3 and (18) below, the optimization of the estimate \( \hat{x}_n \) is also non-convex and there are multiple local optimal estimates.

A. Lower Bound of the Beam Tracking Error

Next, we establish a lower bound of MSE opt defined in (5):

Given the control actions \( (w_1, \ldots, w_n) \), the MSE is lower bounded by the Cramér-Rao lower bound (CRLB) \(\text{RACK ING}\)
\[
\mathbb{E}\left[ (\hat{x}_n - x)^2 \right] \geq \frac{1}{\sum_{i=1}^{n} I(x, w_i)},
\]
(7)
where \( I(x, w_i) \) is the Fisher information \(\text{RACK ING} \) that can be computed by using (4):
\[
I(x, w_i) = \mathbb{E}\left[ -\frac{\partial^2 \log p(y_i | x, w_i)}{\partial x^2} \right] | x, w_i |
\]
\[
= \frac{2\rho}{M} \sum_{m=1}^{M} \frac{2d}{\lambda} (m-1) e^{j[w_{mi}-\frac{\pi d}{M}(m-1)x]}.
\]
Note that the Fisher information \( I(x, w_i) \) is the function of \( w_i \). By optimizing the control actions \( (w_1, \ldots, w_n) \) in the right-hand-side (RHS) of (7), we obtain
\[
\frac{1}{n} \sum_{i=1}^{n} I(x, w_i) \leq \frac{2M(M-1)^2\pi^2 d^2 \rho}{\lambda^2} \triangleq I_{\text{max}},
\]
(8)
where the maximum Fisher information \( I_{\text{max}} \) in (3) is achieved if, and only if, for \( i = 1, \ldots, n \)
\[
w_i = \frac{a(x)}{\sqrt{M}} = \frac{1}{\sqrt{M}} \left[ 1, e^{j2\pi \frac{a}{M}x}, \ldots, e^{j2\pi \frac{a(M-1)x}} \right]^T.
\]
(9)
Hence, MSE opt in (5) is lower bounded by the minimum CRLB
\[
\text{MSE}_{\text{opt}} \geq \frac{1}{n I_{\text{max}}},\quad (10)
\]

V. RECURSIVE ANALOG BEAM TRACKING: ALGORITHM AND ANALYSIS

In this section, we design a recursive analog beam tracking algorithm and prove that its MSE converges to the lower bound on the RHS of (10) with very high probability.

A. Algorithm Design

We develop a recursive analog beam tracking algorithm which consists of two stages: 1) coarse beam sweeping and 2) recursive beam tracking. In Stage 1, an initial estimate \( \hat{x}_{i_{\text{init}}} \) is obtained, which should be in the mainlobe \( \mathcal{B} \), defined by
\[
\mathcal{B} = \left( x - \frac{\lambda}{Md}, x + \frac{\lambda}{Md} \right) \cap [-1, 1],
\]
(11)
with a high probability; one implementation that satisfies this requirement is described below. In Stage 2, the estimate \( \hat{x}_n \) and the control action \( w_n \) are updated recursively to realize an accurate tracking performance. As depicted in Fig. 2, \( M \) pilots will be sent successively in Stage 1, which is assumed to be finished by the end of time-slot \( n_0 \) and obtain the initial estimate \( \hat{x}_{n_0} \). In Stage 2, one pilot is allocated in each time-slot (e.g., at the beginning of each time-slot as in Stage 1).

### Recursive Analog Beam Tracking (Algorithm 1):

1) **Coarse Beam Sweeping:** Transmit \( M \) pilots successively in the first \( n_0 \geq 1 \) time-slots. The analog beamforming vector \( w_m \) for receiving the \( m \)-th pilot is

\[
w_m = \frac{1}{\sqrt{M}} a \left( \frac{2m}{M} - \frac{M+1}{M} \right), m = 1, \ldots, M.
\]

Find the initial estimate \( \hat{x}_{n_0} \) of the beam direction by

\[
\hat{x}_{n_0} = \arg \max_{\chi} a(\hat{x})^H \sum_{m=1}^{M} y_m w_m,
\]

where \( \chi = \left\{ \frac{1-M_0}{M_0}, \frac{2-M_0}{M_0}, \ldots, \frac{M_0-1}{M_0} \right\} \) and \( M_0 \) determines the recovery resolution.

2) **Recursive Beam Tracking:** In each time-slot \( n = n_0 + 1, n_0 + 2, \ldots \), the analog beamforming vector \( w_n \) is

\[
w_n = \frac{1}{\sqrt{M}} a(\hat{x}_{n-1}).
\]

The estimate \( \hat{x}_n \) of the beam direction is updated by

\[
\hat{x}_n = |\hat{x}_{n-1} - a_n \Im \{y_n\}|_1,
\]

where \( |x|_1 = \max \{ \min\{x, c\}, b \} \) and \( a_n > 0 \) is the step-size that will be specified later.

Our simulations suggest that, if the SNR \( \rho \geq 0 \) dB, a good initial estimate \( \hat{x}_{n_0} \) within the mainlobe \( B \) can be obtained with a probability higher than 99.99% in Stage 1 of Algorithm 1.

---

3 The beam direction estimate in [45] is motivated by the recovery algorithm in the compressed sensing method [42]. Its resolution is adapted by the size \( M_0 \) of the dictionary \( \chi \); a larger \( M_0 \) provides a more accurate estimate.

4 One can use more time-slots (pilot resources) to support lower SNR in Stage 1. As Stage 1 is executed only once, this will not increase the total pilot overhead by much.
Hence, for general step-size parameters $\alpha$ and $N_0$ in (19), $\hat{x}_n$ converges to a stable point in $\mathcal{S}(x)$ or a boundary point.

**Theorem 2 (Convergence to the Real Direction $x$).** If (i) the initial point satisfies $\hat{x}_{n_0} \in \mathcal{B}$, (ii) $a_n$ is given by (19) with any $\alpha > 0$, then there exist $N_0 \geq 0$ and $C(\hat{x}_{n_0}) > 0$ such that

$$P(\hat{x}_n \rightarrow x | \hat{x}_{n_0} \in \mathcal{B}) \geq 1 - 2e^{-C(\hat{x}_{n_0})\frac{\rho}{\alpha^2}}.$$  

(20)

**Proof Sketch.** Motivated by Chapter 4 of [43], we will prove this theorem in three steps: in Step 1, we will construct a time-invariant set $\mathcal{I}$ that contains the real direction $x$ within the mainlobe, i.e., $x \in \mathcal{I} \subset \mathcal{B}$; in Step 2, we will get the lower bound for the probability that the sequence $\{\hat{x}_n : n \geq n_0\}$ remains inside $\mathcal{I}$, i.e., $\hat{x}_n \in \mathcal{I}$ for $n \geq n_0$; in Step 3, we will show that this lower bound is also a lower bound for $P(\hat{x}_n \rightarrow x | \hat{x}_{n_0} \in \mathcal{B})$. The details are provided in Appendix B.

By Theorem 2, if the initial point $\hat{x}_{n_0}$ is in the mainlobe $\mathcal{B}$, the probability that $\hat{x}_n$ does not converge to $x$ decreases exponentially with respect to $\rho/\alpha^2$. Hence, one can increase the SNR $\rho$ and reduce the step-size parameter $\alpha$ to ensure $\hat{x}_n \rightarrow x$ with high probability. Typical values of $N_0$ required by the sufficient condition in Theorem 2 are 10-50. However, one can choose any $N_0 \geq 0$ to achieve a sufficiently high probability of $\hat{x}_n \rightarrow x$ in simulations.

**Theorem 3 (Convergence to $x$ with the Minimum MSE).** If (i) $a_n$ is given by (19) with

$$\alpha = \frac{\lambda}{\sqrt{M(M-1)}\sigma d} \triangleq \alpha^*,$$  

(21)

and any $N_0 \geq 0$, and (ii) $\hat{x}_n \rightarrow x$, then

$$\sqrt{n}(\hat{x}_n - x) \xrightarrow{d} \mathcal{N}(0, I_{\max}^{-1}),$$  

(22)

as $n \rightarrow \infty$, where $\xrightarrow{d}$ represents convergence in conditional distribution given $\hat{x}_n \rightarrow x$, and $I_{\max}$ is defined in (9). In addition, we have

$$\lim_{n \rightarrow \infty} n \mathbb{E}[(\hat{x}_n - x)^2 | \hat{x}_n \rightarrow x] = I_{\max}^{-1}. $$  

(23)

**Proof.** See Appendix C.

**Corollary 1.** If the conditions of Theorem 3 are satisfied, then for any second-order differentiable vector function $u(x)$

$$\lim_{n \rightarrow \infty} n \mathbb{E} \left[ \frac{\|u(\hat{x}_n) - u(x)\|^2}{\|u(x)\|^2} | \hat{x}_n \rightarrow x \right] = \frac{\|u(x)\|^2}{\|u(x)\|^2} = 1.$$

**Proof.** See Appendix D.

For example, consider the channel response $h(x) = \beta a(x)$. If $\alpha = \alpha^*$ and $N_0 = 0$, Corollary 1 tells us that, with a high probability, the minimum CRLB of $h(x)$ is achieved in the following limit:

$$\lim_{n \rightarrow \infty} n \mathbb{E} \left[ \frac{\|h(\hat{x}_n) - h(x)\|^2}{\|h(x)\|^2} | \hat{x}_n \rightarrow x \right] = \frac{\|h(x)\|^2}{\|h(x)\|^2} = \frac{(2M-1)\sigma^2}{3(M-1)}. $$  

(24)

**C. To Track the AoA $\theta$ or its Sine $x$?**

We can design the analog beam tracking algorithm by tracking either the AoA $\theta$ or its sine $x$. The algorithm that tracks the sine $x$ is provided in Algorithm 1. The algorithm that directly tracks the AoA $\theta$, called **Algorithm 2**, is described as follows:

1) **Coarse Beam Sweeping:** Transmit $M$ pilots successively in the first $n_0 \geq 1$ time-slots. The analog beamforming vector $\mathbf{w}_m$ for receiving the $m$-th pilot is given by (12). The initial estimate $\hat{\theta}_{n_0}$ of the beam direction is

$$\hat{\theta}_{n_0} = \arcsin \left\{ \max_{\theta \in \mathcal{X}} \left| \mathbf{a}(\hat{x}_{n_0})^H \cdot \sum_{m=1}^{M} y_m \mathbf{w}_m \right| \right\}, $$  

(25)

where $\mathcal{X} = \left\{ \frac{\lambda M_0}{M}, \frac{\lambda M_1}{M}, \ldots, \frac{\lambda M_{M-1}}{M} \right\}$ and $M_0$ determines the recovery resolution.

2) **Recursive Beam Tracking:** In each time-slot $n = n_0 + 1, n_0 + 2, \ldots$, the analog beamforming vector $\mathbf{w}_n$ is

$$\mathbf{w}_n = \frac{1}{\sqrt{M}} \mathbf{a}(\sin(\hat{\theta}_{n-1})).$$  

(26)

The estimate $\hat{\theta}_n$ is updated by

$$\hat{\theta}_n = \hat{\theta}_{n-1} - \frac{a_n}{\cos(\hat{\theta}_{n-1})} \Im \left\{ y_n \right\} \hat{x}_{n},$$  

(27)

The convergence rate of this tracking algorithm is characterized by Corollary 1 with $u(x) = \arcsin x$. In particular, Algorithm 1 and Algorithm 2 share the same asymptotic convergence rate when $\theta_n$ is very close to $\theta$. On the other hand, if $\theta_n = 0$ is close to $-\frac{\pi}{2}$ or $\frac{\pi}{2}$, $\cos(\theta_n)$ in (27) is close to zero. As a result, Algorithm 2 is not stable and may even oscillate when $\theta_n$ is close to $-\frac{\pi}{2}$ or $\frac{\pi}{2}$. However, this oscillation issue does not exist in Algorithm 1.
Algorithm 2 that tracks the AoA $\theta$ starts to oscillate, while Algorithm 1 is stable.

In addition, \( \big| \beta \big| \) and \( \big| \alpha \big| \) in Algorithm 1 are less complicated than (26) and (27) in Algorithm 2 (although both algorithms are of low complexity). Because of these reasons, we choose to track the sine $x$ of the AoA in this paper, instead of directly tracking the AoA $\theta$.

VI. SIMULATION RESULTS

We compare Algorithm 1 with three reference algorithms:

1) **IEEE 802.11ad** [12]: This algorithm contains two stages: beam sweep and beam tracking. In the first stage, sweep the beamforming directions in the DFT codebook (12) and choose the direction with the strongest received signal as the best beam direction. In the second stage, probe the best beam direction and its two adjacent beam directions, then choose the strongest direction as the new best beam direction. The second stage is performed periodically.

2) **Least square** [21]: Sweep all the beamforming directions in the DFT codebook (12) and use the least square algorithm to estimate the channel response $\hat{h}(x_n)$. Then obtain the analog beamforming vector $w_n$ for data transmission by

$$w_{mn} = \vartheta \hat{h}_m(x_n), m = 1, 2, \ldots, M,$$

where $\hat{h}_m(x_n)$ is the $m$-th element of the estimated channel response $\hat{h}(x_n)$.

3) **Compressed sensing** [22]–[24]: Randomly choose the phase shifts $w_{mn}$ from $\{\pm 1, \pm j\}$ to receive pilot signals. Then use the sparse recovery algorithm to estimate the sine of AoA $x_n$, where a DFT dictionary with a size of 1024 is utilized.

Two performance metrics are considered: (i) the MSE of the channel response $\hat{h}(x_n)$, defined by

$$\text{MSE}_{h,n} \triangleq \mathbb{E} \left[ \left| \| \hat{h}(x_n) - h(x) \|_2^2 \right| \right]$$

for the least square algorithm and

$$\text{MSE}_{h,n} \triangleq \mathbb{E} \left[ \left| \| h(x_n) - h(x) \|_2^2 \right| \right]$$

for other algorithms, and (ii) the achievable rate $R_n$, i.e.,

$$R_n \triangleq \log_2 \left( 1 + \rho \| \hat{w}_n^H a(x_n) \|^2 \right).$$

The system parameters are configured as: $\beta = (1+j)/\sqrt{2}, \rho = |\beta|^2/\sigma^2 = 10\, \text{dB}, M = 16, M_0 = 2M, d = 0.5\lambda$. 

A. Static Beam Tracking

For static beam tracking, we assume that one pilot is allocated in each time-slot. Hence, these algorithms have the same pilot overhead. The received pilot signals of all time-slots $1, \ldots, n$ are used for estimating $x_n$ and $h(x_n)$ in the compressed sensing and least square algorithms. The IEEE 802.11ad algorithm was designed for dynamic beam tracking and is not plotted here for the static beam tracking scenario. The step-size $a_n$ is given by (19) with $\alpha = \alpha^*$ and $N_0 = 0$. The simulation results are averaged over 10000 random system realizations, where the beam direction $x$ is randomly generated by a uniform distribution on $[-1, 1]$ in each realization.

Figure 5 plots the convergence performance of $n \times \text{MSE}_{h,n}$ over time. The MSE of Algorithm 1 converges quickly to the minimum CRLB given in (24), which agrees with Corollary 1 and is much smaller than those of least square and compressed sensing algorithms.

B. Dynamic Beam Tracking

For dynamic beam tracking, we assume that continuous pilot training is performed in the first time-slot and an initial estimate is obtained for all the algorithms. After that, one pilot is allocated in each time-slot to ensure that these algorithms have the same amount of pilot overhead.

The last $M/2$ pilot signals are used in the compressed sensing algorithm and the last $M$ pilot signals are used in the least square algorithm. For the IEEE 802.11ad algorithm, the probing period of its beam tracking stage is 3 time-slots. These parameters are chosen to improve the performance of these algorithms. As suggested in [42], the step-size $a_n$ of Algorithm 1 is fixed as

$$a_n = \alpha^* = \frac{\lambda}{\sqrt{M(M-1)\pi d}}, \quad \text{for all } n \geq 1. \quad (28)$$
This step-size is determined by the configuration of the antenna array system, which is independent of the SNR $\rho$.

Figures 6 and 7 depict the AoA tracking and achievable rate performance in dynamic scenarios, where the AoA $\theta_n$ varies according to $\theta_n = \frac{\pi}{3} \sin \left( \frac{2\pi n}{10000} \right) + 0.005 \vartheta_n$ with $\vartheta_n \sim \mathcal{N}(0, 1)$. Algorithm 1 always tracks the actual AoA very well, and achieves the channel capacity $7.33 \text{bits/s/Hz}$ in all the time-slots. The performance of Algorithm 1 is much better than the other three algorithms, and the algorithm used by IEEE 802.11ad is better than the other two.

Figures 8 and 9 illustrate the average AoA tracking and achievable rate performance under a fixed angular velocity model $\theta_n = \theta_{n-1} + s_{n-1} \omega$ where $n = 1, \ldots, 10000$, $\theta_0 = 0$, $s_n \in \{-1, 1\}$ denotes the rotation direction, and $\omega$ is a fixed angular velocity. The rotation direction $s_n$ is chosen such that $\theta_n$ varies within $[-\frac{\pi}{3}, \frac{\pi}{3}]$. The antenna number is 16. One can observe that Algorithm 1 can support higher angular velocities and data rates than the other algorithms when all 16 antennas are used. In addition, by using a subset of antennas, e.g., $M = 4$ or 8, for beam tracking and all 16 antennas for data transmissions, the beam tracking regime of Algorithm 1 can be further enlarged.

According to Fig. 9, Algorithm 1 can achieve 95% of the channel capacity when the angular velocity of the beam direction is $0.064 \text{rad/time-slot}$, the SNR is $\rho = 10 \text{dB}$, and $M = 8$. If each time-slot (TTI) lasts for 0.2ms [25], [26], Algorithm 1 can support an angular velocity of $0.064 \times \frac{1000}{0.2} = 320 \text{rad/s} \approx 51 \text{circles/s}$. Consider a TDMA pilot pattern where 1000 narrow-beam pilots are sent to the antenna array periodically in a round-robin fashion such that 1 pilot is sent in each time-slot. Algorithm 1 can support $0.32 \text{ rad/s}$ per beam for tracking all these 1000 beams, which is $72 \text{mph}$ if the transmitters/reflectors steering these beams are at a distance of 100 meters.

At last, we consider the condition that SNR is $\rho = 0 \text{dB}$ and other parameters are the same as Figs. 8 and 9. As depicted in Figs. 10 and 11, it can be seen that Algorithm 1 can provide higher performance gain than the condition that SNR is $\rho = 10 \text{dB}$, when all 16 antennas are used. Moreover, by using $M = 8$ antennas for tracking and all 16 antennas for data transmissions, the beam tracking regime of Algorithm 1 can still be enlarged. But when $M = 4$ antennas are used for tracking, the performance deterioration is quite significant due to the low antenna gain. Therefore, when SNR is low, more antennas are needed to ensure the good tracking performance.

VII. CONCLUSIONS

We have developed an analog beam tracking algorithm, and established its convergence and asymptotic optimality. Our theoretical and simulation results show that this algorithm can achieve much lower beam tracking error and higher data rate than several state-of-the-art algorithms. In our future work, we will consider hybrid beamforming systems with multiple RF chains and two-dimensional antenna arrays, based on the methodology developed in the current paper.

ACKNOWLEDGEMENT

The authors are grateful to Ashutosh Sabharwal for the helpful discussion of the paper and valuable suggestions.

REFERENCES

[1] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Trans. Wireless Commun., vol. 9, no. 11, Nov. 2010.
[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Process. Mag.*, vol. 30, no. 1, Jan. 2013.

[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014.

[4] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” *IEEE Commun. Mag.*, vol. 49, no. 6, Jun. 2011.

[5] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014.

[6] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An overview of signal processing techniques for millimeter wave MIMO systems,” *IEEE J. Sel. Top. Signal Process.*, Apr. 2016.

[7] T. Ohira and K. Gyoda, “Electronically steerable passive array radiator antennas for low-cost analog adaptive beamforming,” in *IEEE International Conference on Phased Array Systems and Technology*, 2000.

[8] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, “MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?” *IEEE Commun. Mag.*, vol. 52, no. 12, Dec. 2014.

[9] S. Han, C. L. I, Z. Xu, and C. Rowell, “Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G,” *IEEE Commun. Mag.*, vol. 53, no. 1, Jan. 2015.

[10] A. Puglielli, A. Townley, G. LaCaille, V. Milovanovi, P. Lu, K. Trotskovsk, A. Whitcombe, N. Narevsky, G. Wright, T. Courtade, E. Alon, B. Nikoli, and A. M. Niknejad, “Design of energy- and cost-efficient massive MIMO arrays,” *Proc. IEEE*, vol. 104, no. 3, Mar. 2016.

[11] A. F. Molisch, V. V. Ratnam, S. Han, S. L. H. Nguyen, L. Li, and K. Haneda, “Hybrid beamforming for massive MIMO-a survey,” *arXiv preprint arXiv:1609.05076*, 2016.

[12] IEEE standard, “IEEE 802.11ad WLAN enhancements for very high throughput in the 60 GHz band,” Dec. 2012.

[13] “IEEE 802.15.3c WPAN millimeter-wave-based alternative physical layer extension,” Oct. 2009.

[14] METIS Report, “Final performance results and consolidated view on the most promising multi-node/multi-antenna transmission technologies,” Feb. 2015.

[15] ITU Report, “Technical feasibility of IMT in bands above 6GHz,” Jul. 2015.

[16] KEYSIGHT Technologies, “Massive MIMO and mmWave technology insight and challenges,” 2015.

[17] Samsung Electronics, “5G Vision,” Feb. 2015.

[18] A. Ghosh, “Enabling technologies for next generation wireless systems,” *Nokia Bell Labs*, Mar. 2016.

[19] W. Tong, “Bringing 5G into reality,” *Huawei*, Mar. 2016.

[20] G. Brown, O. Koymen, and M. Branda, “The promise of 5G mmWave - How do we make it mobile?” *Qualcomm Technologies*, Jun. 2016.

[21] E. Karami, “Tracking performance of least squares MIMO channel estimation algorithm,” *IEEE Trans. Commun.*, vol. 55, no. 11, Nov. 2007.

[22] B. Gao, Z. Xiao, L. Su, Z. Chen, D. Jin, and L. Zeng, “Multi-device multi-path beamforming training for 60-GHz millimeter-wave communications,” in *2015 IEEE ICC*, Jun. 2015.

[23] A. Alkhateeb, G. Leus, and R. W. Heath, “Hybrid MIMO architectures for millimeter wave communication: potentials and approaches,” *IEEE Commun. Mag.*, vol. 54, no. 5, May 2016.

[24] J. Wang, Z. Lan, C.-W. Pyo, T. Baykas, C.-S. Sum, M. A. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada, and S. Kato, “Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems,” *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, Oct. 2009.

[25] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, Oct. 2014.

[26] Z. Xiao, P. Xia, and X. G. Xia, “Enabling UAV cellular with millimeter-wave communication: potentials and approaches,” *IEEE Commun. Mag.*, vol. 54, no. 5, May 2016.

[27] X. Gao, L. Dai, Y. Zhang, T. Xie, X. Dai, and Z. Wang, “Fast channel tracking for Terahertz beamspace massive MIMO systems,” *IEEE Trans. Veh. Technol.*, vol. PP, no. 99, Oct. 2016.

[28] N. Garcia, H. Wyneersch, and D. Stolk, “Optimal robust precoders for tracking the AoD and AoA of a mm-Wave path,” *arXiv preprint arXiv:1703.10978*, 2017.

[29] J. Bae, S. H. Lim, J. H. Yoo, and J. W. Choi, “New beam tracking technique for millimeter wave-band communications,” *arXiv preprint arXiv:1702.00276*, 2017.

[30] M. B. Nevel’son and R. Z. Has’minskii, *Stochastic approximation and recursive estimation*, 1973.

[31] J. V. Poor, *An introduction to signal detection and estimation*. New York, NY, USA: Springer-Verlag New York, Inc., 1994.

[32] H. Kushner and G. G. Yin, *Stochastic approximation and recursive algorithms and applications*. Springer Science & Business Media, 2003, vol. 35.

[33] V. S. Borkar, *Stochastic approximation: a dynamical systems viewpoint*, 2008.

[34] J. M. Holte, *Discrete Gronwall lemma and applications*, in *MAA-NCS meeting at the University of North Dakota*, vol. 24, 2009.

[35] W. Hoefding, “Probability inequalities for sums of bounded random variables,” *J. Am. Stat. Assoc.*, vol. 58, no. 301, 1963.

**APPENDIX A**

**PROOF OF THEOREM 1**

Recall the recursive procedure (14) and (15):

\[ \hat{x}_n = [\hat{x}_{n-1} - a_n \text{Im}\{y_n\}]_{-1}, \]

where

\[ -\text{Im}\{y_n\} = f(\hat{x}_{n-1}, x) + \tilde{z}_n, \]

(29)

\[ f(\hat{x}_{n-1}, x) \]

is defined in (17) and

\[ \tilde{z}_n \Delta = -\frac{z_n}{\sqrt{\rho}} \sim N(0, \frac{1}{2\rho}), \]

(30)

Let \( G_n : n \geq n_0 \) be an increasing sequence of \( \sigma \)-fields of \( \{\hat{x}_{n-1}, \hat{x}_{n-1} + 1, \ldots, \hat{x}_{n-1} + n_0\} \), i.e., \( G_n \subset G_{n+1} \) where \( G_n \Delta = \sigma(\hat{x}_{n-1}) \) and \( G_n \Delta = \sigma(\hat{x}_{n-1}, \hat{x}_{n-1} + 1, \ldots, \hat{x}_{n-1} + n_0) \) for \( n \geq n_0 + 1 \). Because the \( z_n \)'s are i.i.d. circularly symmetric complex Gaussian random variables with zero mean, \( \tilde{z}_n \) is independent of \( G_{n-1} \) and \( \hat{x}_{n-1} \in G_{n-1} \). Hence, we have

\[ E[-\text{Im}\{y_n\} | G_{n-1}] = E[f(\hat{x}_{n-1}, x) + \tilde{z}_n | G_{n-1}]\]

\[ = E[f(\hat{x}_{n-1}, x) | G_{n-1}] + E[\tilde{z}_n | G_{n-1}] = f(\hat{x}_{n-1}, x), \]

(31)

for \( n \geq n_0 + 1 \).

Theorem 5.2.1 [42] has proposed the sufficient conditions that \( \hat{x}_n \) converges to a unique point within the set of stable...
points with probability one. We will prove that when the step-size \(a_n\) is given by (19) with any \(\alpha > 0\) and \(N_0 \geq 0\), our algorithm satisfies these sufficient conditions:

1) Step-size constraints:

\[
an_n = \frac{\alpha}{n - n_0 + N_0} \to 0,
\]

\[
\sum_{n=n_0+1}^{\infty} a_n = \sum_{n=n_0+1}^{\infty} \frac{\alpha}{n - n_0 + N_0} = \sum_{i=1}^{\infty} \frac{\alpha}{i + N_0} \to \infty,
\]

\[
\sum_{n=n_0+1}^{\infty} a_n^2 = \sum_{n=n_0+1}^{\infty} \left( \frac{\alpha^2}{(n - n_0 + N_0)^2} \right) \leq \sum_{n=n_0+1}^{\infty} \frac{\alpha^2}{n^2} = \sum_{i=1}^{\infty} \frac{\alpha^2}{i^2} < \infty.
\]

2) We need to prove that \(\sup_n \mathbb{E} \left[ \left| - \text{Im} \{y_n\} \right|^2 \right] < \infty\).

From (29), we have

\[
\mathbb{E} \left[ \left| - \text{Im} \{y_n\} \right|^2 \right] \leq \frac{1}{\sqrt{M}} \sum_{m=1}^{M} e^{j \frac{2\pi d}{\lambda} (m-1)(\hat{x}_{n-1} - x)} \leq \frac{1}{\sqrt{M}},
\]

so we get

\[
\mathbb{E} \left[ \left| f(\hat{x}_{n-1}, x) \right|^2 \right] \leq \frac{1}{\sqrt{M}} \sum_{m=1}^{M} e^{j \frac{2\pi d}{\lambda} (m-1)(\hat{x}_{n-1} - x)} \leq \sqrt{M},
\]

Combining (32) and (33), we have

\[
\mathbb{E} \left[ \left| f(\hat{x}_{n-1}, x) \right|^2 \right] \leq \frac{1}{\sqrt{M}} \sum_{m=1}^{M} e^{j \frac{2\pi d}{\lambda} (m-1)(\hat{x}_{n-1} - x)} \leq \sqrt{M}.
\]

3) The function \(f(v, x)\) should be continuous with respect to \(v\).

From (17), \(f(v, x)\) can be rewritten as follows:

\[
f(v, x) = -\frac{1}{\sqrt{M}} \sum_{m=1}^{M} \sin \left[ \frac{2\pi d}{\lambda} (m-1) (v-x) \right] \sin \left[ \frac{2\pi d}{\lambda} (m-1) (\hat{x}_{n-1} - x) \right].
\]

Because \(\sin \left[ \frac{2\pi d}{\lambda} (m-1) (v-x) \right]\) is continuous with respect to \(v\), and \(f(v, x)\) is the summation of a finite amount of \(\sin \left[ \frac{2\pi d}{\lambda} (m-1) (v-x) \right], m = 1, \ldots, M\). Therefore, we can conclude that \(f(v, x)\) is continuous with respect to \(v\).

4) Let \(\gamma_n = \mathbb{E} \left[ -\text{Im} \{y_n\} | G_{n-1} \right] - f(\hat{x}_{n-1}, x)\). We need to prove that \(\sum_{n=n_0+1}^{\infty} |a_n \gamma_n| < \infty\) with probability one.

From (31), we get \(\gamma_n = 0\) for all \(n \geq n_0 + 1\). So we have \(\sum_{n=n_0+1}^{\infty} |a_n \gamma_n| = 0 < \infty\) with probability one.

5) The set of stable points for the ODE (16) should be obtained.

According to (18), \(S(x)\) contains the local optimal stable points for the ODE (16). What’s more, the boundary point 1 (or -1) is a stable point when \(f(1, x) \geq 0\) (or \(f(-1, x) \leq 0\)). Hence, the set of stable points is a subset of \(S(x) \cup \{-1\} \cup \{1\}\).

Therefore, \(\hat{x}_n\) converges to a unique point within \(S(x) \cup \{-1\} \cup \{1\}\) with probability one.

\section*{Appendix B}

\section*{Proof of Theorem 2}

We need to define the following parameters: Let \(t_{n_0} \triangleq 0, t_n \triangleq \sum_{i=n_0+1}^{n} a_i, n \geq n_0 + 1\). We define \(\hat{x}(t), t \geq 0\) as the linear interpolation of the sequence \(\{\hat{x}_n : n \geq n_0\}\), where \(\hat{x}(t_n) = \hat{x}_n, n \geq n_0\) and \(\hat{x}(t)\) is given by

\[
\hat{x}(t) = \hat{x}(t_n) + \frac{(t-t_n)}{\alpha_n} [\hat{x}(t_{n+1}) - \hat{x}(t_n)], t \in [t_n, t_{n+1}].
\]

Also, define \(\hat{x}^n(t)\) as the solution of the ODE (16) for \(t \in [t_n, \infty)\), where \(\hat{x}^n(t_n) = \hat{x}(t_n) = \hat{x}_n, n \geq n_0\). Since we only care about the condition that \(\hat{x}_n \in \mathcal{I} \subset \mathcal{B}\), there exist two cases: (i) if \(\pm 1 \notin \mathcal{I}\), then the solution of the ODE (16) is within \((-1, 1)\), (ii) if 1 (or -1) is in \(\mathcal{I}\), then \(f(1, x) < 0\) (or \(f(-1, x) \geq 0\)). Hence, the ODE is \(\frac{d\hat{x}^n(t)}{dt} = f(\hat{x}^n(t), x)\) and we have

\[
\hat{x}^n(t) = \hat{x}(t_n) + \int_{t_n}^{t} f(\hat{x}^n(v), x) dv, t \geq t_n.
\]

Moreover, let \(\xi_{n_0} \triangleq 0\) and \(\xi_n \triangleq \sum_{m=n_0+1}^{n} a_m \epsilon_m, n \geq n_0 + 1\). We will prove Theorem 2 in three steps.

\textbf{Step 1:} We first construct a time-invariant set \(\mathcal{I}\) that contains the real direction \(x\) within the mainlobe, i.e., \(x \in \mathcal{I} \subset \mathcal{B}\). Pick \(\delta\) such that

\[
\inf_{v \in \partial \mathcal{B}} \{v - \hat{x}_{n_0}\} > \delta > 0,
\]

Then, we construct the invariant set \(\mathcal{I}\) as follows:

\[
\mathcal{I} = \left( x - \left| x - \hat{x}_{n_0} \right| - \delta, x + \left| x - \hat{x}_{n_0} \right| + \delta \right) \subset \mathcal{B}.
\]

\textbf{Step 2:} We would like to obtain the lower bound for the probability that the sequence \(\{\hat{x}_n : n \geq n_0\}\) remains inside \(\mathcal{I}\), i.e., \(\hat{x}_n \in \mathcal{I}\) for \(n \geq n_0\). It can be realized in two sub-steps: in \textbf{Step 2-(a)}, we establish a sufficient condition in Lemma 1 that ensures \(\hat{x}_n \in \mathcal{I}\) for \(n \geq n_0\) with the help of the ODE (16); in \textbf{Step 2-(b)}, we construct the basic equations that are needed to calculate the probability lower bound; in \textbf{Step 2-(c)}, we obtain the probability lower bound by using this sufficient condition.

\textbf{Step 2-(a)}: Pick \(T > 0\) such that the solution \(x(t), t \geq 0\) of the ODE (16) with \(x(0) = \hat{x}_{n_0}\) satisfies \(\inf_{v \in \partial \mathcal{B}} \{v - x(t)\} > 2\delta\) for \(t \geq T\). Due to that the solution \(x(t)\) of the ODE (16) will approach the real direction \(x\) monotonically within the mainlobe \(\mathcal{B}\) as time \(t\) increases, we have \(\left| \hat{x}_n - x(T) \right| > \delta\) and one possible choice of \(T\) given by

\[
T = \max \left\{ \frac{\delta}{\left| f(\hat{x}_{n_0}, x) \right|}, \frac{\delta}{\left| f(\hat{x}_{n_0} - \delta, x, x) \right|} \right\}.
\]

\textsuperscript{5}The boundary of the set \(\mathcal{B}\) is denoted by \(\partial \mathcal{B}\).
Define $T_0 \triangleq 0$ and $T_{m+1} \triangleq \min\{t_i : t_i \geq T_n + T, i \geq n_0\}$ for $m \geq 0$. Then $T_{m+1} - T_m \in [T, T + a_n]$ and $T_m = t_n(m)$ for some $\hat{n}(m) \uparrow \infty$, where $\hat{n}(0) = n_0$. Let $\hat{x}(\hat{n}(m)(t))$ denote the solution of ODE \((16)\) for $t \in I_m \triangleq [T_m, T_{m+1}]$ with $\hat{x}(\hat{n}(m)(T_m)) = \hat{x}(T_m)$, $m \geq 0$.

Hence, we can obtain the following lemma:

**Lemma 1.** If $\sup_{t \in I_m} |\hat{x}(t) - \hat{x}(\hat{n}(m)(t))| \leq \delta$ for $m \geq 0$, then $\hat{x}_n \in I$ for $n \geq n_0$.

**Proof.** When $m = 0$, $\hat{x}(0)(T_0) = \hat{x}_0(T_0) = \hat{x}_n$. There are two cases: (i) $\hat{x}_n \in I \subset B$ and the monotonic property of the ODE \((16)\) within the mainlobe $B$, we get $\hat{x}(\hat{n}(0)(t) \geq 0$, and $x - \hat{x}(\hat{n}(0)(t) 

For $t \in I_0$. Therefore, we can obtain

$$\hat{x}(t) - \hat{x}(\hat{n}(0)(t)) + \delta \geq -\delta + 0 + \delta = 0,$$
and

$$x - \hat{x}(\hat{n}(0)(t)) + \delta \geq 0,$$

which result in $\hat{x}(t) \in I$ for $t \in I_0$. What's more, with the $T$ given by \((38)\), we have

$$x - \hat{x}_n \geq \hat{x}(\hat{n}(0)(T_1) - \hat{x}_n \geq \hat{x}(\hat{n}(0)(T) - \hat{x}_n \geq \delta.$$

Therefore, we get

$$\hat{x}(T_1) - \hat{x}_n \geq \hat{x}(\hat{n}(0)(T_1)) - \hat{x}_n \geq \hat{x}(\hat{n}(0)(T)) - \hat{x}_n \geq \delta.$$

Hence, we have

$$\hat{x}(T_1) - \hat{x}_n \geq \hat{x}(\hat{n}(0)(T_1)) - \hat{x}_n \geq \hat{x}(\hat{n}(0)(T)) - \hat{x}_n \geq \delta.$$

**Step 2-(b):** With \((39), \hat{n}(0)(t) \leq \delta\), and $T$ given by \((38)\), we have for $t_n + m$, $1 \leq m \leq n_T \triangleq \min\{t_i : t_i \geq t_n + T\}$.

$$\hat{x}(t_n + m) = \hat{x}(t_n) + \sum_{i=1}^{m} a_{n+i} f(\hat{x}(t_n+i-1), x) + (\xi_{n+i} - \xi_n),$$
and

$$\hat{x}(t_n + m) = \hat{x}(t_n) + \int_{t_n}^{t_n + m} f(\hat{x}(v), x)dv \geq \hat{x}(t_n) + \sum_{i=1}^{m} a_{n+i} f(\hat{x}(t_n+i-1), x) + \int_{t_n}^{t_n + m} \left[ f(\hat{x}(v), x) - f(\hat{x}(v), x) \right]dv,$$

where $\Delta \triangleq \max\{t_n : t_n \leq v, n \geq n_0\}$ for $v \geq 0$. Note that we only care about $\hat{x}_n \in I \subset B$, so the projection operator does not take effect in \((43)\) and we omit it.

To bound $\int_{t_n}^{t_n + m} \left[ f(\hat{x}(v), x) - f(\hat{x}(v), x) \right]dv$ on the RHS of \((44)\), we obtain the Lipschitz constant of function $f(v, x)$ considering the first variable v, given by

$$L \triangleq \sup_{v \neq v_2} \frac{|f(v_1, x) - f(v_2, x)|}{|v_1 - v_2|}.$$

Plugging \((47)\) into \((45)\), yields $L = \sqrt{M} \{M-1\}$, which is not related to $x$. Similar to \((38)\), for any $t \geq t_n$, we can obtain

$$|f(\hat{x}(T), x)| \leq \sqrt{M}.$$
Step 2-(b): Suppose there are constants $\gamma$ in (36), and step $n$ $(\tilde{\gamma} \leq \gamma \leq \tilde{\gamma} = \sup_{n} a_n)$ due to (46). Then, by subtracting $\tilde{x}_n(t_{n+m})$ in (44) from $\tilde{x}(t_{n+m})$ in (43) and taking norms, the following inequality can be obtained from (45) and (47) for $n \geq n_0$:

$$
\|\tilde{x}(t_{n+m}) - \tilde{x}_n(t_{n+m})\| \\
\leq \frac{\sqrt{M}a_n}{2} + Ce^{L(T+\alpha_n)}
$$

Proof. Apply the discrete Gronwall inequality (44), leading to (50) to

$$
\|\tilde{x}(t_{n+m}) - \tilde{x}_n(t_{n+m})\| \leq Ce^{L(T+\alpha_n)}
$$

Since $1 \leq m \leq n_T$ and $n_T = \inf \{i \in \mathbb{Z}: t_{n+i} \geq t_n + T\}$, we get

$$
\sum_{i=1}^{m} a_{n+i} = t_{n+m} - t_n \leq T + a_n + n_T \leq T + a_n.
$$

By combining (52) and (53), we have

$$
\|\tilde{x}(t_{n+m}) - \tilde{x}_n(t_{n+m})\| \leq Ce^{L(T+\alpha_n)}
$$

For all $t \in [t_{n+m-1}, t_{n+m}]$, we have

$$
\tilde{x}(t) = \tilde{x}(t_{n+m-1}) + \sum_{i=1}^{m} a_{n+i} + (t - t_{n+m-1}) \tilde{x}(t_{n+m}) - \tilde{x}(t_{n+m-1})
$$

where $\gamma = \frac{t_{n+m-1}}{a_n} \in [0, 1]$. Then, we can get (55) on the top of the page, where step (a) is according to the definition of $\tilde{x}_n(t)$ in (36), step (b) is due to (54), step (c) is obtained from (10), and step (d) is obtained by using $\gamma = \frac{t_{n+m-1}}{a_n}$.

Therefore, from (55), we can obtain

$$
\sup_{t \in [t_n, t_{n+m}]} |\tilde{x}(t) - \tilde{x}_n(t)| \leq \frac{\sqrt{M}a_n}{2} + Ce^{L(T+\alpha_n)},
$$

which completes the proof.

Lemma 3. Suppose there are constants $C > 0$ and $T > 0$ such that for $n \geq n_0$ and $m \leq n_T$, we have

$$
|\tilde{x}(t_{n+m}) - \tilde{x}_n(t_{n+m})| \\
\leq L \sum_{i=1}^{m} a_{n+i} |\tilde{x}(t_{n+i-1}) - \tilde{x}_n(t_{n+i-1})| + C.
$$

Following the lemmas are needed to prove Lemma 2.

Lemma 2. If (i) the initial point satisfies $\tilde{x}_{n_0} \in \mathcal{B}$, (ii) $a_n$ is given by (49) with any $\alpha > 0$, then there exist $N_0 \geq 0$ and $C(\tilde{x}_{n_0}) > 0$ such that

$$
P(\tilde{x}_n \in \mathcal{L}, \forall n \geq n_0) \\
\geq P\left(\sup_{t \in [t_n, t_{n+m}]} |\tilde{x}(t) - \tilde{x}_n(t)| \leq \delta, \forall m \geq 0\right) \\
\geq 1 - 2e^{-C(\tilde{x}_{n_0}) \frac{\delta^2}{\alpha^2}}.
$$

The following lemmas are needed to prove Lemma 2.

Lemma 4. Let $\{M_i : i = 1, 2, \ldots\}$ be a Gaussian martingale, then for any $\eta > 0$,

$$
P\left(\sup_{0 \leq i \leq k} |M_i| > \eta\right) \leq 2 \exp\left\{-\frac{\eta^2}{2 \text{Var}[M_k]}\right\}.
$$
Proof. As $M_i$ is a Gaussian martingale in $i$ and the exponential function is positive and convex, $e^{CM_i}$ is a positive submartingale for any $C > 0$. By utilizing the Doob’s inequality \cite{45} for $\eta > 0$, we have

$$P\left( \sup_{0 \leq i \leq k} M_i > \eta \right) \leq \frac{\mathbb{E}[e^{CM_k}]}{e^{C\eta}}.$$ 

Due to the property of Gaussian distribution, we have

$$\mathbb{E}[e^{CM_k}] = \exp \left\{ \frac{C^2}{2} \text{Var}[M_k] \right\}.$$ 

Then we can obtain

$$P\left( \sup_{0 \leq i \leq k} M_i > \eta \right) \leq \exp \left\{ \frac{C^2}{2} \text{Var}[M_k] - C\eta \right\}.$$ 

We choose the $C$ to minimize the upper bound above, which yields $C = \frac{\eta}{\text{Var}[M_k]}$. Therefore, we have

$$P\left( \sup_{0 \leq i \leq k} M_i > \eta \right) \leq \exp \left\{ -\frac{\eta^2}{2 \text{Var}[M_k]} \right\}.$$ 

Because the distribution of $\{M_1, M_2, \ldots, M_k\}$ is symmetric, we get

$$P\left( \sup_{0 \leq i \leq k} |M_i| > \eta \right) = P\left( \sup_{0 \leq i \leq k} M_i > \eta \right) \cup \inf_{0 \leq i \leq k} M_i < -\eta \right) \leq P\left( \sup_{0 \leq i \leq k} M_i > \eta \right) + P\left( \inf_{0 \leq i \leq k} M_i < -\eta \right) = 2P\left( \sup_{0 \leq i \leq k} M_i > \eta \right).$$ 

Hence, we have

$$P\left( \sup_{0 \leq i \leq k} |M_i| > \eta \right) \leq 2 \exp \left\{ -\frac{\eta^2}{2 \text{Var}[M_k]} \right\},$$

which completes the proof.

Lemma 5. Let $C > 0$, then for any $0 < \eta < C$,

$$G(\eta) = \frac{1}{\eta} \exp \left\{ -\frac{C}{\eta} \right\},$$

is increasing.

Proof. The derivative of $G(\eta)$ is

$$G'(\eta) = \frac{C - \eta}{\eta^3} \exp \left\{ -\frac{C}{\eta} \right\}.$$ 

Let $G'(\eta) > 0$ and we can obtain that $G(\eta)$ is increasing for $\eta \in (0, C)$, which completes the proof.

Proof of Lemma 2. Applying Lemma 3 to (48) and letting

$$C = \frac{\sqrt{ML}}{2} \sum_{i=1}^{n_T^2} \eta_{n+i}^2 + \sup_{1 \leq M \leq n_T^1} \eta_{n+m} - \eta_n,$$

yields

$$\sup_{t \in [t, t+n_T]} |\hat{x}(t) - \hat{x}(m)(t)| \leq C \left\{ \frac{\sqrt{ML}}{2} [\eta(n) - \eta(n + n_T)] + \sup_{1 \leq M \leq n_T^1} \eta_{n+m} - \eta_n \right\}.$$

where $C_\epsilon \triangleq e^{L(T^\alpha \eta)}$, and $\eta(n) \triangleq \sum_{i=1}^{n} \eta_i^2$. Letting $n = \tilde{n}(m)$ in (57), we have $n + n_T = \tilde{n}(m + 1)$ due to the definition of $T_{m+1} = \tilde{n}(m+1)$ in Step 2-(a) and

$$\sup_{t \in [\tilde{n}, \tilde{n}+n_T]} |\hat{x}(t) - \hat{x}(m)(t)| \leq C_\epsilon \left\{ \frac{\sqrt{ML}}{2} [\eta_{\tilde{n}(m)} - \eta_{\tilde{n}(m + 1)}] + \frac{\sqrt{M} a_{\tilde{n}(m) + 1}}{2} \right\}.$$

Suppose that the step size $\{\eta_n : n > n_0\}$ satisfies

$$C_\epsilon \frac{\sqrt{ML}}{2} [\eta_{\tilde{n}(m)} - \eta_{\tilde{n}(m + 1)}] + \frac{\sqrt{M} a_{\tilde{n}(m) + 1}}{2} < \frac{\delta}{2}.$$ 

for $m > 0$, and let $\delta_0 = \delta/(2C_\epsilon)$. Then, from (58) and (59), we get (60) on the top of the page, where step (a) is due to that when $\sup_{t \in [\tilde{n}, \tilde{n}+n_T]} |\hat{x}(t) - \hat{x}(m)(t)| > \delta$, we have

$$\sup_{\tilde{n}(m) \leq j \leq \tilde{n}(m + 1)} |\xi_j - \tilde{\xi}_{\tilde{n}(m)}| \geq C_\epsilon \left( \sup_{t \in [\tilde{n}, \tilde{n}+n_T]} |\hat{x}(t) - \hat{x}(m)(t)| \right)$$

$$- C_\epsilon \frac{\sqrt{ML}}{2} [\eta_{\tilde{n}(m)} - \eta_{\tilde{n}(m + 1)}] - \frac{\sqrt{M} a_{\tilde{n}(m) + 1}}{2}$$

$$> \frac{1}{C_\epsilon} \left( \sup_{t \in [\tilde{n}, \tilde{n}+n_T]} |\hat{x}(t) - \hat{x}(m)(t)| - \frac{\delta}{2} \right) > \frac{\delta}{2C_\epsilon} = \delta_0,$$

and step (b) is due to the independence of noise, i.e., $(\xi_j - \tilde{\xi}_{\tilde{n}(m)})$, $\tilde{n}(m) \leq j \leq \tilde{n}(m + 1)$ are independent of $\hat{x}_n$, $n_0 \leq n \leq \tilde{n}(m)$. 


The lower bound of the probability that the sequence \( \{ \hat{x}_n : n \geq n_0 \} \) remains in the invariant set \( \mathcal{I} \) is given by

\[
P ( \hat{x}_n \in \mathcal{I}, \forall n \geq n_0 ) \\
\geq P \left( \sup_{t \in I_m} | \hat{x}(t) - \hat{x}_{\bar{n}(m)}(t) | \leq \delta, \forall m \geq 0 \right) \\
\geq 1 - \sum_{m \geq 0} P \left( \sup_{t \in I_m} | \hat{x}(t) - \hat{x}_{\bar{n}(m)}(t) | > \delta \right) \\
\geq 1 - \sum_{m \geq 0} P \left( \sup_{t \in I_m} | \hat{x}(t) - \hat{x}_{\bar{n}(m)}(t) | \leq 0, \delta \leq 0 \leq i < m \right) \\
\geq 1 - \sum_{m \geq 0} P \left( \sup_{t \in I_m} | \hat{x}(t) - \hat{x}_{\bar{n}(m)}(t) | \leq 0, \delta \leq 0 \leq i < m \right),
\]

where step (a) is due to Lemma \[1\] step (b) is due to Lemma 4.2 in \[43\], and step (c) is due to \[60\].

With the increasing \( \sigma \)-fields \( \mathcal{G}_n : n \geq n_0 \) defined in Appendix \[A\] we have for \( n \geq n_0 \),

1. \( \xi_n = \sum_{n=m+1}^{n} a_m \tilde{x}_m \sim \mathcal{N}(0, \sum_{n=m+1}^{n} \frac{a_m^2}{2^m} \delta^m) \).
2. \( \xi_n \) is \( \mathcal{G}_n \)-measurable, i.e., \( \mathbb{E}[\xi_n | \mathcal{G}_n] = \xi_n \).
3. \( \mathbb{E}[\xi_n^2] = \sum_{n=M+1}^{\infty} \frac{a_m^2}{2^m} < \infty \).
4. \( \mathbb{E}[\xi_n | \mathcal{G}_n] = \xi_n \) for all \( n \leq m < n \).

Therefore, \( \xi_n \) is a Gaussian martingale with respect to \( \mathcal{G}_n \). Letting \( \eta = \delta_0 \), \( M_1 = \xi_{\bar{n}(m+1)} - \xi_{\bar{n}(m)} \) and \( k = \bar{n}(m+1) - \bar{n}(m) \) in Lemma \[4\] then we can obtain

\[
P \left( \sup_{\bar{n}(m) \leq j \leq \bar{n}(m+1)} | \xi_j - \xi_{\bar{n}(m)} | > \delta_0 \right) \\
\leq 2 \exp \left\{ -\frac{\delta_0^2}{2 \text{Var} [\xi_{\bar{n}(m+1)} - \xi_{\bar{n}(m)}]} \right\} \\
= 2 \exp \left\{ -\frac{\rho^2 \delta_0^2}{b(\bar{n}(m)) - b(\bar{n}(m+1))} \right\}. \tag{62}
\]

From \[61\] and \[62\], we have

\[
P ( \hat{x}_n \in \mathcal{I}, \forall n \geq n_0 ) \\
\geq P \left( \sup_{t \in I_m} | \hat{x}(t) - \hat{x}_{\bar{n}(m)}(t) | \leq \delta, \forall m \geq 0 \right) \\
\geq 1 - 2 \sum_{m \geq 0} \exp \left\{ -\frac{\rho^2 \delta_0^2}{b(\bar{n}(m)) - b(\bar{n}(m+1))} \right\}.
\]

Use Lemma \[5\] and assume that the step-size \( a_n \) satisfies

\[ b(n_0) = \sum_{i=M}^{\infty} a_i^2 \leq C = \rho_0^2. \tag{64}\]

Then for \( b(\bar{n}(m)) - b(\bar{n}(m+1)) < b(\bar{n}(m)) \leq b(n_0) \), we can obtain

\[
\exp \left[ -\frac{\rho^2 \delta_0^2}{b(\bar{n}(m)) - b(\bar{n}(m+1))} \right] \leq \exp \left[ -\frac{\rho^2 \delta_0^2}{b(n_0)} \right],
\]

based on the increasing property. Hence, we have

\[
\exp \left[ -\frac{\rho_0^2}{b(\bar{n}(m)) - b(\bar{n}(m+1))} \right] = \left[ b(\bar{n}(m)) - b(\bar{n}(m+1)) \right] \cdot \exp \left[ -\frac{\rho_0^2}{b(n_0)} \right],
\]

and

\[
\leq \sum_{m \geq 0} \left[ b(\bar{n}(m)) - b(\bar{n}(m+1)) \right] \cdot \exp \left[ -\frac{\rho_0^2}{b(n_0)} \right] = b(n_0) \cdot \exp \left[ -\frac{\rho_0^2}{b(n_0)} \right] = \exp \left[ -\frac{\rho_0^2}{b(n_0)} \right]. \tag{65}
\]

As \( \delta_0 = \delta/(2C_c) \), \( C_c = e^{L+\alpha n_0} \), \( b(n_0) = \sum_{i>n_0} a_i^2 \), and \( a_n \) is given by \[19\], we can obtain

\[
\frac{\rho_0^2}{b(n_0)} = \frac{\delta_0^2}{4e^{2L(T+\alpha n_0)} \sum_{i \geq 1} \frac{1}{(1+\alpha N_0)^i}} \cdot \frac{\rho}{\alpha^2}. \tag{66}
\]

In \[66\], \( 0 < \delta < \inf_{\mathcal{B} \in \mathcal{B}} \mathcal{B} \), \[59\] and \[64\] should be satisfied, where a sufficiently large \( N_0 \geq 0 \) can make both \[59\] and \[64\] true.

To ensure that \( \hat{x}_{n_0} + a_{n_0+1} f(\hat{x}_{n_0}, x) \) does not exceed the mainlobe \( \mathcal{B} \), i.e., the first step-size \( a_{n+1} \) satisfies

\[ | \hat{x}_{n_0} + a_{n_0+1} f(\hat{x}_{n_0}, x) - x | < \frac{\lambda}{M d} \]

we can obtain the maximum \( \alpha \) as follows

\[ \alpha_{\text{max}} = \frac{(N_0 + 1)((x - \hat{x}_{n_0}) + \frac{\lambda}{M d})}{| f(\hat{x}_{n_0}, x) |}. \]

Hence, from \[66\], we have

\[
\frac{\rho_0^2}{b(n_0)} \geq C (\hat{x}_{n_0}), \tag{67}
\]

where

\[ C (\hat{x}_{n_0}) \leq \frac{\delta_0^2}{4e^{2L(T+\alpha n_0)} \sum_{i \geq 1} \frac{1}{(1+\alpha N_0)^i}} \cdot \frac{\rho}{\alpha^2}. \]

Combining \[63\], \[65\] and \[67\], yields

\[
P ( \hat{x}_n \in \mathcal{I}, \forall n \geq n_0 ) \\
\geq P \left( \sup_{t \in I_m} | \hat{x}(t) - \hat{x}_{\bar{n}(m)}(t) | \leq \delta, \forall m \geq 0 \right) \\
\geq 1 - 2e^{-C (\hat{x}_{n_0}) - \frac{\lambda}{M d}}.
\]

which completes the proof.

**Step 3**: By applying Lemma \[3\] and Corollary 2.5 in \[43\], we can obtain

\[
P ( \hat{x}_n \to x | \hat{x}_n \to \mathcal{B} ) > P ( \hat{x}_n \in \mathcal{I}, \forall n \geq n_0 ) \tag{68}
\]

\[
> 1 - 2e^{-C (\hat{x}_{n_0}) - \frac{\lambda}{M d}},
\]

which completes the proof of Theorem \[2\].
APPENDIX C

PROOF OF THEOREM [3]

When the step-size $a_n$ is given by (19) with any $\alpha > 0$ and $N_0 \geq 0$, Theorem 6.6.1 [20] has proposed the sufficient conditions to prove the asymptotic normality of $\sqrt{n - n_0 + N_0} (\hat{x}_n - x)$, i.e., $\sqrt{n - n_0 + N_0} (\hat{x}_n - x) \xrightarrow{d} N(0, \Sigma)$, Under the condition that $\hat{x}_n \to x$, we will prove that our algorithm satisfies these sufficient conditions and obtain the variance $\Sigma$:

1) The estimate $\hat{x}_n$ should be within $[-1, 1]$.

The projection operator in (15) ensures that $\hat{x}_n \in [-1, 1]$.

1) Equation (29) should satisfy: (i) there exist an increasing sequence of $\sigma$-fields $\{\mathcal{F}_n : n \geq n_0\}$ such that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for $m < n$, and (ii) the random noise $\hat{z}_n$ is $\mathcal{F}_n$-measurable and independent of $\mathcal{F}_{n-1}$.

As defined in Appendix A, there exist an increasing sequence of $\sigma$-fields $\{G_n : n \geq n_0\}$, such that $\hat{z}_n$ is adapted to $G_n$, i.e., $\mathbb{E} [\hat{z}_n | G_n] = \hat{z}_n$, and is independent of $G_{n-1}$, i.e., $\mathbb{E} [\hat{z}_n | G_{n-1}] = \mathbb{E} [\hat{z}_n] = 0$.

2) $\hat{x}_n$ should converge to $x$ almost surely as $n \to \infty$.

Since $\hat{x}_n \to x$ is assumed, we have that $\hat{x}_n$ converges to $x$ almost surely as $n \to \infty$.

3) The stable condition:

From (17), $f(v, x)$ can be rewritten as follows:

$$f(v, x) = -\frac{1}{\sqrt{M}} \sin \left[ \frac{(M-1)\pi d}{\lambda} (v - x) \right] \sin \left[ \frac{\pi d}{\lambda} (v - x) \right]$$

$$= -\frac{1}{\sqrt{M}} c_1 (v - x) + o(v - x),$$

where $c_1$ is given by

$$c_1 = \left. \left( \sin \left[ \frac{(M-1)\pi d}{\lambda} (v - x) \right] \sin \left[ \frac{\pi d}{\lambda} (v - x) \right] \right) \right|_{v=x}$$

$$= \frac{M(M-1)\pi d}{\lambda}.$$

Then we get the stable condition that

$$A = -\frac{c_1}{\sqrt{M}} + \frac{1}{2} \alpha = \frac{\sqrt{M(M-1)} \pi d \alpha}{\lambda} + \frac{1}{2} < 0,$$

which results in $\alpha > \frac{\lambda}{2 \sqrt{M(M-1)} \pi d}$.

4) The constraints for the random noise:

$$\mathbb{E} \left[ \hat{z}_n^2 \right] = \frac{1}{2} \rho < \infty,$$

and

$$\lim_{n \to \infty} \sup_{n \geq 1} \int_{|\hat{z}_n| > V} |\hat{z}_n|^2 p(\hat{z}_n) d\hat{z}_n = 0.$$

Hence, we have

$$\sqrt{n - n_0 + N_0} (\hat{x}_n - x) \xrightarrow{d} N(0, \Sigma),$$

where

$$\Sigma = \alpha^2 \mathbb{E} \left[ \hat{z}_n^2 \right] = \int_{0}^{\infty} e^{2Av} dv$$

$$= \frac{\alpha^2}{2 \rho \left( 2 \sqrt{M(M-1)} \pi d \alpha \right) - 1}.$$

(69)

Due to that $\lim_{n \to \infty} \sqrt{n - n_0 + N_0}/n = 1$, we have

$$\sqrt{n} (\hat{x}_n - x) \to \mathcal{N}(0, \sqrt{\Sigma}).$$

As $n \to \infty$, By adapting $\alpha$ in (69), we can obtain different $\Sigma$, which achieves the minimum value $\Sigma_{\text{min}} = I_{\text{max}}^{-1}$, i.e., the minimum CRLB in [10], when $\alpha = \frac{\lambda}{\sqrt{M(M-1)} \pi d}$.

By assuming $\alpha = \frac{\lambda}{\sqrt{M(M-1)} \pi d}$, we conclude that

$$\lim_{n \to \infty} n \mathbb{E} \left[ (\hat{x}_n - x)^2 | \hat{x}_n \to x \right] = I_{\text{max}}^{-1}.$$

APPENDIX D

PROOF OF COROLLARY [1]

Let $u(x) = [u_1(x), \ldots, u_N(x)]^T$ be a $N$-dimensional vector function, which is second-order differentiable. Similar to (7)-(10), its MSE is lower bounded by

$$\mathbb{E} \left[ \left\| u(x) - u(m) \right\|^2 \right] = \sum_{m=1}^{N} \mathbb{E} \left[ (u_m(x) - u_m(x))^2 \right]$$

$$\geq \sum_{m=1}^{N} \frac{1}{n I_{\text{max}, m}},$$

(70)

where $I'_{\text{max}, m}$ is given by

$$I'_{\text{max}, m} = \mathbb{E} \left[ -\frac{\partial^2 \log p(y_{m}\mid x, w_i)}{\partial x^2} \right] x, w_i = \frac{a(x)}{\sqrt{M}}.$$

According to Theorem 3 we have

$$\lim_{n \to \infty} n \mathbb{E} \left[ (\hat{x}_n - x)^2 | \hat{x}_n \to x \right] = I_{\text{max}}^{-1},$$

where $I_{\text{max}}$ is given by

$$I_{\text{max}} = -\frac{\partial^2 \log p(y_{m}\mid x, w_i)}{\partial x^2}$$

Since

$$\frac{\partial^2 \log p(y_{m}\mid x, w_i)}{\partial x^2}$$

can be rewritten as

$$\frac{\partial^2 \log p(y_{m}\mid x, w_i)}{\partial x^2} = \frac{\partial^2 \log p(y_{m}\mid x, w_i)}{\partial x^2} \left[ u_m'(x) \right]^2$$

$$+ \frac{\partial^2 \log p(y_{m}\mid x, w_i)}{\partial u_m(x) \partial u_m(x)} u_m''(x),$$

and

$$\mathbb{E} \left[ \frac{\partial \log p(y_{m}\mid x, w_i)}{\partial u_m(x)} u_m''(x) \right] x, w_i = \frac{a(x)}{\sqrt{M}}$$

$$= \int_{-\infty}^{\infty} \frac{\partial \log p(y_{m}\mid x, w_i)}{\partial u_m(x)} u_m''(x) p(y_{m}\mid x, w_i) dy_i$$

$$= u_m''(x) \left. \frac{\partial \int_{-\infty}^{\infty} p(y_{m}\mid x, w_i) dy_i}{\partial u_m(x)} \right|_{x, w_i = a(x)}$$

$$= 0,$$

we get

$$I'_{\text{max}, m} = \frac{I_{\text{max}}}{\left| u_m'(x) \right|^2}.$$
which results in
\[
\lim_{n \to \infty} n \mathbb{E} \left[ |u_m(\hat{x}_n) - u_m(x)|^2 \mid \hat{x}_n \to x \right] = |u'_m(x)|^2 I^{-1}_{\text{max}}.
\]
Then, based on (70), we conclude that
\[
\lim_{n \to \infty} n \mathbb{E} \left[ \|\mathbf{u}(\hat{x}_n) - \mathbf{u}(x)\|_2^2 \mid \hat{x}_n \to x \right] = \|\mathbf{u}'(x)\|_2^2 I^{-1}_{\text{max}}.
\]