Distributed Beamforming with Feedback: Convergence Analysis

Che Lin, Venugopal V. Veeravalli, and Sean Meyn
University of Illinois, Urbana-Champaign, Urbana, IL 61801
Email: {clin, vvv, meyn}@uiuc.edu

Abstract

The focus of this work is on the analysis of transmit beamforming schemes with a low-rate feedback link in wireless sensor/relay networks, where nodes in the network need to implement beamforming in a distributed manner. Specifically, the problem of distributed phase alignment is considered, where neither the transmitters nor the receiver has perfect channel state information, but there is a low-rate feedback link from the receiver to the transmitters. In this setting, a framework for systematically analyzing the performance of a general set of distributed beamforming schemes is proposed. To illustrate the advantage of this framework, a simple adaptive distributed beamforming scheme that was recently proposed by Mudambai et al. is studied. Two important properties for the received signal magnitude function are derived. Using these properties and the systematic framework, it is shown that the adaptive distributed beamforming scheme converges both in probability and in mean. Furthermore, it is established that the time required for the adaptive scheme to converge in mean scales linearly with respect to the number of sensor/relay nodes.

Index Terms

Array signal processing, convergence of numerical methods, detectors, distributed algorithms, feedback communication, networks, relays.

I. Introduction

The problem of distributed beamforming arises quite naturally in wireless sensor/relay networks. In a sensor network, sensors make estimates of a common phenomenon and reach a consensus using a local message passing algorithm. In a relay network, a source node intends to communicate with the destination node by passing the message to all relay nodes. In both settings, sensor/relay nodes then serve as distributed transmitters and seek to convey a common message to the intended receiver. To preserve energy in this stage, transmit beamforming has emerged as a promising scheme due to its potential array gain and low-complexity. However, perfect channel state information (CSI) at the transmitter is required by conventional transmit beamforming schemes to generate beamforming coefficients and achieve phase alignment at the receiver end. This requirement and the distributed

This research was supported in part by the NSF award #CCF 0431088, through the University of Illinois, and by a Vodafone Foundation Graduate Fellowship.

nature of wireless sensor/relay networks make it difficult to implement transmit beamforming schemes in practice. With these issues in mind, one may be tempted to design transmit beamforming schemes that estimate the channel coefficients directly and provide the estimate to the transmitter end. However, the overhead required for channel estimation becomes prohibitively large for a densely populated sensor/relay network, since all channel gains between each transmitter and the receiver need to be measured. Furthermore, sensor/relay nodes may not even be able to estimate the channel due to hardware limitations. These constraints limit us to the setting of non-coherent communications, where both the transmitters and receiver have no knowledge of channel realizations.

Within the class of non-coherent communication strategies, training-based schemes are attractive due to their low-complexity and near-optimal performance. In general, these schemes are separated into the training stage and the communication stage. In the training stage[1], the phases for all transmitters are aligned such that the received signals from all transmitters add coherently at the receiver. The message is sent in the communication stage, after phase alignment have been achieved at the receiver to provide full array gain. Since the communication stage is straightforward once we have achieved phase alignment, the focus of this work is on the design of energy efficient training schemes. Note that in the training stage, it is nearly impossible to achieve phase alignment without at least partial channel knowledge at the transmitter. As a consequence, there has been increasing interest in the design of efficient schemes that achieve phase alignment in the presence of a low-rate feedback link [1], [2]. The low-rate feedback link conveys some form of partial channel knowledge from the receiver to the transmitters. It is hence of interest to investigate the impact of this feedback link on the analysis and design of efficient training schemes. Specifically, our goal is to provide a framework for systematically analyzing the performance of a general set of training schemes with feedback.

To illustrate the advantages of our framework, we focus on the analysis of a recently proposed training scheme for distributed beamforming [1]. The proposed scheme is a simple adaptive algorithm using one bit of feedback information, and is attractive in practice since it is simple to implement. Naturally, one would expect a tradeoff in energy consumption due to slow convergence or other issues. Surprisingly, this scheme converges rapidly and utilizes energy efficiently. Furthermore, the scheme adjusts its phases for all sensors simultaneously for each time slot to achieve phase alignment. This reduces the overhead significantly compared with direct channel estimation between each sensor/relay node and the destination node. In fact, the convergence time of the scheme scales linearly with the number of sensor/relay nodes. Although the scheme has many desirable features, the fundamental reasons behind the effectiveness of the scheme are still unclear. In [1], analysis on the convergence and linear scalability of the scheme has been provided through approximations based on the Central Limit Theorem. A discrete version of the problem has been solved in [2] by considering a simplified model with binary channel and signaling. In this work, we seek to provide a more comprehensive analysis on the fast convergence and linear scalability of the scheme, two of its most desirable properties, based on the framework we establish.

We organize the paper as follows. In Section II, we introduce the system model and the received signal magnitude

[1]Note that the training stage described here is not used to estimate the channel as explained above.

function, which is used as our metric to measure the beamforming array gain throughout the paper. In Section III, we describe the adaptive distributed beamforming scheme proposed in [1]. Furthermore, we establish an equivalence between the distributed beamforming scheme and a local random search algorithm. Based on this observation, we obtain a general framework for systematically analyzing the adaptive distributed beamforming scheme that provides insights into a necessary condition for the convergence of the scheme. These insights lead us to investigate the properties of the received signal magnitude function in Section IV. As a consequence of these properties, we provide intuitive arguments on the fast convergence of the equivalent local random search algorithm. We further use these properties to prove the convergence of the local random search algorithm in probability and in mean. Simulations are provided to illustrate the conclusions of our analysis. In Section V, we study the scaling law for the algorithm and show that the time required for the algorithm to converge in mean scales linearly with the number of sensor/relay nodes. We also provide numerical results that validate our analysis. Finally, we conclude the paper in Section VI and suggest directions for future research.

II. System Setup

We consider the problem of distributed beamforming, where n_s transmitters seek to beamform a common message to one receiver in a distributed manner. We assume that each transmitter and receiver is equipped with one antenna, and that the distributed multi-antenna (MISO) channel experiences frequency-flat, slow fading. That is, the channel coefficients $\{h_i\}_{i=1}^{n_s}$ vary randomly but remain fixed throughout the transmission. The discrete-time, complex baseband model is given by

$$y[t] = \sum_{i=1}^{n_s} h_i g_i[t] s[t] + w[t] = \sum_{i=1}^{n_s} a_i b_i[t] e^{(\phi_i + \psi_i[t])} s[t] + w[t]$$

where $s[t] \in \mathbb{C}$ is the transmitted common message, $y[t] \in \mathbb{C}$ is the received signal, and $w[t] \sim \mathcal{CN}(0, \sigma^2)$ corresponds to the additive white Gaussian noise. For transmitter i, we denote the channel fading gains by $h_i = a_i e^{\phi_i} \in \mathbb{C}$ and beamforming coefficients by $g_i[t] = b_i[t] e^{\psi_i[t]} \in \mathbb{C}$. Note that $a_i \geq 0$, $b_i[t] \geq 0$, and $\phi_i \in [0, 2\pi]$, $\psi_i[t] \in [0, 2\pi]$ for all i and t since they are the corresponding magnitudes and phases of h_i and g_i, respectively. Moreover, there is no time dependency on a_i and ϕ_i due to the slow varying nature of the channel. We assume an average power constraint on $s[t]$ given by $E[|s[t]|^2] \leq P$, and that each transmitter utilizes the same amount of energy for each transmission, i.e., $b_i[t] = 1$ for all i and t.

For channel state information (CSI), we assume a non-coherent communication where the realization of the channel is unknown at both the transmitters and receiver. There is, however, an error-free, zero-delay feedback link from the receiver to all transmitters conveying one bit of information in each time step.

Among non-coherent schemes that have been proposed, we assume a training-based scheme that separates the transmission into a training stage and communication stage. In the training stage, the total phase $\phi_i + \psi_i[t]$ for each transmitter is aligned such that received signals from all transmitters add coherently at the receiver. The message is sent in the communication stage, where a full array gain on the order of n_s for the received SNR is achieved at the receiver. In this work, we focus on the analysis of an efficient training scheme [1] that achieves phase alignment efficiently in the training stage. The goal of the scheme is to consume the least amount of energy while providing the full array gain from distributed beamforming. Since we assume that same amount of energy is used for each transmission for each transmitter, the training scheme that converges with the least amount of time consumes the least amount of energy.

In the training stage, the message $s[t]$ is fixed and known to the receiver. Without loss of generality, we assume $s[t] = \sqrt{P}$. The received signal magnitude can be expressed as

$$\mathrm{Mag}(\theta_1[t], \cdots, \theta_{n_s}[t]) = \sqrt{P}\left|\sum_{i=1}^{n_s} a_i e^{j\theta_i[t]}\right| := \sqrt{P}\left|\sum_{i=1}^{n_s} a_i e^{j(\phi_i+\psi_i[t])}\right| \tag{1}$$

where $\theta_i[t] = \phi_i + \psi_i[t]$ is the total phase for sensor i. Note that we use $\mathrm{Mag}(\cdot)$ as the metric to measure array gain provided by a beamforming scheme. This can be justified since in our setting, the array gain in received SNR directly translates into an array gain in the received magnitude function.[2]

The details of the training scheme are introduced in the following section. To remove redundancy, we term the training scheme as an *adaptive distributed beamforming scheme* with the understanding that this scheme is applied at the training stage and utilizes a one-bit perfect feedback link.

III. Adaptive Distributed Beamforming Scheme

In this section, we introduce the adaptive distributed beamforming scheme proposed in [1] and establish its equivalence with a local random search algorithm. Random search algorithms are well studied in the literature [3], [4], [5] as methods to maximize an unknown function via random sampling. This equivalence allows for a systematic study of the convergence of the distributed beamforming scheme.

A. Description of the Distributed Algorithm

Let $\hat{\theta}_i[t]$ denote the total phase that transmitter i uses to transmit at time t, and $\theta_i[t]$ be the total phase that transmitter i keeps at time t after observing the feedback bit from the receiver. We now describe the adaptive training scheme as follows.

Adaptive Distributed Beamforming Scheme:

- *Step zero: Initialize*. Referring to (1) and noting that the ith transmitter controls its beamforming phase $\psi_i[t]$, we can initialize the algorithm by setting $\psi_i[0] = 0$, and hence $\theta_i[0] = \phi_i$ for transmitter i.
- *Step one: Update and Transmit*. In this step, we randomly perturb the total phase $\theta_i[t-1]$ for transmitter i that is kept in the previous time slot. More precisely, each transmitter updates its total phase by adding the random perturbation $\delta_i[t]$ to its beamforming phase kept in the previous time slot. We assume that $\{\delta_i[t]\}_{i=1}^{n_s}$ are i.i.d. uniform random variables in $[-\delta_0, \delta_0]$ across time and transmitters, where δ_0 is a constant parameter. Let $\boldsymbol{\theta}[t] = [\theta_1[t], \cdots, \theta_{n_s}[t]]^T$, $\hat{\boldsymbol{\theta}}[t] = [\hat{\theta}_1[t], \cdots, \hat{\theta}_{n_s}[t]]^T$, and $\boldsymbol{\delta}[t] = [\delta_1[t], \cdots, \delta_{n_s}[t]]^T$. We have the following update equation:

$$\hat{\boldsymbol{\theta}}[t] = \boldsymbol{\theta}[t-1] + \boldsymbol{\delta}[t] \tag{2}$$

[2] Note that this may not be true in general if there is error in acquiring the common message.

The transmitters then use $\hat{\boldsymbol{\theta}}[t]$ to transmit training symbols to the receiver.

- *Step two: Compare and Select*. After receiving the training symbols, the receiver measures the received signal magnitude[3] and compares it with the signal magnitude received in the previous time slot. If the newly received signal magnitude is larger, the receiver feeds back a "*keep*" beacon to transmitters. Otherwise, a "*discard*" beacon is sent to transmitters. Note that the beacon is a broadcast from the receiver to all transmitters. Clearly, this feedback scheme only requires one bit of feedback information per time step. When a "*keep*" is received at the transmitters, each transmitter selects and keeps its newly updated total phase. Otherwise, the old phase is selected and the new phase discarded. This selection process is determined by whether the random perturbation increases or decreases the array gain for the adaptive distributed beamforming scheme. Specifically, the evolutions of $\boldsymbol{\theta}[t]$ and $\hat{\boldsymbol{\theta}}[t]$ are given by

$$\boldsymbol{\theta}[t] = \begin{cases} \boldsymbol{\theta}[t-1], & \text{if} \quad \mathrm{Mag}(\hat{\boldsymbol{\theta}}[t]) \leq \mathrm{Mag}(\boldsymbol{\theta}[t-1]) \\ \hat{\boldsymbol{\theta}}[t], & \text{if} \quad \mathrm{Mag}(\hat{\boldsymbol{\theta}}[t]) > \mathrm{Mag}(\boldsymbol{\theta}[t-1]) \end{cases} \tag{3}$$

where $\boldsymbol{\theta}[t]$ and $\hat{\boldsymbol{\theta}}[t]$ are in Θ, and $\Theta = [0, 2\pi]^{n_s}$.

B. Equivalence with Random Search Algorithms

The simple adaptive algorithm introduced in the previous section is equivalent to a local random search algorithm, where an unknown function is maximized via local random sampling. To establish this equivalence, we first consider the following problem:

Problem 1: Given a function $f : \Theta \to \mathbb{R}, \Theta \subseteq \mathbb{R}^n$ with an unknown structure, where only samples of $f(\boldsymbol{\theta})$ are available for arbitrary $\boldsymbol{\theta} \in \Theta$. Find the global maxima of f with the fewest function evaluations.

Due to the non-coherent nature of our communication system, the channel realization is not known at both the transmitter and receiver ends. As a consequence, the function $\mathrm{Mag}(\cdot)$ is not known at both the transmitters and the receiver. An estimate of the received signal magnitude function at the receiver is essentially a sample of the unknown function $\mathrm{Mag}(\cdot)$. Thus, from the receiver point-of-view, the problem of phase alignment for distributed beamforming can be considered under the setting of *Problem 1*, a global maximization problem. This allows us to study distributed beamforming schemes in a more systematic manner.

To solve the maximization in *Problem 1*, one may be tempted to use a gradient-based algorithm. Since it is possible for $\mathrm{Mag}(\cdot)$ to possess local maxima, conventional gradient-ascent methods will fail in general. Moreover, acquiring the gradient of the function f may not even be possible. Hence, random search techniques [3], [4], [5] are more appropriate in this setting. Most of these algorithms can be considered under the framework of a conceptual algorithm introduced in [3].

Conceptual Algorithm:

- *Step zero*: Initialize the algorithm by choosing $\boldsymbol{\theta}[0] \in \Theta$.

[3] A good estimate of the received signal magnitude can be obtained directly when the noise is small, or by averaging over several time slots when the noise is not negligible.

- *Step one*: Generate a random perturbation $\boldsymbol{\delta}[t]$ from the sample space $(\mathbb{R}^n, \mathcal{B}, \mu_t)$, where $\mathcal{B}$ is a Borel set on $\mathbb{R}^n$ and μ_t is a probability measure that could be time-varying.
- *Step two*: Update the search point by $\boldsymbol{\theta}[t] = D(\boldsymbol{\theta}[t-1], \boldsymbol{\delta}[t])$, where the map D satisfies the condition $f(D(\boldsymbol{\theta}[t-1], \boldsymbol{\delta}[t])) \geq f(\boldsymbol{\theta}[t-1])$.

In this conceptual algorithm, we require only function evaluations and have control over the probability measure μ_t, which is used to sample the function. The adaptive distributed beamforming algorithm can hence be regarded as a special case of this conceptual algorithm by setting

$$n = n_s \tag{4}$$

$$\mu_t = \mu \tag{5}$$

$$D(\boldsymbol{\theta}[t-1], \boldsymbol{\delta}[t]) = \boldsymbol{\theta}[t-1] + 1_{\{\mathrm{Mag}(\boldsymbol{\theta}[t-1]+\boldsymbol{\delta}[t]) > \mathrm{Mag}(\boldsymbol{\theta}[t-1])\}}\boldsymbol{\delta}[t] \tag{6}$$

where μ is uniform on $[-\delta_0, \delta_0]^{n_s}$. Note that (6) is the same as the evolution described by (2) and (3).

Since the probability measure μ_t is non-zero only within a local hypercube with sides $2\delta_0$ centered around $\boldsymbol{\theta}[t-1]$, the adaptive distributed beamforming scheme can be regarded as a local random search algorithm. It is clear that we can use this framework to study more general adaptive distributed beamforming schemes. For example, the sampling probability measure may be time-varying and with a support that spans the entire space Θ. We can also study distributed beamforming schemes with more than one bit of feedback information. It is also interesting to note the connection between this local random search algorithm and simulated annealing [6]. Simulated annealing is a generic probabilistic algorithm that approximates the global optimal solution of a given function in a large search space. The algorithm uses a time-varying parameter T to control the acceptance probability, i.e., the probability that the current state of the algorithm transitions to a new state. If we let $T \to 0$ and assume that the current state is only allowed to move to neighboring states, the simulated annealing procedure reduces to the local random search algorithm. In this work, we focus on the analysis of the local random search algorithm to illustrate the advantages of our framework.

A local random search algorithm does not necessarily converge in general. For example, if the unknown function possesses local maxima, the sequence $\{\boldsymbol{\theta}[t]\}_{t=0}^{\infty}$ is likely to be trapped in local maxima if δ_0 is not large enough. This phenomenon is illustrated in Fig. 1, where the local perturbation is too small for the sequence to leave the neighborhood of a local maximum. Thus, a necessary condition for the convergence of local random search algorithms for arbitrary δ_0 is that there is no local maximum point for $\mathrm{Mag}(\cdot)$. Since it is still unclear whether $\mathrm{Mag}(\cdot)$ satisfies this necessary condition, two questions arise naturally: $a)$ Does the equivalent local random search algorithm converge? $b)$ If it does, is there a fundamental reason behind the convergence? In the following section, we investigate properties of the function $\mathrm{Mag}(\cdot)$ towards the goal of addressing these questions.

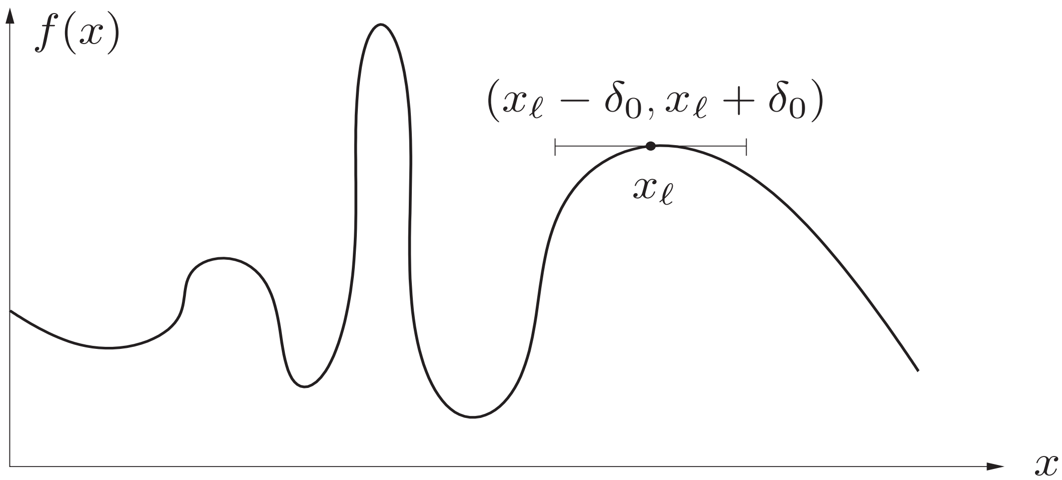

Fig. 1. An example of a local random search algorithm that fails to converge to the global maximum.

IV. Convergence of the Distributed Beamforming Scheme

A. Properties of Received Signal Magnitude Function

The properties of the received signal magnitude function $\mathrm{Mag}(\cdot)$ do not depend on the time evolution of its arguments. We hence omit the time dependence of $\theta[t]$ in this section. The following proposition states the first property of $\mathrm{Mag}(\cdot)$.

Proposition 1: For the received signal magnitude function $\mathrm{Mag}(\cdot)$ defined in (1), all local maxima are global maxima.

Proof: To facilitate analysis, we introduce a change of variables

$$\mathbf{x}_i := \begin{bmatrix} x_i^R \\ x_i^I \end{bmatrix} = \begin{bmatrix} \cos\theta_i \\ \sin\theta_i \end{bmatrix}$$

Eqn. (1) can be rewritten as

$$\mathrm{Mag}(\mathbf{x}_1, \cdots, \mathbf{x}_{ns}) = \sqrt{P} \left\| \sum_{i=1}^{n_s} a_i \mathbf{x}_i \right\|$$

where $\|\mathbf{x}_i\|^2 = 1$ for all $i = 1, \cdots, n_s$. The maximization of $\mathrm{Mag}(\cdot)$ can be rewritten as:

$$\max_{\|\mathbf{x}_i\|^2=1, i=1,\cdots,n_s} \left\| \sum_{i=1}^{n_s} a_i \mathbf{x}_i \right\|^2 \tag{7}$$

In the following, we will show that all local maxima of this objective function correspond to complete phase alignment for all transmitters. That is, all local maximum points are global maximum points.

By relaxing the equality constraints to inequality constraints, the optimization problem in (7) is equivalent to

$$\max_{\|\mathbf{x}_i\|^2\leq 1, i=1,\cdots,n_s} \left\| \sum_{i=1}^{n_s} a_i \mathbf{x}_i \right\|^2 \tag{8}$$

This equivalence can be seen as follows: if $\mathbf{x}^*$ is a local maximum with an inactive constraint $\|\mathbf{x}_k^*\|^2 < 1$, by fixing all other variables $\{\mathbf{x}_j^*\}_{j\neq k}$, we obtain

$$\left\|\sum_{i=1}^{n_s} a_i \mathbf{x}_i^*\right\|^2 = \|a_k \mathbf{x}_k^* + \mathbf{c}\|^2 = (a_k x_k^{R*} + c^R)^2 + (a_k x_k^{I*} + c^I)^2$$

where $\mathbf{c} = [c^R c^I]^T$ is a constant vector depending on $\{\mathbf{x}_j^*\}_{j\neq k}$. Obviously, the above function can be improved by appropriately perturbing $\|\mathbf{x}_k^*\|$ according to the signs of c^R and c^I. This contradicts with the fact that $\mathbf{x}^*$ is a maximum. Thus, all constraints are active if $\mathbf{x}^*$ is a maximum point. This shows that the optimization problems (7) and (8) are equivalent.

Focusing on the optimization problem with relaxed constraints, the Lagrangian of (8) reads

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \left\|\sum_{i=1}^{n_s} a_i \mathbf{x}_i\right\|^2 + \sum_{i=1}^{n_s} \lambda_i (\|\mathbf{x}_i\|^2 - 1) =: \|\mathbf{w}\|^2 + \sum_{i=1}^{n_s} \lambda_i (\|\mathbf{x}_i\|^2 - 1)$$

where $\mathbf{x} = [\mathbf{x}_1^T, \cdots, \mathbf{x}_{n_s}^T]^T$, $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_{n_s}]^T$, and $\lambda_i \geq 0$ for all $i = 1, \cdots, n_s$. By the Lagrange Multiplier Theorem, all local maxima satisfy

$$\nabla_{\mathbf{x}_i} L(\mathbf{x}, \boldsymbol{\lambda}) = 2a_i \mathbf{w}^T + 2\lambda_i \mathbf{x}_i^T = \mathbf{0}^T, \forall i = 1, \cdots, n_s \quad (9)$$

$$\sum_{i=1}^{n_s} \lambda_i (\|\mathbf{x}_i\|^2 - 1) = 0 \quad (10)$$

$$\|\mathbf{x}_i\|^2 - 1 \leq 0, \forall i = 1, \cdots, n_s \quad (11)$$

Let $\mathbf{x}^*$ be a local maximum and $\boldsymbol{\lambda}^*$ be the corresponding Lagrange multiplier. If $\lambda_i^* = 0$, Eqn. (9) implies that $\mathbf{w} = \mathbf{0}$ since[4] $a_i > 0$. In this case, $\mathrm{Mag}(\mathbf{x}^*) = 0$ and this contradicts the fact that $\mathbf{x}^*$ is a local maximum, since we can always improve $\mathrm{Mag}(\cdot)$ by letting $\mathbf{x}_i^* = [\xi \ 0]^T$, $\xi \leq 1$, and $\mathbf{x}_j = \mathbf{0}$ for all $j \neq i$. This leads to $\lambda_i > 0$ for all i. We hence have

$$\mathbf{x}_i^* = -\frac{a_i}{\lambda_i^*}\mathbf{w} \quad (12)$$

$$\lambda_i^* = a_i \|\mathbf{w}\| \quad (13)$$

The optimal solutions described by (12) and (13), however, also satisfy

$$\mathrm{Mag}(\mathbf{x}^*) = \sqrt{P}\left\|\sum_{i=1}^{n_s} a_i \frac{\mathbf{w}}{\|\mathbf{w}\|}\right\| = \sqrt{P}\sum_{i=1}^{n_s} a_i$$

and hence are global maxima. This completes our proof.

■

Proposition 1 implies that the local random search algorithm cannot be trapped in a suboptimal local maximum since all local maxima are global maxima. Furthermore, it also suggests that the necessary condition for the convergence of random search algorithms is satisfied. While it is intuitively clear that the local random search algorithm should converge according to *Proposition* 1, it is to be noted that the condition is only necessary and

[4]Note that the case where $a_i = 0$ is not interesting since we can always reduce the dimension of the problem by ignoring $\mathbf{x}_i$

may not be sufficient. We will provide a rigorous proof of the convergence of the local random search algorithm later. Now, we explore an additional property of $\mathrm{Mag}(\cdot)$ that explains the efficiency of the algorithm.

Another interesting property of $\mathrm{Mag}(\cdot)$ is that it is invariant under a common phase shift to all transmitters. That is,

$$\mathrm{Mag}(\boldsymbol{\theta}+\theta_c\mathbf{e}) = \sqrt{P}\left|\sum_{i=1}^{n_s} a_i e^{j(\theta_i+\theta_c)}\right| = \sqrt{P}\left|e^{j\theta_c}\sum_{i=1}^{n_s} a_i e^{j\theta_i}\right| = \mathrm{Mag}(\boldsymbol{\theta})$$

where $\mathbf{e}$ is a $n_s \times 1$ vector and all of its elements are one, and θ_c is a common phase shift that can depend on $\{\theta_i\}_{i=1}^{n_s}$. One possible choice for the common phase shift is to let $\theta_c(\theta_1,\cdots,\theta_{n_s})$ be such that the imaginary part within the modulus function is canceled, i.e.,

$$\mathrm{Mag}(\boldsymbol{\theta}) = \mathrm{Mag}(\boldsymbol{\theta}+\theta_c(\theta_1,\cdots,\theta_{n_s})\mathbf{e}) = \sqrt{P}\sum_{i=1}^{n_s} a_i\cos\left(\theta_i+\theta_c(\theta_1,\cdots,\theta_{n_s})\right) = \sqrt{P}\sum_{i=1}^{n_s} a_i\cos\theta_i' = \mathrm{Mag}(\boldsymbol{\theta}')$$

where $\boldsymbol{\theta}' = [\theta_1',\cdots,\theta_{n_s}']^T$. Note that in the shifted $\boldsymbol{\theta}'$ domain, the global maxima occur only when $\theta_i' = 0$ or 2π for all i. The shift-invariant property results in multiple global maxima for the function $\mathrm{Mag}(\cdot)$. In fact, all global maxima form a one-dimensional "ridge" since if $\boldsymbol{\theta}^*$ is a global maximum, $\bar{\boldsymbol{\theta}}$ with $\bar{\theta}_i = \theta_i^* + \theta_c$ is also a global maximum. This property hints on the rapid convergence of the local random search algorithm since converging to any of these global maximum points is adequate.

We conclude this section by summarizing these two important properties of $\mathrm{Mag}(\cdot)$ as follows:

1) all local maxima are global maxima, and
2) a common shift to its arguments does not change its value.

B. *Proof of Convergence*

Intuitively, *Property* 1 guarantees the convergence of any local random search algorithm. We adopt convergence in probability as our notion of convergence. To define this, we introduce the ϵ-convergence region

$$R_\epsilon = \{\boldsymbol{\theta}\in\Theta : \mathrm{Mag}(\boldsymbol{\theta}) > \mathrm{Mag}\left(\boldsymbol{\theta}^*\right) - \epsilon\} \tag{14}$$

where $\boldsymbol{\theta}^*$ is the optimal total phase and satisfies $\mathrm{Mag}(\boldsymbol{\theta}^*) = \sqrt{P}\sum_{i=1}^{n_s} a_i$.

Definition 1: A sequence $\{\boldsymbol{\theta}[t]\}_{t=0}^{\infty}$ generated by a random search algorithm is said to be *convergent in probability* if, given $\epsilon > 0$,

$$\lim_{t\to\infty} \Pr\left[\boldsymbol{\theta}[t]\in R_\epsilon\right] = 1$$

In other words, $\mathrm{Mag}(\boldsymbol{\theta}[t])$ converges to $\mathrm{Mag}(\boldsymbol{\theta}^*)$ in probability.

For the proof of convergence, we derive a proposition stating that for any $\boldsymbol{\theta}$ outside of R_ϵ, there is a non-zero probability of improving $\mathrm{Mag}(\cdot)$ by applying a local perturbation to $\boldsymbol{\theta}$.

Proposition 2: For any given $\boldsymbol{\theta}\in\Theta\setminus R_\epsilon$ and $\delta_0 > 0$, there correspond $\gamma > 0$ and $0 < \eta \le 1$ such that

$$\Pr\left[\mathrm{Mag}(\boldsymbol{\theta}+\boldsymbol{\delta}) - \mathrm{Mag}(\boldsymbol{\theta}) \ge \gamma\right] \ge \eta$$

where $\boldsymbol{\delta}$ is a random vector with i.i.d. elements uniformly distributed over $[-\delta_0, \delta_0]$.

Proof: From *Proposition 1*, all local maxima are global maxima for the function $\mathrm{Mag}(\cdot)$. This implies that for all $\boldsymbol{\theta} \notin R_\epsilon$ and all $\delta_0 > 0$, there exists a point $\boldsymbol{\theta}_u \in S_{\boldsymbol{\theta}}$ and a constant $\gamma(\boldsymbol{\theta}) > 0$ such that

$$\mathrm{Mag}(\boldsymbol{\theta}_u) - \mathrm{Mag}(\boldsymbol{\theta}) \geq 2\gamma(\boldsymbol{\theta}) \tag{15}$$

where the set $S_{\boldsymbol{\theta}}$ is a hypercube of length $2\delta_0$ centered around $\boldsymbol{\theta}$ given by

$$S_{\boldsymbol{\theta}} = \{\boldsymbol{\omega} \in \Theta : \boldsymbol{\omega} = \boldsymbol{\theta} + \boldsymbol{\delta},\ \ \boldsymbol{\delta} \in [-\delta_0, \delta_0]^{n_s}\}$$

The continuity of $\mathrm{Mag}(\cdot)$ implies that there exists $\sigma(\boldsymbol{\theta}_u) > 0$ such that for all $\boldsymbol{\xi} \in T := \{\boldsymbol{\omega} \in \Theta : \|\boldsymbol{\omega}\| \leq \sigma(\boldsymbol{\theta}_u)\}$, we have

$$|\mathrm{Mag}(\boldsymbol{\theta}_u + \boldsymbol{\xi}) - \mathrm{Mag}(\boldsymbol{\theta}_u)| \leq \gamma(\boldsymbol{\theta}) \tag{16}$$

Combining (15) and (16), we arrive at a lower bound

$$\mathrm{Mag}(\boldsymbol{\theta}_u + \boldsymbol{\xi}) - \mathrm{Mag}(\boldsymbol{\theta}) = \mathrm{Mag}(\boldsymbol{\theta}_u + \boldsymbol{\xi}) - \mathrm{Mag}(\boldsymbol{\theta}_u) + \mathrm{Mag}(\boldsymbol{\theta}_u) - \mathrm{Mag}(\boldsymbol{\theta}) \geq -\gamma(\boldsymbol{\theta}) + 2\gamma(\boldsymbol{\theta}) = \gamma(\boldsymbol{\theta})$$

Referring to (4) for the definition of μ, the above lower bound leads to

$$\Pr\left[\mathrm{Mag}(\boldsymbol{\theta} + \boldsymbol{\delta}) - \mathrm{Mag}(\boldsymbol{\theta}) \geq \gamma(\boldsymbol{\theta})\right] \geq \mu\left(T\right) =: \eta(\boldsymbol{\theta})$$

Note that $\mu(T)$ is a function of $\boldsymbol{\theta}$, since $\boldsymbol{\theta}_u$ is a function of $\boldsymbol{\theta}$. We complete the proof of the proposition by letting

$$\begin{aligned} \gamma &= \inf_{\boldsymbol{\theta} \in \Theta \setminus R_\epsilon} \gamma(\boldsymbol{\theta}) \\ \eta &= \inf_{\boldsymbol{\theta} \in \Theta \setminus R_\epsilon} \eta(\boldsymbol{\theta}) \end{aligned}$$

■

Since before the sequence reaches the ϵ-convergence region, there is always a non-zero probability of improving $\mathrm{Mag}(\cdot)$ for each time step, the convergence of the sequence is to be expected. A simple deterministic analogue is the convergence of a monotonically non-decreasing function. The probabilistic nature of the algorithm complicates the proof. This will become clear in the proof of our next theorem.

Theorem 1: For the function $\mathrm{Mag}(\cdot)$ defined in (1), let $\{\boldsymbol{\theta}[t]\}_{t=1}^{\infty}$ be a sequence generated by the local random search algorithm described in Eqn. (4)-(6). Then the resulting sequence converges in probability, i.e., for any given $\epsilon > 0$,

$$\lim_{t \to \infty} \Pr\left[\boldsymbol{\theta}[t] \in R_\epsilon\right] = 1$$

Proof: By *Proposition* 2, we know that given any time t

$$\Pr\left[\{\mathrm{Mag}(\boldsymbol{\theta}[t-1] + \boldsymbol{\delta}[t]) - \mathrm{Mag}(\boldsymbol{\theta}[t-1]) \geq \gamma\}\ \text{ or }\ \{\boldsymbol{\theta} \in R_\epsilon\}\right] \geq \bar{\eta}$$

where $\bar{\eta} = \min\left\{\Pr[\boldsymbol{\theta} \in R_\epsilon], \eta\right\}$. Since Θ is compact and $\mathrm{Mag}(\cdot)$ is continuous, there always exists a positive integer p such that

$$p\gamma > \mathrm{Mag}(\boldsymbol{\theta}_1) - \mathrm{Mag}(\boldsymbol{\theta}_2),\ \ \forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$$

The probability that the sequence lies in R_ϵ after p time steps is hence lower bounded by

$$\Pr\left[\boldsymbol{\theta}[p] \in R_\epsilon\right] \geq \bar{\eta}^p$$

since $\{\boldsymbol{\delta}[t]\}_{t=0}^{\infty}$ are independent across time. This leads to $\Pr[\boldsymbol{\theta}[p] \notin R_\epsilon] \le 1 - \bar{\eta}^p$ and

$$\Pr[\boldsymbol{\theta}[pm] \in R_\epsilon] = 1 - \Pr[\boldsymbol{\theta}[pm] \notin R_\epsilon] \ge 1 - (1 - \bar{\eta}^p)^m, \; m = 1, 2, \cdots$$

The lower bound is still valid if we let the sequence progress ℓ time steps further, i.e.,

$$\Pr[\boldsymbol{\theta}[pm + \ell] \in R_\epsilon] \ge 1 - (1 - \bar{\eta}^p)^m, \; m = 1, 2, \cdots, \; \ell = 0, \cdots, p - 1$$

We complete the proof by letting $m \to \infty$.

■

Theorem 1 states that the local random search algorithm in (4)-(6) converges in probability, and hence also provides a proof of convergence for the adaptive distributed beamforming algorithm in (2)-(3). In fact, *Theorem 1* also implies the convergence of the sequence $\{\mathrm{Mag}(\boldsymbol{\theta}[t])\}_{t=0}^{\infty}$ in probability. Since the sequence is non-negative and monotonically non-decreasing, we conclude that $\{\mathrm{Mag}(\boldsymbol{\theta}[t])\}_{t=0}^{\infty}$ also converges in mean by the Monotone Convergence Theorem [7].

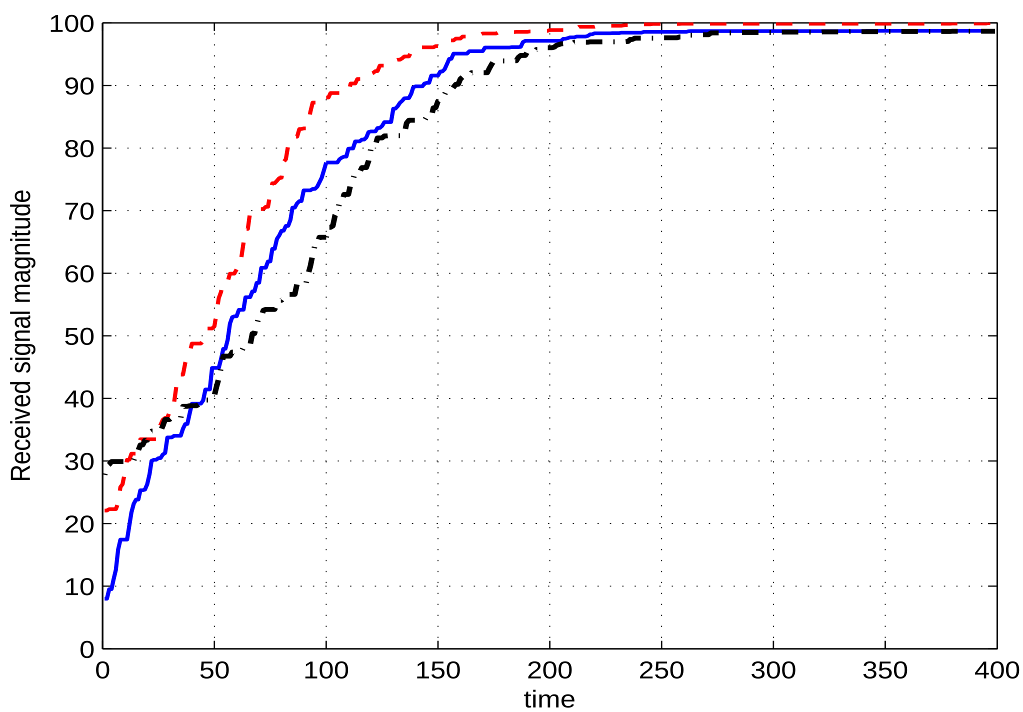

Fig. 2. Evolutions of sequences generated by the adaptive distributed beamforming scheme.

The evolution of the sequences generated by the local random search algorithm is illustrated in Fig. 2. Various initial points were generated randomly from a uniform distribution over Θ. Only three sample paths of the sequence are included in the figure since similar behaviors were observed for other sample paths. For each iteration, the random perturbation δ_i for the ith transmitter is a uniform random variable over $[-\delta_0, \delta_0]$, where $\delta_0 = \pi/30$. Note that we use the same channel coefficients to generate these sequences since the focus here is on the effect of different initial points. In particular, the channel coefficients are randomly generated from i.i.d. $\mathcal{CN}(0,1)$ in the beginning of the simulation, and remain fixed afterwards.

From the figure, we observe the rapid convergence of the local random search algorithm, irrespective of where it is initialized. We emphasize again that the fast convergence results follow from the two important properties for the function $\mathrm{Mag}(\cdot)$ as discussed in Section IV-A. *Property* 1 guarantees the convergence of the local search algorithm; *Property* 2 results in multiple global maxima for the function $\mathrm{Mag}(\cdot)$ and hence the fast convergence of the algorithm. The simulations provide a partial validation of our proof since we would expect the convergence to fail from some initial points if there were non-optimal local maxima for $\mathrm{Mag}(\cdot)$. It is to be noted that the convergence of the local random search algorithm only shows that the algorithm is a feasible solution to *Problem 1* but not that it is the most efficient scheme in terms of the number of function evaluations. In other words, the analysis does not suggest that the algorithm requires the fewest number of function evaluations among all random search algorithms that converge. However, the algorithm does have a desirable scaling property, i.e., the time required for the algorithm to converge in mean scales linearly with the number of transmitters. This is the topic of the following section.

V. Scaling Law

Due to the random nature of the local random search algorithm, we showed in Section IV-B that the local random search algorithm converges in probability. For the analysis of the scaling law, we require an alternative definition of convergence.

Definition 2: The sequence $\{\boldsymbol{\theta}[t]\}$ is said to be *convergent in mean* if for any given $\epsilon > 0$ there exists $t_N \geq 0$ such that

$$E_{\{\boldsymbol{\delta}[\tau]\}_{\tau=0}^{t}|\mathbf{a},\boldsymbol{\theta}[0]}\left[\left|\mathrm{Mag}\left(\boldsymbol{\theta}[t]\right) - \mathrm{Mag}\left(\boldsymbol{\theta}^*\right)\right|\right] \leq \epsilon \tag{17}$$

for all $t \geq t_N$, where $\mathbf{a} = [a_1, \cdots, a_{n_s}]^T$. That is, $\mathrm{Mag}(\boldsymbol{\theta}[t])$ converges to $\mathrm{Mag}(\boldsymbol{\theta}^*)$ in mean. Note that since $\mathrm{Mag}\left(\boldsymbol{\theta}^*\right)$ is not random and $\mathrm{Mag}\left(\boldsymbol{\theta}^*\right) \geq \mathrm{Mag}\left(\boldsymbol{\theta}[t]\right)$ for all t, we can rewrite (17) as

$$E_{\{\boldsymbol{\delta}[\tau]\}_{\tau=0}^{t}|\mathbf{a},\boldsymbol{\theta}[0]}\left[\mathrm{Mag}\left(\boldsymbol{\theta}[t]\right)\right] \geq \mathrm{Mag}\left(\boldsymbol{\theta}^*\right) - \epsilon = \sqrt{P}\sum_{i=1}^{n_s} a_i - \epsilon \tag{18}$$

In this section, our goal is then to find the time required for the local random search algorithm to converge in mean, starting from any initial point. In other words, we are interested in finding the *hitting time*[5] of the random search algorithm, and determining its behavior as a function of the number of transmitters. Specifically, we derive an upper bound on the hitting time of the local random search algorithm as a function of n_s. Note that the study of the hitting time makes sense only if the sequence indeed converges in mean, which we established in Section IV-B.

[5]The hitting time in this work is defined as the time required for the algorithm to converge in mean as described in *Definition* 2.

To facilitate analysis, we define and lower bound the increment function of $\mathrm{Mag}(\cdot)$ at time τ as

$$\begin{aligned} I[\tau] &= \mathrm{Mag}\left(\boldsymbol{\theta}[\tau]\right) - \mathrm{Mag}\left(\boldsymbol{\theta}[\tau-1]\right) && (19)\\ &\overset{(a)}{=} \sqrt{P}\left[\left|\sum_{i=1}^{n_s} a_i e^{j(\theta_i[\tau-1]+\delta_i[\tau])}\right| - \sum_{i=1}^{n_s} a_i \cos\theta_i[\tau-1]\right]^+ && (20)\\ &\overset{(b)}{\geq} \sqrt{P}\left[\sum_{i=1}^{n_s} a_i \cos\left(\theta_i[\tau-1]+\delta_i[\tau]\right) - \sum_{i=1}^{n_s} a_i \cos\theta_i[\tau-1]\right]^+ && (21)\\ &\overset{(c)}{\geq} \sqrt{P}\sum_{i=1}^{n_s} a_i \left[\cos\left(\theta_i[\tau-1]+\delta_i[\tau]\right) - \cos\theta_i[\tau-1]\right]^+ =: \sum_{i=1}^{n_s} I_i[\tau] && (22) \end{aligned}$$

where (a) follows from properly shifting the phase according to $\boldsymbol{\theta}[\tau-1]$[6], (b) follows from expanding the magnitude in (20) and ignoring the imaginary part, and (c) from the sub-additivity of $[\cdot]^+$ and the positivity of a_i. From the definition of the increment function $I[\tau]$, we can rewrite the received signal magnitude function at any given time $k_0 n_s$ as

$$\mathrm{Mag}\left(\boldsymbol{\theta}[k_0 n_s]\right) = \sum_{\tau=1}^{k_0 n_s} I[\tau] + \mathrm{Mag}\left(\boldsymbol{\theta}[0]\right) =: \sum_{\tau=1}^{k_0 n_s} I[\tau] + c_0 \tag{23}$$

where k_0 is a positive integer and $c_0 \geq 0$. The continuity of $\mathrm{Mag}(\cdot)$ in each of its arguments allows us to rewrite the ϵ-convergence region as $R_\epsilon = \{\boldsymbol{\theta} : |\theta_i - \theta_i^*| \leq \sigma_i(\epsilon), \forall i = 1, \cdots, n_s\}$. To lower bound the conditional expectation of individual increment function $I_i[\tau]$, we define the following three sets

$$A_1[\tau-1] := \{ i = 1, \cdots, n_s \mid \pi \leq \theta_i[\tau-1] \leq 2\pi - \delta_0 \} \tag{24}$$

$$A_2[\tau-1] := \{ i = 1, \cdots, n_s \mid \delta_0 \leq \theta_i[\tau-1] \leq \pi \} \tag{25}$$

$$A_3[\tau-1] := \{ i = 1, \cdots, n_s \mid \sigma_i(\epsilon) \leq \theta_i[\tau-1] \leq \delta_0 \ \text{ or } \ 2\pi - \delta_0 \leq \theta_i[\tau-1] \leq 2\pi - \sigma_i(\epsilon) \} \tag{26}$$

Recall from Section IV-A that $\theta_i^* = 0$ or 2π for all i in the shifted domain. It is also important to note that for all $\boldsymbol{\theta}[\tau-1] \notin R_\epsilon$, there exists at least one i such that $i \in A_1[\tau-1] \cup A_2[\tau-1] \cup A_3[\tau-1]$. Now let us lower bound $E_{\boldsymbol{\delta}[\tau]|\mathbf{a},\boldsymbol{\theta}[\tau-1]}[I_i[\tau]]$ in each of these three different sets. To simplify notations, we omit the time dependency in the following discussions.

A. Lower bound on $E_{\boldsymbol{\delta}|\mathbf{a},\boldsymbol{\theta}}[I_i]$ for $i \in A_1$

If $i \in A_1$, it is clear that $\delta_i > 0$ implies that

$$\cos\left(\theta_i + \delta_i\right) - \cos\theta_i > 0 \tag{27}$$

We can lower bound the conditional expectation of I_i as

$$\begin{aligned} E_{\boldsymbol{\delta}|\mathbf{a},\boldsymbol{\theta}}\left[I_i\right] &\overset{(a)}{\geq} a_i\sqrt{P}\int_0^{\delta_0} \frac{1}{2\delta_0}\left(\cos\left(\theta_i+\delta_i\right) - \cos\theta_i\right) d\delta_i && (28)\\ &\overset{(b)}{\geq} a_i\sqrt{P}\int_0^{\delta_0} \frac{1}{2\delta_0}\left(1 - \cos\left(\delta_i\right)\right) d\delta_i && (29)\\ &= \sqrt{P}\frac{a_i}{2\delta_0}\left(\delta_0 - \sin\delta_0\right) && (30) \end{aligned}$$

[6] Note that all $\{\theta_i[\tau-1]\}$ are now in the shifted domain as discussed in Section IV-A

where (a) follows from (27), and (b) from the fact that the integrand in (28) as a function of θ_i attains its minimum at $\theta_i = \pi$ for all $i \in A_1$.

B. Lower bound on $E_{\boldsymbol{\delta}|\mathbf{a},\boldsymbol{\theta}}[I_i]$ for $i \in A_2$

Similarly, if $i \in A_2$, it is clear that $\delta_i < 0$ implies (27). We then obtain the lower bound of the conditional expectation of $I_i[\tau]$ as

$$\begin{aligned} E_{\boldsymbol{\delta}|\mathbf{a},\boldsymbol{\theta}}\left[I_i\right] &\geq a_i\sqrt{P}\int_{-\delta_0}^{0}\frac{1}{2\delta_0}\left(\cos\left(\theta_i+\delta_i\right)-\cos\theta_i\right)d\delta_i & (31)\\ &\overset{(a)}{\geq} a_i\sqrt{P}\int_{-\delta_0}^{0}\frac{1}{2\delta_0}\left(1-\cos\left(\delta_i\right)\right)d\delta_i & (32)\\ &= \sqrt{P}\frac{a_i}{2\delta_0}(\delta_0-\sin\delta_0) & (33)\end{aligned}$$

where (a) follows from the fact that the integrand in (31) as a function of θ_i attains its minimum at $\theta_i = \pi$.

C. Lower bound on $E_{\boldsymbol{\delta}|\mathbf{a},\boldsymbol{\theta}}[I_i]$ for $i \in A_3$

For all $i \in A_3$, if $\sigma_i(\epsilon) \leq \theta_i \leq \delta_0$, we have

$$\begin{aligned} E_{\boldsymbol{\delta}|\mathbf{a},\boldsymbol{\theta}}\left[I_i\right] &\geq a_i\sqrt{P}\int_{-\theta_i}^{0}\frac{1}{2\delta_0}\left(\cos\left(\theta_i+\delta_i\right)-\cos\theta_i\right)d\delta_i & (34)\\ &= \sqrt{P}\frac{a_i}{2\delta_0}\left(\sin\theta_i-\theta_i\cos\theta_i\right) & (35)\\ &\geq \sqrt{P}\frac{a_i}{2\delta_0}\left(\sin\sigma_i(\epsilon)-\sigma_i(\epsilon)\cos\sigma_i(\epsilon)\right) & (36)\end{aligned}$$

where the last inequality follows since $\sin x - x\cos x$ is a increasing function of x for all $x \leq \pi$.

Similarly, if $2\pi - \delta_0 \leq \theta_i \leq 2\pi - \sigma_i(\epsilon)$,

$$\begin{aligned} E_{\boldsymbol{\delta}|\mathbf{a},\boldsymbol{\theta}}\left[I_i\right] &\geq a_i\sqrt{P}\int_{0}^{2\pi-\theta_i}\frac{1}{2\delta_0}\left(\cos\left(\theta_i+\delta_i\right)-\cos\theta_i\right)d\delta_i & (37)\\ &= \sqrt{P}\frac{a_i}{2\delta_0}\left(-\sin\theta_i-(2\pi-\theta_i)\cos\theta_i\right) & (38)\\ &\geq \sqrt{P}\frac{a_i}{2\delta_0}\left(\sin\sigma_i(\epsilon)-\sigma_i(\epsilon)\cos\sigma_i(\epsilon)\right) & (39)\end{aligned}$$

where the second inequality follows since $-\sin x-(2\pi-x)\cos x$ is a decreasing function of x for all $\pi \leq x \leq 2\pi$.

Now for all τ and $i \in A_1[\tau-1] \cup A_2[\tau-1] \cup A_3[\tau-1]$,

$$E_{\boldsymbol{\delta}[\tau]|\mathbf{a},\boldsymbol{\theta}[\tau-1]}\left[I_i[\tau]\right] \geq \sqrt{P}\frac{a_i}{2\delta_0}\min\left\{\sin\sigma_i(\epsilon)-\sigma_i(\epsilon)\cos\sigma_i(\epsilon), \delta_0-\sin\delta_0\right\} =: c_1 \quad (40)$$

where c_1 is independent of n_s, i.e., the number of transmitters. Recall that for all $\boldsymbol{\theta}[\tau-1] \notin R_\epsilon$, there exists at least one i such that $i \in A_1[\tau-1] \cup A_2[\tau-1] \cup A_3[\tau-1]$. Thus, we have $E_{\boldsymbol{\delta}[\tau]|\mathbf{a},\boldsymbol{\theta}[\tau-1]}\left[I[\tau]\right] \geq c_1$ for any τ. Referring to (19)-(23), we obtain

$$E_{\{\boldsymbol{\delta}[\tau]\}_{\tau=0}^{k_0 n_s}|\mathbf{a},\boldsymbol{\theta}[0]}\left[\mathrm{Mag}\left(\boldsymbol{\theta}[k_0 n_s]\right)\right] = \sum_{\tau=1}^{k_0 n_s} E_{\boldsymbol{\delta}[\tau]|\mathbf{a},\boldsymbol{\theta}[\tau-1]}\left[I[\tau]\right] + c_0 \geq k_0 n_s c_1 + c_0 \geq \sqrt{P}\sum_{i=1}^{n_s} a_i \quad (41)$$

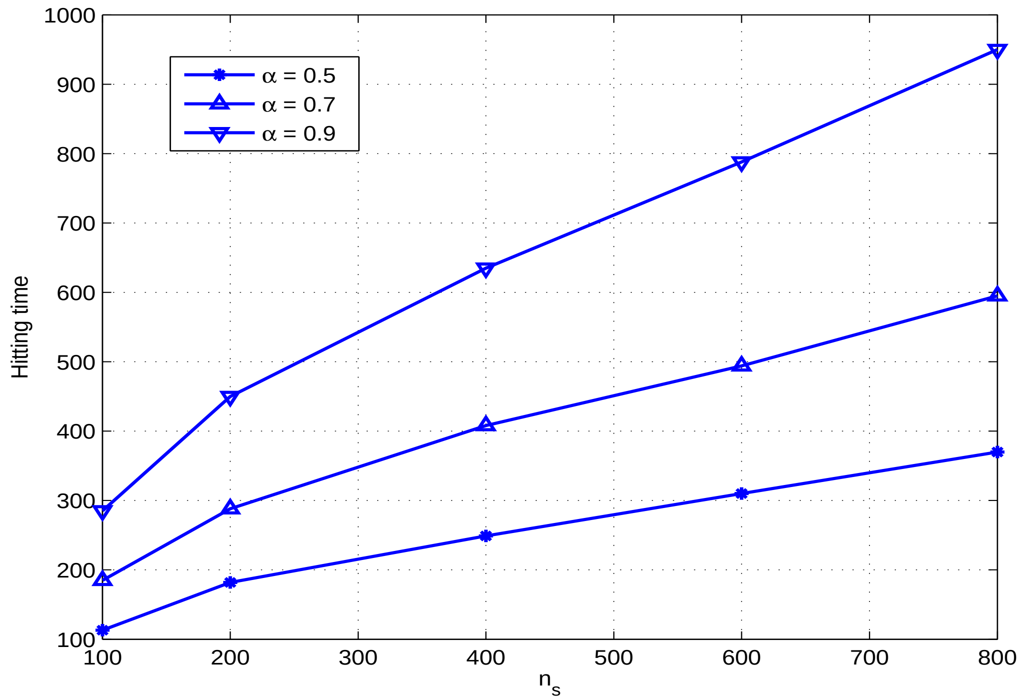

Fig. 3. Hitting time for the adaptive distributed beamforming scheme with different values of α.

where the last inequality follows by choosing $k_0 = \left\lceil \frac{\sqrt{P}\max_i\{a_i\}}{c_1} \right\rceil$. According to (18), this implies that the hitting time for the local random search algorithm is at most $k_0 n_s$, from any initial point. Hence, the hitting time for the algorithm scales linearly with the number of transmitters.

Although ϵ-convergence region is useful in defining convergence in an absolute sense in theory, the definition of convergence in a relative sense is more widely adopted in practice. In our simulations, we use this definition and say that the sequence converges to the α fraction of the global maxima if $\mathrm{Mag}(\boldsymbol{\theta}[t]) \geq \alpha\mathrm{Mag}(\boldsymbol{\theta}^*)$. We assume that channel coefficients are i.i.d. complex Gaussian variables $\mathcal{CN}(0,1)$, and use the origin as our initial point. We set $\delta_0 = \pi/90$ for all our simulations. Fig. 3 demonstrates the hitting time required for the adaptive distributed beamforming scheme to converge in a relative sense when $\alpha = 0.5$, 0.7, and 0.9. It is clear that the hitting time increases as α increases. The scaling law for the hitting time with respect to n_s, however, is the same for all values of α. Indeed, we observe linear scaling for all values of α. This observation confirms our theoretical analysis. Fig. 4 shows the average convergence time for the adaptive distributed beamforming scheme to within a fraction of the globally maximum value $\alpha\mathrm{Mag}(\boldsymbol{\theta}^*)$, for different values of α. It is important to note the difference between the hitting time and the average convergence time. Given the number of transmitters, we obtain the average convergence time by averaging over the convergence time for a hundred sample paths of the sequence $\{\boldsymbol{\theta}[t]\}$ generated by the algorithm, while the hitting time is obtained by comparing $E[\mathrm{Mag}(\boldsymbol{\theta}[t])]$ with

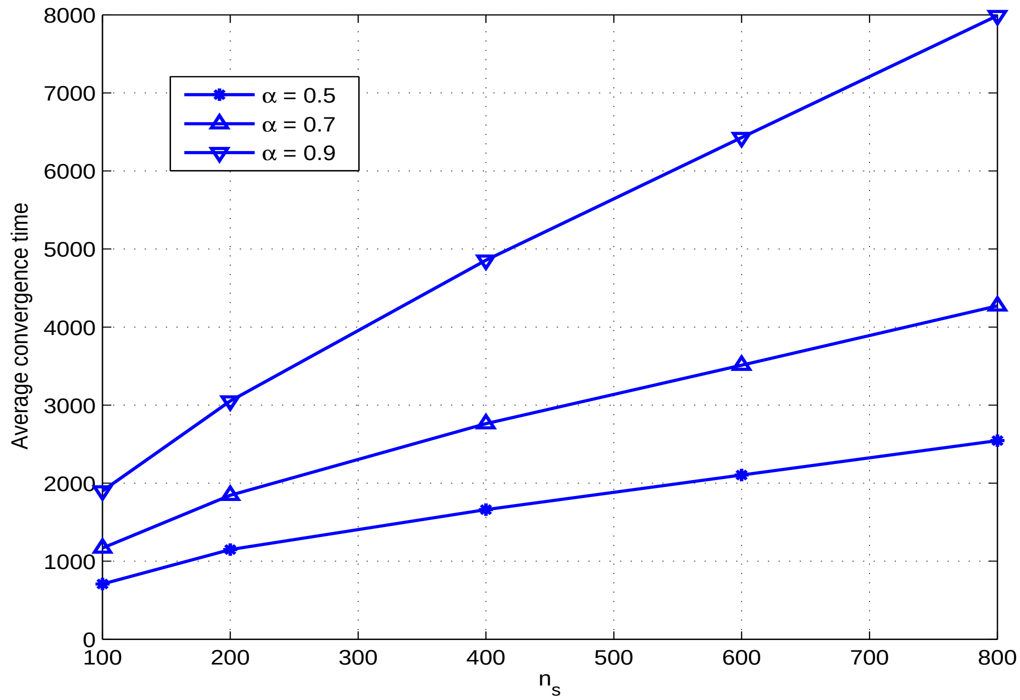

Fig. 4. Average convergence time for the adaptive distributed beamforming scheme with different values of α.

$\alpha\mathrm{Mag}(\boldsymbol{\theta}^*)$. From Fig. 4, we observe the same linear scaling behavior for the average convergence time. We expect this property for the average convergence time can be shown in a similar manner as in (19)-(41).

VI. Concluding Remarks and Future Work

In this work, we have studied the convergence and scaling law of a recently proposed [1] adaptive distributed beamforming scheme for sensor/relay networks. We first established an equivalence between the distributed scheme and a local random search algorithm. The equivalence provided insights into the convergence of the distributed beamforming scheme, and led us to investigate the fundamental properties for the received signal magnitude function $\mathrm{Mag}(\cdot)$. We found two important properties of the function that contribute to the rapid convergence of the algorithm. First, all local maxima are global maxima. This prevents any local random search algorithm from being trapped in non-optimal local maximum points. The second property is that $\mathrm{Mag}(\cdot)$ is invariant under a common shift to its arguments. This property results in multiple global maximum points for the function and hence the rapid convergence of the algorithm. Based on these properties, we have shown the convergence of the algorithm, both in probability and in mean. We further provided an upper bound on the hitting time of the algorithm, and demonstrated that the hitting time scales linearly with the number of sensor/relay nodes. This linear scaling is desirable, especially when the sensor/relay network is densely populated. We have also provided simulations that validate our analysis.

It is important to note that the effectiveness of the adaptive distributed beamforming scheme highly depends on the properties of the function $\mathrm{Mag}(\cdot)$. While maximizing $\mathrm{Mag}(\cdot)$ is equivalent to maximizing the received SNR if there is no error in obtaining the common message, we may need to consider a more complicated function when there are errors in the common message. In this case, the new metric function to consider may not possess the same desirable properties as $\mathrm{Mag}(\cdot)$. For example, there may be local maxima for the new metric function. Much work needs to be done to understand how our results can be applied in this more sophisticated scenario. One thing that is clear, however, is that we will need to develop new algorithms that exploit the global structure of the new metric function since local algorithms can be trapped in local maxima. In this scenario, we believe that it will be useful to formulate the problem in our framework using the equivalence we established since it connects the problem to a well-studied field of global optimization algorithms. This is a topic for future research.

REFERENCES

[1] R. Mudumbai, B. Wild, U. Madhow, and K. Ramchandran, “Distributed Beamforming using 1 Bit Feedback: from Concept to Realization,” *Proc. Annual Allerton Conf. Commun. Control and Computing*, 2006.

[2] J. Thukral and H. Blcskei, “Distributed Spatial Multiplexing with 1-bit Feedback,” *Allerton Conf. Commun. Cont. and Comp.*, 2007.

[3] F. J. Solis and R. J.-B. Wets, “Minimization by Random Search Techniques,” *Journal of Mathematics of Operations Research*, vol. 6, pp. 19–30, Feb. 1981.

[4] L. Shi and S. Olafsson, “Nested Partitions Method for Global Optimization,” *Journal of Operations Research*, vol. 48, pp. 390–407, May 2000.

[5] Z. B. Tang, “Adaptive Partitioned Random Search to Global Optimization,” *IEEE Trans. Automatic Control*, vol. 39, pp. 2235–2244, Nov. 1994.

[6] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by Simulated Annealing,” *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[7] R. A. Durrett, *Probability: Theory and Examples*. Duxbury Press, 2nd ed., 1995.