A GENERAL PROOF OF CONVERGENCE FOR ADAPTIVE DISTRIBUTED BEAMFORMING SCHEMES

Chang-Ching Chen, Chia-Shiang Tseng, and Che Lin

Institute of Communication Engineering & Department of Electrical Engineering
National Tsing Hua University,
Hsinchu, Taiwan 30013

E-mail: s9864511@m98.nthu.edu.tw, s9964518@m99.nthu.edu.tw, clin@ee.nthu.edu.tw

ABSTRACT
This work focuses on the convergence analysis of adaptive distributed beamforming schemes that can be reformulated as local random search algorithms via a random search framework. Once reformulated as local random search algorithms, it is proved that under two sufficient conditions: a) the objective function of the algorithm is continuous and all its local maxima are global maxima, and b) the origin is an interior point within the range of the considered transformation of the random perturbation, the corresponding adaptive distributed beamforming schemes converge both in probability and in mean. This proof of convergence is general since it can be applied to analyze randomized adaptive beamforming schemes with any type of objective functions and probability measures as long as both the sufficient conditions are satisfied. Further, this framework can be generalized to analyze an asynchronous scheme where distributed transmitters can only update their beamforming coefficients asynchronously. Simulation results are also provided to validate our analyses.

Index Terms— Beamforming, convergence analysis, distributed algorithms, feedback communications.

1. INTRODUCTION
In a distributed network, distributed beamforming is considered as a promising scheme that allows distributed transmitters to convey common information efficiently in energy due to its potential array gain and low-complexity. However, to achieve distributed beamforming, or phase alignment, channel state information (CSI) at the transmitters is required and the cost of obtaining perfect CSI is too expensive in practice. Therefore, instead of obtaining perfect CSI, an adaptive scheme that uses a one-bit feedback link to acquire partial CSI at the transmitter was proposed in [1], and the analyses of this one-bit feedback adaptive scheme have been extensively studied in [2]-[6]. In [3], [4], a discrete version of the one-bit adaptive scheme with binary signaling is considered. To analyze the behavior of the one-bit adaptive scheme, the authors in [5] applied stochastic approximations to show the convergence of the scheme. It has been shown that the sample path of the one-bit adaptive scheme approximately follows an ordinary differential equation under suitable conditions, and the convergence of the scheme is established accordingly. Alternatively, the one-bit adaptive scheme was reformulated as a local random search algorithm via a random search framework and analyzed in our previous work [6]. This reformulation allows us to use the techniques studied in the mature field of random search for systematically analyzing the characteristics of the one-bit adaptive scheme. With the help of this framework, it has been shown, without any approximation, that the one-bit adaptive scheme converges both in probability and in mean, and its convergence time scales linearly with the number of distributed transmitters.

In this paper, we further generalize the framework in [6] and provide a more general proof of convergence and linear scalability for adaptive distributed beamforming schemes. We show that any adaptive distributed beamforming scheme that can be reformulated as a local random search algorithm converges both in probability and in mean if the algorithm satisfies two sufficient conditions for the objective function and the random perturbation. This framework is more general since once the sufficient conditions are satisfied, there is no need to specify the objective functions and probability measures. Instead of focusing on a particular objective function and symmetric probability measures as in [1], [5], [6], our framework can be applied to analyze a much broader set of adaptive distributed beamforming schemes and hence provide more potential to unify theoretical analysis in this field. Furthermore, since the transmitters are deployed in a distributed fashion, it is possible that they experience different environments that cause their local clocks to be asynchronous. Our framework can be further extended to analyze adaptive distributed beamforming schemes in such asynchronous setting by a straightforward application of Bayes’ rule.

2. SYSTEM SETUP
We consider a network with $N$ distributed transmitters that attempt to convey a common message to the receiver. We assume that all transmitters and the receiver are equipped with one antenna and the channels between each transmitter and the receiver are slow faded and frequency flat. Furthermore, we assume a noncoherent communication model where both the transmitters and the receiver do not have CSI. However, there is an error-free, zero-delay, and low-rate feedback link from the receiver to each transmitter so that the receiver can help aligning the beamforming phases through this reverse feedback link. The discrete-time, complex baseband received signal can be expressed as

$$y[n] = \sum_{i=1}^{N} h_i g_i[n] s[n] + w[n] = \sum_{i=1}^{N} a_i b_i[n] e^{j\phi_i + \psi_i[n]} s[n] + w[n]$$

where $y[n] \in \mathbb{C}$ is the received signal, $s[n] \in \mathbb{C}$ is the common message with the average power constraint $E[|s[n]|^2] \leq P_s$ for all $n$, and $w[n] \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise. For transmitter $i$, we denote the time-invariant channel fading gain by $h_i = a_i e^{j\phi_i} \in \mathbb{C}$, and the beamforming coefficient by $g_i[n] =$
same space, above, we further provide two sufficient conditions for our proof. This makes our proofs much more general than those in [6].

Further, if all the local maxima of the objective function are global maxima, there is no knowledge on the objective function and its gradient at the transmitter end in our case. However, from our previous work [6], it can be solved by global random search algorithms in general. Further, if all the local maxima of the objective function are global maxima, local random search algorithms, which are more efficient, can also be applied.

3. GENERAL PROOF OF CONVERGENCE AND LINEAR SCALABILITY

In this section, we provide a more general proof for the convergence and linear scalability of a set of adaptive distributed beamforming problems that can be reformulated as local random search algorithms. To this end, we briefly describe the local random search algorithm in [6] as follows:

- **Step zero:** Initialize the algorithm by choosing \( \theta[0] \in \Theta \).
- **Step one:** Generate a random perturbation \( \delta[n] \) from the probability measure \( \mu_n \) that could be time-varying and has the support \( \text{supp}(\mu_n) \).
- **Step two:** Update the search point by \( \theta[n] = D(\theta[n-1], \delta[n]) \), where the mapping \( D : \Theta \times \text{supp}(\mu_n) \to \Theta \) satisfies the condition that

\[
f(D(\theta[n-1], \delta[n])) \geq f(\theta[n-1])
\]

In most cases, the mapping \( D \) can be expressed as

\[
D(\theta[n-1], \delta[n]) = \theta[n-1] + G_n(\delta[n]) 1( f(\theta[n-1]) + G_n(\delta[n])) > f(\theta[n-1]) \)

where \( 1(\cdot) \) is the indicator function and \( G_n : \text{supp}(\mu_n) \to \Gamma_n \subseteq \mathbb{R}^N \) can be a general transformation. Note that \( \Theta \) and \( \Gamma_n \) are in the same space, i.e., \( \mathbb{R}^N \).

Now, referring to the local random search algorithm described above, we further provide two sufficient conditions for our proof.

(S1) The objective function \( f \) is continuous and all its local maxima are global maxima.

(S2) The origin is an interior point of the range of the transformation \( G_n \), i.e., \( \Gamma_n \), for all \( n \).

In the following analyses, we consider local random search algorithms that satisfy both the sufficient conditions. Note that we do not need to specify the exact expressions of the objective function and the probability measure once the sufficient conditions are satisfied. This makes our proofs much more general than those in [6].

3.1. Convergence

We first define the convergence in probability as follows.

**Definition 1** A sequence \( \{\theta[n]\}_{n=0}^{\infty} \) generated by a random search algorithm is said to converge in probability if given \( \epsilon > 0 \),

\[
\lim_{n \to \infty} \Pr[\theta[n] \in R_{\epsilon}] = 1
\]

where \( R_{\epsilon} := \{\theta : f(\theta) > f(\theta^*) - \epsilon\} \) is the \( \epsilon \)-convergence region. In other words, \( f(\theta[n]) \) converges to \( f(\theta^*) \) in probability, where \( \theta^* \) is a global maximum point.

Now, we further derive two propositions for the proof of convergence. For notational simplicity, we omit the time indices and write \( \theta, \delta \) instead of \( \theta[n], \delta[n] \) in the following propositions and the corresponding proofs.

**Proposition 1** Given an objective function \( f \) and a transformation \( G_n \) with domain \( \text{supp}(\mu_n) \) that satisfy the sufficient conditions (S1) and (S2). Let \( D \) be the mapping as defined in [6], then for any \( \theta \notin \Theta \), there exists a set \( A(\theta) \subseteq \text{Range}(D(\theta, \cdot)) \) with nonempty interior points, such that \( f(a) > f(\theta) \) for all \( a \in A(\theta) \), where \( \text{Range}(D(\theta, \cdot)) \) is the range of the mapping \( D \) at \( \theta \).

**Proof:** We begin the proof by defining a \( f(\theta) \)-superlevel set

\[
U_\theta := \{\psi : f(\psi) \geq f(\theta)\}
\]

which is the subset of \( \Theta \) with all the elements in \( U_\theta \) mapping to function values no less than \( f(\theta) \). Note that the assumption \( \Theta \not\subseteq U_\theta \) implies that \( R_{\epsilon} \not\subseteq U_\theta \) and since \( R_{\epsilon} \) is nonempty, \( U_\theta \) is nonempty. Furthermore, since \( f \) is continuous, the boundary points are those points that satisfy \( f(\psi) = f(\theta) \).

We claim that for any \( \xi > 0 \), there exists a set \( A(\theta) \subseteq B(\theta, \xi) \cap U_\theta \) with nonempty interior points, such that \( f(a) > f(\theta) \) for all \( a \in A(\theta) \), where \( B(\theta, \xi) \) is an open ball centered at \( \theta \) with radius \( \xi \). To this end, we first rule out the case where there are flat regions outside \( R_{\epsilon} \), that is, for some \( \theta \notin R_{\epsilon} \), there exists a \( \xi > 0 \) such that \( f(\psi) = f(\theta) \) for all \( \psi \in B(\theta, \xi) \). This case cannot happen since if it does, then \( f(\psi) \leq f(\theta) \) for all \( \psi \in B(\theta, \xi) \) and this implies that \( \theta \) is a local maxima point and thus a global maxima point by the sufficient condition (S1), and this contradicts to the assumption that \( \theta \notin R_{\epsilon} \). Therefore, there are no flat regions outside \( R_{\epsilon} \).

Excluding the existence of flat regions outside \( R_{\epsilon} \), we now consider two cases where \( \theta \) is either a local minimum point or not. Note that the arguments leading to the absence of flat regions outside \( R_{\epsilon} \) implies that \( \theta \) cannot be a local maxima point.

Case I: If \( \theta \) is a local minimum point, we can show that it is an interior point of \( U_\theta \) since the fact that \( \theta \) is a local minimum point but not a local maxima point implies that there exists a \( \xi > 0 \) such that \( f(\psi) > f(\theta) \) for all \( \psi \in B(\theta, \xi) \setminus \{\theta\} \). This means that \( B(\theta, \xi) \subseteq U_\theta \) and thus \( \theta \) is an interior point of \( U_\theta \). Note that in this case, the entire \( B(\theta, \xi) \) are in the interior of \( U_\theta \) and we choose \( A(\theta) = B(\theta, \xi) \setminus \{\theta\} \), which obviously has nonempty interior points.

Case IIf: If \( \theta \) is not a local minimum point, it can be shown that it is a boundary point of \( U_\theta \) but not an isolated point. First, we show that \( \theta \) is not an isolated point. If \( \theta \) is an isolated point, it means that there exists a \( \zeta > 0 \) such that for all \( \psi \in B(\theta, \zeta) \setminus \{\theta\} \), \( \psi \not\in U_\theta \), or equivalently, \( f(\psi) < f(\theta) \). This means that \( \theta \) is a local maxima point and thus a global maximum point by the sufficient condition (S1). Again, we obtain a contradiction to the assumption that \( \theta \notin R_{\epsilon} \). Now, since \( \theta \) is neither a local maximum point nor
a local minimum point, then for any $\xi > 0$, we can always find a point $\psi_1 \in \mathcal{B}(\Theta, \xi)$ such that $f(\psi_1) > f(\Theta)$, i.e., $\psi_1 \in U_0$ and a point $\psi_2 \in \mathcal{B}(\Theta, \xi)$ such that $f(\psi_2) < f(\Theta)$, i.e., $\psi_2 \notin U_0$. Thus, for any $\xi > 0$, $\mathcal{B}(\Theta, \xi)$ contains both points inside and outside $U_0$. This means that $\Theta$ is a boundary point. Note that $\psi_1$ is an interior point in $U_0$ since it is not, then it is a boundary point of $U_0$ and it means $f(\psi_1) = f(\Theta)$, which contradicts to the statement that $f(\psi_1) > f(\Theta)$. In this case, we choose $A(\Theta) = \mathcal{B}(\Theta, \xi) \cap U_\Theta \cap C$, where $C := \{\psi \in U_\Theta : f(\psi) = f(\Theta)\}$ and there are nonempty interior points in $A(\Theta)$. Therefore, our claim is indeed true.

Now we define

$$\Omega := \{\omega \in \Theta : \omega = \Theta + \mu_n(\delta), \delta \in \text{supp}(\mu_n)\}$$

as the set by shifting $\Gamma_\Theta$ with respect to $\Theta$. Then the range of $D(\Theta, \cdot)$ for any $\Theta \notin R_\epsilon$ is given by

$$\text{Range}(D(\Theta, \cdot)) = \Omega \cap U_\Theta$$

Note that the sufficient condition (S2) implies that $\Theta$ is an interior point of $\Omega$. Since $\Theta$ is also a boundary or interior point of $U_\Theta$, there always exist $\xi > 0$ such that $\mathcal{B}(\Theta, \xi) \cap U_\Theta \cap \text{Range}(D(\Theta, \cdot))$. Therefore, we can conclude that $A(\Theta) \subset \text{Range}(D(\Theta, \cdot)).$}

The following proposition states that for any $\Theta$ outside $R_\epsilon$, there is a non-zero probability to improve $f$ by applying a local perturbation to $\Theta$.

**Proposition 2** For any given $\Theta \in \Theta \setminus R_\epsilon$, there correspond $\gamma > 0$ and $0 < \eta \leq 1$ such that

$$\Pr[f(\Theta + \mu_n(\delta)) - f(\Theta) \geq \gamma] \geq \eta$$

where $\delta$ is a random vector generated with the probability measure $\mu_n$.

**Proof:** Proposition 7 implies that for any $\Theta \in \Theta \setminus R_\epsilon$, there exists an interior point $\Theta \in A(\Theta)$ and $\xi > 0$ such that $\forall \psi \in T := B(\Theta, \xi)$, $f(\psi) - f(\Theta) \geq \gamma(\Theta)$. Then,

$$\Pr[f(\Theta + \mu_n(\delta)) - f(\Theta) \geq \gamma(\Theta)] \geq \mu_n(T) =: \eta(\Theta) > 0$$

Note that $T$ is a function of $\Theta$ since $\Theta$ depends on $\Theta$. We complete the proof by letting

$$\gamma = \inf_{\Theta \in \Theta \setminus R_\epsilon} \gamma(\Theta)$$

$$\eta = \inf_{\Theta \in \Theta \setminus R_\epsilon} \eta(\Theta)$$

Since there is always a non-zero probability to improve $f(\cdot)$ for each time step before the sequence reaches $R_\epsilon$, the convergence is expected. We describe this more precisely in the following theorem.

**Theorem 1** For an objective function $f$ and a transformation $G_n$ with domain $\text{supp}(\mu_n)$ that satisfy the sufficient conditions (S1) and (S2), let $\{\Theta[n]\}_n^{\infty}$ be a sequence generated from the local random search algorithm. Then the resulting sequence converges in probability, i.e., given $\epsilon > 0$,

$$\lim_{n \to \infty} \Pr[\Theta[n] \in R_\epsilon] = 1$$

With Proposition 7, the proof directly follows the one provided in our previous work [6] by replacing the $\text{Mag}(\cdot)$ in [6] by a general objective function $f(\cdot)$. We hence omit the details. Note that since $\{f(\Theta[n])\}_n^{\infty}$ is non-negative and monotonically non-decreasing, this sequence also converges in mean by the Monotone Convergence Theorem 7.

### 3.2. Linear Scalability

For the analysis of the scaling law, we use an alternative definition of convergence.

**Definition 2** A sequence $\{\Theta[n]\}_n^{\infty}$ generated by a random search algorithm is said to converge in mean if there exists $M \geq 0$ such that

$$\mathbb{E}_n(\delta[n]) = 0$$



for all $n \geq M$. That is, $f(\Theta[n])$ converges to $f(\Theta^*)$ in mean. Furthermore, the iterations required to converge in mean is defined as the hitting time.

With the above definition, we introduce the following theorem for linear scalability.

**Theorem 2** Given $\Theta \subseteq \mathbb{R}^N$, the hitting time of the random search algorithm scales linearly with $N$. That is, there exists $k < \infty$, such that Eqn. 4 holds for $M \leq kN$.

By substituting the $\text{Mag}(\cdot)$ in [6] by a more general objective function $f(\cdot)$ which satisfies the stated sufficient conditions, the proof directly follows the one provided in [6].

Even though the derivation of the proofs for the above two theorems directly follow those in [6], with only little modification, we emphasize again that we do not need to specify the objective function and the probability measure used as in [6] once the sufficient conditions are satisfied. Therefore, our proofs are more general.

### 4. Asynchronous Scheme

In this section, we model an asynchronous adaptive distributed beamforming scheme and analyze its convergence by extending our random search framework. We consider the case where at each time instance based on the global clock, only $\rho\%$ of the distributed transmitters update their phases and the rest of them keep their phases unchanged. This scenario can be equivalently modeled by assuming that each transmitter perturbs its phase independently with probability $p = \rho\%$ at each time slot. This is similar to but different from the $\rho\%$ algorithm proposed in [5] since in our case, we assume that all the transmitters transmit at all times but only on average $\rho\%$ of them change their phases at each time slot.

Now, we show that this adaptive distributed beamforming scheme still converges even when the phases are updated asynchronously. Similar to the analysis in Sec. 3, we first derive a proposition stating that for each perturbation, there is a non-zero probability to improve the objective function $f$.

**Proposition 3** Consider the asynchronous scheme described above, then for any given $\Theta \in \Theta \setminus R_\epsilon$, there corresponds $\gamma > 0$ and $0 < \lambda \leq 1$ such that

$$\Pr[f(\Theta + \mu_n(\delta)) - f(\Theta) \geq \gamma] \geq \lambda$$

where $\delta$ is a random vector generated with the probability measure $\mu_n$.

**Proof:** We prove the proposition by Bayes’ rule. Let $Z$ be the event that all transmitters update their phases. Then

$$\Pr[f(\Theta + \mu_n(\delta)) - f(\Theta) \geq \gamma]
\geq \Pr[f(\Theta + \mu_n(\delta)) - f(\Theta) \geq \gamma | Z] \Pr[Z]
\geq \eta p^N \geq \lambda$$

The last inequality follows from Proposition 7.
In this section, numerical results are provided to validate our analyses. We consider a network with $N = 200$ distributed transmitters and assume the channel coefficients are i.i.d. $\mathcal{CN}(0, 1)$. Unless otherwise specified, we use the SNR function described by (2) as our objective function, set $G_n(\cdot)$ to be an identity function, and use a random perturbation with the time-invariant probability measure $\mu \sim U[-\Delta_0, \Delta_0]^N$, where $U[\cdot, \cdot]$ denotes the uniform distribution and $\Delta_0 = 5^\circ$ in the simulation. Besides, for simplicity, we say that the algorithm converges once $\text{SNR}(\theta) \geq \alpha \text{SNR}(\theta^*)$, where $\alpha = 0.9$. It has been shown in [6] that the SNR function, the probability measure $\mu$, and the transformation $G_n(\cdot)$ considered above satisfy both the sufficient conditions. We hence use this setting to verify our analyses.

With Proposition 3 we can apply Theorem 2 to show that this asynchronous scheme converges both in probability and in mean. Note that by applying the simple Bayes’ rule, we are able to prove the convergence for a much broader set of adaptive distributed beamforming schemes. This demonstrates again the generality and flexibility of our framework.

5. SIMULATION RESULTS AND CONCLUSION

In this section, numerical results are provided to validate our analyses. We consider a network with $N = 200$ distributed transmitters and assume the channel coefficients are i.i.d. $\mathcal{CN}(0, 1)$. Unless otherwise specified, we use the SNR function described by (2) as our objective function, set $G_n(\cdot)$ to be an identity function, and use a random perturbation with the time-invariant probability measure $\mu \sim U[-\Delta_0, \Delta_0]^N$, where $U[\cdot, \cdot]$ denotes the uniform distribution and $\Delta_0 = 5^\circ$ in the simulation. Besides, for simplicity, we say that the algorithm converges once $\text{SNR}(\theta) \geq \alpha \text{SNR}(\theta^*)$, where $\alpha = 0.9$. It has been shown in [6] that the SNR function, the probability measure $\mu$, and the transformation $G_n(\cdot)$ considered above satisfy both the sufficient conditions. We hence use this setting to verify our analyses.

Fig. 1 shows the convergence of (a) an arbitrary objective function satisfying the sufficient condition (S1) and (b) a random perturbation with probability measure which is not symmetric but satisfies the sufficient condition (S2). For Fig. 1(a), we choose the objective function to be $f(\theta) = \sum_{i=1}^{N} \left[ \left( \theta_i \mod \pi - \frac{\pi}{2} \right)^2 + \frac{\pi}{2} \right]^2$, where mod is the modulo function. It can be shown that this function satisfies the sufficient condition (S1) and we omit the details due to space constraint. For Fig. 1(b), we generate the asymmetric distribution by randomly shifting $\mu$ around zero. For both settings, we demonstrate that the modified algorithm either with new objective function or asymmetric probability measure converges from different initial points. Note that more simulations with different random shifts of $\mu$ have been done and show similar behaviors but are not included here due to space constraint. The asynchronous scheme described in Sec. 4 is considered in Fig. 2. In Fig. 2(a), we show the difference between the asynchronous scheme with different $\rho\%$ and the synchronous scheme in terms of the convergence time. This figure also shows the average iterations necessary for the asynchronous scheme to converge for different $\rho\%$. It is obvious that the performance of the asynchronous scheme is worse than that of the synchronous scheme. However, we emphasize that even when only $\rho\%$ of the distributed transmitters update their phases at a given time instance, the algorithm still converges. Since the convergence behavior from different initial points of the asynchronous scheme for any $\rho\%$ is also similar to those in Fig. 1, we omit it due to space limitation. In Fig. 2(b), we demonstrate the linear scalability of the asynchronous scheme with $\rho\% = 50\%$. We observe that the convergence time increases with $\alpha$, and that for any $\alpha$, the average number of iterations necessary for the convergence scales linearly with the number of distributed transmitters. These simulation results verify our theoretical analyses.

In this paper, we generalized the proof of convergence for adaptive distributed beamforming schemes that can be reformulated as local random search algorithms and satisfy two sufficient conditions. Specifically, we have shown that such adaptive distributed beamforming schemes converge both in probability and in mean and their convergence time scales linearly with the number of distributed transmitters. We further extended our framework and analyzed the case where distributed transmitters can only update their beamforming coefficients asynchronously. Simulations were provided to validate our analyses.

6. REFERENCES

[1] R. Mudumbai, B. Wild, U. Madhow, and K. Ramchandran, “Distributed Beamforming Using 1 Bit Feedback: from Concept to Realization,” Allerton Conf. on Commun., Cont. and Comp., vol. 44, pp. 1020–1027, 2006.

[2] R. Mudumbai, “Energy efficient wireless communication using distributed beamforming,” Ph. D thesis, Santa Barbara, Dec. 2007.

[3] M. Johnson, M. Mitzenmacher, and K. Ramchandran, “Distributed Beamforming with Binary Signaling,” IEEE Intl. Symp. on Info. Theory, 2008.

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

[5] J. Bucklew and W. Sethares, “Convergence of a Class of Decentralized Beamforming Algorithms,” IEEE Trans. on Sig. Processing, vol. 56, no. 6, pp. 2280–2288, June 2006.

[6] C. Lin, V. V. Veeravalli, and S. P. Meyn, “A random search framework for convergence analysis of distributed beamforming with feedback,” IEEE Trans. Info. Theory, vol. 56, no. 12, Dec 2010.

[7] R. A. Durrett, “Probability: Theory and examples,” Duxbury Press, 2nd ed., 1995.