On the Performance of MIMO-Based Ad Hoc Networks under Imperfect CSI

Jia Liu and Y. Thomas Hou
The Bradley Department of Electrical and Computer Engineering
Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

***Abstract*—In this paper, we investigate the performance of MIMO-based ad hoc networks under imperfect channel state information (CSI). Most existing results on MIMO ad hoc networks assume perfect CSI. This assumption is usually unrealistic in practice because the feedback overhead of perfect CSI is prohibitively high. Thus, it is of great interest to study how to optimize the performance of MIMO-based ad hoc networks under imperfect CSI. In this paper, we propose a solution based on stochastic approximation technique to address imperfect CSI. To analyze the performance gap due to the existence of CSI error, we first show that our proposed solution is stable and converges with probability one to a neighborhood of an optimal solution obtained under perfect CSI. Based on this stability result, we characterize the performance gap due to imperfect CSI. We show that the boundary of the stability neighborhood is constituted by points for which the excess capacity of each link is equal to the ergodic link capacity loss due to imperfect CSI. Thus, as the amount of CSI feedback increases, the performance gap of our proposed solution vanishes.**

I. Introduction

The pioneering works by Foschini and Gans [1], and Teletar [2] have shown that much higher spectral efficiency and capacity gain can be achieved by the use of multiple antenna systems, now known as multiple-input multiple-output (MIMO) technology. The potential of significant improvement in channel capacity has positioned MIMO technology as a major breakthrough in modern wireless communications. As a result, there is a great interest on applying MIMO to ad hoc networks.

However, applying MIMO in practical ad hoc networks is far from trivial. One major obstacle is due to channel state information (CSI) feedback. It is well understood that with or without CSI knowledge at the transmitter can make a big difference in MIMO channel capacity [3]. Usually, CSI knowledge at transmitter side requires CSI feedback on the reverse link from the receiver. Unfortunately, in practice, acquiring full CSI appears to be difficult even in point-to-point or single-hop systems. For example, a 4×4 matrix channel has 32 complex parameters that need to be quantized each time when the channel changes. Compared with conventional single-antenna systems, this is a factor of 30 increase in feedback overhead on the reverse link [4]. Apparently, such feedback overhead is not acceptable. In more complex multi-hop MIMO networks, assuming perfect CSI becomes even more unrealistic because the amount of CSI feedback grows not only with the product of transmit antennas and receive antennas but also with the number of links and queuing delay spread in the network. The bandwidth consumed for CSI feedback becomes prohibitively large even for a moderate sized networks.

To optimize system performance while still keeping feedback overhead acceptable, researchers have proposed to use various limited feedback (LFB) techniques for point-to-point systems (see [4] for an overview) and single-hop systems [5]–[7]. The basic idea of these LFB schemes is that, instead of full CSI, only a limited number of feedback bits representing certain state of CSI are transmitted by the receiver back to the transmitter. The transmitter can then use this imperfect CSI to adapt its signals to approximate the current channel eigen-structure with certain error. However, compared with the research on imperfect CSI (due to LFB) in point-to-point and single-hop systems, research on multi-hop ad hoc network performance under imperfect CSI remains in its infancy. This is perhaps because imperfect CSI brings further challenges to cross-layer design, which is now widely considered a necessary approach due to the high complexity of MIMO [8], [9]. Imperfect CSI at the physical layer may have a chain effect to upper layers as follows. Due to power control decision error under imperfect CSI, the transmitted signals are no longer adapted to the eigen-direction of the channel, thus resulting in suboptimal link capacities. The loss in link capacity will in turn cause other network-wide issues, such as erroneous end-to-end rate control and routing decisions. Moreover, many existing cross-layer optimization algorithms can no longer be applied if the perfect CSI assumption is violated.

Our goal in this paper is to fill this important gap by developing a methodology to optimize the performance of MIMO ad hoc networks under imperfect CSI. The main contributions of this paper are the following:

- We propose a solution to the cross-layer optimization problem for MIMO networks under imperfect CSI. We show that our proposed solution converges with probability one to a neighborhood of the solution obtained under perfect CSI. The mathematical technique for this result is based on stochastic approximation.
- Based on the above result, we further analyze the size of the neighborhood (performance gap) of our proposed solution under imperfect CSI. We show that the size of the performance gap is determined by the ergodic capacity loss and power control error, which depend on the number of feedback bits. We further show that the performance

978-1-4244-2677-5/08/$25.00

gap of our proposed solution vanishes as the number of feedback bits goes to infinity.

The remainder of this paper is organized as follows. In Section II, we discuss the network model and the problem under study. Section III gives a solution under perfect CSI. In Section IV, we propose a solution under imperfect CSI. In Section V, we show the stability of our proposed solution and analyze the performance gap. Section VI presents some numerical results. Section VII concludes this paper.

II. Network Model and Problem Formulation

We consider a joint rate and power control problem for a MIMO-based ad hoc network. Our goal is to jointly optimize the transmission rate of each session and the power allocation of each node such that some utility function (a function of session rates) is maximized.

We first introduce notation for matrices, vectors, and complex scalars in this paper. We use boldface to denote matrices and vectors. For a matrix $\mathbf{A}$, $\mathbf{A}^{\dagger}$ denotes the conjugate transpose, $\mathrm{Tr}\{\mathbf{A}\}$ denotes the trace of $\mathbf{A}$, and $|\mathbf{A}|$ denotes the determinant of $\mathbf{A}$. We denote $\mathbf{I}$ the identity matrix with dimension determined from context. $\mathbf{A} \succeq 0$ represents that $\mathbf{A}$ is Hermitian and positive semidefinite (PSD). $\mathbf{1}$ and $\mathbf{0}$ denote vectors whose elements are all ones and zeros, respectively, and their dimensions are determined from context. For matrix $\mathbf{A}$, A_{ij} denotes the entry on the i-th row and the j-th column. For a real vector $\mathbf{v}$ and a real matrix $\mathbf{A}$, $\mathbf{v} \geq \mathbf{0}$ and $\mathbf{A} \geq \mathbf{0}$ mean that all entries in $\mathbf{v}$ and $\mathbf{A}$ are nonnegative, respectively. The operator "$\langle \cdot, \cdot \rangle$" represents the inner product operation for vectors or matrices.

The topology of a MIMO-based ad hoc network is modeled as a set of nodes, denoted by $\mathcal{N}$, and a set of links, denoted by $\mathcal{L}$. In our model, a link connecting a pair of nodes exists if the distance between the nodes is less than or equal to the maximum transmission range $d_{\max}$, i.e., $\mathcal{L} = \{(i,j) : d_{ij} \leq d_{\max}, i,j \in \mathcal{N}, i \neq j\}$, where d_{ij} represents the distance between node i and node j. The maximum transmission range $d_{\max}$ is determined by a node's maximum transmission power. Suppose that the cardinalities of the sets $\mathcal{N}$ and $\mathcal{L}$ are $|\mathcal{N}| = N$ and $|\mathcal{L}| = L$, respectively. For convenience, we identify the links in the network numerically as $1 \ldots L$ instead of using node pairs (i,j). Also, we use $\mathcal{O}(n)$ to represent the set of outgoing links from node n and $tx(l)$ to represent the transmitting node of link l. In this paper, we assume that each node has been assigned non-overlapping (possibly reused) frequency bands for its incoming and outgoing links. There is a vast amount of literature (see, e.g., [10] and references therein) on how to perform channel assignments and its discussion is beyond the scope of this paper.

Let $\mathbf{H}_l \in \mathbb{C}^{n_r \times n_t}$ represent the channel gain matrix of link l, where n_t and n_r are the numbers of transmitting and receiving antennas, respectively. The channel is assumed to be block faded, i.e., the channel remains static within each block and varies independently from block to block. The received complex base-band signal vector for MIMO link l is given by $\mathbf{y}_l = \sqrt{\rho_l}\mathbf{H}_l\mathbf{x}_l + \mathbf{n}_l$, where $\mathbf{y}_l$ and $\mathbf{x}_l$ represent the received and transmitted signal vectors, $\mathbf{n}_l$ is complex additive white Gaussian noise vector with zero mean and unit variance, ρ_l denotes signal-to-noise ratio (SNR).

Let matrix $\mathbf{Q}_l$ represent the covariance matrix of input symbol vector $\mathbf{x}_l$ at link l, i.e., $\mathbf{Q}_l = \mathbb{E}\left\{\mathbf{x}_l \cdot \mathbf{x}_l^{\dagger}\right\}$. We normalize $\mathbf{x}_l$ such that $\mathbf{Q}_l \succeq 0$ and $\mathrm{Tr}(\mathbf{Q}_l) \leq 1$. It is well known that the optimal input covariance matrix for link l is achieved by "water-filling" its power along the eigen-direction of the effective channel [2]. Let $P_{\mathbf{H}_l} \leq 1$ represent the total power of link l when the channel realization is $\mathbf{H}_l$ and denote the water-filling solution under total power $P_{\mathbf{H}_l}$ as $\mathbf{Q}^*(P_{\mathbf{H}_l})$. Denote the capacity of link l as C_l. The ergodic mutual information can be computed by [3]

$$C_l = \mathbb{E}_{\mathbf{H}_l}\left[\log_2\left|\mathbf{I} + \rho_l\mathbf{H}_l\mathbf{Q}^*(P_{\mathbf{H}_l})\mathbf{H}_l^{\dagger}\right|\right], \quad (1)$$

where $\mathbb{E}_{\mathbf{H}_l}[\cdot]$ represents the expectation taken over the fading distribution of $\mathbf{H}_l$. During a unit time interval, since the power sum (averaged over fading distribution) of all outgoing links at each node cannot be larger than the node's maximum transmission power, we have, for all n,

$$\sum_{l \in \mathcal{O}(n)} \mathbb{E}_{\mathbf{H}_l}\left[P_{\mathbf{H}_l}\right] \leq 1.$$

We assume there are S communication sessions in the network. We assume that the source of the s-th session emits one flow with rate f_s, using a pre-assigned route to its destination. We can describe the routing paths compactly by using a routing matrix $\mathbf{R} \in \mathbb{R}^{L \times S}$, defined as

$$R_{ls} = \begin{cases} 1 & \text{if session } s \text{ passes through link } l \\ 0 & \text{otherwise.} \end{cases}$$

Since the total traffic on a link cannot exceed its link capacity, we have $\sum_{s=1}^{S} R_{ls}f_s \leq C_l$, for all $l \in \mathcal{L}$.

For the problem under study, we aim to perform joint rate and power control such that some network utility is maximized. We adopt weighted proportional fairness as the utility function [11], i.e., $w_s \ln(f_s)$ for flow rate f_s, where w_s is the given weight for session s. Coupling the network layer and the physical layer models, we have the problem formulation as follows:

$$\begin{array}{lll} \text{Maximize} & \sum_{s=1}^{S} w_s \ln(f_s) & \\ \text{subject to} & \sum_{s=1}^{S} R_{ls}f_s \leq C_l & \forall\, l \\ & C_l = \mathbb{E}_{\mathbf{H}_l}\left[\log_2\left|\mathbf{I} + \rho_l\mathbf{H}_l\mathbf{Q}^*(P_{\mathbf{H}_l})\mathbf{H}_l^{\dagger}\right|\right] & \forall\, l \\ & \sum_{l \in \mathcal{O}(n)} \mathbb{E}_{\mathbf{H}_l}\left[P_{\mathbf{H}_l}\right] \leq 1 & \forall\, n \\ & P_{\mathbf{H}_l} \geq 0 \quad \forall\, \mathbf{H}_l, \quad m_s \leq f_s \leq M_s \quad \forall\, s & \end{array} \quad (2)$$

III. Solution under Perfect CSI

When perfect CSI is available, (2) can be solved rather cleanly. This is because the formulation in (2) is convex, which can be solved efficiently. We employ dual decomposition (DD) technique, which solves (2) in its dual domain rather than attacking it directly. This is because, for a convex problem, the optimal dual objective value is equal to that of the original primal problem due to strong duality [12] (i.e., zero duality gap). The main purpose of solving the problem in the dual domain is due to the availability of nice decomposable structure so that each subproblem can be solved with reasonably low complexity.

In this paper, we focus on PD to solve (2) in its dual domain because PD generally has a faster convergence rate compared with other algorithms [13]. Denote vector $\mathbf{u} \triangleq [u_1\, u_2\, \ldots\, u_L]^T$ the collection of dual variables associated with link capacity coupling constraints and $\mathbf{v} \triangleq [v_1\, v_2\, \ldots\, v_N]^T$ the collection of dual variables associated with power constraints. Then, the Lagrangian dual function can be written as

$$\Theta(\mathbf{u}, \mathbf{v}) \triangleq \max_{\mathbf{f}, \mathbf{P}} \{L(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v}) | (\mathbf{f}, \mathbf{P}) \in \Psi\}, \tag{3}$$

where $\mathbf{f}$ and $\mathbf{P}$ denotes the collections of all flow rates and power variables, respectively. In (3), the Lagrangian $L(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})$ can be computed as

$$\begin{aligned} L(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v}) &= \sum_{s=1}^{S} w_s \ln f_s + \sum_{l=1}^{L} u_l \times \\ &\left[\mathbb{E}_{\mathbf{H}_l}\left[\log_2\left|\mathbf{I} + \rho_l \mathbf{H}_l \mathbf{Q}^*(P_{\mathbf{H}_l})\mathbf{H}_l^\dagger\right|\right] - \sum_{s=1}^{S} R_{ls} f_s\right] \\ &+ \sum_{n=1}^{N} v_n \left[1 - \sum_{l \in \mathcal{O}(n)} \mathbb{E}_{\mathbf{H}_l}\left[P_{\mathbf{H}_l}\right]\right] \end{aligned}$$

and $\Psi \triangleq \{(\mathbf{f}, \mathbf{P}) | P_{\mathbf{H}_l} \geq 0\ \forall \mathbf{H}_l,\ m_s \leq f_s \leq M_s\ \forall s\}$. Then, the Lagrangian dual problem can be written as:

$$\text{Minimize } \Theta(\mathbf{u}, \mathbf{v}) \quad \text{subject to } \mathbf{u} \geq \mathbf{0},\ \mathbf{v} \geq \mathbf{0}. \tag{4}$$

After re-arranging terms, the Lagrangian function can be rewritten as

$$\begin{aligned} L(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v}) &= \underbrace{\sum_{s=1}^{S}\left[w_s \ln(f_s) - \sum_{l=1}^{L} R_{ls} u_l f_s\right]}_{\text{Network Layer Subproblem}} + \sum_{n=1}^{N} v_n \\ &+ \underbrace{\sum_{l=1}^{L} \mathbb{E}_{\mathbf{H}_l}\left[u_l \log_2\left|\mathbf{I} + \rho_l \mathbf{H}_l \mathbf{Q}^*(P_{\mathbf{H}_l})\mathbf{H}_l^\dagger\right| - v_{tx(l)} P_{\mathbf{H}_l}\right]}_{\text{Physical Layer Subproblem}}. \end{aligned}$$

As a result, the Lagrangian dual function can be written as $\Theta(\mathbf{u}, \mathbf{v}) = \max_{\mathbf{f}, \mathbf{P}} L(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v}) = \Theta_{\text{net}}(\mathbf{u}) + \Theta_{\text{phy}}(\mathbf{u}, \mathbf{v}) + \langle \mathbf{1}, \mathbf{v} \rangle$, where $\Theta_{\text{net}}(\mathbf{u}) \triangleq \{\max \sum_{s=1}^{S} w_s \ln(f_s) - \sum_{l=1}^{L} \sum_{s=1}^{S} R_{ls} u_l f_s | m_s \leq f_s \leq M_s, \forall s\}$; and $\Theta_{\text{phy}}(\mathbf{u}, \mathbf{v}) \triangleq \{\max \sum_{l=1}^{L} \mathbb{E}_{\mathbf{H}_l}\left[u_l \log_2\left|\mathbf{I} + \rho_l \bar{\mathbf{H}}_l \mathbf{Q}^*_{P_{\mathbf{H}_l}} \bar{\mathbf{H}}_l^\dagger\right| - v_{tx(l)} P_{\mathbf{H}_l}\right] | P_{\mathbf{H}_l} \geq 0, \forall \mathbf{H}_l\}$. For the physical layer subproblem, we have the following lemma.

Lemma 1. *In the k-th iteration, the physical layer subproblem Θ_{phy} can be solved in closed form as*

$$P_{\mathbf{H}_l}^{(k)} = \left[\frac{u_l^{(k)}}{v_{tx(l)}^{(k)}} D - \sum_{d=1}^{D} \sigma_d^{-1}(\rho_l \mathbf{H}_l^{(k)\dagger} \mathbf{H}_l^{(k)})\right]_+, \tag{5}$$

where $[\cdot]_+ = \max(\cdot, 0)$, D is the rank of $\mathbf{H}_l^{(k)\dagger}\mathbf{H}_l^{(k)}$, $\sigma_d^{-1}(\rho_l \mathbf{H}_l^{(k)\dagger}\mathbf{H}_l^{(k)})$ represents the inverse of the d-th non-zero eigenvalue of $(\rho_l \mathbf{H}_l^{(k)\dagger}\mathbf{H}_l^{(k)})$.

The basic idea of the proof is based on the argument that power should be water-filled according to the eigen-direction of $(\rho_l \mathbf{H}_l^{(k)\dagger}\mathbf{H}_l^{(k)})$ with water level $(u_l^{(k)}/v_{tx(l)}^{(k)})D$. Due to space limitation, we omit the details of the proof in this paper.

When using PD to solve (4), we update primal variables $\mathbf{f}$ and dual variables $\mathbf{u}$ and $\mathbf{v}$ on the same time-scale. We summarize the PD algorithm in Algorithm 1. In Algorithm 1, M_u and m_v are appropriate upper and lower bounds for dual variables $\mathbf{u}$ and $\mathbf{v}$, respectively. In general, as long as step size λ_k satisfies $\lambda_k \to 0$ as $k \to \infty$, $\sum_{k=1}^{\infty} \lambda_k \to \infty$, and $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$, PD converges to a dual optimal solution [12].

Algorithm 1 The Primal-Dual Algorithm

1. Choose the initial starting points $\mathbf{f}^{(0)}$, $\mathbf{u}^{(0)}$, $\mathbf{v}^{(0)}$. Compute $\mathbf{P}^{(0)}$ according to (5)
2. In the k^{th} iteration, choose an appropriate step size λ_k. Update primal variables $\mathbf{f}$ as:

$$f_s^{(k+1)} = \left[f_s^{(k)} + \lambda_k\left(\frac{w_s}{f_s^k} - \sum_{l=1}^{L} R_{ls} u_l^{(k)}\right)\right]_{m_s}^{M_s}, \ \forall s. \tag{6}$$

Update dual variables $\mathbf{u}$ as:

$$\begin{aligned} u_l^{(k+1)} &= \left[u_l^{(k)} - \lambda_k\left(\mathbb{E}_{\mathbf{H}_l}\left[\log_2\left|\mathbf{I} + \mathbf{H}_l^{(k)}\mathbf{Q}^*(P_{\mathbf{H}_l}^{(k)})\mathbf{H}_l^{(k)\dagger}\right|\right]\right.\right. \\ &\left.\left. - \sum_{s=1}^{S} R_{ls} f_s^{(k)}\right)\right]_0^{M_u}, \ \forall l. \end{aligned} \tag{7}$$

Update dual variables $\mathbf{v}$ as:

$$v_n^{(k+1)} = \left[v_n^{(k)} - \lambda_k\left(1 - \sum_{l \in \mathcal{O}(n)} \mathbb{E}_{\mathbf{H}_l}\left[P_{\mathbf{H}_l}^{(k)}\right]\right)\right]_{m_v}^{\infty}, \ \forall n. \tag{8}$$

3. Compute $\mathbf{P}^{(k+1)}$ according to (5).
4. If $\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|^2 + \|\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}\|^2 + \|\mathbf{v}^{(k+1)} - \mathbf{v}^{(k)}\|^2 < \epsilon$, stop; otherwise, let $k \leftarrow k+1$ and go to 2.

IV. Solution under Imperfect CSI

It can be observed that the updates on dual variables $\mathbf{u}$ and $\mathbf{v}$ in Algorithm 1 (Eqs. (7) and (8)) require perfect CSI as well as channel distribution information (CDI). However, as discussed earlier, obtaining full CSI and CDI may not be feasible in practice as the amount of overhead is prohibitively large. So the result by Algorithm 1 can only be used as an ideal reference.

In this section, we strive to answer the following two questions: *Is it possible to develop a solution to (2) based on imperfect CSI? Can the performance gap of the proposed solution (under imperfect CSI) to the ideal solution (under perfect CSI) be characterized?* We show that answers to both questions are yes. In this section, we show the answer to the first question and in the next section, we show the answer to the second. We propose a revised PD algorithm based on Algorithm 1 using imperfect CSI. Our key technique in this revised algorithm is stochastic approximation (SA). We refer to this approach as PD-SA algorithm.

The purpose of using SA is to avoid computing expectations over fading distributions, thus eliminating the requirement of perfect CSI and CDI. The main idea is that, instead of evaluating the subgradients by averaging over all possible channel realizations, we only use an imperfect observation of the current channel realization to compute $\mathbf{u}$ and $\mathbf{v}$ in each iteration. Although only one channel realization is used in each iteration, it can be shown that the decreasing step sizes would provide an implicit averaging over the observed channel realizations across iterations. Therefore, our proposed algorithm can adapt to unknown fading distributions on the fly. More precisely, we revise the dual updates in (7) and (8) as follows.

$$\hat{u}_l^{(k+1)} = \Big[\hat{u}_l^{(k)} - \lambda_k \big(\log_2 \Big|\mathbf{I} + \rho_l \mathbf{H}_l^{(k)} \mathbf{Q}(\hat{P}_{\mathbf{H}_l}^{(k)}) \mathbf{H}_l^{(k)\dagger}\Big| - \textstyle\sum_{s=1}^{S} R_{ls} f_s^{(k)} \big)\Big]_0^{M_u}, \ \forall l, \tag{9}$$

$$\hat{v}_n^{(k+1)} = \Big[\hat{v}_n^{(k)} - \lambda_k \big(1 - \sum_{l \in \mathcal{O}(n)} \hat{P}^{(k)} \big)\Big]_{m_v}^{\infty}, \ \forall n, \tag{10}$$

where $\mathbf{Q}(\hat{P}_{\mathbf{H}_l}^{(k)})$ and $\hat{P}_{\mathbf{H}_l}^{(k)}$ denotes the suboptimal covariance matrix and power level due to imperfect CSI. From (9) and (10), we find that the differences to the original PD are: 1) $\mathbf{Q}^*(P_{\mathbf{H}_l}^{(k)})$ and $P_{\mathbf{H}_l}^{(k)}$ are replaced by $\mathbf{Q}(\hat{P}_{\mathbf{H}_l}^{(k)})$ and $\hat{P}_{\mathbf{H}_l}^{(k)}$; and 2) In each iteration, there is no expectation computation over fading distributions in PD-SA. Although the changes in (9) and (10) appear minor, they have profound impacts on the algorithm's behavior. First, the running time of each iteration becomes much shorter and since no expectation computation is needed. Second, whether or not such a SA is stable and how far it deviates from the original optimal solution remain unclear and will be the main topic in the next section.

V. Performance Gap Analysis

Due to the existence of CSI error, a performance gap between the solutions under imperfect and perfect CSI is inevitable. In this section, we study this performance gap mathematically.

A. A Closer Look at Stochastic Approximation

Before we study the stability of PD-SA, we need to take a closer look at our proposed SA. Recall that the subgradients $L_{u_l}^{(k)}$ and $L_{v_n}^{(k)}$ of the Lagrangian function (with respect to u_l and v_n, respectively) under perfect CSI during the k-th iteration are as follows:

$$L_{u_l}^{(k)} = \mathbb{E}_{\mathbf{H}_l}\left[\log_2 \left|\mathbf{I} + \mathbf{H}_l^{(k)} \mathbf{Q}^*(P_{\mathbf{H}_l}^{(k)}) \mathbf{H}_l^{(k)\dagger}\right|\right] - \sum_{s=1}^{S} R_{ls} f_s^{(k)},$$

$$L_{v_n}^{(k)} = 1 - \sum_{l \in \mathcal{O}(n)} \mathbb{E}_{\mathbf{H}_l}\left[P_{\mathbf{H}_l}^{(k)}\right] = \sum_{l \in \mathcal{O}(n)} L_{v_{n,l}}^{(k)},$$

where $L_{v_{n,l}}^{(k)} \triangleq \frac{1}{|\mathcal{O}(n)|} - \mathbb{E}_{\mathbf{H}_l}\left[P_{\mathbf{H}_l}^{(k)}\right]$. Note that in our proposed PD-SA, the subgradients are changed to the following SA:

$$\hat{L}_{u_l}^{(k)} = \log_2 \left|\mathbf{I} + \mathbf{H}_l^{(k)} \mathbf{Q}(\hat{P}_{\mathbf{H}_l}^{(k)}) \mathbf{H}_l^{(k)\dagger}\right| - \sum_{s=1}^{S} R_{ls} f_s^{(k)},$$

$$\hat{L}_{v_n}^{(k)} = 1 - \sum_{l \in \mathcal{O}(n)} \hat{P}_{\mathbf{H}_l}^{(k)} \triangleq \sum_{l \in \mathcal{O}(n)} \hat{L}_{v_{n,l}}^{(k)},$$

where $\hat{L}_{v_{n,l}}^{(k)} \triangleq \frac{1}{|\mathcal{O}(n)|} - \hat{P}_{\mathbf{H}_l}^{(k)}$. Observe that $\hat{L}_{u_l}^{(k)}$ can be decomposed as follows:

$$\hat{L}_{u_l}^{(k)} = L_{u_l}^{(k)} + \mu_l^{(k)} + \zeta_l^{(k)}, \tag{11}$$

where $\mu_l^{(k)}$ is the error of biased estimation of L_{u_l}, defined as

$$\begin{aligned}\mu_l^{(k)} &\triangleq \mathbb{E}_{\mathbf{H}_l}\left[\hat{L}_{u_l}^{(k)}\right] - L_{u_l}^{(k)} = \\ &\mathbb{E}_{\mathbf{H}_l}\left[\log_2 \left|\mathbf{I} + \mathbf{H}_l^{(k)} \mathbf{Q}(\hat{P}_{\mathbf{H}_l}^{(k)}) \mathbf{H}_l^{(k)\dagger}\right| - \log_2 \left|\mathbf{I} + \mathbf{H}_l^{(k)} \mathbf{Q}^*(P_{\mathbf{H}_l}^{(k)}) \mathbf{H}_l^{(k)\dagger}\right|\right],\end{aligned} \tag{12}$$

and $\zeta_l^{(k)}$ is defined as

$$\begin{aligned}\zeta_l^{(k)} &\triangleq \hat{L}_{u_l}^{(k)} - \mathbb{E}_{\mathbf{H}_l}\left[\hat{L}_{u_l}^{(k)}\right] = \\ &= \log_2 \left|\mathbf{I} + \mathbf{H}_l^{(k)} \mathbf{Q}(\hat{P}_{\mathbf{H}_l}^{(k)}) \mathbf{H}_l^{(k)\dagger}\right| - \mathbb{E}_{\mathbf{H}_l}\left[\log_2 \left|\mathbf{I} + \mathbf{H}_l^{(k)} \mathbf{Q}(\hat{P}_{\mathbf{H}_l}^{(k)}) \mathbf{H}_l^{(k)\dagger}\right|\right].\end{aligned} \tag{13}$$

Similarly, the stochastic subgradient approximation $\hat{L}_{v_n}^{(k)}$ can be decomposed as follows:

$$\hat{L}_{v_n}^{(k)} = \sum_{l \in \mathcal{O}(n)} \left[L_{v_{n,l}}^{(k)} + \nu_{n,l}^{(k)} + \xi_{n,l}^{(k)}\right], \tag{14}$$

where $\nu_{n,l}^{(k)}$ is the error of biased estimation of $L_{v_{n,l}}^{(k)}$, defined as

$$\nu_{n,l}^{(k)} \triangleq \mathbb{E}_{\mathbf{H}_l}\left[\hat{L}_{v_{n,l}}^{(k)}\right] - L_{v_{n,l}}^{(k)} = \mathbb{E}_{\mathbf{H}_l}\left[P_{\mathbf{H}_l}^{(k)} - \hat{P}_{\mathbf{H}_l}^{(k)}\right], \tag{15}$$

and $\xi_{n,l}^{(k)}$ is defined by

$$\begin{aligned}\xi_{n,l}^{(k)} &\triangleq \hat{L}_{v_{n,l}}^{(k)} - \mathbb{E}_{\mathbf{H}_l}\left[\hat{L}_{v_{n,l}}^{(k)}\right] \\ &= \left(\frac{1}{|\mathcal{O}(n)|} - \hat{P}_{\mathbf{H}_l}^{(k)}\right) - \mathbb{E}_{\mathbf{H}_l}\left[\frac{1}{|\mathcal{O}(n)|} - \hat{P}_{\mathbf{H}_l}^{(k)}\right] \\ &= \mathbb{E}_{\mathbf{H}_l}\left[\hat{P}_{\mathbf{H}_l}^{(k)}\right] - \hat{P}_{\mathbf{H}_l}^{(k)}.\end{aligned} \tag{16}$$

A closer look at the biased estimation error $\mu_l^{(k)}$ in (12) reveals that $\mu_l^{(k)}$ is precisely the *ergodic capacity loss* on link l due to imperfect CSI. This ergodic capacity loss will not diminish as $k \to \infty$. Thus, $\hat{L}_{u_l}^{(k)}$ can be viewed as a biased estimator of the true subgradient $L_{u_l}^{(k)}$. Likewise, we can also see that the biased estimation error $\nu_{n,l}^{(k)}$ in (15) can be interpreted as the *ergodic power allocation error* of outgoing link l at node n due to imperfect CSI.

B. Stability Study

We now formally define the stability of a stochastic approximation algorithm.

Definition 1. *Denote the iterates generated by an approximation algorithm $\mathcal{A}$ as x_k, $k = 1, 2, \ldots$. Suppose that at iteration k_0, the current solution x_{k_0} is in a neighborhood $\mathbb{N}$ around the original optimal solution x^*. Let $p_{k_0}(n)$ denote the probability of the first reentry to $\mathbb{N}$ at iterate $k_0 + n$, i.e., $p_{k_0}(n) = \Pr(x_{k_0+n} \in \mathbb{N}, x_{k_0+n-1} \notin \mathbb{N}, \ldots, x_{k_0+1} \notin$*

$\mathbb{N}|x_{k_0} \in \mathbb{N})$. *We say $\mathcal{A}$ is stable if $\sum_{n=1}^{\infty} p_{k_0}(n) \to 1$. That is, the iterates generated by $\mathcal{A}$ is recurrent to $\mathbb{N}$.*

The stability of PD-SA can be shown by the following steps. First, we will construct a Lyapunov function and derive some appropriate upper bounds for one-step dynamics of the stochastic iterates. Second, we will identify some neighborhood and establish the stability with respect to this neighborhood.

Let $(\mathbf{f}^*, \mathbf{P}^*, \mathbf{u}^*, \mathbf{v}^*)$ be an optimal point under PD. Now, we define a Lyapunov function as follows.

$$V(\mathbf{f},\mathbf{P},\mathbf{u},\mathbf{v}) \triangleq \|\mathbf{f}-\mathbf{f}^*\|^2+\|\mathbf{P}-\mathbf{P}^*\|^2+\|\mathbf{u}-\mathbf{u}^*\|^2+\|\mathbf{v}-\mathbf{v}^*\|^2,$$

For the stochastic iterates $\{(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{P}}^{(k)}, \hat{\mathbf{u}}^{(k)}, \hat{\mathbf{v}}^{(k)}) : k = 1, 2, \ldots\}$ generated by PD-SA, we have the following lemma.

Lemma 2. *The average distance between $\hat{\mathbf{u}}^{(k+1)}$ and $\mathbf{u}^*$ in the $(k+1)$-th iteration is bounded as follows:*

$$\mathbb{E}_{\mathbf{H}}\left[\|\hat{\mathbf{u}}^{(k+1)} - \mathbf{u}^*\|^2\right] \le \|\hat{\mathbf{u}}^{(k)} - \mathbf{u}^*\|^2 - 2\lambda_k(\hat{\mathbf{u}}^{(k)} - \mathbf{u}^*)^T \times L_{\mathbf{u}}^{(k)} - 2\lambda_k(\hat{\mathbf{u}}^{(k)} - \mathbf{u}^*)^T \mathbb{E}_{\mathbf{H}}\left[\boldsymbol{\mu}^{(k)}\right] + O(\lambda_k^2).$$

Proof: Substituting (11) into (9), we have

$$\hat{u}_l^{(k+1)} = \hat{u}_l^{(k)} - \lambda_k\left[L_{u_l}^{(k)} + \mu_l^{(k)} + \zeta_l^{(k)}\right] + \lambda_k Z_l^{(k)}, \quad (17)$$

where $Z_l^{(k)}$ is a correction term that projects the iterate $\hat{u}_l^{(k+1)}$ back to the non-negative orthant. Thus, in vector form, we have

$$\|\hat{\mathbf{u}}^{(k+1)} - \mathbf{u}^*\|^2 \le \|\hat{\mathbf{u}}^{(k)} - \mathbf{u}^*\|^2 - 2\lambda_k(\hat{\mathbf{u}}^{(k)} - \mathbf{u}^*)^T \times \left[L_{\mathbf{u}}^{(k)} + \boldsymbol{\mu}^{(k)} + \boldsymbol{\zeta}^{(k)}\right] + \lambda_k^2\|L_{\mathbf{u}}^{(k)} + \boldsymbol{\mu}^{(k)} + \boldsymbol{\zeta}^{(k)}\|^2, \quad (18)$$

where the inequality follows because the projection term $\lambda_k Z_l^{(k)}$ is non-expansive. Since the Lagrangian function $L(\mathbf{f},\mathbf{P},\mathbf{u},\mathbf{v})$ is twice-differentiable, $L_{\mathbf{u}}^{(k)}$ is bounded. Also, $\|\mathbf{u}^{(k)} - \mathbf{u}^*\|$ is bounded and $\mathbb{E}_{\mathbf{H}_l}\left[\zeta_l^{(k)}\right] = 0$. From the iteration update in (9), $|\boldsymbol{\zeta}^{(k)}|$ must be bounded. As a result, $\|\hat{L}_{\mathbf{u}}^{(k)}\|$ is bounded. Thus, taking expectation of both sides of (18) yields the result stated in the lemma. ■

Following the same token, we can also derive the upper bounds for $\mathbb{E}_{\mathbf{H}}\left[\|\hat{\mathbf{v}}^{(k+1)} - \mathbf{v}^*\|^2\right]$ and $\mathbb{E}_{\mathbf{H}}\left[\|\hat{\mathbf{f}}^{(k+1)} - \mathbf{f}^*\|^2\right]$. We omit the derivations in here for brevity.

Now, we define a family of neighborhoods $\mathbb{B}_\phi$ parameterized by a variable ϕ as follows:

$$\mathbb{B}_\phi \triangleq \left\{ (\mathbf{f},\mathbf{P},\mathbf{u},\mathbf{v}) \;\middle|\; \begin{array}{l} \bar{\mu}_l \ge \phi|L_{u_l}|, \text{ for all } l \\ \bar{\nu}_n \ge \phi|L_{v_n}|, \text{ for all } n \\ \phi \ge 0. \end{array} \right\}, \quad (19)$$

where $\bar{\mu}_l \triangleq \limsup_{k\to\infty} \mu_l^{(k)}$ and $\bar{\nu}_n \triangleq \limsup_{k\to\infty} \nu_n^{(k)}$, respectively. In essence, $\mathbb{B}_\phi$ defines a neighborhood around the optimal point $(\mathbf{f}^*, \mathbf{P}^*, \mathbf{u}^*, \mathbf{v}^*)$. This is because $L_{u_l}(\mathbf{f},\mathbf{P},\mathbf{u},\mathbf{v})$ and $L_{v_n}(\mathbf{f},\mathbf{P},\mathbf{u},\mathbf{v})$ are continuous at $(\mathbf{f}^*, \mathbf{P}^*, \mathbf{u}^*, \mathbf{v}^*)$ and by KKT conditions, we have $L_{u_l}(\mathbf{f}^*, \mathbf{P}^*, \mathbf{u}^*, \mathbf{v}^*) = 0$ for all l and $L_{v_n}(\mathbf{f}^*, \mathbf{P}^*, \mathbf{u}^*, \mathbf{v}^*) = 0$ for all n, respectively. The constraints in (19) means that the points in $\mathbb{B}_\phi$ satisfy $|L_{u_l}(\mathbf{f},\mathbf{P},\mathbf{u},\mathbf{v})| \le \frac{\bar{\mu}_l}{\phi}$ and $|L_{v_n}(\mathbf{f},\mathbf{P},\mathbf{u},\mathbf{v})| \le \frac{\bar{\nu}_n}{\phi}$, which clearly includes $(\mathbf{f}^*, \mathbf{P}^*, \mathbf{u}^*, \mathbf{v}^*)$.

The size of $\mathbb{B}_\phi$ depends on the value of parameter ϕ, $\bar{\mu}_l$, and $\bar{\nu}_n$. For a given LFB scheme (with a certain number of feedback bits), $\bar{\mu}_l$ and $\bar{\nu}_n$ can be determined. In this case, the smaller the value of ϕ is, the larger the size of $\mathbb{B}_\phi$ is. In the extreme cases, $\mathbb{B}_\phi$ becomes the whole space of $(\mathbf{f},\mathbf{P},\mathbf{u},\mathbf{v})$ when $\phi = 0$ and $\mathbb{B}_\phi$ collapses to the optimal solution $(\mathbf{f}^*, \mathbf{P}^*, \mathbf{u}^*, \mathbf{v}^*)$ when $\phi \to \infty$. When we characterize the recurrence region of PD-SA later, we will specify what condition ϕ should satisfy.

On the other hand, for a fixed ϕ, the larger the values of $\bar{\mu}_l$ and $\bar{\nu}_n$ are (i.e., due to poor LFB and/or fewer number of feedback bits), the larger the size of $\mathbb{B}_\phi$. For convenience, let $\mathbf{a}$ denote the collection of all dual variables $\mathbf{u}$ and $\mathbf{v}$. Then, for $\mathbb{B}_\phi$ and the stochastic iterates $\{(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{P}}^{(k)}, \hat{\mathbf{a}}^{(k)}) : k = 1, 2, \ldots\}$, we have the following lemma.

Lemma 3. *If the current iteration $(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{P}}^{(k)}, \hat{\mathbf{a}}^{(k)})$ is not in $\mathbb{B}_\phi$, then the following inequality holds:*

$$\mathbb{E}_{\mathbf{H}}\left[\|\hat{\mathbf{f}}^{(k+1)} - \mathbf{f}^*\|^2 + \|\hat{\mathbf{a}}^{(k+1)} - \mathbf{a}^*\|^2\right] - \left(\|\mathbf{f}^{(k)} - \mathbf{f}^*\|^2 + \|\hat{\mathbf{a}}^{(k)} - \mathbf{a}^*\|^2\right) \le 2\lambda_k[2 + (\text{sgn}[L_{\mathbf{f}}^{(k)}] + \text{sgn}[L_{\mathbf{a}}^{(k)}])\phi]J^{(k)} + O(\lambda_k^2),$$

where $\text{sgn}[x] = 1$ if $x \ge 0$ and -1 otherwise; $J^{(k)}$ is defined as $J^{(k)} \triangleq (\hat{\mathbf{f}}^{(k)} - \mathbf{f}^*)^T L_{\mathbf{f}}^{(k)} - (\hat{\mathbf{a}}^{(k)} - \mathbf{a}^*)^T L_{\mathbf{a}}^{(k)}$.

Proof: If $\hat{\mathbf{u}}^{(k)}$ is not in $\mathbb{B}_\phi$, then we have $\limsup_k |\boldsymbol{\mu}^{(k)}| \le \phi|L_{\mathbf{u}}^{(k)}|$. From Lemma 2, it follows that

$$\mathbb{E}_{\mathbf{H}}\left[\|\hat{\mathbf{u}}^{(k+1)} - \mathbf{u}^*\|^2\right] \le \|\hat{\mathbf{u}}^{(k)} - \mathbf{u}^*\|^2 - 2\lambda_k(\hat{\mathbf{u}}^{(k)} - \mathbf{u}^*)^T \times \left[L_{\mathbf{u}}^{(k)} + \phi|L_{\mathbf{u}}^{(k)}|\right] + O(\lambda_k^2) = \|\hat{\mathbf{u}}^{(k)} - \mathbf{u}^*\|^2 - 2\lambda_k \times [1 + \text{sgn}[L_{\mathbf{u}}^{(k)}]\phi](\hat{\mathbf{u}}^{(k)} - \mathbf{u}^*)^T L_{\mathbf{u}}^{(k)} + O(\lambda_k^2). \quad (20)$$

Likewise, we can obtain

$$\mathbb{E}_{\mathbf{H}}\left[\|\hat{\mathbf{v}}^{(k+1)} - \mathbf{v}^*\|^2\right] \le \|\hat{\mathbf{v}}^{(k)} - \mathbf{v}^*\|^2 - 2\lambda_k \times [1 + \text{sgn}[L_{\mathbf{v}}^{(k)}]\phi](\hat{\mathbf{v}}^{(k)} - \mathbf{v}^*)^T L_{\mathbf{v}}^{(k)} + O(\lambda_k^2), \quad (21)$$

$$\mathbb{E}_{\mathbf{H}}\left[\|\hat{\mathbf{f}}^{(k+1)} - \mathbf{f}^*\|^2\right] \le \|\hat{\mathbf{f}}^{(k)} - \mathbf{f}^*\|^2 + 2\lambda_k \times [1 + \text{sgn}[L_{\mathbf{f}}^{(k)}]\phi](\hat{\mathbf{f}}^{(k)} - \mathbf{f}^*)^T L_{\mathbf{f}}^{(k)} + O(\lambda_k^2). \quad (22)$$

Combining (20), (21), and (22) and using the notation $\mathbf{a}$ and the definition of $J^{(k)}$, the proof is complete. ■

Now, we are ready to prove the stability result, which is stated in the following theorem.

Theorem 1. *If the step size selection satisfies $\lambda_k \to 0$, $\sum_{k=1}^{\infty} \lambda_k \to \infty$, and $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$, then the iterates $\{(\mathbf{f}^{(k)}, \mathbf{P}^{(k)}, \mathbf{u}^{(k)}, \mathbf{v}^{(k)}) : k = 1, 2, \ldots\}$ generated by PD-SA returns with probability one to a neighborhood $\mathbb{B}_\phi$ with $0 \le \phi < 1$.*

Proof: Since $0 \leq \phi < 1$ and $\lambda_k > 0$, we have $2\lambda_k[2 + (\text{sgn}[L_{\mathbf{a}}^{(k)}] + \text{sgn}[L_{\mathbf{f}}^{(k)}])\phi] > 0$. As a result, by Lemma 3 and according to [14, Theorem 5.4], the proof of this theorem boils down to showing that $J^{(k)} < 0$ for $(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{P}}^{(k)}, \hat{\mathbf{a}}^{(k)}) \notin \mathbb{B}_\phi$. Note that the Lagrangian function $L(\mathbf{f}, \mathbf{P}, \mathbf{a})$ is strictly concave in $\mathbf{f}$ and convex in $\mathbf{a}$. We have

$$L(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{P}}^{(k)}, \hat{\mathbf{a}}^{(k)}) - L(\mathbf{f}^*, \hat{\mathbf{P}}^{(k)}, \hat{\mathbf{a}}^{(k)}) > (\hat{\mathbf{f}}^{(k)} - \mathbf{f}^*)^T L_{\mathbf{f}}^{(k)},$$
$$L(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{P}}^{(k)}, \hat{\mathbf{a}}^{(k)}) - L(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{P}}^{(k)}, \mathbf{a}^*) \leq (\hat{\mathbf{a}}^{(k)} - \mathbf{a}^*)^T L_{\mathbf{a}}^{(k)},$$

Combining these two inequalities, we have

$$\begin{aligned} J^{(k)} &< L(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{P}}^{(k)}, \mathbf{a}^*) - L(\mathbf{f}^*, \hat{\mathbf{P}}^{(k)}, \hat{\mathbf{a}}^{(k)}) \\ &= \left[L(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{P}}^{(k)}, \mathbf{a}^*) - L(\mathbf{f}^*, \mathbf{P}^*, \mathbf{a}^*)\right] \\ &+ \left[L(\mathbf{f}^*, \mathbf{P}^*, \mathbf{a}^*) - L(\mathbf{f}^*, \mathbf{P}^*, \hat{\mathbf{a}}^{(k)})\right] \\ &+ \left[L(\mathbf{f}^*, \mathbf{P}^*, \hat{\mathbf{a}}^{(k)}) - L(\mathbf{f}^*, \hat{\mathbf{P}}^{(k)}, \hat{\mathbf{a}}^{(k)})\right]. \end{aligned}$$

Since $(\mathbf{f}^*, \mathbf{P}^*, \mathbf{a}^*)$ is an optimal solution, from saddle point optimality condition, we have that

$$L(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{P}}^{(k)}, \mathbf{a}^*) \leq L(\mathbf{f}^*, \mathbf{P}^*, \mathbf{a}^*) \leq L(\mathbf{f}^*, \mathbf{P}^*, \hat{\mathbf{a}}^{(k)}),$$

which indicates that $\left[L(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{P}}^{(k)}, \mathbf{a}^*) - L(\mathbf{f}^*, \mathbf{P}^*, \mathbf{a}^*)\right]$ and $\left[L(\mathbf{f}^*, \mathbf{P}^*, \mathbf{a}^*) - L(\mathbf{f}^*, \mathbf{P}^*, \hat{\mathbf{a}}^{(k)})\right]$ are non-positive. On the other hand, when $L(\cdot, \cdot, \cdot)$ is fixed at $\mathbf{f}^*$ and $\hat{\mathbf{a}}^{(k)}$, we have that $\hat{\mathbf{P}}^{(k)}$ is the unique maximizer from (5), which means $\left[L(\mathbf{f}^*, \mathbf{P}^*, \hat{\mathbf{a}}^{(k)}) - L(\mathbf{f}^*, \hat{\mathbf{P}}^{(k)}, \hat{\mathbf{a}}^{(k)})\right]$ is also non-positive. Thus, we can conclude that $J^{(k)} < 0$. ∎

The proof of Theorem 2 implies that the stochastic iterates may not return to $\mathbb{B}_{\phi \geq 1}$ with probability one. Thus, the recurrence region of PD-SA includes the whole space of $(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})$ minus the union of all $\mathbb{B}_{\phi \geq 1}$'s. Since the iterates is recurrent with probability one to $\mathbb{B}_{\phi = 1-\delta}$ where $\delta > 0$ is arbitrarily small, the iterates path of PD-SA will be arbitrarily close to the boundary of $\mathbb{B}_{\phi=1}$ as $k \to \infty$.

C. Characterizing Performance Gap

Now that we have shown that PD-SA is stable and will get arbitrarily close to the boundary of $\mathbb{B}_{\phi=1}$, the next question is how large the performance gap is. It is apparent that the performance gap depends on the size of $\mathbb{B}_{\phi=1}$. As mentioned earlier, when ϕ is fixed, the size of $\mathbb{B}_\phi$ depends on the values of $\bar{\mu}_l$, and $\bar{\nu}_n$. In the following theorem, we characterize the performance gap using ergodic capacity loss and ergodic power allocation error.

Theorem 2. *Each point* $(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})$ *on the boundary of the stability region satisfies* $|L_{u_l}(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})| = \Delta C_l$ *for all* l *and* $|L_{v_n}(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})| = \Delta P_n$ *for all* n*, where* ΔC_l *and* ΔP_n *denote the largest ergodic capacity loss for link* l *and largest ergodic power allocation error due to imperfect CSI, respectively.*

Proof: Recall that any point $(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v}) \in \mathbb{B}_\phi$ satisfies $\phi|L_{u_l}(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})| \leq \bar{\mu}_l$. Also by definition, $\bar{\mu}_l$ is the supremum limit of $\mu_l^{(k)}$, which is precisely the largest ergodic capacity loss on link l due to imperfect CSI. Therefore, from Theorem 1, we know that there exists a $\phi \in [0, 1)$ such that

$$\phi|L_{u_l}(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})| \leq \Delta C_l = \bar{\mu}_l, \text{ for } (\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v}) \in \mathbb{B}_\phi,$$

It then follows that

$$|L_{u_l}(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})| \leq \frac{\Delta C_l}{\phi}. \tag{23}$$

From our earlier discussion, we know that for a given LFB scheme, the iterates of PD-SA will move toward $\mathbb{B}_{\phi=1}$ arbitrarily close and infinitely often. Thus, we have that the neighborhood is an open region inner-bounded by points satisfying $|L_{u_l}(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})| = \Delta C_l$ for all l. Along the same line, we can also show that $|L_{v_n}(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})| = \Delta P_n$ for all n. ∎

Recall that the gradients $L_{\mathbf{u}}$ can be interpreted as the excess link capacity of every link in the network (i.e., how much link capacity is under-utilized) and $\boldsymbol{\mu}^{(k)}$ can be interpreted as the ergodic capacity loss due to imperfect CSI. When $(\mathbf{f}^{(k)}, \mathbf{P}^{(k)}, \mathbf{u}^{(k)}, \mathbf{v}^{(k)})$ is too far away from the saddle point, the excess link capacity $L_{\mathbf{u}}$ will dominate the ergodic capacity loss $\boldsymbol{\mu}^{(k)}$. Thus, the subgradient approximation $\hat{L}_{\mathbf{u}}^{(k)}$ follows approximately the correct subgradient direction $L_{\mathbf{u}}^{(k)}$ and will drive the iterates toward $(\mathbf{f}^*, \mathbf{P}^*, \mathbf{u}^*, \mathbf{v}^*)$. On the other hand, since $\boldsymbol{\mu}^{(k)}$ is dictated by imperfect CSI, it will not diminish as k increase. If $(\mathbf{f}^{(k)}, \mathbf{P}^{(k)}, \mathbf{u}^{(k)}, \mathbf{v}^{(k)})$ gets too close to the $(\mathbf{f}^*, \mathbf{P}^*, \mathbf{u}^*, \mathbf{v}^*)$, the ergodic capacity loss will dominate the excess link capacity and drive the iterates away from the $(\mathbf{f}^*, \mathbf{P}^*, \mathbf{u}^*, \mathbf{v}^*)$. Theorem 2 says that, as $k \to \infty$, the iterates will reach with high probability to an equilibrium state where the excess link capacity of each link is equal to the ergodic capacity loss caused by imperfect CSI. A similar conclusion can also be drawn for the interaction between the excess power at each node (i.e., how much power is unused, represented by $|L_{v_n}(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})|$) and the ergodic power allocation error at each node due to imperfect CSI (represented by $\bar{\nu}_n$).

Note also that if the number of feedback bits $B \to \infty$, we have $\Delta C_l \to 0$ and $\Delta P_n \to 0$ because the quality of CSI becomes better and better. In this case, we can see from (23) that $|L_{u_l}(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})| \to 0$ (so does $|L_{v_n}(\mathbf{f}, \mathbf{P}, \mathbf{u}, \mathbf{v})| \to 0$). This means that the size of the neighborhood $\mathbb{B}_{\phi=1}$ becomes smaller and smaller and PD-SA will converges to $(\mathbf{f}^*, \mathbf{P}^*, \mathbf{u}^*, \mathbf{v}^*)$ asymptotically. In other words, the performance gap vanishes as the quality of CSI feedback improves. Clearly, there exists a tradeoff between the amount of CSI feedback and the performance gap.

VI. Numerical Results

In this section, we provide some numerical results to better understand on our theoretical analysis. Fig. 1 shows a network example consisting of 15 nodes. Each node in the network is equipped with 4 antennas. There are 4 sessions in the network and their flow routes are shown in Fig. 1. The maximum transmit power of each node are all 20dBm and the path loss index is assumed to be 3. The weights of all sessions are set to 1.

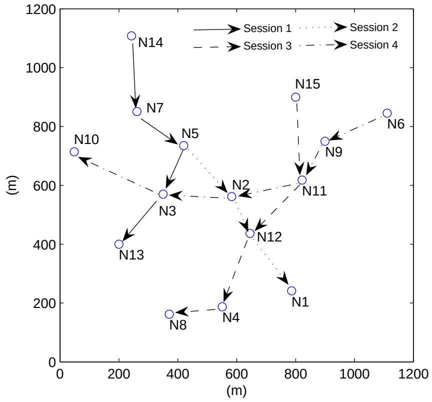

Fig. 1. A network example of 15 nodes and 4 sessions.

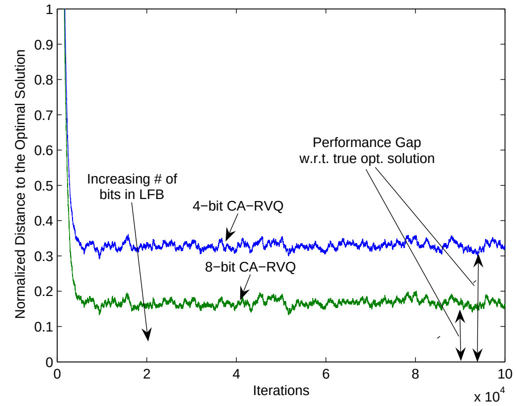

Fig. 2. Normalized Distance to the Optimal Solution

For this example setting, we first use the PD method with perfect CSI to solve the cross-layer rate control for this network. We found that the optimal rates (in (b/s/Hz)) for this network example are $f_1^* = 3.9903$, $f_2^* = 3.9633$, $f_3^* = 3.8322$, $f_4^* = 4.1232$.

Now, we use PD-SA with a LFB scheme. The LFB scheme we use is covariance adaption with random vector quantization (CA-RVQ). The RVQ codebook generation follows the uniformly distributed random covariance codebook design in [15]. In Fig. 2, we plot the normalized distance of the iterates to the optimal flow rates, which is defined as $\frac{\|\mathbf{f}^{(k)} - \mathbf{f}^*\|}{\|\mathbf{f}^*\|}$ and represents how much link capacity resources are underutilized in the network. In Fig. 2, 4-bit CA-RVQ and 8-bit CA-RVQ mean that using 4 and 8 bits to convey the information of input covariance matrices, respectively. For both 4-bit and 8-bit CA-RVQ, it can be seen that the algorithm returns to a neighborhood of the optimal solution (the horizontal axis) infinitely often. It is also seen that the normalized distances (performance gap) for 4-bit CA-RVQ and 8-bit CA-RVQ are bounded approximately by 0.39 and 0.18, respectively, which are in accordance with the simulation results of the ergodic capacity loss in [15]. This result corroborates our theoretical analysis in Theorem 2.

VII. Conclusion

In this paper, we investigated the performance of MIMO-based ad hoc networks under imperfect CSI. We first proposed an solution based on stochastic approximation. We showed that the proposed solution converges with probability one to a neighborhood of an optimal solution obtained under perfect CSI. We further analyzed the performance gap due to the existence of imperfect CSI. This paper offers a theoretical understanding on the performance of MIMO-based ad hoc networks under imperfect CSI.

Acknowledgement

The work of Y.T. Hou and J. Liu has been supported in part by ONR Grant N00014-08-1-0084 and NSF Grant CNS-0721421.

References

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–355, Mar. 1998.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[3] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 1, pp. 684–702, Jun. 2003.

[4] D. J. Love, R. W. Heath, W. Santipach, and M. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

[5] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[6] C. Swannack, E. Uysal-Biyikoglu, and G. Wornell, "MIMO broadcast scheduling with limited channel state information," in *Proc. Allerton Conf. on Commun., Control, and Comput.*, Oct. 2005.

[7] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5059, Nov. 2006.

[8] M. Zorzi, J. Zeidler, A. Anderson, B. Rao, J. Proakis, A. L. Swindlehurst, M. Jensen, and S. Krishnamurthy, "Cross-layer issues in MAC protocol design for MIMO ad hoc networks," *IEEE Wireless Commun. Mag.*, vol. 13, no. 4, pp. 62–76, Aug. 2006.

[9] J. H. Winters, "Smart antenna techniques and their application to wireless ad hoc networks," *IEEE Wireless Commun. Mag.*, vol. 13, no. 4, pp. 77–83, Aug. 2006.

[10] R. Vedantham, S. Kakumanu, S. Lakshmanan, and R. Sivakumar, "Component based channel assignment in single radio, multi-channel ad hoc networks," in *Proc. ACM Mobicom*, Las Angeles, CA, USA, Sep.23-26, 2006, pp. 378–389.

[11] F. P. Kelly, A. K. Malullo, and D. K. H. Tan, "Rate control in communications networks: Shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[12] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 3rd ed. New York, NY: John Wiley & Sons Inc., 2006.

[13] S. Low and R. Srikant, "A mathematical framework for designing a low-loss low-delay internet," *Network and Spatial Economics*, vol. 4, no. 1, pp. 75–101, 2004.

[14] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. Net York: Springer, 2003.

[15] A. D. Dabbagh and D. J. Love, "Feedback rate-capacity loss tradeoff for limited feedback MIMO systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2190–2202, May 2006.