Distributed and Optimal Reduced Primal-Dual Algorithm for Uplink OFDM Resource Allocation

Xiaoxin Zhang, Liang Chen, Jianwei Huang, Minghua Chen, and Yuping Zhao

***Abstract*— Orthogonal frequency division multiplexing (OFDM) is the key component of many emerging broadband wireless access standards. The resource allocation in OFDM uplink, however, is challenging due to heterogeneity of users' quality of service requirements, channel conditions, and individual resource constraints. We formulate the resource allocation problem as a non-strictly convex optimization problem, which typically has multiple global optimal solutions. We propose a reduced primal-dual algorithm, which is distributed, requires simple local updates, and probably globally converges to a global optimal solution under easily satisfied sufficient technical conditions. The performance of the algorithm is studied through a realistic OFDM simulator based on field measurements. Compared with the previously proposed standard primal-dual algorithm, the reduced algorithm decreases the total number of iterations by 80% and the variance by 85%.**

I. INTRODUCTION

Orthogonal Frequency Division Multiplexing (OFDM) is a promising technology for future broadband wireless networks. In OFDM, the entire frequency band is divided into a large number of subchannels, and network resource can be allocated flexibly over each of the subchannels. In this paper, we consider the resource allocation problem in a single cell OFDM uplink system, where multiple end users transmit data to the same base station. This is motivated by several practical wireless systems, such as WiMAX/802.16e, LTE for 3GPP, and UMB for 3GPP2. Given the channel conditions of users at a particular time, we need to determine which subset of users to schedule (i.e., transmit with positive rates), how to allocate subchannels to the scheduled users, and the power allocation across these subchannels.

Most previous work on resource allocation in OFDM systems focused on the downlink case, where the base station sends traffic to multiple end users subject to a total power constraint. The optimization problem in the downlink case is easier to solve and a centralized algorithm is reasonable to implement [1]. Due to different resource constraints in the uplink case, however, the algorithms proposed for the downlink case can not be directly applied to the uplink case.

Uplink OFDM resource allocation only receives limited attention recently [2]–[8]. In [2], the problem was formulated in the framework of Nash Bargaining with a focus of fair resource allocation. The authors of [3] proposed a heuristic algorithm that tries to minimize each user's transmission power while satisfying the individual rate constraints. In [4], the author considered the sum-rate maximization problem and derived algorithms based on Rayleigh fading on each subchannel. The authors in [5]–[8] proposed several heuristic algorithms to solve a problem similar as the one considered here with additional integer channel allocation constraints. None of the previous literature focused on solving the uplink resource allocation problem optimally.

We formulate the resource allocation problem as a weighted rate maximization problem, which is motivated by the gradient-based scheduling framework in [9]–[11]. This problem, however, is quite challenging to solve due to the heterogeneity of users' quality of service requirements, channel conditions, and individual resource constraints. In this paper, we propose a distributed primal-dual algorithm that achieves the optimal resource allocation in uplink OFDM systems. Our key contributions are:

- *Optimal algorithm with global convergence*: the proposed algorithm is provably globally convergent to one of the global optimal solutions of the resource allocation problem, despite non-strict convexity of the problem under which setting primal-dual algorithms may not be able to converge [12]–[14].
- *Distributed algorithm with low complexity*: the proposed algorithm is distributed, requires simple local updates, and demands only limited message passing.
- *Simpler algorithm with better convergence*: the proposed algorithm only needs to iteratively update a subset of all decision variables, and thus requires much fewer iterations to converge compared with the previously proposed standard primal-dual algorithm.
- *OFDM model with self-noise*: we consider an OFDM model where the achievable data rate is calculated by taking the "self-noise" into consideration. We demonstrate how this realistic model will affect the optimal solution and the corresponding algorithm design.

II. PROBLEM STATEMENT

We consider a single OFDM cell, where there is a set $\mathcal{M} = \{1, \ldots, M\}$ of users transmitting to the same base

X. Zhang and Y. Zhao are with the State Key Laboratory of Advanced Optical Communication Systems & Networks, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, P.R.China {xiaoxin.zhang,yuping.zhao}@pku.edu.cn

L. Chen, J. Huang and M. Chen are with the Dept. of Information Engineering, Chinese University of Hong Kong, Hong Kong, P.R.China {lchen6,jwhuang,minghua}@ie.cuhk.edu.hk

J. Huang is supported by the General Research Fund (Project Number 412308 and 412509) established under the University Grant Committee of the Hong Kong Special Administrative Region, China, the National Key Technology R&D Program (Project Number 2007BAH17B04) established by the Ministry of Science and Technology of China, and the Microsoft-CUHK Joint Laboratory for Human-centric Computing & Interface Technologies. Part of the work was done when X. Zhang visited Chinese University of Hong Kong in 2008.

station. Each user $i \in \mathcal{M}$ has a priority weight w_i.[1] The total frequency band is divided into a set $\mathcal{N} = \{1, \ldots, N\}$ of subchannels (e.g., tones/carriers). A user $i \in \mathcal{M}$ can transmit over a subset of the subchannels (not necessarily adjacent), with transmission power p_{ij} over subchannel $j \in \mathcal{N}$ satisfying the individual power constraint, i.e., $\sum_j p_{ij} \leq P_i$. For channel j, it is allocated to user i with fraction $x_{ij} \geq 0$, and the total allocation across all users should be no larger than 1, i.e., $\sum_i x_{ij} \leq 1$.

We define e_{ij} as the received signal-to-noise ratio (SNR) per unit power for user i on subchannel j. We further assume that the channel conditions do not change within the time of interests, i.e., we are looking at a resource allocation period smaller than the channel coherence time.[2]

With perfect channel estimation, user i's achievable rate on subchannel j is $r_{ij} = x_{ij} B \log\left(1 + \frac{p_{ij} e_{ij}}{x_{ij}}\right)$, which corresponds to the Shannon capacity of a Gaussian noise channel with bandwidth $x_{ij}B$ and received SNR $p_{ij}e_{ij}/x_{ij}$. This SNR arises from viewing p_{ij} as the average power user i is allowed to use on subchannel j; the corresponding instantaneous transmission power is p_{ij}/x_{ij} when only a time fraction x_{ij} of the subchannel is allocated. For notation simplicity we normalize the bandwidth to be $B = 1$ in the analysis.[3]

In a real OFDM system, imperfect carrier synchronization and inaccurate channel estimation may result in "self-noise" [15], [16]. We follow a similar approach as in [15] to model self-noise and use an estimate value β to represent the level of self-noise. With self-noise, user i's feasible rate on subchannel j becomes $r_{ij} = x_{ij} \log\left(1 + \frac{p_{ij} e_{ij}}{x_{ij} + \beta p_{ij} e_{ij}}\right)$, where $p_{ij}e_{ij}/(x_{ij} + \beta p_{ij} e_{ij})$ depicts the effective SNR.

The key notations used throughout this paper are listed in Table I. We use bold symbols to denote vectors and matrices of these quantities, e.g., $\boldsymbol{p} = \{p_{ij}, \forall i, j\}$ and $\boldsymbol{x} = \{x_{ij}, \forall i, j\}$.

Our objective is to maximize the weighted sum of the users' rates over the feasible rate-region defined as follows,

$$\mathcal{R}(\boldsymbol{e}) = \left\{ \boldsymbol{r} \in \Re_+^M : r_i = \sum_{j \in \mathcal{N}} x_{ij} \log\left(1 + \frac{p_{ij} e_{ij}}{x_{ij} + \beta p_{ij} e_{ij}}\right), \ \forall i \in \mathcal{M} \right\}, \tag{1}$$

where $(\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X}$ are chosen subject to

$$\sum_i x_{ij} \leq 1, \forall j \in \mathcal{N}, \tag{2}$$

$$\sum_j p_{ij} \leq P_i, \forall i \in \mathcal{M}, \tag{3}$$

and the set

$$\mathcal{X} := \left\{ (\boldsymbol{x}, \boldsymbol{p}) \geq \mathbf{0} : 0 \leq x_{ij} \leq 1, p_{ij} \geq 0 \right\}. \tag{4}$$

[1] The priority weights w_i's are motivated by the gradient-based scheduling framework in [9]–[11]. Assume each user i has a utility function $U_i(W_{i,t})$ depending on its average throughput $W_{i,t}$ up to time t. To maximize the total network utility $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_i U_i(W_{i,t})$, it is enough to solve the weighted rate maximization problem during each time slot t with $w_i = \partial U_i(W_{i,t})/\partial W_{i,t}$.

[2] This is particularly suitable for fixed broadband wireless access (part of the IEEE 802.16 standard), where users are relatively static and the corresponding coherence time is long.

[3] A realistic value of B will be considered in the simulations (Section V).

TABLE I
KEY NOTATIONS

Notation	Physical Meaning
N	total number of subchannels
$\mathcal{N}$	set of all subchannels
M	total number of users
$\mathcal{M}$	set of all users
i	user index
j	subchannel index
w_i	user i's (dynamic) weight
e_{ij}	normalized SNR on subchannel j for user i
p_{ij}	power allocated on subchannel j for user i
x_{ij}	fraction of subchannel j allocated to user i
P_i	maximum transmit power for user i
β	self-noise coefficient

In many OFDM standards, x_{ij} is constrained to be an integer, in which case we can add the additional constraint $x_{ij} \in \{0, 1\}$ for all i, j. The integer constraint makes the resource allocation very difficult to solve, and various heuristic algorithms dealing with such constraint are proposed in [6]–[8], [17]. In this paper, we focus on the rate region defined by (1) to (4), i.e., no integer constraints are considered. The corresponding optimal solution typically contains fractional values of x_{ij}'s. There are several practical methods of achieving these fraction allocations. For example, if resource allocation is done in blocks of OFDM symbols, then fractional values of x_{ij} can be implemented by time-sharing the symbols in a block. Likewise, if there are several tones in a subchannel, then fractional values of x_{ij}'s can also be implemented by frequency-sharing the tones in a subchannel [18].

To summarize, we want to solve the following problem

$$\max_{\boldsymbol{r} \in \mathcal{R}(\boldsymbol{e})} \sum_i w_i r_i, \tag{5}$$

where rate r_i and rate region $\mathcal{R}(\boldsymbol{e})$ are given in (1).

III. A Reduced Primal-Dual Algorithm

We rewrite problem (5) in variables $\boldsymbol{x}$ and $\boldsymbol{p}$ as follows:

Problem 1 (Weighted Rate Maximization):

$$\max_{(\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X}} \sum_{i \in \mathcal{M}} w_i \sum_{j \in \mathcal{N}} x_{ij} \log\left(1 + \frac{p_{ij} e_{ij}}{x_{ij} + \beta p_{ij} e_{ij}}\right), \tag{6}$$

subject to the per subchannel assignment constraints in (2) and the per user power constraints in (3). Set $\mathcal{X}$ is given in (4).

Although the objective function in (6) is concave, its derivative is not well defined at the origin ($\boldsymbol{x} = 0, \boldsymbol{p} = 0$). This motivates us to look at the following ϵ-relaxed version of Problem 1:

Problem 2 (ϵ-relaxed Weighted Rate Maximization):

$$\max_{(\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X}} \sum_{i \in \mathcal{M}} w_i \sum_{j \in \mathcal{N}} (x_{ij} + \epsilon_{ij}) \log\left(1 + \frac{p_{ij} e_{ij}}{x_{ij} + \beta p_{ij} e_{ij} + \epsilon_{ij}}\right), \tag{7}$$

where constants ϵ_{ij} take small positive value for all i and j. The constraint set remains the same as in Problem 1.

By such relaxation, the objective function in (7) now has derivative defined everywhere in the constraint set $\mathcal{X}$. Thanks to the continuity of the objective function, the optimal value to Problem 2 can be arbitrarily close to that of Problem 1 if $\boldsymbol{\epsilon} = [\epsilon_{ij}, \forall i, j]$ is chosen to be small enough.

The constraint set of Problem 2 is convex, and the objective function in (7) is continuous and *non-strictly* concave.[4] As such, Problem 2 has multiple optimal solutions, and there is no duality gap between it and its dual problem.

The existence of derivatives allows us to write down a primal-dual algorithm to pursue the optimal solution to Problem 2. The Lagrangian for Problem 2 is as follows,

$$L(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{x}, \boldsymbol{p}) := \sum_{i,j} w_i(x_{ij} + \epsilon_{ij}) \log\left(1 + \frac{p_{ij}e_{ij}}{x_{ij} + \beta p_{ij}e_{ij} + \epsilon_{ij}}\right) + \sum_i \lambda_i\Big(P_i - \sum_j p_{ij}\Big) + \sum_j \mu_j\Big(1 - \sum_i x_{ij}\Big). \quad (8)$$

By strong duality theorem, the optimal primal and dual solutions must satisfy KKT conditions, i.e., for all i and j,

$$\mu_j \geq 0, \ \sum_i x_{ij} \leq 1, \ \mu_j\Big(\sum_i x_{ij} - 1\Big) = 0, \quad (9)$$

$$\lambda_i \geq 0, \ \sum_j p_{ij} \leq P_i, \ \lambda_i\Big(\sum_j p_{ij} - P_i\Big) = 0, \quad (10)$$

$$x_{ij} \geq 0, \ p_{ij} \geq 0, \quad (11)$$

$$x_{ij}\left(f_{ij}(x_{ij}, p_{ij}) - \mu_j\right) \leq 0, \quad (12)$$

$$p_{ij}\left(g_{ij}(x_{ij}, p_{ij}) - \lambda_i\right) \leq 0, \quad (13)$$

where $f_{ij}(\cdot)$ and $g_{ij}(\cdot)$ are gradients of the objective function in (7) with respect to x_{ij} and p_{ij}, respectively, and are given by

$$f_{ij}(x_{ij}, p_{ij}) = w_i \log\left(1 + \frac{p_{ij}e_{ij}}{x_{ij} + \beta p_{ij}e_{ij} + \epsilon_{ij}}\right) - \frac{w_i(x_{ij} + \epsilon_{ij})p_{ij}e_{ij}}{(x_{ij} + \beta p_{ij}e_{ij} + \epsilon_{ij})[x_{ij} + (\beta+1)p_{ij}e_{ij} + \epsilon_{ij}]},$$

and

$$g_{ij}(x_{ij}, p_{ij}) = \frac{w_i e_{ij}(x_{ij} + \epsilon_{ij})^2}{(x_{ij} + \beta p_{ij}e_{ij} + \epsilon_{ij})[x_{ij} + (\beta+1)p_{ij}e_{ij} + \epsilon_{ij}]}.$$

The last two KKT conditions in (12) and (13) become equalities if $x_{ij} > 0$ and $p_{ij} > 0$, respectively. It can be verified that the optimal solutions of Problem 2, satisfying above KKT conditions, are exactly the saddle points of the Lagrangian function in (8). Since the primal problem has at least one solution, the saddle point exists.

For notation simplicity, we define $(a)^+ = \max(a, 0)$ and

$$(a)_b^+ = \begin{cases} a, & b > 0, \\ \max(a, 0), & \text{otherwise.} \end{cases}$$

To pursue saddle points of the Lagrangian function, we start by reviewing the *standard* primal-dual algorithm introduced in [13]. Then we derive a new *reduced* primal-dual algorithm which converges faster than the standard one.

[4]This can be verified by showing that the Hessian of the objective function in Problem 2 is negative-semidefinite but not negative-definite [19].

For each i and j, we consider the following standard primal-dual algorithm:

Algorithm SPD: Standard Primal-Dual Algorithm

$$\dot{x}_{ij} = k_{ij}^x \left(f_{ij}(x_{ij}, p_{ij}) - \mu_j\right)_{x_{ij}}^+, \quad (14)$$

$$\dot{p}_{ij} = k_{ij}^p \left(g_{ij}(x_{ij}, p_{ij}) - \lambda_i\right)_{p_{ij}}^+, \quad (15)$$

$$\dot{\mu}_j = k_j^\mu \Big(\sum_i x_{ij} - 1\Big)_{\mu_j}^+, \quad (16)$$

$$\dot{\lambda}_i = k_i^\lambda \Big(\sum_j p_{ij} - P_i\Big)_{\lambda_i}^+, \quad (17)$$

where $k_{ij}^x, k_{ij}^p, k_j^\mu$ and k_i^λ are constants representing update stepsizes. Here the derivatives on the left hand sides of (14) to (17) are defined with respect to time.

We call a point $(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\lambda})$ an *equilibrium* of Algorithm SPD if and only if the corresponding derivatives in (14) to (17) are zero for all i and j. We can show that the set of equilibria of Algorithm SPD is equivalent to the set of global optimal solutions of Problem 2 [13].

In Algorithm SPD, all variables x_{ij}, p_{ij}, μ_j, and λ_i are dynamically adapted, which might lead to slow convergence. One way to address this is to reduce the number of dynamically adapting variables. To achieve this, we will constrain the algorithm trajectories onto a manifold that includes all optimal primal and dual solutions.

We study the following manifold by setting (15) to zero, i.e. $\forall i, j$,

$$0 = \left(g_{ij}(x_{ij}, p_{ij}) - \lambda_i\right)_{p_{ij}}^+, \quad (18)$$

which in turn implies

$$\begin{cases} g_{ij}(x_{ij}, p_{ij}) = \lambda_i, & \text{if } p_{ij} > 0, \\ g_{ij}(x_{ij}, p_{ij}) \leq \lambda_i, & \text{if } p_{ij} = 0. \end{cases} \quad (19)$$

Based on the KKT conditions, the optimal primal and dual solutions must lie on the above manifold.

After simplification we get the following expression of the manifold:

$$p_{ij} = h_{ij}(x_{ij}, \lambda_i), \quad (20)$$

where $h_{ij}(\cdot)$ is denoted by

$$h_{ij}(x_{ij}, \lambda_i) = \left(\frac{\sqrt{1 + 4\beta(\beta+1)\frac{w_i e_{ij}}{\lambda_i}} - (2\beta+1)}{2\beta(\beta+1)e_{ij}}(x_{ij} + \epsilon_{ij})\right)_{w_i e_{ij} - \lambda_i}^+$$

when $\beta \neq 0$, and

$$h_{ij}(x_{ij}, \lambda_i) = \left(\frac{w_i e_{ij} - \lambda_i}{\lambda_i e_{ij}}(x_{ij} + \epsilon_{ij})\right)_{w_i e_{ij} - \lambda_i}^+$$

when $\beta = 0$.

Substituting (20) into (14) to (17), we obtain a new reduced primal-dual algorithm as follows:

Algorithm RPD: Reduced Primal-Dual Algorithm

$$\dot{x}_{ij} = k^x_{ij}\Big(f_{ij}(x_{ij}, p_{ij}) - \mu_j\Big)^+_{x_{ij}}, \quad (21)$$

$$\dot{\mu}_j = k^\mu_j\Big(\sum_i x_{ij} - 1\Big)^+_{\mu_j}, \quad (22)$$

$$\dot{\lambda}_i = k^\lambda_i\Big(\sum_j p_{ij} - P_i\Big)^+_{\lambda_i}, \quad (23)$$

$$p_{ij} = h_{ij}(x_{ij}, \lambda_i). \quad (24)$$

Proposition 1: The set of equilibria of Algorithm RPD is the same as the set of global optimal solutions of Problem 2.

This means that if Algorithm RPD converges, it reaches a global optimal solution of the ϵ-relaxed weighted rate maximization problem.

Compared to Algorithm SPD in (14) to (17) in which $\boldsymbol{p}$ is dynamically adapted, $\boldsymbol{p}$ in the new Algorithm RPD is directly computed from $\boldsymbol{x}$ and $\boldsymbol{\lambda}$. Consequently, Algorithm RPD has fewer dynamically adapting variables, hence is expected to converge faster.

Similar as Algorithm SPD, Algorithm RPD can also be implemented in a distributed fashion by end users and the base station. End user i is responsible of updating x_{ij}'s and p_{ij}'s as well as dual variables λ_i's locally. During each iteration, it sends the latest values of x_{ij}'s to the base station, but not the p_{ij}'s or λ_i's. The base station is responsible of updating dual variables μ_j's for all subchannels and broadcasting to the users. In particular, the base station does not need to know users' priority weights, power constraints, or the power allocation. Both the communication complexity and computation complexity per iteration are $O(MN)$.

Next we will show that trajectories of Algorithm RPD converge to its equilibria, and thus the global optimal solution of Problem 2.

IV. CONVERGENCE OF THE REDUCED PRIMAL-DUAL ALGORITHM

The key challenge of the convergence proof is the *non-strict concavity* of the objective function in (7). It has been well observed in literature that although primal-dual algorithms globally converge to the optimal solution of strictly concave optimization problem, they may oscillate indefinitely and fail to converge when applying to non-strictly concave optimization problem [12]–[14].

In this section, we study convergence of Algorithm RPD. We first show that the trajectories converge to an invariant set that contains all global optimal solutions of Problem 2.

Theorem 1: All trajectories of Algorithm RPD converge to an invariant set V_0 globally and asymptotically. Furthermore, let $(\boldsymbol{x}^*, \boldsymbol{p}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ be a global optimal solution of Problem 2 and $(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\lambda})$ be any point in set V_0, the following is true for all i and j,

1) $(\boldsymbol{x}^*, \boldsymbol{p}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ is contained in V_0;
2) μ_j is nonzero only if $\sum_i x^*_{ij} = 1$;
3) $\sum_j p^*_{ij} = P_i$, and λ_i is a positive constant;
4) Over set V_0, $f_{ij}(x_{ij}, p_{ij}) = f_{ij}(x^*_{ij}, p^*_{ij}) = \mu^*_j$, and $g_{ij}(x_{ij}, p_{ij}) = g_{ij}(x^*_{ij}, p^*_{ij}) = \lambda^*_i$;
5) $p_{ij}/(x_{ij} + \epsilon_{ij}) = p^*_{ij}/(x^*_{ij} + \epsilon_{ij})$.

The detailed proof can be found in Appendix A of the online technical report [20]. Results 1) to 4) will be used in later analysis.

Result 5) is of independent interest. It implies that although there can be multiple global optimal solutions to Problem 2, the effective SNR achieved by user i on subchannel j is the same in all solutions.

Although all trajectories of Algorithm RPD may converge to the desired equilibria in V_0, they may also converge to non-equilibrium points in V_0 (if there are any). We now study the conditions for V_0 to contain *only* the desired equilibria, under which Theorem 1 guarantees the convergence of Algorithm RPD to a global optimal solution of Problem 2.

Plug result 4) of Theorem 1 into Algorithm RPD, and recall that M is the total number of users and N is the total number of subchannels. We find that V_0 is exactly the set that contains all trajectories of the following linear system in (25) to (27) over set $\{\boldsymbol{x} \geq 0, \boldsymbol{\mu} \geq 0\}$.

$$\dot{\boldsymbol{x}} = K^x A_1^T \boldsymbol{\mu}^* - K^x A_1^T \boldsymbol{\mu}, \quad (25)$$

$$\dot{\boldsymbol{\mu}} = K^\mu A_1 \boldsymbol{x} - K^\mu \mathbf{1}, \quad (26)$$

$$\dot{\boldsymbol{\lambda}} = K^\lambda A_2 B(\boldsymbol{x} + \boldsymbol{\epsilon}) - K^\lambda \boldsymbol{P} = 0, \quad (27)$$

where K^x is an $MN \times MN$ diagonal matrix with diagonal terms equal to k^x_{ij}'s, K^μ is an $N \times N$ diagonal matrix with diagonal terms equal to k^μ_j's, and K^λ is an $M \times M$ diagonal matrix with diagonal terms equal to k^λ_i's. B is an $MN \times MN$ diagonal matrix given by $B = diag(b_{ij},\ \forall i, j)$, where $b_{ij} = \frac{p^*_{ij}}{x^*_{ij} + \epsilon_{ij}}$. The matrix A_1 has a dimension of $N \times MN$, and is given by $A_1 = [I_N, \cdots, I_N]$, where I_N is an identity matrix with dimension N. The matrix A_2 has a dimension of $M \times MN$ and is given by

$$A_2 = \begin{bmatrix} \mathbf{1}_{1\times N} & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{1\times N} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_{1\times N} \end{bmatrix},$$

where $\mathbf{1}_{1\times N}$ is an all one vector with dimension 1 by N.

We observe the following for the above linear system:

Lemma 1: For the linear system in (25) to (27), we have

1) every order Lie derivative of $A_2 B x$ is constant, that is $\forall n \geq 0$:
$$\frac{d^n}{dt^n} A_2 B \boldsymbol{x} = \text{constant},$$
where t denotes time;
2) starting from a non-equilibrium point, trajectories of $\boldsymbol{x}$ and $\boldsymbol{\mu}$, following (25) and (26) respectively, do not converge and form limit cycles.

Proof: (Sketch) From (27), we have

$$A_2 B \boldsymbol{x} = P - A_2 B \boldsymbol{\epsilon} = \text{constant}.$$

Result 1 can be derived by taking derivatives (with respect to time) on both sides of the above equation. For result 2, it can be verified that the transfer function matrix of the linear system (25)-(26) is a product of positive diagonal matrix and a skew-symmetric matrix. Hence, all eigenvalues of the transfer matrix are purely imaginary. ■

Result 1 in Lemma 1 states that every order Lie derivative of $A_2 B\boldsymbol{x}$ is constant. By linear system theory, if the system state $\boldsymbol{\mu}$ is completely observable from $A_2 B\boldsymbol{x}$, then constant $A_2 B\boldsymbol{x}$ will lead to $\dot{\boldsymbol{\mu}}$ equal to 0 and $\boldsymbol{\mu}$ being constant. When $\boldsymbol{\mu}$ is constant, $\dot{\boldsymbol{x}}$ is constant according to (25). Combining with the constraint that $\boldsymbol{x} \geq 0$ and $A_2 B\boldsymbol{x}$ is constant, we can show that $\dot{\boldsymbol{x}}$ is also zero if $\dot{\boldsymbol{\mu}}$ is zero over the set $\{\boldsymbol{x} \geq 0, \boldsymbol{\mu} \geq 0\}$.

In the following theorem, we state conditions for $\boldsymbol{\mu}$ to be completely observable from $A_2 B\boldsymbol{x}$, and summarize its consequence on convergence of Algorithm RPD.

Theorem 2: All trajectories of Algorithm RPD converge globally and asymptotically to the system equilibria if the following condition holds:

$$\begin{bmatrix} A_2 B K^x A_1^T \\ K^\mu A_1 K^x A_1^T - \sigma I \end{bmatrix} \text{ has rank } N, \qquad (28)$$

where σ denotes any eigenvalue of matrix $K^\mu A_1 K^x A_1^T$.

Proof: By linear system theory, $\boldsymbol{\mu}$ is completely observable from the constant $A_2 B\boldsymbol{x}$ if and only if the complete observability (28) holds [21]. Then the invariant set V_0 contains only the equilibria of the linear system in (25) to (27), which are the global optimal solutions of Problem 2. Consequently, all trajectories of Algorithm RPD converge globally and asymptotically to the global optimal solutions. ■

For the problem we studied in this paper, we can choose properly the update stepsizes of Algorithm RPD to satisfy the conditions in (28).

Corollary 1: Conditions (28) in Theorem 2 are satisfied if both of the following are true

- $K^x = kI$ (diagonal terms of K^x take the same value k);
- all diagonal elements of K^μ take different values.

The proof of Corollary 1 can be found in Appendix B in the online technical report [20].

In this section, we have investigated the convergence of Algorithm RPD by combining both La Salle principle from nonlinear stability theory and complete observability from linear system theory. The proof shows that Algorithm RPD can globally and asymptotically converge to one of the global optimal solutions of Problem 2 when satisfying the conditions in Corollary 1.

V. SIMULATION RESULTS

We show the convergence and optimality of Algorithm RPD over a realistic OFDM uplink simulator. Each user's subchannel gains e_{ij}'s are the product of two terms: a constant location-based term picked using an empirically obtained distribution, and a fast fading term generated using a block-fading model and a standard mobile delay-spread model with a delay spread of 10μsec. The system bandwidth is 5.12MHz consisting of 512 tones, which is further grouped into 64 subchannels.[5] The symbol duration is 100μsec with a cyclic prefix of 10μsec.

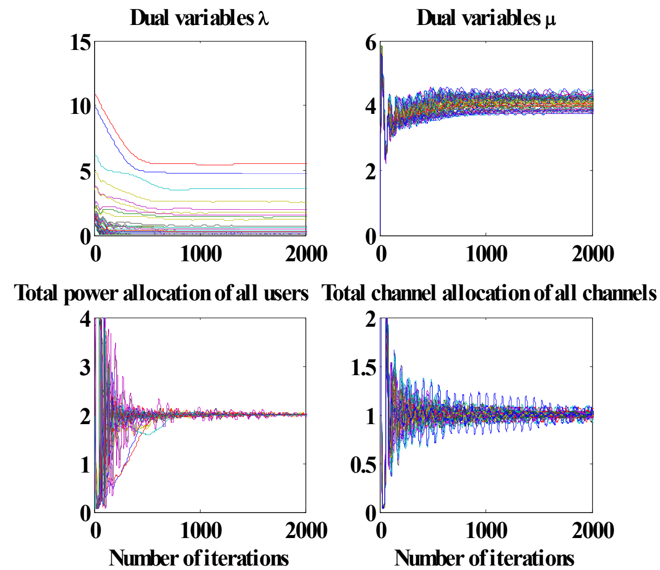

Fig. 1. Primal and dual variable convergence of Algorithm RPD

Unless otherwise specified, we assume the following parameter setting throughout all simulations. The variables are initialized as $x_{ij} = 1/M$, $p_{ij} = P_i/N$, $\mu_j = 0$, and $\lambda_i = 0.01 \max_j(w_i e_{ij})$ for all i and j. The update stepsizes in Algorithm RPD are chosen as $k_{ij}^x = 10^{-2}$, $k_j^\mu = 10^{-1} + \varepsilon_j$, and $k_i^\lambda = 10^{-2}$ for all i and j. Here, ε_j's for all channels are chosen to be very small values and diverse from each other, in order to meet the requirement of Corollary 1 as the sufficient condition for system convergence. Each user has a total transmission power constraint $P_i = 2$Watts. Users' channel conditions are randomly generated from the simulator, and users have equal weights $w_i = 1$ for all i.

A. Algorithm Convergence

We first show the convergence of Algorithm RPD with 40 users and 64 subchannels. Here we assume $\beta = 0$. Fig. 1 shows the convergence of dual variables (upper two subgraphs, λ_i for 40 users and μ_j for 64 subchannels) and primal variables (lower two subgraphs, $\sum_j p_{ij}$ for 40 users and $\sum_i x_{ij}$ for 64 subchannels).

In Fig. 2, the upper subgraph shows how the dual value and primal feasible value change with iterations. The dual value is an upper bound of the global optimal solution of Problem 2. The primal feasible value is a lower bound and is calculated as follows: given the primal values of $\boldsymbol{p}(t)$ and $\boldsymbol{x}(t)$ at iteration t, normalize so that they are feasible and the resources are fully utilized (i.e., $\tilde{p}_{ij}(t) = p_{ij}(t)P_i/(\sum_j p_{ij}(t))$ and $\tilde{x}_{ij}(t) = x_{ij}(t)/(\sum_i x_{ij}(t))$), and calculate the achievable rate accordingly. The bottom subfigure shows the relative

[5] Every 8 adjacent tones are grouped into one subchannel. This corresponds to the "Band AMC mode" of 802.16 d/e and can help to reduce the feedback overhead. For discussions on various ways of subchannelization, see [1].

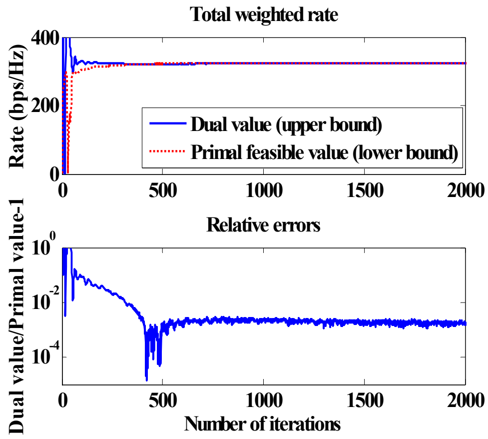

Fig. 2. Total weighted rate convergence of Algorithm RPD

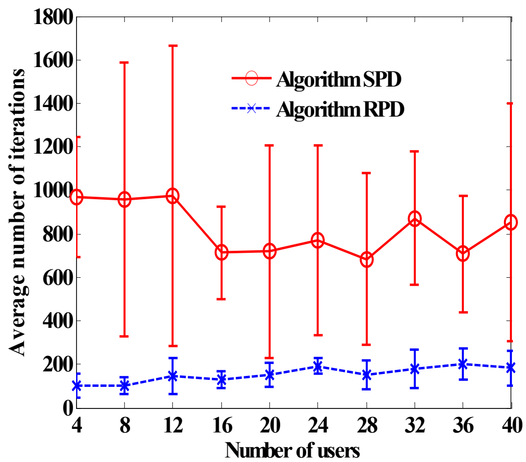

Fig. 3. Average number of iterations and standard deviation of Algorithm RPD and Algorithm SPD

errors of two curves plotted in the upper subfigure. If we define the stopping criterion to be the relative error less than 5×10^{-3}, then Algorithm RPD converges in 364 iterations.

B. Comparison with the Standard Primal-Dual Algorithm

In Fig. 3, we compare the convergence speed of Algorithm RPD with Algorithm SPD. The self-noise coefficient is $\beta = 0.01$. We vary the number of users from 4 to 40. For a fixed user population size, we randomly generated 10 sets of different weights and channel conditions. We plot both the average and the standard deviation (i.e., error bar) of the number of iterations for both algorithms as the number of users changes. In all cases, Algorithm RPD converges with much fewer iterations (about 80% less than Algorithm SPD in average) and a much smaller variance (about 85% less than Algorithm SPD in average).

VI. CONCLUSIONS AND FUTURE WORKS

We presented a distributed optimal primal-dual resource allocation algorithm for uplink OFDM systems. The key features of the proposed algorithm include: (a) distributed implementation by end users and base station with simple local updates, (b) global convergence despite the existence of multiple global optimal solutions, (c) reduced primal-dual algorithm which eliminates unnecessary variable updates and hence converges faster than the standard algorithm, and (d) incorporating self-noise observed in the practical OFDM systems. The absolute convergence speed of the proposed algorithm, on the other hand, is still slow for real-time implementation. Designing fast and near optimal heuristic algorithms based on the proposed algorithm would be an interesting future research direction.

REFERENCES

[1] J. Huang, V. G. Subramanian, R. Agrawal, and R. A. Berry, "Downlink scheduling and resource allocation for OFDM systems," *IEEE Trans. on Wireless Communications*, vol. 8, no. 1, pp. 288–296, Jan. 2009.

[2] Z. Han, Z. Ji, and K. Liu, "Fair Multiuser Channel Allocation for OFDMA Networks Using Nash Bargaining Solutions and Coalitions," *IEEE Trans. Comm.*, vol. 53, no. 8, pp. 1366–1376, 2005.

[3] S. Pfletschinger, G. Muenz, and J. Speidel, "Efficient subcarrier allocation for multiple access in OFDM systems," in *7th International OFDM Workshop*, 2002.

[4] Y. Ma, "Constrained Rate-Maximization Scheduling for Uplink OFDMA," *Proc. IEEE MILCOM*, pp. 1–7, Oct. 2007.

[5] K. Kim, Y. Han, and S. Kim, "Joint subcarrier and power allocation in uplink OFDMA systems," *IEEE Comms. Letters*, vol. 9, no. 6, pp. 526–528, 2005.

[6] C. Ng, and C. Sung, "Low complexity subcarrier and power allocation for utility maximization in uplink OFDMA systems," *IEEE Trans. Wireless Comm.*, vol. 7, no. 5 Part 1, pp. 1667–1675, 2008.

[7] K. Kwon, Y. Han, and S. Kim, "Efficient Subcarrier and Power Allocation Algorithm in OFDMA Uplink System," *IEICE Trans. Comm.*, vol. 90, no. 2, pp. 368–371, 2007.

[8] L. Gao, and S. Cui, "Efficient subcarrier, power, and rate allocation with fairness consideration for OFDMA uplink," *IEEE Trans. Wireless Comm.*, vol. 7, no. 5 Part 1, pp. 1507–1511, 2008.

[9] R. Agrawal, and V. Subramanian, "Optimality of Certain Channel Aware Scheduling Policies," *Proc. of Allerton Conference*, Oct. 2002.

[10] A. L. Stolyar, "On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation," *Operations Research*, vol. 53, No. 1, pp. 12–25, 2005.

[11] H. Kushner, and P. Whiting, "Asymptotic properties of proportional-fair sharing algorithms," in *Proc. 40th Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2002.

[12] T. Voice, "Stability of Congestion Control Algorithms with Multi-Path Routing and Linear Stochastic Modelling of Congestion Control," PhD Thesis, University of Cambridge, May, 2006.

[13] M. Chen, and J. Huang, "Optimal Resource Allocation for OFDM Uplink Communication: A Primal-Dual Approach," *CISS conference*, 2008.

[14] M. Chen, M. Ponec, S. Sengupta, J. Li, and P. A. Chou, "Utility Maximization in Peer-to-peer Systems," *ACM SIGMETRICS*, 2008.

[15] H. Jin, R. Laroia, and T. Richardson, "Superposition by position," *IEEE Information Theory Workshop*, March 2006.

[16] J. Lee, H. Lou, and D. Toumpakaris, "Analysis of Phase Noise Effects on Time-Direction Differential OFDM Receivers," *Proc. of IEEE GlobeCom*, 2005.

[17] J. Huang, V. Subramanian, R. Berry, and R. Agrawal, "Joint Scheduling and Resource Allocation in Uplink OFDM Systems for Broadband Wireless Access Networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 2, pp. 226–234, Feb. 2009.

[18] W. Yu, R. Lui, and R. Cendrillon, "Dual optimization methods for multiuser orthogonal frequency division multiplex systems," *Proc. of IEEE GlobeCom*, vol. 1, 2004, pp. 225–229.

[19] D. Bertsekas, *Nonlinear Programming*, 2nd ed., Massachusetts: Athena Scientific, 1999.

[20] X. Zhang, L. Chen, J. Huang, M. Chen, and Y. Zhao, "Optimal Primal-Dual Algorithm for Uplink OFDM Resource Allocation," technical report, http://www.ie.cuhk.edu.hk/~jwhuang/publication/PD_TechReport.pdf.

[21] F. Callier and C. Desoer, *Linear System Theory*, Springer-Verlag, 1991.