Optimal Resource Allocation for OFDM Uplink Communication: A Primal-Dual Approach

Minghua Chen and Jianwei Huang

***Abstract*—Orthogonal Frequency Division Multiplexing (OFDM) with dynamic resource allocation is widely considered to be a key component of most emerging broadband wireless access networks. However, resource allocation in an OFDM system is complicated, especially in the uplink due to the heterogeneity of the users' subchannel conditions, individual resource constraints and application requirements. We formulate the resource allocation problem as a convex optimization problem, which has a unique optimal objective value but might have multiple corresponding optimal solutions. We then present a primal-dual based algorithm that is distributed, low complexity, and is provably global convergent to the optimal solutions. The convergence and optimality of the algorithm is studied through a realistic OFDM simulator.**

I. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) is a promising technology for future broadband wireless networks, due to many of its advantages such as robustness against intersymbol interference and multipath fading, and no need for complex equalizations. It is the core technology for a number of wireless data systems, such as IEEE 802.16 (WiMAX), IEEE 802.11a/g (Wireless LANs), and IEEE 802.20 (Mobile Broadband Wireless Access). In this paper, we consider the problem of uplink resource allocation for OFDM wireless access networks. This problem is motivated by the WiMAX/802.16e standard[1]. During each time slot, the scheduler at the base station needs to make the following resource allocation decision: which subset of users to schedule, how to allocate the subchannels to the users and the corresponding power allocation across these subchannels.

Using OFDM on the uplink of an access network with dynamic resource allocation has only recently attracted significant attention. Thus the literature on this subject is still in a nascent state (e.g., [1], [10], [11]). The authors of [1] considered the same problem as in this paper and designed centralized optimal and various heuristic algorithms accordingly. In particular, the optimal algorithm requires the scheduler to solve a linear programming problem to determine the fractional channel allocations after the multi-dimensional subgradient search converges. Furthermore, since the algorithm is centralized, various information such as the QoS classes, queue-lengths and delays of the packets queued on each mobile device needs to be communicated to the scheduler, and the scheduler needs to convey the scheduling decisions back to the mobiles, all with short delays.

There are two disadvantages of the above centralized approach: (i) The mobiles are forced to reveal all local information to the centralized schedular, which may not be desirable in certain applications due to privacy concerns. (ii) The centralized schedular needs to have sophisticated computational capability to finish complicated optimizations with a short delay. Due to these two reasons, in this paper we will discuss how to design a distributed resource allocation algorithm for uplink OFDM systems. The proposed algorithm is based on a primal-dual approach. It only requires simple updates at the mobile users and the base station, thus is scalable with network size. Furthermore, it is guaranteed to converge to the optimal solutions under proper update rules, despite of the non-strict concavity of the optimization problem.

The rest of the paper is organized as follows. The precise problem is stated in Section II. The primal-dual based algorithm is proposed in Section III, and its convergence behavior is analyzed in Section IV. Simulations results on convergence and optimality are given in Section V and we finally conclude in Section VI.

II. Problem Statement

We consider a system model similar as in [1]. The key notations are listed in Table I[2]. We consider a single OFDM cell, where there is a set of $\mathcal{M} = \{1, \ldots, M\}$ users transmitting to the same base station. Each user $i \in \mathcal{M}$ has a total transmission power constraint P_i and priority weight w_i. We will later discuss how these weights are derived in practice. The total frequency band is divided into a set of $\mathcal{N} = \{1, \ldots, N\}$ subchannels (e.g., frequency bands), and a user i can transmit over a subset of the subchannels, with transmit power p_{ij} over subchannel j. For channel j, it is allocated to user i with fraction x_{ij}, and the total allocation across all users should be no larger than 1, i.e., $\sum_i x_{ij} \leq 1$.

In every scheduling epoch, the scheduler seeks to maximize a (time-varying) weighted sum of the users' rates over a given (time-varying) rate-region. Next we will describe this this rate-region. We represent the time-varying subchannel quality vector at time t as $\boldsymbol{e}_t$, where e_{ij} is the received Signal-to-noise ratio (SNR) per unit power for user i on subchannel j. As in [4], this model can also incorporate various sub-subchannelization schemes where the resource allocation is

Both authors are with the Dept. of Information Engineering, Chinese University of Hong Kong, e-mail: {minghua,jwhuang}@ie.cuhk.edu.hk.
J. Huang is supported in part by Direct Grant of the Chinese Univ. of Hong Kong under Grant 2050398.

[1]LTE for 3GPP and 3GPP2 and the FLASH OFDM system from Qualcomm Flarion also fit the model we consider in this paper. Furthermore, this model is applicable for both FDD and TDD systems.

[2]we use bold symbols to denote vectors and matrices of these quantities, e.g., $\boldsymbol{w} = \{w_i, \forall i\}$, $\boldsymbol{e} = \{e_{ij}, \forall i, j\}$, $\boldsymbol{p} = \{p_{ij}, \forall i, j\}$, and $\boldsymbol{x} = \{x_{ij}, \forall i, j\}$.

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
w_i	user i's (dynamic) weight
e_{ij}	normalized SINR 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

performed in terms sets of frequency bands in the frequency domain or with a granularity of multiple symbols in the time domain.

Let $\mathcal{R}(\boldsymbol{e}_t)$ denote the feasible rate region at time t, i.e.,

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

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

$$\mathcal{X} := \{(\boldsymbol{x}, \boldsymbol{p}) \geq \boldsymbol{0} : 0 \leq x_{ij} \leq 1, \forall i, j\}. \tag{4}$$

In practical OFDM systems, x_{ij} is constrained to be an integer, in which case we 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 to deal with such constraint are proposed in [1]. In this paper, we will ignore this integer constraint and focus on the rate region defined by (1) to (4). The corresponding solution will typically contain 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 the number of subchannels are large enough so that the subchannel SNRs do not change dramatically among adjacent subchannels, then the fractional value of x_{ij} can also implemented by frequency sharing (e.g., [11]).

Next we formulate the resource allocation problem, which is essentially a weighted rate maximization problem. In particular, the priority weights are motivated by the gradient-based scheduling framework presented in [3], [7], [8]. Each user i is assigned a utility function $U_i(W_{i,t}, Q_{i,t})$ depending on their average throughput $W_{i,t}$ up to time t and their queue-length $Q_{i,t}$ at time t. This is used to quantify fairness and ensure stability of the queues. During each scheduling epoch t, the system objective is to choose a rate vector $\boldsymbol{r}_t$ in $\mathcal{R}(\boldsymbol{e}_t)$ that maximizes a (dynamic) weighted sum of the users' rates, where the weights are determined by the gradient of the sum utility across all users at time t. More precisely, the scheduler seeks to maximize the projection of $\boldsymbol{r}_t$ onto the gradient

$$\nabla_{\boldsymbol{w}} U(\boldsymbol{W}_t, \boldsymbol{Q}_t) - \nabla_{\boldsymbol{q}} U(\boldsymbol{W}_t, \boldsymbol{Q}_t),$$

where

$$U(\boldsymbol{W}_t, \boldsymbol{Q}_t) = \sum_{i=1}^{K} U_i(W_{i,t}, Q_{i,t}).$$

We further assume that for each user i,

$$U_i(W_{i,t}, Q_{i,t}) = u_i(W_{i,t}) - \frac{d_i}{p}(Q_{i,t})^p,$$

where $u_i(W_{i,t})$ is an increasing concave function, $d_i \geq 0$ is a QoS weight for user i's queue length, and $p > 1$ is a fairness parameter associated with the queue length. Hence, the resource allocation decision is the solution to

$$\begin{aligned} &\max_{\boldsymbol{r}_t \in \mathcal{R}(\boldsymbol{e}_t)} (\nabla_{\boldsymbol{w}} U(\boldsymbol{W}_t, \boldsymbol{Q}_t)^T - \nabla_{\boldsymbol{q}} U(\boldsymbol{W}_t, \boldsymbol{Q}_t)^T) \cdot \boldsymbol{r}_t \\ &= \max_{\boldsymbol{r}_t \in \mathcal{R}(\boldsymbol{e}_t)} \sum_i \left(\frac{\partial u_i(W_{i,t})}{\partial W_{i,t}} + d_i (Q_{i,t})^{p-1} \right) r_{i,t}. \end{aligned} \tag{5}$$

Several variations of the policy in (5) have been studied. If $d_i = 0$ for all $i \in \mathcal{M}$, the resulting policy has been shown to yield utility maximizing solutions [3], [7], [8]. A specific choice of d_i for "usual" utility functions $u_i(\cdot)$ has been shown to produce utility maximizing solutions subject to stability [6].

As a concrete example, one class of utility functions typically used (e.g. [2], [9]) for $u_i(\cdot)$ is

$$u_i(W_{i,t}) = \begin{cases} \frac{c_i}{\alpha}(W_{i,t})^\alpha, & \alpha \leq 1, \ \alpha \neq 0 \\ c_i \log(W_{i,t}), & \alpha = 0, \end{cases} \tag{6}$$

where $\alpha \leq 1$ is a fairness parameter and $c_i \geq 0$ is a QoS weight. In this case, the objective in (5) becomes

$$\sum_i \left(c_i (W_{i,t})^{\alpha - 1} + d_i (Q_{i,t})^{p-1} \right) r_{i,t}.$$

With zero queue weights d_i and equal throughput weights c_i, setting $\alpha = 1$ results in a "maximum throughput" scheduling rule that maximizes the total throughput during each slot. For $\alpha = 0$, this results in the proportional fair rule [5].

The optimization in (5) can be written as

$$\max_{\boldsymbol{r}_t \in \mathcal{R}(\boldsymbol{e}_t)} \sum_i w_{i,t} r_{i,t}, \tag{7}$$

where $w_{i,t} \geq 0$ is a time-varying weight assigned to the ith user at time t. Our focus in this paper is on solving such a problem for an uplink OFDM system, i.e., when $\mathcal{R}(\boldsymbol{e}_t)$ is given by (1). For simplicity we will drop the time index t.

III. Primal-Dual Algorithm to Find Optimal Solution

Before solving problem in (7), let us first rewrite it in variables $\boldsymbol{x}$ and $\boldsymbol{p}$ directly instead of in rate $\boldsymbol{r}$. Problem (7) can be written as

$$\max_{(\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X}} \sum_{i \in \mathcal{M}} w_i \sum_{j \in \mathcal{N}} x_{ij} \log \left(1 + \frac{p_{ij} e_{ij}}{x_{ij}}\right) \tag{8}$$

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

It can be shown that the objective function of problem in (8) is continuous and concave over the constraint set $\mathcal{X}$, thus there is no duality gap between it and its dual problem. However, the objective's derivative is not well defined at the origin. This motivates us to look at the following ϵ-relaxed version of the problem in (8):

$$\max_{(\boldsymbol{x},\boldsymbol{p})\in\mathcal{X}} \sum_{i\in\mathcal{M}} w_i \sum_{j\in\mathcal{N}} (x_{ij}+\epsilon_{ij}) \log\left(1+\frac{p_{ij}e_{ij}}{x_{ij}+\epsilon_{ij}}\right), \quad (9)$$

where constants ϵ_{ij} takes small positive value for all i and j. The constraint set remains the same.

By such relaxation, the objective function in the new problem in (9) now has derivative defined everywhere in the constraint set $\mathcal{X}$. Thanks to the continuity of the objective function, the optimal value to the relaxed problem in (9) can be arbitrarily close to that of the original problem in (8), if $\epsilon = [\epsilon_{ij}, \forall i, j]$ is chosen to be small enough.

The existence of derivatives allows us to write down a primal-dual algorithm to pursue the optimal solution to the relaxed problem in (9). The Lagrangian for the relaxed problem in (9) is as follows,

$$L(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{x},\boldsymbol{p}) := \sum_{i,j} w_i (x_{ij}+\epsilon_{ij}) \log\left(1+\frac{p_{ij}e_{ij}}{x_{ij}+\epsilon_{ij}}\right) + \sum_i \lambda_i \left(P_i - \sum_j p_{ij}\right) + \sum_j \mu_j \left(1-\sum_i x_{ij}\right). \quad (10)$$

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

$$\begin{aligned}
u_j \ge 0, \ \sum_i x_{ij} \le 1, \ u_j\left(\sum_i x_{ij} - 1\right) &= 0,\\
\lambda_i \ge 0, \ \sum_j p_{ij} \le P_i, \ \lambda_i\left(\sum_j p_{ij} - P_i\right) &= 0,\\
x_{ij} \ge 0, \ p_{ij} &\ge 0,\\
x_{ij}\left(w_i \log\left(1+\frac{p_{ij}e_{ij}}{x_{ij}+\epsilon_{ij}}\right) - \frac{w_i p_{ij} e_{ij}}{x_{ij}+\epsilon_{ij}+p_{ij}e_{ij}} - u_j\right) &\le 0,\\
p_{ij}\left(\frac{w_i e_{ij}(x_{ij}+\epsilon_{ij})}{x_{ij}+\epsilon_{ij}+p_{ij}e_{ij}} - \lambda_i\right) &\le 0.
\end{aligned}$$

The last two inequalities become equalities if $x_{ij} > 0$ and $q_{ij} > 0$, respectively. The points satisfying above KKT conditions are exactly the saddle points of the Lagrangian function in (10). Since the primal problem has at least one solution, these saddle points exist.

Define $(a)^+ = \max(a, 0)$, and

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

For ease of use in following sections, we define

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

and

$$g_{ij}(x_{ij}, p_{ij}) = \frac{w_i(x_{ij}+\epsilon_{ij})e_{ij}}{x_{ij}+\epsilon_{ij}+p_{ij}e_{ij}}.$$

To pursue the saddle points of the Lagrangian function, we consider the following primal-dual algorithm: $\forall i, j$,

$$\begin{aligned}
\dot{x}_{ij} &= k^x_{ij}\left(f_{ij}(x_{ij},p_{ij}) - u_j\right)^+_{x_{ij}}, && (11)\\
\dot{p}_{ij} &= k^p_{ij}\left(g_{ij}(x_{ij},p_{ij}) - \lambda_i\right)^+_{p_{ij}}, && (12)\\
\dot{u}_j &= k^u_j\left(\sum_i x_{ij} - 1\right)^+_{u_j}, && (13)\\
\dot{\lambda}_i &= k^\lambda_i\left(\sum_j p_{ij} - P_i\right)^+_{\lambda_i}, && (14)
\end{aligned}$$

where $k^x_{ij}, k^p_{ij}, k^u_j$ and k^λ_i are constants representing adaption rates. It is easy to verify that equilibria of the above system are the wanted saddle points.

Applying primal-dual algorithms to solve *non-strictly* concave optimization problem in general encounters challenge in guaranteeing its convergence. For instance, it has been shown that primal-dual algorithms can fail to converge in a non-strictly concave optimization setting in a communication network setting [14]. In the next section, we study dynamics of the above primal-dual algorithm and show it does not converge in general. Then we derive a sufficient condition for its convergence. Utilizing this condition, we put constraints on adaptation rates of our proposed primal-dual algorithm to warranty its convergence.

IV. Convergence of the Primal-dual Algorithm

In this section, we study convergence of the proposed primal-dual algorithm. We first show the trajectories converge to an invariant set that contains all wanted saddle points.

Theorem 1: All trajectories of the non-linear primal-dual system in (11) to (14) converge to an invariant set V_0 globally asymptotically. Furthermore, let $(x^*, p^*, \lambda^*, u^*)$ be any saddle point of the Lagrangian function in (10), the following is true on V_0 and any point $(x, p, \lambda, u) \in V_0$,

- $(x^*, p^*, \lambda^*, u^*)$ is contained in V_0;
- λ_i is nonzero only if $\sum_j p^*_{ij} = P_i$;
- u_j is nonzero only if $\sum_i x^*_{ij} = 1$;
- $\frac{p_{ij}}{x_{ij}+\epsilon_{ij}} = \frac{p^*_{ij}}{x^*_{ij}+\epsilon_{ij}}$.

Proof: Let $(x^*, p^*, u^*, \lambda^*)$ be one point satisfying the KKT condition. Motivated by the Lyapunov function used in [12], we consider the following La salle function

$$\begin{aligned}
&V(x,p,u,\lambda)\\
&= \sum_{i,j} \frac{1}{k^x_{ij}} \int_0^{x_{ij}} (\xi - x^*_{ij})\, d\xi + \sum_{i,j} \frac{1}{k^p_{ij}} \int_0^{p_{ij}} (\xi - p^*_{ij})\, d\xi\\
&+ \sum_j \frac{1}{k^u_j} \int_0^{u_j} (\xi - u^*_j)\, d\xi + \sum_i \frac{1}{k^\lambda_j} \int_0^{\lambda_i} (\xi - \lambda^*_i)\, d\xi.
\end{aligned}$$

It is straightforward to verify that V is semi-positive definite. Its Lee derivative over an invariant set $\{(x, p, u, \lambda) | x \ge 0, p \ge$

$0, u \geq 0, \lambda \geq 0\}$ is given by

$$\dot{V} = \sum_{i,j} \frac{\partial V}{\partial x_{ij}} \dot{x}_{ij} + \sum_{i,j} \frac{\partial V}{\partial p_{ij}} \dot{p}_{ij} + \sum_{j} \frac{\partial V}{\partial u_j} \dot{u}_j + \sum_{i} \frac{\partial V}{\partial \lambda_i} \dot{\lambda}_i$$

$$\leq \sum_i (\lambda_i - \lambda_i^*) \left(\sum_j p_{ij}^* - P_i \right) + \sum_j (u_j - u_j^*) \left(\sum_i x_{ij}^* - 1 \right) + \sum_{i,j} \left([f_{ij}^*, g_{ij}^*] - [u_j^*, \lambda_i^*] \right) \begin{bmatrix} x_{ij} - x_{ij}^* \\ p_{ij} - p_{ij}^* \end{bmatrix} + \sum_{i,j} \left([f_{ij}, g_{ij}] - [f_{ij}^*, g_{ij}^*] \right) \begin{bmatrix} x_{ij} - x_{ij}^* \\ p_{ij} - p_{ij}^* \end{bmatrix}.$$

The first sum is non-positive, since $\sum_j p_{ij}^* \leq P_i$ and $\lambda_i^* = 0$ when $\sum_j p_{ij}^* < P_i$. Following the same logic, we determine the second sum to be non-positive.

The third sum is also non-positive since $f_{ij}(x_{ij}^*, p_{ij}^*) \leq u_j^*$, and the inequality holds only when $x_{ij}^* = 0$; similarly, $g_{ij}(x_{ij}^*, p_{ij}^*) \leq \lambda_i^*$, and the inequality holds only when $p_{ij}^* = 0$. For $x_{ij}^* \neq 0$, we have $p_{ij}^* \neq 0$ and

$$[f_{ij}(x_{ij}^*, p_{ij}^*), g_{ij}(x_{ij}^*, p_{ij}^*] = [u_j^*, \lambda_i^*],$$

and $[f_{ij}(x_{ij}, p_{ij}), g_{ij}(x_{ij}, p_{ij}]^T$ is the gradient of following function

$$H_{ij}(x_{ij}, p_{ij}) = w_i (x_{ij} + \epsilon_{ij}) \log \left(1 + \frac{p_{ij} e_{ij}}{x_{ij} + \epsilon_{ij}} \right).$$

By concavity of the function H_{ij}, we have

$$H_{ij}(x_{ij}, p_{ij}) \leq H_{ij}(x_{ij}^*, p_{ij}^*) + \nabla H_{ij}(x_{ij}^*, p_{ij}^*) \begin{bmatrix} x_{ij} - x_{ij}^* \\ p_{ij} - p_{ij}^* \end{bmatrix}, \quad (15)$$

$$H_{ij}(x_{ij}^*, p_{ij}^*) \leq H_{ij}(x_{ij}, p_{ij}) + \nabla H_{ij}(x_{ij}, p_{ij}) \begin{bmatrix} x_{ij}^* - x_{ij} \\ p_{ij}^* - p_{ij} \end{bmatrix}. \quad (16)$$

Consequently, the inner product of the gradient difference and the variable difference is non-positive, i.e.

$$\left([f_{ij}(x_{ij}, p_{ij}), g_{ij}(x_{ij}, p_{ij})] - [f_{ij}^*, g_{ij}^*] \right) \begin{bmatrix} x_{ij} - x_{ij}^* \\ p_{ij} - p_{ij}^* \end{bmatrix} \leq 0.$$

Hence, the fourth sum is also non-positive.

As such, we have $\dot{V} \leq 0$. According to La Salle principle [13], trajectories of the system in (11) to (14) converge to the set $V_0 = \{(x, p, \lambda, u) : \dot{V} = 0\}$ globally asymptotically. Over set V_0, we have the following observations:

- λ_i is nonzero only if $\sum_j p_{ij}^* = P_i$;
- u_j is nonzero only if $\sum_i x_{ij}^* = 1$;
- $\left([f_{ij}, g_{ij}] - [f_{ij}^*, g_{ij}^*] \right) \begin{bmatrix} x_{ij} - x_{ij}^* \\ p_{ij} - p_{ij}^* \end{bmatrix} = 0.$

Combining the third observation with Eqn. (15) and (16), we further know that $\forall (x, p, \lambda, u) \in V_0$,

$$H_{ij}(x_{ij}, p_{ij}) = H_{ij}(x_{ij}^*, p_{ij}^*) + \nabla H_{ij}(x_{ij}^*, p_{ij}^*) \begin{bmatrix} x_{ij} - x_{ij}^* \\ p_{ij} - p_{ij}^* \end{bmatrix}.$$

Taking the derivative of both sides, we have $\forall (x, p, \lambda, u) \in V_0$,

$$\begin{bmatrix} f_{ij}(x_{ij}, p_{ij}) \\ g_{ij}(x_{ij}, p_{ij}) \end{bmatrix} = \nabla H_{ij}(x_{ij}^*, p_{ij}^*) = \text{constant}.$$

Multiplying a vector $[1, -1/e_{ij}]$ on both sides of the above equation, we get

$$f_{ij}(x_{ij}, p_{ij}) - \frac{1}{e_{ij}} g_{ij}(x_{ij}, p_{ij}) = w_i \log \left(1 + \frac{p_{ij} e_{ij}}{x_{ij} + \epsilon_{ij}} \right) - w_i = f_{ij}(x_{ij}^*, p_{ij}^*) - \frac{1}{e_{ij}} g_{ij}(x_{ij}^*, p_{ij}^*) = \text{constant}.$$

At the end, we get $p_{ij} = (x_{ij} + \epsilon_{ij}) \frac{p_{ij}^*}{x_{ij}^* + \epsilon_{ij}}$. ■

Although all trajectories of the primal-dual system in (11) to (14) converge to V_0 that contains the wanted equilibria, they might not be able to converge to any of these equilibria. This is because V_0 may also contain non-equilibrium points, in particular, limit cycles. If a trajectory converges to limit cycles in V_0, then it never has chance to converge to the equilibria. Next we make this precise and also give sufficient conditions under which this non-convergent behavior will not happen.

Recall that M is the total number users and N is the total number of subchannels. Over set V_0, the primal-dual system in (11) to (14) turns into a linear one as follows:

$$\begin{bmatrix} \dot{x} \\ \dot{p} \end{bmatrix} = K_1 A^T \begin{bmatrix} u^* \\ \lambda^* \end{bmatrix} - K_1 A^T \begin{bmatrix} u \\ \lambda \end{bmatrix}, \quad (17)$$

$$\begin{bmatrix} \dot{u} \\ \dot{\lambda} \end{bmatrix} = K_2 A \begin{bmatrix} x \\ p \end{bmatrix} - K_2 \begin{bmatrix} \mathbf{1} \\ P \end{bmatrix}, \quad (18)$$

$$B \begin{bmatrix} x + \epsilon \\ p \end{bmatrix} = 0, \quad (19)$$

where K_1 and K_2 are $2MN \times 2MN$ diagonal matrices given by

$$K_1 = \begin{bmatrix} K^x & 0 \\ 0 & K^p \end{bmatrix}, \quad K_2 = \begin{bmatrix} K^u & 0 \\ 0 & K^\lambda \end{bmatrix},$$

B is a $MN \times 2MN$ matrix given by

$$B = [C, -D], \quad C = diag(c_{ij}, \forall i, j),$$

where $c_{ij} = \frac{p_{ij}^*}{x_{ij}^* + \epsilon_{ij}}$,

$$D = diag(Ind(c_{ij}), \forall i, j),$$

where $Ind(x)$ takes value 1 if $x > 0$ and 0 otherwise, and A is a $(2MN) \times (M + N)$ matrix given by

$$A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix},$$

where A_1 is an $N \times MN$ matrix and is given by $A_1 = [I_N, \ldots, I_N]$, and A_2 is an $M \times MN$ matrix given by

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

Here I_N is an identity matrix with dimension N, and $\mathbf{1}_{1\times N}$ is an all one vector with dimension 1 by N.

For the above linear system, we have the following.

Lemma 1: For the linear system in (17) to (19), the following is true:

1) Every order Lee derivative of $B[x,p]^T$ is constant, i.e.,

$$BK_1A^T \begin{bmatrix} \frac{d^q x}{dt^q} \\ \frac{d^q p}{dt^q} \end{bmatrix} = \text{const},\ \forall q = 1, 2, \cdots .$$

2) Its trajectories do not converge.

Proof: (Sketch) The first observation can be easily derived from (17) to (19). For the second observation, it is straightforward to verify the transfer function matrix of the linear system is a product of positive diagonal matrix and a skew-symmetric matrix. Hence, all eigenvalues of the transfer matrix are purely imaginary. ■

According to Lemma 1, if a trajectory of the primal-dual system converges to limit cycles in V_0, then then it never has a chance to converge to the wanted equilibria. Therefore, trajectories of the primal-dual system in (11) to (14) converge to equilibria of the system, if and only if V_0 only contains equilibria of the system.

On the other hand, Lemma 1 states that every order Lee derivative of $B[x,p]^T$ is constant. Consequently, if λ and u in V_0 are completely observable from the constant $B[x,p]^T$ through the linear system in (17) to (19), then $\dot{\lambda} = 0$ and $\dot{u} = 0$. If all λ and u in V_0 satisfy $\dot{\lambda} = 0$ and $\dot{u} = 0$, then all x and p in V_0 satisfy $\dot{x} = 0$ and $\dot{p} = 0$, and V_0 contains only equilibria of the primal-dual system. We state conditions for λ and u to be completely observable from $B[x,p]^T$ and its consequence on system convergence in the following theorem.

Theorem 2: All trajectories of the primal-dual system in (11) to (14) converge to equilibria of the system globally asymptotically if the following conditions hold: for any eigenvalue of matrix $K_2AK_1A^T$, denoted by v,

$$\begin{bmatrix} BK_1A^T \\ K_2AK_1A^T - vI \end{bmatrix} \text{ has rank } M+N. \qquad (20)$$

Proof: By linear system theory, λ and u are completely observable from $B[x,p]^T$ if and only if the complete observability conditions, expressed by (20), hold [13]. If λ and u are completely observable from the constant $B[x,p]^T$, then $\dot{\lambda} = 0$ and $\dot{u} = 0$. According to (17), x and p in V_0 satisfy $\dot{\lambda} = 0$ and $\dot{u} = 0$. As such, V_0 contains only equilibria of the primal-dual system, and Theorem 1 guarantees convergence of its trajectories. ■

For the problem we studied in this paper, we verify that above conditions in (20) can be satisfied, by designing the adaptation rates of the algorithm.

In particular, we choose K_1 to be kI where k is a positive constant, and K_2 to be of some form that will be clear later.

By direct computation, we get

$$BK_1A^T = k\left[CA_1^T, -A_2^T\right].$$

We observe that BK_1A^T has at least rank $M+N-1$. This is because CA_1^T has N linearly independent columns, $-A_2^T$ has M linearly independent columns, and any $M+N-1$ columns of the matrix are linearly independent. $K_2AK_1A^T = kK_2AA^T$ is a diagonal matrix given by:

$$k\begin{bmatrix} K^u & 0 \\ 0 & K^\lambda \end{bmatrix} \begin{bmatrix} M & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & 0 & \vdots & \cdots & \vdots \\ 0 & \cdots & M & 0 & \cdots & 0 \\ 0 & \cdots & 0 & N & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & N \end{bmatrix}.$$

As long as either K^u or K^λ is not the product of a constant and an identity matrix, then $kK_2AA^T - vI$ has at least one non-zero row for any eigenvalue v of matrix $K_2AK_1A^T$.

We combine this non-zero row with BK_1A^T to form a new matrix. By doing so, it is straightforward to verify the $M+N$ columns of this new matrix is linearly independent; hence, it has rank $M+N$. We summarize the above analysis into the following Corollary.

Corollary 1: Conditions in (20) in Theorem 2 are satisfied if the following is true:

- $K^x = kI$ and $K^p = kI$ for some positive constant k (diagonal terms of K^x and K^p take the same value k);
- K^u or K^λ is not product of a constant and an identity matrix.

Following the above choice on on adaptation rates, trajectories of (x,p) of the primal-dual system in (11) to (14) converge to the optimal solutions of the problem in (9).

Finally, we comment on the message passing needed in the primal-dual algorithm. The primal-dual algorithm can be implemented in a distributed fashion by mobile users and the base station. A mobile user i is responsible of updating x_{ij} and p_{ij} for all channel j as well as dual variable λ_i locally. It also needs to send the latest value of x_{ij}'s to the base station, but not the p_{ij}'s and λ_i. The base station is responsible for updating dual variables μ_j for all channel j locally and broadcasting to the users. The total communication overhead per iteration would be $(M+1)N$ messages. In particular, there is no need for the base station to know the weights and power constraints of the individual users.

V. Simulation Results

We test the convergence and optimality of the primal-dual algorithm over a realistic OFDMA uplink simulator. We consider a single OFDM cell model. Each user's subchannel gains are the product of 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 5MHz corresponding to 512 OFDM

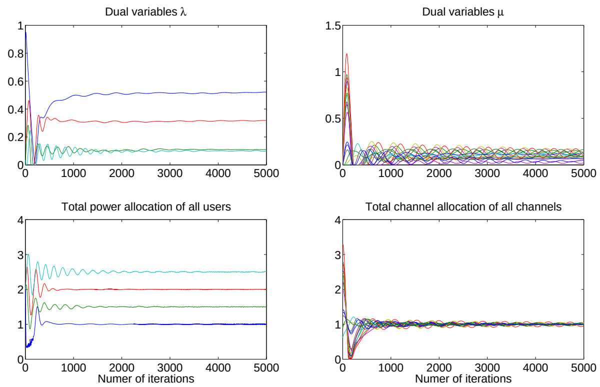

Fig. 1. Primal and Dual Variable Convergence of Primal-Dual Algorithm

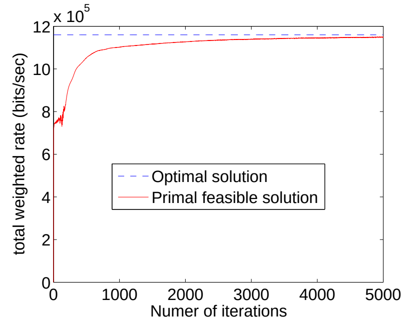

Fig. 2. Total Weighted Rate Convergence of Primal-Dual Algorithm

tones. Resource allocation is performed using adjacent groups of 32 tones[3], thus the total number of subchannels is 16. The symbol duration is $100\mu\text{sec}$ with a cyclic prefix of $10\mu\text{sec}$.

Next we show the simulation results the primal-dual algorithm with an example of 4 users. We also simulate the algorithm with more users (40) and subchannels (64), but will not show the results here due to space limitations. Here the weights of the users are randomly generated from $[0, 1]$, and the channel conditions are randomly generated from the above simulator. Four users have total power constraints of 1w, 1.5w, 2w and 2.5w, respectively. The update stepsizes in (11) to (14) are chosen as $k_{ij}^x = k_{ij}^p = k_{ij}^\lambda = k_{ij}^\mu = 0.01$ for all i and j. The initial values of primal and dual variables are randomly generated. We choose $\epsilon_{ij} = 10^{-4}$ for all i and j. Figure 1 shows the convergence of the dual variables (λ_i and μ_j) as well as the total power allocation of each user ($\sum_j p_{ij}$ for each i) and total channel allocation on each channel ($\sum_i x_{ij}$ for each j). It is clear that the system has converged to a neighborhood of the optimal solution. Figure 2 shows the convergence of the total weighted sum rate computed by the primal-dual algorithm (i.e., the primal feasible solution) under the ϵ-approximation model and the optimal value (calculated by the centralized optimal algorithm in [1]). The primal-dual algorithm achieves 90% of the optimal performance within 500 iterations and 95% within 1000 iterations. We have observed in our simulations that the convergence time is heavily dependent on the choice of stepsizes. Larger stepsizes can increase the convergence speed while leading to more fluctuations of the variables around the optimal solution.

VI. Conclusion and Future Work

OFDM has become a key technology for various wireless broadband access systems. In this paper, we presented the first distributed and low complexity optimal resource allocation algorithm for uplink OFDM systems. The key features of the proposed algorithm include: (i) distributed implementation at the mobile users and the base station schedular, (ii) simple local updates with low message passing overhead, and (iii) global convergence despite of the existence of multiple optimal solutions. From simulations we observed that the actual convergence time of the proposed algorithm can be long and is heavily dependent on the choices of stepsizes. One future work direction is to design good stepsize choice rules to achieve faster and more robust convergence.

[3]This corresponds to the "Band AMC mode" of 802.16 d/e.

Acknowledgement

We want to acknowledge helpful discussions with Professor Randall Berry from Northwestern University, Dr. Vijay G. Subramanian from National University of Ireland, and Dr. Rajeev Agrawal from Motorola Inc.

References

[1] J. Huang, V. Subramanian, R. Berry and R. Agrawal, “”Joint Scheduling and Resource Allocation in Uplink OFDM Systems,” *Proc. of Asilomar Conference*, Pacific Grove, CA, Nov. 2007.

[2] R. Agrawal, A. Bedekar, R. La and V. Subramanian, “A Class and Channel-Condition based Weighted Proportionally Fair Scheduler,” *Proc. of ITC 2001*, Salvador, Brazil, Sept. 2001.

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

[4] J. Huang, V. G. Subramanian, R. Agrawal and R. Berry, “Downlink scheduling and resource allocation for OFDM systems,” under revision of *IEEE Trans. on Wireless*, 2008.

[5] A. Jalali, R. Padovani, R. Pankaj, “Data throughput of CDMA-HDR a high efficiency - high data rate personal communication wireless system.,” in *Proc. VTC ’2000*, Spring, 2000.

[6] A. L. Stolyar, “Maximizing Queueing Network Utility subject to Stability: Greedy Primal-Dual Algorithm,” *Queueing Systems*, Vol. 50, pp. 401–457, 2005.

[7] A. L. Stolyar, “On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation,” Operations Research, Vol. 53, No. 1, pp. 125, 2005.

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

[9] J. Mo and J. Walrand, “Fair end-to-end window-based congestion control,” *IEEE/ACM Transactions on Networking*, Vol. 8, Vo. 5, pp. 556-567, October 2000.

[10] S. Pfletschinger, G. Muenz, and J. Speidel, “Efficient subcarrier allocation for multiple access in OFDM systems,” in *7th International OFDM-Workshop 2002 (InOWo’02)*, 2002.

[11] W. Yu, R. Lui, and R. Cendrillon, “Dual optimization methods for multiuser orthogonal frequency division multiplex systems,” in *Proceedings of IEEE Globecom*, vol. 1, 2004, pp. 225–229.

[12] S. Liu, T. Basar, and R. Srikant, “Exponential-RED: A Stabilizing AQM Scheme for Low- and High-Speed TCP Protocols,” in *IEEE Trans. on Networking*, vol. 5, 2005.

[13] H. Khalil, “Nonlinear Systems (3rd edition),” Prentice Hall, 2001.

[14] 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.