CONVERGENCE ANALYSIS OF SADDLE POINT PROBLEMS IN TIME VARYING WIRELESS SYSTEMS - CONTROL THEORETICAL APPROACH

Junting Chen and Vincent K. N. Lau

Department of Electronic and Computer Engineering
The Hong Kong University of Science and Technology, Hong Kong

ABSTRACT

Saddle point problems arise from many wireless applications, and primal-dual iterative algorithms are widely applied to find the saddle points. In the existing literature, the convergence results of such algorithms are established assuming the problem specific parameters remain unchanged during the iterations. However, this assumption is unrealistic in time varying wireless systems, as explicit message passing is usually involved in the iterations and the channel state information (CSI) may change in a time scale comparable to the algorithm update period. This paper investigates the convergence behavior and the tracking error of primal-dual iterative algorithms under time varying CSI. The convergence results are established by studying the stability of an equivalent virtual dynamic system derived in the paper, and the Lyapunov theory is applied for the stability analysis. We show that the average tracking error is proportional to the time variation rate of the CSI. We also derive an adaptive primal-dual algorithm by introducing a compensation term to reduce the tracking error under the time varying CSI.

***Index Terms*—** Saddle Point, Convergence Analysis, Time-Varying, Lyapunov Stability, Convex Optimization

1. INTRODUCTION

Saddle point problems arise in a number of wireless communication applications such as competitive games and resource allocations. Resource allocation problems can be formulated as a constrained maximization of some utility functions. By constructing a Lagrangian function, the constrained problem can be reformulated into an unconstrained one and be solved by computing the saddle point of the Lagrangian function. Most remarkably, primal-dual gradient methods have been widely used for computing the saddle points of general Lagrangian functions. The primal-dual gradient methods update the primal and dual variables simultaneously by evaluating the gradient of both the primal function and dual function at the same time. A classical study of primal-dual algorithms has been done by Arrow *et al.* in work [1]. Recently, Feijer *et al.* [2] have studied the stability of these primal-dual algorithm dynamics and extended the results to various network resource allocations problems.

In the above literature, when people discuss the convergence behavior of the primal-dual gradient algorithm, all the problem specific parameters are considered to be time invariant. However, this assumption is not always realistic, especially in wireless communication scenarios. For instance, the operating environment in terms of the channel state information (CSI) may be changing frequently, and, as a result, the optimization problem varies from time to time. On the other hand, as *explicit message passing* may be involved in the iterations, the optimization algorithms cannot always converge fast enough to catch up with time varying effects, especially for large scale problems. As is shown in our numerical example, convergence errors would lead to performance loss in a wireless communication system. However, it is yet unknown whether the algorithms converge or not when the static assumption is dropped, even for strictly convex problems. Therefore, it is highly important to the study convergence behavior, or robustness, for primal-dual gradient algorithms under time varying CSI.

However, towards this end there are a lot of technical challenges, such as how to quantify the performance penalty due to the time varying parameters, how to evaluate the cost-performance tradeoff, and how to enhance the algorithm. These difficulties are highly nontrivial due to the stochastic nature of wireless communication problems as well as the complexities of the algorithms that solve them. Although there are some preliminary works studying the effects caused by time varying parameters [3, 4], the authors in [3] and [4] did not consider the CSI being time varying and hence their problems have static equilibrium points.

In this paper, we shall investigate the convergence behavior of primal-dual algorithms for solving general saddle point problems under time varying CSI. We first define an equivalent virtual dynamic system, following which we study its stability based on the Lyapunov theory [5]. We model the dynamics of the time varying CSI as an auto-regressive system and derive the convergence properties of primal-dual algorithms under time varying CSI. We also quantify the average tracking errors in terms of the average exogenous excitations induced to the CSI dynamics. Furthermore, we propose a novel adaptive algorithm to enhance the tracking performance under time varying CSI.

2. SYSTEM MODEL AND VIRTUAL DYNAMIC SYSTEMS

In this section, we shall introduce a general saddle point problem setup and the notion of *virtual dynamic system* which will be used for stability analysis for the primal-dual algorithm.

2.1. General Saddle Point Problem

We consider a min-max optimization problem

$$\min_{\lambda \in \mathbb{R}_+^m} \max_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x},\, \lambda;\, \mathbf{h}) \tag{1}$$

The objective function $\mathcal{L}(\mathbf{x},\, \lambda;\, \mathbf{h})$ is strongly concave in $\mathbf{x} \in \mathbb{R}^n$ and convex in $\lambda \in \mathbb{R}_+^m$. $\mathbf{h} \in \mathcal{H} \subseteq \mathbb{R}^q$ is a vector parameter that arises from specific optimization problems. In the context of wireless communication optimizations, the problem parameter $\mathbf{h}$ can be the CSI. Under the above convexity assumption, the min-max problem has a unique optimal solution $(\mathbf{x}^*, \lambda^*) = (\psi_x(\mathbf{h}), \psi_\lambda(\mathbf{h}))$ where $\psi_x : \mathcal{H} \mapsto \mathbb{R}^n$ and $\psi_\lambda : \mathcal{H} \mapsto \mathbb{R}^m_+$ are $\mathcal{C}^1$ functions.

It is known that solving the above min-max optimization problem (1) is equivalent to computing the saddle point of $\mathcal{L}(\mathbf{x}, \lambda; \mathbf{h})$ [6]. For a given $\mathbf{h} \in \mathcal{H}$, a saddle point $(\mathbf{x}^*, \lambda^*)$ of $\mathcal{L}(\mathbf{x}, \lambda; \mathbf{h})$ is defined to be a vector that satisfies $\mathcal{L}(\mathbf{x}, \lambda^*; \mathbf{h}) \leq \mathcal{L}(\mathbf{x}^*, \lambda^*; \mathbf{h}) \leq \mathcal{L}(\mathbf{x}^*, \lambda; \mathbf{h})$ for all $\mathbf{x} \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m_+$. It is unique under the above convexity assumption of the function $\mathcal{L}(\centerdot)$ [6].

The classical primal-dual gradient algorithm dynamics for solving (1) are given in the following [1]

$$\dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt} = \kappa \left[\frac{\partial}{\partial \mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda; \mathbf{h}(t))\right]^T \quad (2)$$

$$\dot{\lambda} = \frac{d\lambda}{dt} = \kappa \left[\left(-\frac{\partial}{\partial \lambda} \mathcal{L}(\mathbf{x}, \lambda; \mathbf{h}(t))\right)^T\right]^+_\lambda \quad (3)$$

for some step size parameter $\kappa > 0$. The projection $[\centerdot]^+_\lambda$ is to restrict the dynamics $\lambda(t)$ in the nonnegative domain $\mathbb{R}^m_+$. For scalars u_i and λ_i, the projection is defined to be $[u_i]^+_{\lambda_i} := u_i$ if $u_i > 0$ or $\lambda_i > 0$, and $[u_i]^+_{\lambda_i} := 0$ otherwise. For the vector case, it is defined entry-wise.

In most of the existing works, the convergence results for the primal-dual algorithms solving saddle point problems were established based on the assumption that the wireless channel state $\mathbf{h}$ stays time invariant before the algorithm converges [1, 2]. In this paper, we address the situation where the CSI $\mathbf{h}(t)$ changes in a similar time scale as the primal-dual algorithm in (2)-(3).

2.2. Time Varying CSI Model

We model the CSI $\mathbf{h}(t)$ as a solution to the following ordinary differential equation (ODE),

$$\dot{\mathbf{h}}(t) = \frac{d\mathbf{h}(t)}{dt} = A(\mathbf{h}(t) - \bar{\mathbf{h}}) + u(t) \quad (4)$$

where A is a real symmetric negative definite matrix, $u(t)$ is a vector valued complex Gaussian process, and $\bar{\mathbf{h}}$ is a constant vector, corresponding to the line-of-sight (LOS) component in the channel model. Note that the dynamic model (4) resembles an AR(1) process and $|\mathbf{h}(t)|$ has a stationary Rician distribution. Similar CSI models are also studied in [7, 8].

2.3. Virtual Dynamic Systems

Let $\widetilde{\mathbf{x}} = (\mathbf{x}, \lambda)$ be the joint state of the primal and dual variables in algorithm (2)-(3) and denote the algorithm dynamics as $\widetilde{f} : \mathbb{R}^n \times \mathbb{R}^m_+ \mapsto \mathbb{R}^{n+m}$, where

$$\widetilde{f}(\widetilde{\mathbf{x}}; \mathbf{h}(t)) = \begin{bmatrix} \kappa \left(\frac{\partial}{\partial \mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda; \mathbf{h}(t))\right)^T \\ \kappa \left[\left(-\frac{\partial}{\partial \lambda} \mathcal{L}(\mathbf{x}, \lambda; \mathbf{h}(t))\right)^T\right]^+_\lambda \end{bmatrix}. \quad (5)$$

Definition 1 (Virtual Dynamic System and Equilibrium Point) *We refer to the following ODE as a* virtual dynamic system

$$\widetilde{\mathcal{X}} : \quad \dot{\widetilde{\mathbf{x}}} = \frac{d}{dt}\widetilde{\mathbf{x}}(t) = \widetilde{f}(\widetilde{\mathbf{x}}; \mathbf{h}(t)). \quad (6)$$

Moreover $\widetilde{\mathbf{x}}^*$ is called an equilibrium point of the vector field in (6)*, if $\widetilde{f}(\widetilde{\mathbf{x}}^*; \mathbf{h}(t)) = 0$.*

The virtual dynamic system (6) describes the algorithm dynamics under time varying parameters $\mathbf{h}(t)$. Note that the equilibrium point $\widetilde{\mathbf{x}}^*(\mathbf{h}(t))$ is a function of $\mathbf{h}(t)$ and hence time varying. As a result, we are interested in the instantaneous algorithm tracking error $\widetilde{\mathbf{x}}_e(t) = \widetilde{\mathbf{x}}(t) - \widetilde{\mathbf{x}}^*(t)$. Let $\psi(\mathbf{h}) := (\mathbf{x}^*(\mathbf{h}), \lambda^*(\mathbf{h})) = \widetilde{\mathbf{x}}^*(\mathbf{h})$ be a mapping from the CSI $\mathbf{h} \in \mathcal{H}$ to the equilibrium point $\widetilde{\mathbf{x}}^*(\mathbf{h}(t))$. As $\dot{\widetilde{\mathbf{x}}}_e = \dot{\widetilde{\mathbf{x}}} - \dot{\widetilde{\mathbf{x}}}^*$, we have an virtual error dynamic system $\widetilde{\mathcal{X}}_e$ defined as $\dot{\widetilde{\mathbf{x}}}_e = \widetilde{f}_e(\widetilde{\mathbf{x}}_e; \mathbf{h}(t)) - \varphi(\mathbf{h})\dot{\mathbf{h}}(t)$ where $\widetilde{f}_e(\widetilde{\mathbf{x}}_e; \mathbf{h}(t)) := \widetilde{f}(\widetilde{\mathbf{x}}_e + \psi(\mathbf{h}); \mathbf{h}(t))$ and $\varphi(\mathbf{h}) := \frac{\partial}{\partial \mathbf{h}}\psi(\mathbf{h})$.

It is straight forward to see that studying the convergence behavior of the primal-dual algorithm is equivalent to studying the stability of the corresponding virtual error dynamic system $\widetilde{\mathcal{X}}_e$.

3. CONVERGENCE ANALYSIS OF DEGRADED SADDLE POINT PROBLEMS UNDER TIME VARYING CSI

In this paper, we study saddle point functions $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}; \mathbf{h})$ that are strongly concave in primal variables $\mathbf{x}$ and convex in dual variables $\boldsymbol{\lambda}$. Specifically, we assume $\nabla^2_x \mathcal{L} \preceq -M_x \mathbf{I}$ for $M_x > 0$ and $\nabla^2_\lambda \mathcal{L} \succeq \mathbf{0}$. We call such a saddle point problem the *degraded saddle point problem* and the associated virtual dynamic systems the *degraded virtual dynamic systems*. Examples of degraded saddle point problems include the primal-dual iterations of constrained convex optimization problems. For instance, the Lagrangian function (16) of the NUM problem (15) is of this type.

We first consider the virtual dynamic system $\dot{\widetilde{\mathbf{x}}}_e = \widetilde{f}_e(\widetilde{\mathbf{x}}_e; \mathbf{h}(t)) - \varphi(\mathbf{h})\dot{\mathbf{h}}(t)$ with a quasi-time varying CSI $\mathbf{h}(t)$ in (4) for $u(t) = 0$. We define a partial state as $\mathbf{z}_e = (\mathbf{x}_e, \mathbf{h}_e)$. It satisfies the following virtual dynamic system $\mathcal{Z}_e$

$$\dot{\mathbf{z}}_e = \begin{bmatrix} f(\mathbf{x}_e + \mathbf{x}^*, \lambda_e = 0; \mathbf{h}_e + \bar{\mathbf{h}}) + \varphi_x(\mathbf{h}_e + \bar{\mathbf{h}})A\mathbf{h}_e \\ A\mathbf{h}_e \end{bmatrix} \quad (7)$$

where $\varphi_x(\mathbf{h}) = \frac{\partial}{\partial \mathbf{h}}\mathbf{x}^*(\mathbf{h}) = \frac{\partial}{\partial \mathbf{h}}\psi_x(\mathbf{h})$ and $f(\centerdot) = \kappa\left(\nabla_{\mathbf{x}}\mathcal{L}(\centerdot)\right)^T$ is the vector field of the algorithm dynamics of the primal variables $\mathbf{x}$. The following lemma summarizes the stability results for the above virtual system $\dot{\mathbf{z}}_e = \mathcal{Z}(\mathbf{z}_e)$.

Lemma 1 (Partial Exponential Stability) *Suppose the following inequality holds,*

$$\|\varphi_x(\mathbf{h})A\| < \kappa \min\{2M_x, -\lambda_{\max}(A)\} \quad (8)$$

for all $\mathbf{h} \in \mathcal{H}$, where $\lambda_{\max}(A)$ denotes the largest eigenvalue of CSI coefficient matrix A. Then the virtual dynamic system $\mathcal{Z}_e$ in (7) *is* partially exponentially stable*, i.e., there exists some positive constants k and a such that $\|\mathbf{z}_e\| \leq k\|\mathbf{z}_e(t_0)\|e^{-a(t-t_0)}$, for all $t \geq t_0$, and there exists a Lyapunov function for the joint state $\mathbf{z}_e$ satisfying*

$$a_1\|\mathbf{z}_e\|^2 \leq V(\mathbf{z}_e) \leq a_2\|\mathbf{z}_e\|^2, \quad \left\|\frac{\partial V}{\partial \mathbf{z}_e}\right\| \leq a_4\|\mathbf{z}_e\|, \quad (9)$$

$$\dot{V}(\mathbf{z}_e) = \frac{\partial V}{\partial \mathbf{z}_e} Z(\mathbf{z}_e) \leq -a_3\|\mathbf{z}_e\|^2. \quad (10)$$

Lemma 1 suggests that as long as the transient of quasi-time varying $\mathbf{h}(t)$ is not changing too fast (i.e. A has small eigenvalues) and the sensitivity of the primal part of the equilibrium $\mathbf{x}^*_e$ w.r.t. the change of $\mathbf{h}(t)$ is small (i.e. small $\|\varphi_x(\mathbf{h})\|$), the virtual dynamic system $\mathcal{Z}_e$ in (7) possesses globally exponential stability on the partial state $\mathbf{z}_e = (\mathbf{x}_e, \mathbf{h}_e)$. Please refer to [9] for the proof.

We now consider the degraded virtual dynamic system under the time varying CSI model in (4) as follows

$$\begin{aligned} \dot{\mathbf{z}}_e &= \begin{bmatrix} f(\boldsymbol{.}) + \varphi_x(\mathbf{h}_e + \bar{\mathbf{h}})A\mathbf{h}_e \\ A\mathbf{h}_e \end{bmatrix} + \begin{bmatrix} \varphi_x(\mathbf{h}_e + \bar{\mathbf{h}}) \\ \mathbf{I} \end{bmatrix} u(t) \\ &\triangleq \mathcal{Z}(\mathbf{z}_e) + \Phi_x(\mathbf{h}_e)u(t). \end{aligned} \tag{11}$$

The stability result of the degraded virtual dynamic system is summarized in the following theorem.

Theorem 1 (Stability of $\mathcal{Z}_e(u)$ for Time Varying CSI) *Given $\overline{\|\Phi_x(\mathbf{h}_e)\|} \le \gamma_x$, and $\overline{\|u(t)\|^2} \le \alpha^2$, the average trajectory $\overline{\|\mathbf{z}_e\|^2}$ of the degraded virtual dynamic system $\mathcal{Z}_e(u)$ satisfies*

$$\overline{\|\mathbf{z}_e\|^2} = \frac{1}{T}\int_0^T \|\mathbf{z}_e(t)\|^2 dt \le \frac{a_4^2\gamma_x^2}{a_3^2}\alpha^2.$$

The proof can be found in [9]. As a result, we summarize the convergence performance of the primal-dual iterative algorithm under time varying CSI for a degraded saddle point problem in the following corollary.

Corollary 1 *Suppose $\varphi_x(\mathbf{h}) \le \bar{\gamma}_x$ for all $\mathbf{h} \in \mathcal{H}$ and $\overline{\|u(t)\|^2} \le \alpha^2$, the average tracking error $\overline{\|\mathbf{x}_e\|^2}$ for the primal-dual algorithm under time varying CSI satisfies $\overline{\|\mathbf{x}_e\|^2} \le a_4^2(\bar{\gamma}_x^2 + 1)\alpha^2/a_3^2$.*

Together with Lemma 1, the above results establish sufficient conditions for the convergence of the a primal-dual algorithm under time varying CSI. It shows that the tracking error of the primal-dual algorithm under time varying CSI is bounded and scaled according to $\mathcal{O}\left(\overline{\|u(t)\|^2}\right)$, which specifies the variation and the time varying rate of the CSI $\mathbf{h}(t)$.

4. ADAPTIVE COMPENSATION FOR THE PRIMAL DUAL ALGORITHMS IN TIME VARYING CHANNELS

We have shown in Theorem 1 that the average tracking error of a primal-dual algorithm under time varying CSI is $\mathcal{O}(\alpha^2\gamma^2)$ where $\alpha^2\gamma^2$ is the square average of the norm of the exogenous input $\Phi(\mathbf{h}_e)u(t)$ to the virtual dynamic system $\mathcal{Z}_e$ in (7). As a result, one way to reduce the tracking error is to compensate the disturbance from the exogenous input $\Phi(\mathbf{h}_e)u(t)$. We thus introduce a compensation term $\widehat{\Phi}(\widetilde{\mathbf{z}}_e)u(t)$ to $\mathcal{Z}_e(u)$ in (7) as follows,

$$\widetilde{\mathcal{Z}}_e(\widehat{u}): \quad \dot{\widetilde{\mathbf{z}}}_e = \widetilde{\mathcal{Z}}(\widetilde{\mathbf{z}}_e) + \Phi(\mathbf{h}_e)u(t) - \widehat{\Phi}(\widetilde{\mathbf{z}}_e)u(t) \tag{12}$$

where $\widehat{\Phi}(\widetilde{\mathbf{z}}_e) = \begin{bmatrix} \widehat{\varphi}(\widetilde{\mathbf{x}}_e,\, \mathbf{h}_e) \\ \mathbf{I} \end{bmatrix}$, and $\widehat{\varphi}(\widetilde{\mathbf{x}}_e, \mathbf{h})$ is the compensation term to be derived. Obviously, if we could set $\widehat{\Phi}(\widetilde{\mathbf{z}}_e) = \Phi(\mathbf{h}_e)$, the impact of the exogenous input $\Phi(\mathbf{h}_e)u(t)$ can be totally suppressed. However, as we do not have the closed form expression for the saddle point $\widetilde{\mathbf{x}}^*(\mathbf{h})$, $\Phi(\mathbf{h}_e)$ cannot be obtained during the iteration.

Fortunately, for a convex optimization problem, we can always explicitly derive the optimality conditions $F(\widetilde{\mathbf{x}}^*; \mathbf{h}) = \mathbf{0}$ [6], where $F : \mathbb{R}^n \times \mathbb{R}_+^m \mapsto \mathbb{R}^{n+m} \in \mathcal{C}^1$. For example, we have $\widetilde{f}(\widetilde{\mathbf{x}}^*; \mathbf{h}(t)) = 0$ by the definition of equilibrium point, where $\widetilde{f}(\boldsymbol{.})$ is given in (5). Suppose $\frac{\partial}{\partial \widetilde{\mathbf{x}}}F(\widetilde{\mathbf{x}}^*; \mathbf{h})$ is non-singular. Using the implicit function theorem, $\widehat{\varphi}(\widetilde{\mathbf{x}}_e = \mathbf{0}, \mathbf{h}_e) \triangleq -\left(\frac{\partial}{\partial \widetilde{\mathbf{x}}^*}F(\widetilde{\mathbf{x}}^*; \mathbf{h})\right)^{-1}\frac{\partial}{\partial \mathbf{h}}F(\widetilde{\mathbf{x}}^*; \mathbf{h})$ represents the movement of the equilibrium point $\widetilde{\mathbf{x}}^*(\mathbf{h}(t))$. This can be further derived into $\Phi(\mathbf{h}_e)$, and hence we choose the compensation as $\widehat{\varphi}(\widetilde{\mathbf{x}}_e, \mathbf{h}) = -\left(\frac{\partial}{\partial \widetilde{\mathbf{x}}}F(\widetilde{\mathbf{x}}; \mathbf{h})\right)^{-1}\frac{\partial}{\partial \mathbf{h}}F(\widetilde{\mathbf{x}}; \mathbf{h})$, where we use the instantaneous algorithm state $\widetilde{\mathbf{x}}(t)$ as an approximation of the optimal target $\widetilde{\mathbf{x}}^*(t)$. As a result, the corresponding primal-dual algorithm iterations with compensation is given by

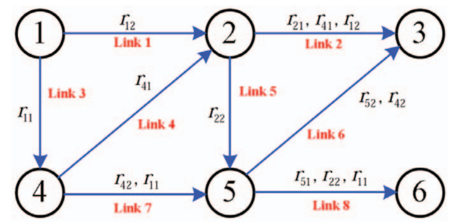

Fig. 1. A specific example of wireless ad hoc network. The interference at each receiving node is handled by multiuser detection (MUD) and successive decoding techniques.

$$\dot{\mathbf{x}} = \left[\frac{\partial}{\partial \mathbf{x}}\mathcal{L}(\boldsymbol{.};\, \mathbf{h}(t)) + \widehat{\varphi}_{\mathbf{x}}(\mathbf{x},\, \lambda;\, \mathbf{h}(t))^T\dot{\mathbf{h}}(t)\right]^T \tag{13}$$

$$\dot{\lambda} = \left[\left(-\frac{\partial}{\partial \lambda}\mathcal{L}(\boldsymbol{.};\, \mathbf{h}(t)) + \widehat{\varphi}_{\lambda}(\mathbf{x},\, \lambda;\, \mathbf{h}(t))^T\dot{\mathbf{h}}(t)\right)^T\right]_{\lambda}^{+} \tag{14}$$

where $\widehat{\varphi}_{\mathbf{x}}(\mathbf{x},\, \lambda;\, \mathbf{h}(t))$ and $\widehat{\varphi}_{\lambda}(\mathbf{x},\, \lambda;\, \mathbf{h}(t))$ are the primal and dual parts of the compensation term $\widehat{\varphi}(\widetilde{\mathbf{x}};\, \mathbf{h}(t))$.

The compensation terms in (13)-(14) can also be interpreted as a predictor on where the *saddle point* $\widetilde{\mathbf{x}}^*(\mathbf{h})$ moves. We summarize the performance of the proposed algorithm in the following theorem.

Theorem 2 (Compensation Algorithm Performance) *Suppose the compensation term $\widehat{\varphi}_{\mathbf{x}}(\mathbf{x},\, \lambda;\, \mathbf{h}(t))$ is Lipschitz continuous on $\mathbf{x}$ satisfying $\|\widehat{\varphi}_{\mathbf{x}}(\mathbf{x}(t),\, \lambda(t);\, \mathbf{h}(t) - \widehat{\varphi}_{\mathbf{x}}(\mathbf{x}^*(t),\, \lambda(t);\, \mathbf{h}(t))\| \le L_x\|\mathbf{x}_e(t)\|$ for all $\mathbf{h}(t) \in \mathcal{H}$ and $\boldsymbol{\lambda}(t) \in \mathbb{R}_+^m$, and $\overline{\|u(t)\|} = \beta < \frac{a_3}{a_4 L_x}$. Then the average tracking error $\overline{\|\mathbf{x}_e\|^2}$ of the proposed algorithm* (13)-(14) *asymptotically converges to 0.*

Note that the convergence results here are derived under continuous algorithm trajectories. For a discrete time algorithm, additional errors are introduced due to the time lag of the algorithm time slots. Hence the compensation algorithm in discrete time scale may not track the moving target with no errors. However, it still has significant performance gains against the traditional algorithms as demonstrated in the numerical results in the following section.

5. RESULTS AND DISCUSSIONS

In this section, we simulate the performance of the proposed compensation algorithm under a specific network utility maximization (NUM) problem where the network topology is depicted in Fig.1. The problem can be formulated as [3],

$$\begin{aligned} \max_{\mathbf{r}\succeq 0,\, \mathbf{p}\in\mathcal{P}} \quad & \textstyle\sum_{(s,\, d)\in\mathcal{C}} U_{sd}(r_{sd}) \\ \text{subject to} \quad & \textstyle\sum_{(s,\, d):\, l\in L(s,\, d)} r_{sd} \le c_l(\mathbf{p};\, \mathbf{h}) \quad \forall l \end{aligned} \tag{15}$$

where the ordered pairs $(s,\, d)$ denote the traffics, r_{sd} are the corresponding data rates, $U_{sd}(\boldsymbol{.})$ are the utility functions, $c_l(\boldsymbol{.})$ are link capacity functions and $L(s, d)$ is the set of links that traffic (s, d) goes through. The Lagrangian of (15) can be written as $\mathcal{L}(\mathbf{r}, \mathbf{p},\, \lambda;\, \mathbf{h}) =$

$$\sum_{(s,d)\in\mathcal{C}} U_{sd}(r_{(s,d)}) - \sum_{l}\lambda_l\left(\sum_{(s,d):\, l\in L(s,d)} r_{sd} - c_l(\mathbf{p};\, \mathbf{h})\right) \tag{16}$$

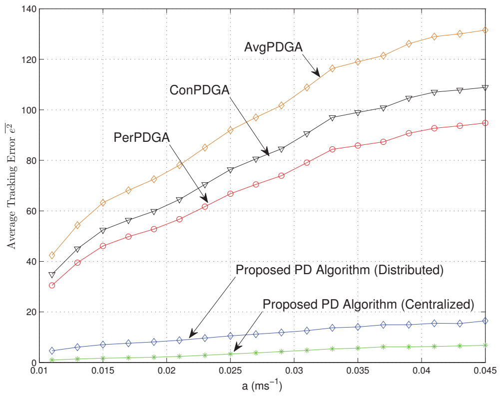

Fig. 2. The convergence performance comparison of the proposed algorithms and the baseline algorithms.

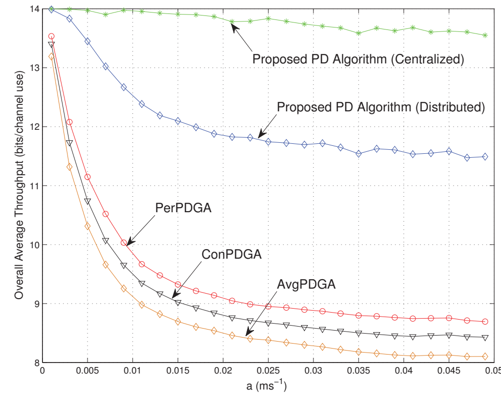

Fig. 3. Average network throughput versus the channel fading rate parameter a. The average throughput decreases when the fading rate a increases.

subjected to the power constraint $\mathbf{p} \in \mathcal{P}$. From the Lagrangian theory [6], solving (15) is equivalent to finding the saddle point of the Lagrangian (16).

The proposed compensated algorithm is given by (13)-(14). We also evaluate our proposed algorithm with a distributed implementation manner, in which we impose a block diagonal structure based on the network topology to the compensation term $\widehat{\varphi}(\widetilde{\mathbf{x}}; \mathbf{h}(t))$, as discussed in details in [9]. The three baselines are: (I) conventional primal-dual gradient algorithm (ConPDGA) [1, 2], i.e., the primal and dual variables are updated simultaneously in (2)-(3); (II) averaging primal-dual gradient algorithm (AvgPDGA) [10], i.e., the approximate primal solutions are generated by averaging over the past primal solutions; and (III) perturbed primal-dual gradient algorithm (PerPDGA) [11], i.e., the primal and dual variables are updated in the gradients evaluated at perturbed points that are generated via auxiliary mappings.

The comparison of the average tracking error $\overline{e^2}$ versus a in the time varying CSI model is shown in Fig.2. The tracking errors have been reduced greatly after introducing the compensation terms in both of the proposed algorithms.

Fig.3 shows the average network throughput versus the channel fading rate a. Due to the tracking errors under time varying CSI, transmission rates may exceed the channel capacity region and packet drops may occur leading to a degradation of throughput. The average throughput decreases when the fading rate a increases. The results also show that the proposed algorithms significantly outperform over all the other baselines.

6. CONCLUSIONS

In this paper, we have analyzed the convergence behavior of the primal-dual algorithm for solving a saddle point problem under time varying CSI. The convergence results have been derived by studying the stabilities of the equivalent virtual dynamic systems based on the Lyapunov theory from the control theoretical approach. We showed that the average tracking errors were given by $\mathcal{O}(\alpha^2)$, where α^2 represents the variation and the fading rate of the CSI dynamics. Based on these analyses, we have proposed a novel adaptive primal-dual algorithm with a predictive compensation to counteract the effects of the time varying CSI. We showed that the average tracking error of the proposed algorithm converges to zero despite time varying CSI. Numerical results were consistent with our analyses and the proposed algorithm demonstrated significantly better convergence performance over the baseline schemes.

7. REFERENCES

[1] Kenneth. J Arrow, Leonid Hurwicz, and H. Uzawa, *Studies in linear and non-linear programming*, Stanford University Press, California, Stanford, 1958.

[2] Diego Feijer and Fernando Paganini, “Stability of primal-dual gradient dynamics and applications to network optimization,” *Automatica*, vol. 46, pp. 1974–1981, Dec. 2010.

[3] FP Kelly, A Maulloo, and D Tan, “Rate control for communication networks: Shadow prices, proportional fairness and stability,” *Journal of Operation Research*, vol. 49, no. 3, pp. 237–252, 1998.

[4] Junshan Zhang, Junshan Zhang, Senior Member, Dong Zheng, and Mung" Chiang, “Impact of stochastic noisy feedback on distributed network utility maximization,” *IN INFOCOM 2007*, vol. 2008, pp. 222–230, 2007.

[5] Hassan K. Khalil, *Nonlinear Systems*, Prentice-Hall, 1996.

[6] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[7] K.E. Baddour and N.C. Beaulieu, “Autoregressive models for fading channel simulation,” in *Global Telecommunications Conference, 2001. IEEE GLOBECOM ’01.*, 2001, vol. 2, pp. 1187 –1192 vol.2.

[8] Tao (Stephen) Feng, Timothy R. Field, and Simon Haykin, “Stochastic Differential Equation Theory Applied to Wireless Channels,” *IEEE Transactions on Communications*, vol. 55, no. 8, pp. 1478–1483, Aug. 2007.

[9] Junting Chen and Vincent K.N. Lau, “Convergence analysis of saddle point problems in time varying wireless systems - control theoretical approach,” *to appear in IEEE Transactions on Signal Processing*, [Online] http://arxiv.org/abs/1109.2313.

[10] A. Nedić and A. Ozdaglar, “Subgradient methods for saddle-point problems,” *Journal of Optimization Theory and Applications*, vol. 142, pp. 205–228, 2009.

[11] M.J. Kallio and Andrzej Ruszczynski, “Perturbation methods for saddle point computation,” Working papers, International Institute for Applied Systems Analysis, 1994.