This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2011 proceedings.

Optimal Transmission over a Fading Channel with Imperfect Channel State Information

Yichuan Hu and Alejandro Ribeiro
Department of Electrical and Systems Engineering, University of Pennsylvania

***Abstract*—This paper considers a wireless channel when the probability distribution of fading is unknown and the transmitter has access to imperfect channel state information (CSI). Transmitted power and coding mode are adapted to the available imperfect CSI through the use of a backoff function in order to maximize the expected transmission rate subject to an average power constraint. Determination of the optimal power allocation and backoff function is a non-convex stochastic optimization problem with infinitely many variables. Despite its non-convexity, the duality gap of this problem has been shown to be null. Exploiting this property, we show that the optimal power allocation and channel backoff functions are uniquely determined by the optimal dual variable. This affords considerable simplification because the dual optimization problem is convex and one-dimensional – whereas the original primal problem is non-convex and infinite-dimensional. Iterative algorithms that find the optimal power and backoff function based on imperfect CSI without having access to the channel probability distribution are further developed. Numerical results are provided to corroborate theoretical findings.**

I. INTRODUCTION

Although accurate channel state information (CSI) is essential to achieve high spectral efficiency, perfect CSI is rarely available in practice due to estimation errors or feedback delay. Algorithms to handle imperfect CSI at the transmitter side are the subject matter of this paper. To exploit favorable channel conditions, the transmitter adapts its power and coding mode to the measured CSI. Due to the inaccuracy of CSI, channel outages occur when the rate selected turns out too aggressive for the actual channel realization. To reduce the chance of outages a backoff function is also employed by the transmitter. Instead of adapting the code to the observed imperfect CSI, the code is adapted to a backed-off version to provide a larger security margin for channel outages. Our goal in this paper is to characterize the optimal power allocation and channel backoff functions that maximize the average transmission rate subject to an average power constraint. The proposed algorithms are also adaptive in that they find the optimal operating point using channel observations without knowledge of the channel probability density function (pdf).

Significant effort has been devoted to analyze effects of imperfect CSI on different wireless channels; e.g. [1]–[3]. From a practical perspective, it is recognized that to mitigate the negative effect of capacity outages due to imperfect CSI, a rate backoff function is needed in addition to power control. Ideally, power allocation and rate backoff should be jointly optimized but this results in a non convex problem that is difficult to solve; see Section II-A. By imposing additional restrictions, the problem can be simplified to more tractable formulations. E.g., when power is fixed and only rate adaptation is considered the problem is reduced to the determination of the optimal backoff function; e.g., [4]. A second possibility is to fix a target outage probability and separate the optimization into the determination of a backoff function for target outage, followed by optimal power allocation over estimated channels [5]. A third possible restriction is to assume that the backoff function takes a certain parametric form and proceed to optimize the corresponding parameters, e.g. [6]. While yielding tractable formulations, the resultant transmission rates are not optimal. The work on this paper differs from these contributions in that we do not impose any additional constraints or assumptions and develop algorithms to find jointly optimal power allocations and channel backoff functions.

Jointly optimal power and rate adaptation to imperfect CSI can be formulated as a stochastic optimization problem which has infinite dimensionality and lacks convexity (Section II). Rather than reformulating the original problem into a suboptimal tractable alternative, we work with the Lagrangian dual problem which is convex and one-dimensional. Despite the lack of convexity, the Lagrangian dual is equivalent to the primal problem, since the structure of the latter makes it part of a class of non convex problems with null duality gap [7]. This result is leveraged to show that the optimal power allocation and channel backoff functions are determined by the optimal dual variable (Section III). We further develop stochastic subgradient descent algorithms to find optimal power allocation and channel backoff functions without requiring access to the channel's pdf (Section IV). Numerical results and conclusions are presented in Sections V and VI, respectively.

II. PROBLEM FORMULATION

Consider a block wireless channel with time slots indexed by t. The channel at time t is denoted as $h(t)$ and modeled as stationary block fading in that channel coefficients are assumed independent realizations of a complex random variable h. Corresponding channel gains are defined as $\gamma(t) := |h(t)|^2$ and are independent realizations of a random variable that we denote as γ. The pdfs $f_h(h)$ of the fading coefficient h and $f_\gamma(\gamma)$ of its gain are unknown. In each time slot the transmitter computes an estimate $\hat{\gamma}(t)$ of the current gain $\gamma(t)$ to adapt transmitted power and code selection to the channel state. The accuracy of estimates $\hat{\gamma}(t)$ is characterized through the conditional probability distribution $f_{\gamma|\hat{\gamma}}(\gamma|\hat{\gamma})$ that determines the probability of the actual channel being γ when the estimate is $\hat{\gamma}$. The probability distribution $f_{\gamma|\hat{\gamma}}(\gamma|\hat{\gamma})$ depends on the channel estimation method and is assumed known, although we make no assumptions on its specific form – see also Remark 1.

Based on the value of the channel estimate $\hat{\gamma}$, the transmitter decides on a power allocation $P(\hat{\gamma}) : \mathbf{R}_+ \rightarrow [0, P_{\mathsf{max}}]$, where $P_{\mathsf{max}} > 0$ is the maximum instantaneous power the transmitter can use. The amount of information that can be delivered through the channel is jointly determined by the transmitted power $P(\hat{\gamma})$ and the *actual* channel gain γ as determined by a given channel capacity function $C(P(\hat{\gamma}), \gamma)$. The function $C(P(\hat{\gamma}), \gamma)$ depends on how signals are coded and modulated in the physical layer. E.g., if capacity-achieving codes are used,

$$C(P(\hat{\gamma}), \gamma) = \log \left[1 + \frac{P(\hat{\gamma})\gamma}{N_0} \right], \tag{1}$$

where N_0 is the noise power at the receiver. If adaptive modulation and coding (AMC) with M transmission modes is considered, communication rate τ_m is supported when the signal to noise ratio (SNR) $P(\hat{\gamma})\gamma/N_0$ is between η_m and η_{m+1}. The channel capacity function is therefore

$$C(P(\hat{\gamma}), \gamma) = \sum_{m=1}^{M} \tau_m \mathbb{I} \left\{ \eta_m \leq \frac{P(\hat{\gamma})\gamma}{N_0} \leq \eta_{m+1} \right\}, \tag{2}$$

where $\mathbb{I}(\cdot)$ stands for the indicator function. We do not restrict $C(P(\hat{\gamma}), \gamma)$ to a specific form. We only assume that $C(P(\hat{\gamma}), \gamma)$ is a nonnegative nondecreasing function of the product $P(\hat{\gamma})\gamma$.

Whatever its specific form, capacity $C(P(\hat{\gamma}), \gamma)$ cannot be achieved. Indeed, to achieve $C(P(\hat{\gamma}), \gamma)$ the transmitter has to select an appropriate code adapted to the received SNR $P(\hat{\gamma})\gamma/N_0$ which is unknown at the transmitter side because it depends on the unknown channel gain γ. Furthermore, if the transmitter observes $\hat{\gamma}$ and chooses a transmission code based on it – i.e., it transmits at a rate $C(P(\hat{\gamma}), \hat{\gamma})$ – a channel outage will occur when the transmitted rate $C(P(\hat{\gamma}), \hat{\gamma})$ exceeds the maximum rate $C(P(\hat{\gamma}), \gamma)$ the channel can afford – i.e. when $C(P(\hat{\gamma}), \hat{\gamma}) > C(P(\hat{\gamma}), \gamma)$ or simply when $\hat{\gamma} > \gamma$. The instantaneous transmission rate achieved in the channel is therefore given by

$$R(\gamma, \hat{\gamma}) = C(P(\hat{\gamma}), \hat{\gamma}) \cdot \mathbb{I} \{ \hat{\gamma} \leq \gamma \}, \tag{3}$$

which corrects for lost packets through the indicator $\mathbb{I} \{ \hat{\gamma} \leq \gamma \}$.

To alleviate the negative effect of outages, a channel backoff function $B(\hat{\gamma}) : \mathbf{R}_+ \rightarrow \mathbf{R}_+$ is used to determine a backed-off channel gain $B(\hat{\gamma})$. The code is then adapted to the received SNR $P(\hat{\gamma})B(\hat{\gamma})/N_0$ – as opposed to $P(\hat{\gamma})\hat{\gamma}/N_0$ – and communication proceeds at a rate $C(P(\hat{\gamma}), B(\hat{\gamma}))$. With codes adapted to $P(\hat{\gamma})B(\hat{\gamma})/N_0$, an outage occurs if $B(\hat{\gamma}) < \gamma$. Thus, the instantaneous transmission rate can be written as

$$R(\gamma, \hat{\gamma}) = C(P(\hat{\gamma}), B(\hat{\gamma})) \cdot \mathbb{I} \{ B(\hat{\gamma}) \leq \gamma \}. \tag{4}$$

The idea is that making $B(\hat{\gamma}) < \hat{\gamma}$ reduces the chance of an outage thereby increasing the effective rate $R(\gamma, \hat{\gamma})$ [cf. (3) and (4)]. However, as we shall show later, making $B(\hat{\gamma}) > \hat{\gamma}$ is optimal in some cases.

A. Ergodic rate optimization

The goal of this paper is to find the optimal power allocation function $P(\hat{\gamma})$ and channel backoff function $B(\hat{\gamma})$ such that the expected transmission rate is maximized subject to average power constraint P_0, i.e.

$$\begin{aligned} \mathsf{P} = \max \quad & \mathbb{E}_{\gamma, \hat{\gamma}} \left[C(P(\hat{\gamma}), B(\hat{\gamma})) \cdot \mathbb{I} \{ B(\hat{\gamma}) \leq \gamma \} \right] \\ \text{s.t.} \quad & \mathbb{E}_{\hat{\gamma}} \left[P(\hat{\gamma}) \right] \leq P_0. \end{aligned} \tag{5}$$

Solving (5) is challenging because: (I1) The objective includes an expectation over the random channel gain γ, whose realizations are not available at the transmitter and whose pdf is unknown. (I2) Variables in this optimization problem are functions $P(\hat{\gamma})$ and $B(\hat{\gamma})$ defined on $\mathbf{R}_+$, implying the dimensionality of the problem is infinite. (I3) The objective and the constraint involve expectations over channel estimates $\hat{\gamma}$, whose realizations are observed in each time slot but whose pdf is unknown. (I4) The channel capacity function $C(P(\hat{\gamma}), B(\hat{\gamma}))$ may be nonconvex or even discontinuous as in the case of AMC [cf. (2)].

To overcome issue (I1), we rewrite the expectation in the objective of (5) as a conditional expectation over γ with $\hat{\gamma}$ given, followed by an expectation over $\hat{\gamma}$, i.e.

$$\begin{aligned} & \mathbb{E}_{\gamma, \hat{\gamma}} \left[C(P(\hat{\gamma}), B(\hat{\gamma})) \cdot \mathbb{I} \{ B(\hat{\gamma}) \leq \gamma \} \right] \\ & \qquad = \mathbb{E}_{\hat{\gamma}} \left[C(P(\hat{\gamma}), B(\hat{\gamma})) \mathbb{E}_{\gamma | \hat{\gamma}} \left[\mathbb{I} \{ B(\hat{\gamma}) \leq \gamma \} \right] \right]. \end{aligned} \tag{6}$$

Note that the inner expectation in (6) is just the probability $\Pr(\gamma \geq B(\hat{\gamma}) | \hat{\gamma})$ of the backed off channel being smaller than the actual channel γ for a given estimate $\hat{\gamma}$. This probability can be written in terms of the complementary cumulative distribution function (ccdf) $F^c_{\gamma|\hat{\gamma}}(\cdot)$ of γ given $\hat{\gamma}$ as $\Pr(\gamma \geq B(\hat{\gamma}) | \hat{\gamma}) = F^c_{\gamma|\hat{\gamma}}(B(\hat{\gamma}))$. Since $f_{\gamma|\hat{\gamma}}(\cdot)$ is known – see Remark 1 – the ccdf $F^c_{\gamma|\hat{\gamma}}(B(\hat{\gamma}))$ is available. This allows us to simplify (6) to

$$\begin{aligned} & \mathbb{E}_{\gamma, \hat{\gamma}} \left[C(P(\hat{\gamma}), B(\hat{\gamma})) \cdot \mathbb{I} \{ B(\hat{\gamma}) \leq \gamma \} \right] \\ & \qquad = \mathbb{E}_{\hat{\gamma}} \left[C(P(\hat{\gamma}), B(\hat{\gamma})) \cdot F^c_{\gamma|\hat{\gamma}}(B(\hat{\gamma})) \right]. \end{aligned} \tag{7}$$

Using (7) the objective in (5) can be written as a single expectation over $\hat{\gamma}$ yielding the equivalent formulation

$$\begin{aligned} \mathsf{P} = \max \quad & \mathbb{E}_{\hat{\gamma}} \left[C(P(\hat{\gamma}), B(\hat{\gamma})) \cdot F^c_{\gamma|\hat{\gamma}}(B(\hat{\gamma})) \right] \\ \text{s.t.} \quad & \mathbb{E}_{\hat{\gamma}} \left[P(\hat{\gamma}) \right] \leq P_0. \end{aligned} \tag{8}$$

Problems (5) and (8) are equivalent. Our goal is to find optimal power allocation $P^*(\hat{\gamma})$ and backoff $B^*(\hat{\gamma})$ functions that solve problem (8). Since actual channel gains γ are not present in (8), issue (I1) has been resolved. Issues (I2)-(I4), however, still hold for problem (8). Sections III and IV discuss a method to solve (8) that overcomes these issues. We pursue this after the following remark.

Remark 1: The probability distribution $f_{\gamma|\hat{\gamma}}(\gamma|\hat{\gamma})$ depends on the channel estimation method. A typical way of estimating the channel is to send a training signal that is known to both the transmitter and the receiver and get feedback from the receiver on the measured channel gain. Due to estimation error and/or feedback delays, estimated channels $\hat{h}$ are different from actual channels h and are modeled as

$$\hat{h} = h + e, \tag{9}$$

where e is a complex Gaussian random noise $\mathcal{CN}(0, \sigma_e^2)$. For the model in (9) it holds that the pdf of γ given $\hat{\gamma}$ is a noncentral chi-square given by [8]

$$f_{\gamma|\hat{\gamma}}(\gamma|\hat{\gamma}) = \frac{1}{\sigma_e^2} \exp\left(-\frac{\gamma+\hat{\gamma}}{\sigma_e^2}\right) I_0\left(\frac{2\sqrt{\gamma\hat{\gamma}}}{\sigma_e^2}\right), \qquad (10)$$

where $I_0(x) = \sum_{i=0}^{\infty} \frac{(x^2/4)^i}{(i!)^2}$ is the zeroth order modified Bessel function of the first kind. This particular form for the conditional pdf $f_{\gamma|\hat{\gamma}}(\gamma|\hat{\gamma})$ is used to provide numerical results in Section V. The rest of the development in the paper holds independently of the particular form of this pdf.

III. Optimal Power Allocation & Channel Backoff

The optimization problem in (8) has only one constraint, implying that while the primal problem is infinite dimensional, the dual problem is one-dimensional. More importantly, it has been shown that problems like (8), where the non-convex functions appear inside expectations, have null duality gap as long as the pdf of the random variable with respect to which we take the expected value has no points of striclty positive probability, [7]. Thus, working in the dual domain is simpler and equivalent. To introduce the dual function associate Lagrange multiplier λ with the power constraint and define the Lagrangian as

$$\begin{aligned}&\mathcal{L}(P(\hat{\gamma}), B(\hat{\gamma}), \lambda)\\ &= \mathbb{E}_{\hat{\gamma}}\left[C(P(\hat{\gamma}), B(\hat{\gamma})) \cdot F^c_{\gamma|\hat{\gamma}}(B(\hat{\gamma}))\right] + \lambda\left[P_0 - \mathbb{E}_{\hat{\gamma}}\left[P(\hat{\gamma})\right]\right]\\ &= \mathbb{E}_{\hat{\gamma}}\left[C(P(\hat{\gamma}), B(\hat{\gamma})) \cdot F^c_{\gamma|\hat{\gamma}}(B(\hat{\gamma})) - \lambda P(\hat{\gamma})\right] + \lambda P_0, \quad (11)\end{aligned}$$

where in the second equality we rearranged terms. The dual function is then defined as the maximum of the Lagrangian over the sets of feasible $P(\hat{\gamma})$ and $B(\hat{\gamma})$, i.e.,

$$g(\lambda) = \max_{P(\hat{\gamma}), B(\hat{\gamma})} \mathcal{L}(P(\hat{\gamma}), B(\hat{\gamma}), \lambda). \qquad (12)$$

We now can write the dual problem as the minimum of $g(\lambda)$ over nonnegative λ, i.e.,

$$\mathsf{D} = \min_{\lambda \geq 0} g(\lambda) \qquad (13)$$

Since the problem (8) and its dual (13) have been shown to have null gap we have that $\mathsf{P} = \mathsf{D}$. This property can be exploited to characterize the optimal power allocation and channel backoff functions as is done in the following theorem.

Theorem 1: *The optimal power allocation function $P^*(\hat{\gamma})$ and channel backoff function $B^*(\hat{\gamma})$ for solving problem* (8) *are uniquely determined by the optimal dual variable λ^* of the dual problem* (13)*. In particular,*

$$\{P^*(\hat{\gamma}), B^*(\hat{\gamma})\} \in \underset{b\in[0,\infty), p\in[0,P_{\mathsf{max}}]}{\operatorname{argmax}} \left\{C(p,b) \cdot F^c_{\gamma|\hat{\gamma}}(b) - \lambda^* p\right\}. \qquad (14)$$

Proof: According to the definition of the dual function [cf. (12)], $g(\lambda^*)$ is the maximum of the Lagrangian $\mathcal{L}(P(\hat{\gamma}), B(\hat{\gamma}), \lambda^*)$ for all $P(\hat{\gamma})$ and $B(\hat{\gamma})$. Since $P^*(\hat{\gamma})$ and $B^*(\hat{\gamma})$ are feasible to the primal problem, it follows that $\mathcal{L}(P^*(\hat{\gamma}), B^*(\hat{\gamma}), \lambda^*)$ must be bounded above by $g(\lambda^*)$, i.e.,

$$\mathsf{D} = g(\lambda^*) \geq \mathcal{L}(P^*(\hat{\gamma}), B^*(\hat{\gamma}), \lambda^*). \qquad (15)$$

According to (11), we have

$$\begin{aligned}&\mathcal{L}(P^*(\hat{\gamma}), B^*(\hat{\gamma}), \lambda^*)\\ =&\mathbb{E}_{\hat{\gamma}}\left[C(P^*(\hat{\gamma}), B^*(\hat{\gamma})) \cdot F^c_{\gamma|\hat{\gamma}}(B^*(\hat{\gamma}))\right] + \lambda^*\left[P_0 - \mathbb{E}_{\hat{\gamma}}\left[P^*(\hat{\gamma})\right]\right].\end{aligned}$$

Since $B^*(\hat{\gamma})$ and $P^*(\hat{\gamma})$ are feasible to the primal problem, the average power constraint must be satisfied, i.e., $P_0 - \mathbb{E}_{\hat{\gamma}}\left[P^*(\hat{\gamma})\right] \geq 0$. Moreover, we know that $\lambda^* \geq 0$. Therefore, $\lambda^*\left[P_0 - \mathbb{E}_{\hat{\gamma}}\left[P^*(\hat{\gamma})\right]\right] \geq 0$ and as a result

$$\begin{aligned}&\mathcal{L}(P^*(\hat{\gamma}), B^*(\hat{\gamma}), \lambda^*) \geq\\ &\qquad \mathbb{E}_{\hat{\gamma}}\left[C(P^*(\hat{\gamma}), B^*(\hat{\gamma})) \cdot F^c_{\gamma|\hat{\gamma}}(B^*(\hat{\gamma}))\right] = \mathsf{P}. \quad (16)\end{aligned}$$

Substituting (16) into (15) gives us

$$\mathsf{D} = g(\lambda^*) \geq \mathcal{L}(P^*(\hat{\gamma}), B^*(\hat{\gamma}), \lambda^*) \geq \mathsf{P}. \qquad (17)$$

Since the duality gap is null, i.e. $\mathsf{D} = \mathsf{P}$, the inequalities in (17) must be satisfied with equalities, i.e.

$$\mathsf{D} = g(\lambda^*) = \mathcal{L}(P^*(\hat{\gamma}), B^*(\hat{\gamma}), \lambda^*) = \mathsf{P}.$$

This implies that $P^*(\hat{\gamma})$ and $B^*(\hat{\gamma})$ are maximizers of the Lagrangian $\mathcal{L}(P(\hat{\gamma}), B(\hat{\gamma}), \lambda^*)$.

$$\{P^*(\hat{\gamma}), B^*(\hat{\gamma})\} \in \underset{P(\hat{\gamma}), B(\hat{\gamma})}{\operatorname{argmax}} \mathcal{L}(P(\hat{\gamma}), B(\hat{\gamma}), \lambda^*). \qquad (18)$$

Note that in (18) we used $\in$ instead of $=$ because the maximizer may not be unique. Using the definition of the Lagrangian [cf. (11)], we can rewrite (18) as

$$\begin{aligned}&\{P^*(\hat{\gamma}), B^*(\hat{\gamma})\} \in\\ &\underset{P(\hat{\gamma}), B(\hat{\gamma})}{\operatorname{argmax}} \mathbb{E}_{\hat{\gamma}}\left[C(P(\hat{\gamma}), B(\hat{\gamma})) \cdot F^c_{\gamma|\hat{\gamma}}(B(\hat{\gamma})) - \lambda^* P(\hat{\gamma})\right], \quad (19)\end{aligned}$$

where we ignored the term $\lambda^* P_0$ since it does not depend on $P(\hat{\gamma})$ or $B(\hat{\gamma})$. Due to the linearity of the expectation operator, the maximization in (19) can be carried out inside the expectation for each channel state $\hat{\gamma}$, i.e.,

$$\{P^*(\hat{\gamma}), B^*(\hat{\gamma})\} \in \underset{b\in[0,\infty), p\in[0,P_{\mathsf{max}}]}{\operatorname{argmax}} \left\{C(p,b) \cdot F^c_{\gamma|\hat{\gamma}}(b) - \lambda^* p\right\}. \qquad (20)$$

(20) coincides with (14). This completes the proof. ■

Provided that λ^* is available, Theorem 1 states that $P^*(\hat{\gamma})$ and $B^*(\hat{\gamma})$ can be obtained by solving the maximization in (14). Although the problem in (14) might be nonconvex, solving it is by no means a difficult task as it only involves two variables. This provides a great advantage because the problem dimensionality is reduced from infinity to 1. Also, we remark that Theorem 1 is true no matter what the capacity function is and how the underlying channel is distributed. Next, we shall develop online algorithms that find the optimal solutions for problem (8) using only instantaneous imperfect CSI.

IV. Online Learning Algorithms

Unlike the nonconvex primal problem, the dual problem in (13) is always convex. This suggests that gradient descent algorithms in the dual domain are guaranteed to converge to the optimal multiplier λ^*. In particular, we use stochastic subgradient descent algorithms that iteratively compute primal

Algorithm 1: Optimal power control and channel backoff

Initialize Lagrangian multiplier $\lambda(0)$;
for $t = 0, 1, 2, \cdots$ **do**
  Compute primal variables as per (21):
  $\{p(t), b(t)\} \in \underset{b\in[0,\infty),p\in[0,P_{\max}]}{\operatorname{argmax}} \left\{C(p,b)\cdot F^c_{\gamma|\hat{\gamma}(t)}(b) - \lambda(t)p\right\}$;
  Transmit with power $p(t)$ and rate $C(p(t), b(t))$;
  Update dual variables as per (22):
  $\lambda(t+1) = [\lambda(t) - \epsilon(t)\,[P_0 - p(t)]]^+$;
end

and dual variables; see Algorithm 1. Given dual variable $\lambda(t)$, the algorithm proceeds to a primal iteration in which it computes power allocation $p(t)$ and backoff function $b(t)$ as

$$\{p(t), b(t)\} \in \underset{b\in[0,\infty),p\in[0,P_{\max}]}{\operatorname{argmax}} \left\{C(p,b)\cdot F^c_{\gamma|\hat{\gamma}(t)}(b) - \lambda(t)p\right\}. \tag{21}$$

Multipliers $\lambda(t+1)$ are then updated based on $\lambda(t)$ and $p(t)$ as

$$\lambda(t+1) = [\lambda(t) - \epsilon(t)\,[P_0 - p(t)]]^+, \tag{22}$$

where $\epsilon(t) > 0$ is a step size which might be time dependent and $[x]^+ = \max\{0, x\}$. It can be shown that the expected value of the quantity $P_0 - p(t)$ in (22) is a subgradient of the dual function [9], [10]. This property implies that $P_0 - p(t)$ points to λ^* on an average sense and can be exploited to prove convergence in the dual domain.

Particular convergence properties depend on whether constant or time varying step sizes are used. We first consider diminishing step sizes. If $\epsilon(t)$ is nonsummable but square summable, i.e., $\sum_{t=0}^{\infty}\epsilon(t) = \infty$ and $\sum_{t=0}^{\infty}\epsilon^2(t) < \infty$, then using standard stochastic approximation techniques [11] it can be shown that $\lambda(t)$ converges to λ^* almost surely. As a consequence of Theorem 1, this indicates that the primal variables almost surely converge to the optimal values as time grows, i.e., $p(t) = P^*(\hat{\gamma}(t))$ and $b(t) = B^*(\hat{\gamma}(t))$ almost surely as t goes to infinity.

The drawback of using diminishing step size is the lack of adaptivity. Constant step sizes are more desirable as the underlying channel distribution might be time-varying in a dynamical wireless environment. However, with constant step size the dual iterates $\lambda(t)$ no longer converge to the optimal value almost surely. Instead, $\lambda(t)$ stays within a small distance of λ^* with probability close to 1 as t goes to infinity [12] and the convergence is established in an ergodic sense [9, Theorem 1]. In particular, if $\epsilon(t) = \epsilon > 0$ is a constant, then the average power constraint is almost surely satisfied in an ergodic sense

$$\lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t} p(u) \le P_0 \quad \text{a.s.}, \tag{23}$$

and the ergodic limit of the transmission rates almost surely converges to a value within $\kappa\epsilon$ of optimal,

$$\mathsf{P} - \lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t} C(p(u), b(u))\cdot F^c_{\gamma|\hat{\gamma}}(b(u)) \le \kappa\epsilon, \tag{24}$$

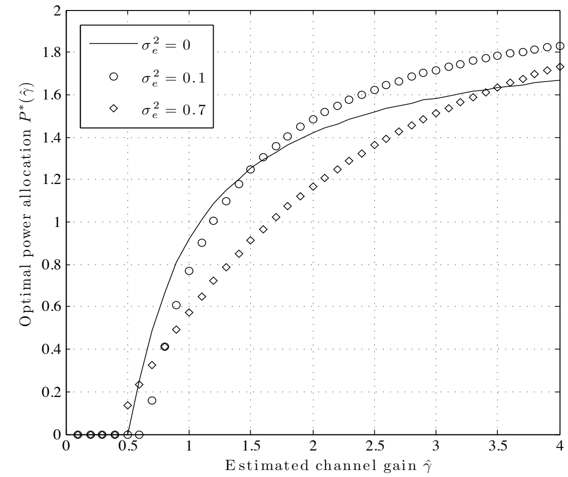

Fig. 1. Optimal power allocation function $P^*(\hat{\gamma})$ for $\sigma_e^2 = 0.1$ and 0.7, respectively. For comparison purpose, optimal power allocation function when there is not channel error is also depicted.

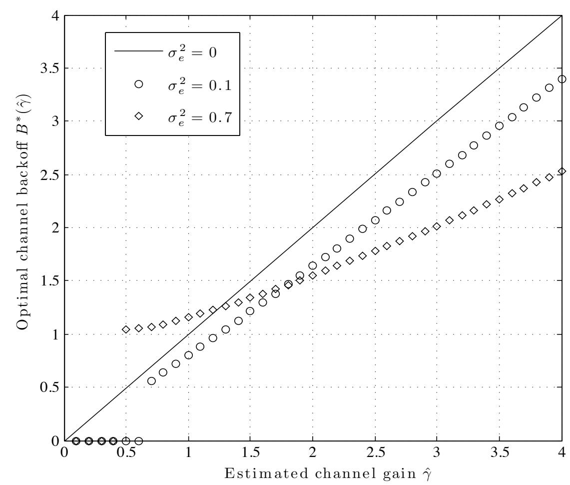

Fig. 2. Optimal channel backoff function $B^*(\hat{\gamma})$ for $\sigma_e^2 = 0.1$ and 0.7, respectively. For comparison purpose, a line (solid) corresponding to no channel backoff is also depicted.

where the constant κ is defined as $\kappa := \min\{P_0^2, (P_{\max} - P_0)^2\}$. Although the algorithm with constant step size does not find $P^*(\hat{\gamma}(t))$ and $B^*(\hat{\gamma}(t))$, it generates $p(t)$ and $b(t)$ that are almost surely feasible and almost surely near optimal. In addition, (24) indicates that the optimality gap can be made arbitrarily small by reducing the step size.

V. Numerical Results

To evaluate the performance of the proposed algorithms, we conduct several numerical tests. Specifically, we assume that the actual channel coefficient h follows a complex Gaussian distribution $\mathcal{CN}(0, 2)$ and that the channel estimation error is modeled by (9). The average power budget is 1, i.e. $P_0 = 1$ and the channel capacity function takes the form of (1). Without loss of generality, we assume normalized noise power $N_0 = 1$.

In the first set of tests, two channel estimation error variances $\sigma_e^2 = 0.1$ and $\sigma_e^2 = 0.7$, corresponding to small and big channel errors, are simulated. We apply diminishing step size

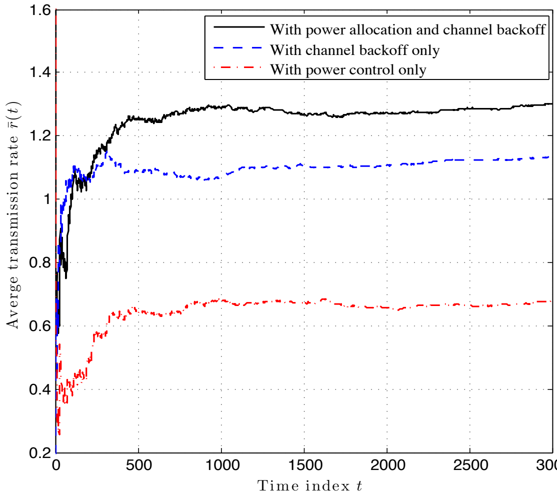

Fig. 3. Comparison of the average transmission rate $\bar{r}(t)$ achieved by the proposed algorithm with and without channel backoff function. The step size is a constant $\epsilon = 0.01$ and the channel estimation error $\sigma_e^2 = 0.1$.

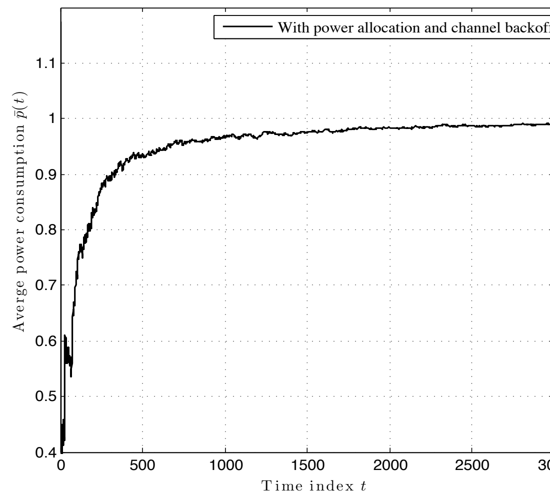

Fig. 4. Average transmission power $\bar{p}(t)$ used by the proposed algorithm with constant step size $\epsilon = 0.01$. The channel estimation error $\sigma_e^2 = 0.1$.

$\epsilon(t) = 1/\sqrt{t}$ to obtain the optimal dual variable λ^* for both cases and then find optimal power allocation function $P^*(\hat{\gamma})$ and optimal channel backoff function $B^*(\hat{\gamma})$ according to Theorem 1, as shown in Fig. 1 and Fig. 2, respectively. For comparison purposes, $P^*(\hat{\gamma})$ and $B^*(\hat{\gamma})$ for $\sigma_e^2 = 0$ are also depicted. For both small and big channel errors, the optimal power allocation functions are water-filling, but power is allocated more conservatively when channel is small. The difference between the channel backoff functions for both cases is more obvious. When $\sigma_e^2 = 0.1$, the channel backoff is almost linear and $B^*(\hat{\gamma}) < \hat{\gamma}$ for all $\hat{\gamma}$, i.e. making $B^*(\hat{\gamma})$ smaller is always beneficial. On the contrary, when $\sigma_e^2 = 0.7$ the channel backoff function exhibits nonlinearity, and it is interesting to note that $B^*(\hat{\gamma}) > \hat{\gamma}$ for relatively small $\hat{\gamma}$ ($0.5 \leq \hat{\gamma} \leq 1.2$). The intuition is that when σ_e^2 is comparable to $\hat{\gamma}$, it is very likely that γ is greater than $\hat{\gamma}$. Therefore, making $B(\hat{\gamma})$ a little bigger than γ is not likely to result in an outage.

In our next simulation, we test the algorithm with constant step size $\epsilon = 0.01$ and assuming channel error $\sigma_e^2 = 0.1$. Other parameters remain the same as before. We define the average transmission rate $\bar{r}(t) = (1/t)\sum_{u=1}^{t} C(p(u), b(u)) \cdot \mathbb{I}\{b(u) \leq \gamma(u)\}$ and average power consumption $\bar{p}(t) = (1/t)\sum_{u=1}^{t} p(u)$. Fig. 3 compares average rates achieved when we use: (1) Both, power allocation and channel backoff. (2) Channel backoff only, i.e., $P^*(\hat{\gamma}) = P_0$. (3) Power allocation only, i.e., $B^*(\hat{\gamma}) = \hat{\gamma}$. There is a considerable improvement in average transmission rate when power allocation and channel backoff are jointly optimized. We further show in Fig. 4 that the average power constraint is always satisfied, ratifying the almost sure feasibility result in (23).

VI. Conclusion

We considered power and code adaptation to imperfect CSI in order to maximize expected transmission rates subject to average power constraints. We showed that the optimal power allocation and channel backoff functions are determined by a single parameter in the form of the optimal multiplier of the Lagrange dual problem. We further developed stochastic subgradient descent algorithms on the dual domain that operate without knowledge of the channel's probability distribution. For vanishing step sizes these dual stochastic descent algorithms converge to the optimal multiplier λ^*. With constant step sizes optimal multipliers are not found but a policy that is optimal in an ergodic sense is determined. Numerical results showed significant performance gains of jointly optimal backoff and power allocation[1].

References

[1] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933 –946, May 2000.

[2] G. Ganesan, Y. Li, and A. Swami, "Channel aware aloha with imperfect csi," in *Proc. GLOBECOM'06*, San Francisco, CA, Dec. 2006.

[3] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading mimo channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203 –2214, May 2006.

[4] A. Vakili, M. Sharif, and B. Hassibi, "The effect of channel estimation error on the throughput of broadcast channels," in *Proc. ICASSP'06*, Toulouse, France, May 2006.

[5] R. Wang and V. Lau, "Cross layer design of downlink multi-antenna ofdma systems with imperfect csit for slow fading channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2417 –2421, Jul. 2007.

[6] D. Zheng, M.-O. Pun, W. Ge, J. Zhang, and V. H. Poor, "Distributed opportunistic scheduling for ad hoc communications with imperfect channel information," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5450 –5460, Dec. 2008.

[7] A. Ribeiro and G. B. Giannakis, "Separation principles of wireless networking," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4488 – 4505, Sep. 2010.

[8] J. Proakis and M. Salehi, *Digital communications, 5th ed.* McGraw-Hill, 2007.

[9] A. Ribeiro, "Ergodic stochastic optimization algorithms for wireless communication and networking," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6369–6386, Dec. 2010.

[10] Y. Hu and A. Ribeiro, "Adaptive distributed algorithms for optimal random access channels," *IEEE Trans. Wireless Commun.*, Feburary 2010, submitted.

[11] H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications, 2nd ed.* Springer-Verlag, 2003.

[12] Y. Hu and A. Ribeiro, "Optimal wireless networks based on local channel state information," *IEEE Trans. Signal Process.*, Oct. 2010, submitted.

[1]Work in this paper is supported by the Army Research Office grant P-57920-NS and the National Science Foundation CAREER award CCF-0952867. Email at {yichuan, aribeiro}@seas.upenn.edu