No More Tears: A No-Regret Approach to Power Control in Dynamically Varying MIMO Networks

Ioannis Stiakogiannakis[*], Panayotis Mertikopoulos[†‡§], Corinne Touati[‡§†]

[*] Mathematical and Algorithmic Sciences Lab, France Research Center, Huawei Technologies Co. Ltd.

[†] CNRS, LIG, F-38000 Grenoble, France

[‡] Inria

[§] Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France

Email: ioannis.steiakogiannakis@huawei.com, panayotis.mertikopoulos@imag.fr, corinne.touati@inria.fr

Abstract—In this paper, we address the trade-off between radiated power and achieved throughput in wireless multiple-input and multiple-output (MIMO) systems that evolve over time in an unpredictable fashion (e.g. due to changes in the wireless medium or the users' QoS requirements). Contrary to the static/stationary channel regime, there is no optimal power allocation profile to converge to (either static or in the mean), so the system's users must adapt to changes in the environment "on the fly", without being able to predict the system's evolution ahead of time. In this dynamic context, we formulate the users' power/throughput trade-off as an online optimization problem and we provide a matrix exponential learning algorithm that leads to *no regret* – i.e. the proposed transmit policy is asymptotically optimal in hindsight, irrespective of how the system varies with time. As a result, users are able to track the evolution of their individually optimum transmit profiles remarkably well, even in arbitrarily changing wireless environments.

I. Introduction

Current and emerging wireless systems are facing a crucial trade-off between transmit power and achieved throughput: in many applications (such as e-mail and voice calls), radiated power must be reduced to the bare minimum in order to preserve battery life; by contrast, in rate-hungry applications (such as multimedia streaming and video calling), it is crucial to optimize the allocation of the users' available power so as to maximize their throughput. Consequently, coupled with the prolific deployment of multiple-input and multiple-output (MIMO) technologies and the anticipated impact of massive MIMO, next-generation wireless networks call for flexible power control (PC) algorithms tailored to systems with several degrees of freedom.

In its most basic form, power control allows wireless links to achieve their required throughput while minimizing radiated power and the induced interference (individually or globally), thus increasing spatial spectrum reuse and battery life [1–3]. That being said, while the benefits of power control are relatively easy to assess in networks with static channel conditions, it is much harder to analyze the associated performance gains (if any) in networks that vary with time. In the ergodic regime (where channels follow a stationary ergodic process), the

authors of [4, 5] provided power control algorithms that minimize the users' transmit power while achieving a minimum ergodic rate requirement, while [6] studied the problem of ergodic rate maximization in fast-fading multi-carrier systems. Beyond this ergodic case however, when the wireless medium does not evolve according to an independent and identically distributed (i.i.d.) sequence of random variables, power control remains a very open issue.

In this paper, we drop all stationarity/i.i.d. assumptions and we focus squarely on wireless systems that evolve *arbitrarily* over time in terms of both channel conditions and user quality of service (QoS) requirements. In this framework, standard approaches based on linear programming (for static channels) and/or stochastic convex optimization (for the ergodic case) are no longer relevant because there is no underlying optimization problem to solve – either static or in the mean. Instead, we treat power control as a dynamically evolving optimization problem and we employ techniques and ideas from *online* optimization to quantify how well the system's users adapt to changes in the wireless medium (and/or track their individually optimum transmit powers as they change over time).

The most widely used performance criterion in this setting is that of *regret minimization*, a concept which was first introduced by Hannan [7] and which has since given rise to a vigorous literature at the interface of machine learning, optimization, statistics and game theory – for a comprehensive survey, see e.g. [8, 9]. Specifically, in the language of game theory, the notion of regret compares the cumulative payoff of an agent over a given time horizon to the cumulative payoff that the agent would have obtained by employing the *a posteriori* best possible action over the time horizon in question. Accordingly, in the context of power control, regret minimization corresponds to dynamic transmit policies that are asymptotically optimal in hindsight, *irrespective* of how the user's environment and/or requirements evolve over time.

Regret minimization methodologies were recently used in [10] to study the transient phase of the original Foschini–Miljanic (FM) power control algorithm [2] in static environments, while [11] focused on the regret minimization properties of the algorithm in dynamic single-input and single-output (SISO), single-carrier systems. In [12], the authors considered a poten-

tial game formulation for the joint power control and channel allocation problem in cognitive radio (CR) networks and they employed a regret minimizing algorithm [13] to reach a Nash equilibrium state. The same problem was also examined in the context of infrastructureless wireless networks by the authors of [14] who provided a potential game formulation and derived a power control algorithm that minimizes the users' internal regret and converge to the game's unique correlated – and, hence, Nash – equilibrium. Finally, in a very recent paper, the authors of [15] employed online optimization methodologies to derive efficient power allocation policies for online rate maximization in dynamic cognitive radio networks.

In this paper, we focus on wireless MIMO systems that evolve arbitrarily over time (for instance, due to fading, intermittent user activity, changing QoS requirements, etc.), and we seek to provide an efficient power control scheme that allows users to balance their radiated power against their achieved throughput "on the fly", based only on locally available channel state information (CSI). In particular, we formulate the users' power/rate trade-off as a nonlinear online optimization problem and we derive an adaptive *no-regret* power control policy based on the method of matrix exponential learning (MXL) [16–18].

The proposed MXL algorithm is provably asymptotically optimal against the system's evolution in hindsight; furthermore, it also enjoys the following desirable properties:

- *Distributedness:* users update their power allocation profiles based only on local channel state information.
- *Asynchronicity:* there is no need for a global update timer to coordinate the users' updates.
- *Statelessness:* transmitters do not need to know the network's topology or overall state.
- *Reinforcement:* each connection tends to move towards better power/rate trade-offs.

Our theoretical analysis is supplemented by extensive numerical simulations in Section IV where we illustrate the power and throughput gains of the proposed power control algorithm under realistic fading conditions.

II. System Model and Problem Formulation

Consider a set $\mathcal{U} = \{1, \dots, U\}$ of wireless point-to-point connections corresponding to the *users* of the wireless system. Each connection $u \in \mathcal{U}$ consists of a transmit-receive pair (t_u, r_u) with M_u antennas at the transmitter and N_u antennas at the receiver. Thus, if $\mathbf{x}_u \in \mathbb{C}^{M_u}$ (resp. $\mathbf{y}_u \in \mathbb{C}^{N_u}$) denotes the signal transmitted (resp. received) over connection $u \in \mathcal{U}$, we obtain the familiar signal model:

$$\mathbf{y}_u = \mathbf{H}_{uu}\mathbf{x}_u + \sum_{v \neq u} \mathbf{H}_{vu}\mathbf{x}_v + \mathbf{z}_u \quad (1)$$

where $\mathbf{z}_u \in \mathbb{C}^{N_u}$ stands for the ambient noise in the channel (including thermal, atmospheric and other peripheral interference effects) and $\mathbf{H}_{vu} \in \mathbb{C}^{N_u \times M_v}$ denotes the channel matrix between t_v and r_u. Unavoidably, the received signal $\mathbf{y}_u$ is affected by ambient noise and interference due to the transmissions of
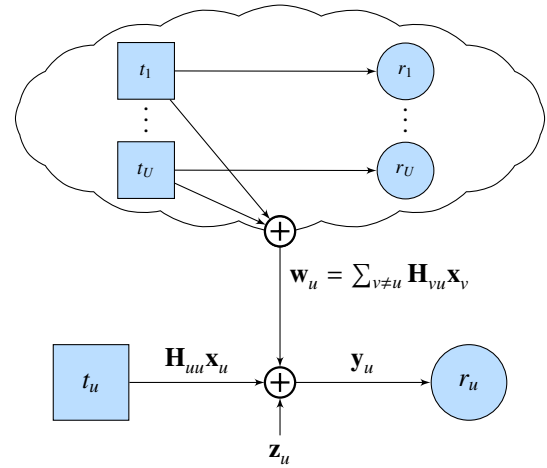

Fig. 1. Example of a wireless network with several active connections where we focus on a particular connection u between transmitter t_u and receiver r_u. The network's other active connections cause co-channel interference to the connection under consideration, which is is treated as additive color noise. The focal connection is also subject to ambient white noise.

other connections on the same subcarrier, so we will write

$$\mathbf{w}_u = \sum_{v \neq u} \mathbf{H}_{vu}\mathbf{x}_v + \mathbf{z}_u \quad (2)$$

for the multi-user interference-plus-noise at the receiver r_u of connection u (for a schematic representation, see Fig. 1). In this way, (1) attains the simpler form

$$\mathbf{y}_u = \mathbf{H}_{uu}\mathbf{x}_u + \mathbf{w}_u. \quad (3)$$

In what follows, we will focus on a specific connection $u \in \mathcal{U}$, so, for clarity, we will drop the index u altogether and we will write (3) more simply as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \quad (4)$$

In this context, assuming Gaussian input and noise and single user decoding (SUD) at the receiver (i.e. the multi-user interference $\mathbf{w}$ is treated as additive color noise), the transmission rate of the focal connection will be [19, 20]

$$R(\mathbf{Q}) = \log \det[\mathbf{W} + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger] - \log \det \mathbf{W}, \quad (5)$$

where $\mathbf{Q} = \mathbb{E}[\mathbf{x}\mathbf{x}^\dagger]$ is the user's input signal covariance matrix, and $\mathbf{W} = \mathbb{E}[\mathbf{w}\mathbf{w}^\dagger]$ denotes the covariance matrix of the multi-user interference-plus-noise in the channel. Thus, if we write

$$\widetilde{\mathbf{H}} = \mathbf{W}^{-1/2}\mathbf{H} \quad (6)$$

for the user's *effective channel matrix*, Eq. (5) can be written more compactly as:

$$R(\mathbf{Q}) = \log \det[\mathbf{I} + \widetilde{\mathbf{H}}\mathbf{Q}\widetilde{\mathbf{H}}^\dagger]. \quad (7)$$

Throughout this paper, we focus on wireless users who seek to minimize their radiated power on one hand while maximizing their transmission rate on the other. Thus, to account for this power minimization/rate maximization trade-off, we will consider the general power control objective:

$$\ell(\mathbf{Q}) = \text{tr}(\mathbf{Q}) - \phi(R(\mathbf{Q})) \quad (8)$$

where $\phi: \mathbb{R}_+ \to \mathbb{R}_+$ is a nondecreasing function of the user's achievable transmission rate. In this way, $\ell(\mathbf{Q})$ can be interpreted as a "*loss function*": higher values of $\ell(\mathbf{Q})$ indicate that the user is transmitting at very high power or at very low rate (or both), so he is incurring a "loss" in his power/rate trade-off. Accordingly, we will only assume that ϕ is Lipschitz and concave: the first assumption is a mild technical requirement which we only make for simplicity; the second one reflects the effects of "diminishing returns" on high data rates (a rate increase from 1 bps to 2 bps is much more significant than an increase from $1,000$ bps to $1,001$ bps).

Example 1. As a special case of the objective (8), consider the scenario where the focal user seeks to minimize his transmit power $\mathrm{tr}(\mathbf{Q})$ subject to achieving a target transmission rate R^*. This classical formulation of power control can be recovered by considering a cost function ϕ of the form $\phi(R(\mathbf{Q}) - R^*)$ with $\phi(r) = 0$ if $r \geq 0$: when the target transmission rate is achieved (i.e. $R(\mathbf{Q}) \geq R^*$), the only term in the user's loss function (8) is the user's total transmit power $\mathrm{tr}(\mathbf{Q})$; otherwise, if the target transmission rate is not met, the user incurs a positive loss of at least $\phi'(0^-) \cdot (R^* - R(\mathbf{Q}))$.[1] By this token, the quantity $\lambda = \phi'(0^-)$ measures the tolerance of the connection with respect to transmission rate deficits: smaller values of λ correspond to softer rate requirements, while, in the large λ limit, the loss function (8) stiffens to a hard constraint where no violations are tolerated.

Of course, when the user's effective channel matrices vary with time, the user's transmission rate will be given by

$$R(\mathbf{Q}; t) = \log \det[\mathbf{I} + \widetilde{\mathbf{H}}(t) \, \mathbf{Q} \, \widetilde{\mathbf{H}}^\dagger(t)], \qquad (9)$$

where $\widetilde{\mathbf{H}}(t)$ denotes the user's effective channel matrix at time t.[2] With this in mind, the user's loss function at time t will be

$$\ell(\mathbf{Q}; t) = \mathrm{tr}(\mathbf{Q}) - \phi(R(\mathbf{Q}; t); t), \qquad (10)$$

thus leading to the *online power control problem:*

$$\begin{aligned} \text{minimize} \quad & \ell(\mathbf{Q}; t) \\ \text{subject to} \quad & \mathbf{Q} \in \mathcal{X} \end{aligned} \qquad \text{(OPC)}$$

where

$$\mathcal{X} = \{\mathbf{Q} : \mathbf{Q} \succcurlyeq 0, \; \mathrm{tr}(\mathbf{Q}) \leq P\} \qquad (11)$$

is the problem's state space and $P > 0$ denotes the user's maximum transmit power. As such, given that the user has no control over the effective channel matrices $\widetilde{\mathbf{H}}$, we obtain the following sequence of events:

1) At each instance $t \geq 0$, the user selects a transmit power profile $\mathbf{Q}(t) \in \mathcal{X}$.

2) The user's loss $\ell(\mathbf{Q}(t); t)$ is determined by the state of the network and the behavior of all other users via the effective channel matrix $\widetilde{\mathbf{H}}(t)$.

3) The user employs some update rule to select a new transmit power profile, and the process repeats.

The key challenge in this dynamic framework is that the user does not know his objective function $\ell(\mathbf{Q}; t)$ ahead of time (recall that $\ell(\mathbf{Q}; t)$ depends at each stage t on the evolution of the environment and the transmit power profiles of all other users), so he must try to predict his optimum transmit profile "on the fly". Consequently, static solution concepts (such as Nash or correlated equilibria) are no longer relevant because there is no stationary optimization criterion to achieve – either static or in the mean.

Instead, given a time horizon T, we will compare the cumulative loss incurred by the user's chosen transmit power profile to the hypothetical loss that the user would have incurred if he had chosen the best possible transmit profile in hindsight. More precisely, focusing on continuous time for simplicity, we define the user's cumulative (*external*) regret as:

$$\mathrm{Reg}(T) = \max_{\mathbf{Q}^* \in \mathcal{X}} \int_0^T [\ell(\mathbf{Q}(t); t) - \ell(\mathbf{Q}^*; t)] \, dt. \qquad (12)$$

The notion of regret was first introduced in a game-theoretic setting by Hannan [7] and it has since given rise to an extremely active field of research at the interface of optimization, statistics and theoretical computer science – for a survey, see e.g. [8, 9].[3] The user's *average regret* is then defined as $T^{-1} \mathrm{Reg}(T)$ and the goal of *regret minimization* is to devise a dynamic transmit policy $\mathbf{Q}(t)$ which is asymptotically optimal in hindsight, i.e. that leads to *no regret*:

$$\limsup_{T \to \infty} \mathrm{Reg}(T)/T \leq 0, \qquad (13)$$

or, equivalently:

$$\mathrm{Reg}(T) = o(T), \qquad (14)$$

irrespective of how the objective function (8) evolves over time.

Obviously, if the user could somehow predict the solution of (OPC) in an oracle-like fashion, we would have $\mathrm{Reg}(T) \leq 0$ in (12) for all T. In particular, if the user's objective (8) does not vary with time (or if it varies in a stochastic fashion, following some i.i.d. process), a no-regret policy converges to the problem's static (or, respectively, average) solution [9]; as such, the no-regret requirement (13) is an indicator that $\mathbf{Q}(t)$ tracks the solution of (OPC) as it evolves over time.

Remark 1. In the machine learning literature, there exist more sophisticated notions of regret (such as adaptive [21] or shifting [22] regret) to further quantify the quality of this tracking; however, due to space limitations, we will focus our theoretical analysis almost exclusively on external regret minimization (which requires less technical language to describe).

---

[1] Recall that ϕ is assumed concave, so the user's loss grows at least linearly with the rate deficit $R^* - R(\mathbf{Q})$.

[2] In what follows, we will tacitly assume that $\widetilde{\mathbf{H}}(t)$ is measurable and bounded with respect to t. This assumption is justified by factors such as the minimum distance between transmitters and receivers, antenna directivity, RF circuit losses, etc. which make it impossible for the user's (effective) channel gains to become arbitrarily high. Furthermore, we assume that the variability of $\widetilde{\mathbf{H}}(t)$ is such that standard results from information theory remain valid [19].

[3] The terminology stems from the fact that large positive values of $\mathrm{Reg}(T)$ indicate that the user would have achieved a better power/rate trade-off in the past by employing some fixed $\mathbf{Q}^*$ instead of $\mathbf{Q}(t)$, making him "regret" his choice.

III. Adaptive Power Control via Exponential Learning

A key element in the derivation of a no-regret transmit policy for the online problem (OPC) will be the gradient $\mathbf{V} = \nabla_{\mathbf{Q}} \ell$ of the user's objective function (8). Specifically, we have:

$$\mathbf{V} = \nabla_{\mathbf{Q}} \ell = \mathbf{I} - \phi'(R) \cdot \nabla_{\mathbf{Q}} R, \tag{15}$$

where, after some matrix calculus:

$$\nabla_{\mathbf{Q}} R = \widetilde{\mathbf{H}}^{\dagger} [\mathbf{I} + \widetilde{\mathbf{H}} \mathbf{Q} \widetilde{\mathbf{H}}^{\dagger}]^{-1} \widetilde{\mathbf{H}}. \tag{16}$$

In this way, the gradient of $\ell(\mathbf{Q}(t); t)$ evaluated at $\mathbf{Q}(t)$ will be:

$$\mathbf{V}(t) = \mathbf{I} - \phi'(R(\mathbf{Q}(t); t)) \cdot \widetilde{\mathbf{H}}^{\dagger}(t) \left[\mathbf{I} + \widetilde{\mathbf{H}}(t) \mathbf{Q}(t) \widetilde{\mathbf{H}}^{\dagger}(t)\right]^{-1} \widetilde{\mathbf{H}}(t). \tag{17}$$

Importantly, the user's gradient matrix $\mathbf{V}(t)$ at time t is a simple function of his signal covariance matrix $\mathbf{Q}(t)$ and his effective channel matrix $\widetilde{\mathbf{H}}^{\dagger}(t)$. The former is obviously known to the receiver, while the latter can be measured at the receiver and then fed back to the transmitter (e.g. during the downlink phase of a time-division duplexing (TDD) scheme); in this way, $\mathbf{V}(t)$ can be calculated based on purely local CSI, so any algorithm relying on $\mathbf{V}(t)$ will be likewise distributed.

In view of the above, a first idea would be to update the user's power profile $\mathbf{Q}(t)$ along the direction of steepest descent indicated by $\mathbf{V}(t)$, that is take $\dot{\mathbf{Q}} = -\mathbf{V}$ [23]. However, this approach would invariably violate the users' semidefiniteness constraints ($\mathbf{Q} \succcurlyeq 0$), so it is not a viable transmit policy. Instead, inspired by the matrix regularization methods of [16–18], we will consider a learning scheme that tracks the direction of steepest descent in a dual, unconstrained space and then maps the result back to the problem's state space via matrix exponentiation. More precisely, we will concentrate on the matrix exponential learning (MXL) process:

$$\dot{\mathbf{Y}} = -\mathbf{V},$$
$$\mathbf{Q} = P \frac{\exp(\eta \mathbf{Y})}{1 + \operatorname{tr}[\exp(\eta \mathbf{Y})]}, \tag{MXL}$$

where $\eta > 0$ is a parameter that controls the user's learning rate (for an algorithmic implementation, see Algorithm 1 below).

The learning process (MXL) will be the main focus of our paper, so some remarks are in order:

Remark 1. Intuitively, the exponentiation step in (MXL) assigns more power to the spatial directions that perform well; the trace normalization then ensures that $\mathbf{Q}(t)$ satisfies the feasibility constraints of (OPC) for all $t \geq 0$, while the learning parameter η sharpens the method's reinforcement effect.[4] In particular, (MXL) can be seen as a "primal-dual" online mirror descent (OMD) method [9] with exponential regularization; for an in-depth discussion, see e.g. [9, 16–18, 24] and references therein.

Remark 2. From an implementation perspective, Algorithm 1 has the following desirable properties:

(P1) It is *distributed*: each transmitter updates his power profile based only on local CSI.

---

[4]For large η, (MXL) assigns almost all power to the transmit direction which corresponds to the highest eigenvalue of $\mathbf{V}$.

Algorithm 1: Matrix exponential learning (MXL)

parameter: $\eta > 0$

```
/* Initialization                              */
```
1 $t \leftarrow 0; \mathbf{Y} \leftarrow \mathbf{0}$;

2 **repeat**

3 $\quad t \leftarrow t + 1$;
```
   /* Pre-transmission: Set Power             */
```
4 $\quad \mathbf{Q} \leftarrow P \dfrac{\exp(\eta \mathbf{Y})}{1 + \operatorname{tr}[\exp(\eta \mathbf{Y})]}$;
```
   /* Transmission                            */
   /* Post-Transmission Measurements          */
```
5 $\quad R \leftarrow \log \det (\mathbf{I} + \widetilde{\mathbf{H}} \mathbf{Q} \widetilde{\mathbf{H}}^{\dagger})$;

6 $\quad \mathbf{V} \leftarrow \mathbf{I} - \phi'(R) \cdot \widetilde{\mathbf{H}}^{\dagger} (\mathbf{I} + \widetilde{\mathbf{H}} \mathbf{Q} \widetilde{\mathbf{H}}^{\dagger})^{-1} \widetilde{\mathbf{H}}$;

7 $\quad \mathbf{Y} \leftarrow \mathbf{Y} - \mathbf{V}$;

8 **until** *transmission ends*;

(P2) It is *asynchronous*: power updates are performed without signaling/coordination between connections.

(P3) It is *agnostic*: transmitters do not need to know the topology (or state) of the wireless network.

(P4) It is *reinforcing*: each connection tends to optimize its individual power vs. rate objective function.

Remark 3. In terms of feedback, Algorithm 1 requires that *a)* the transmitter measures his achieved rate R; and *b)* the receiver feeds back to the transmitter the received signal covariance $\mathbb{E}[\mathbf{y}\mathbf{y}^{\dagger}] = \mathbf{W} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger}$ (e.g. via broadcasting or over a downlink duplex pilot). From a computational standpoint, it is then easy to see that the comlplexity of each iteration of Algorithm 1 is polynomial in the number of transmit antennas M (typically of the order of $M^{2.373}$ if users employ fast matrix multiplication methods).

In this context, our main theoretical result regarding the learning scheme (MXL) is as follows:

Theorem 1. *The learning scheme* (MXL) *leads to no regret in the online power control problem* (OPC). *Specifically,* (MXL) *enjoys the regret bound:*

$$\frac{1}{T} \operatorname{Reg}(T) \leq P \frac{\log(1 + M)}{\eta T} = \mathcal{O}(1/T), \tag{18}$$

irrespective of the system's evolution over time.

Proof: See Appendix A. ∎

We close this section with a few remarks on optimizing the performance of the learning scheme (MXL) (and, respectively, Alg. 1):

a) Multi-user interference: Even though Theorem 1 focuses on a given connection, the focal connection is still subject to interference from other connections in the network (captured by the effective channel matrices $\widetilde{\mathbf{H}}$ which depend on the interfering users' transmit policies). In this light, Theorem 1 provides a worst-case performance guarantee which holds even in the presence of malicious users (jammers).

b) Initialization: The initialization $\mathbf{Y}(0) = 0$ is a conservative choice reflecting the worst-case scenario where the user begins with no information regarding his channel. Indeed, if $\mathbf{Y}(0) = 0$, the user's initial transmit power will be $P \cdot M / (M+1)$, which is asymptotically equal to P in the large M limit (corresponding to massive MIMO transmitters); as such, the user's transmit power will likely be reduced under Algorithm 1 in the presence of good channel conditions. In particular, if the transmitter can estimate his initial channel conditions, it would be preferable to initialize power accordingly: if the user expects good channel conditions, initial power should be set lower (to save battery life); otherwise, if bad channel conditions are expected, the user should transmit with high power so as to avoid very low transmission rates during the first frames.

c) The role of the learning parameter η: As we mentioned before, larger values of η tend to enhance the reinforcement effect of (MXL) because the user's power tends to be allocated only along the maximum eigen-directions of $\mathbf{V}$. On the other hand, if η is chosen very large with respect to the channels' characteristic time scale, the exponent of (MXL) might reach very high levels very quickly. This can lead (MXL) to become too greedy in discrete-time implementations (cf. Algorithm 1), in which case a slowly decreasing learning parameter would be preferable; due to space limitations however, these considerations are delegated to a future paper.

IV. NUMERICAL RESULTS

To validate the theoretical analysis of Section III, we conducted extensive numerical simulations over a wide range of design parameters and specifications. In what follows, we present a representative subset of these results, but the conclusions drawn remain valid in most typical mobile wireless environments.

Throughout this section, we consider a wireless network cell with $U = 4$ connections, each with $M = 2$ transmit and $N = 2$ receive antennas. We focus on the uplink (UL) case, and the receivers are assumed stationary whereas the transmitters may be either stationary or mobile, depending on the simulated scenario. The connections operate at a central frequency $f_c = 2.5\,\mathrm{GHz}$, and communication occurs over a time-division duplexing (TDD) scheme with frame duration $T_f = 5\,\mathrm{ms}$. Specifically, transmission occurs during the UL subframe while receivers process the transmitted signal and provide feedback during the downlink (DL) subframe: upon reception of the feedback, transmitters update their transmit powers according to Algorithm 1, and the process repeats until transmission ends. For demonstration purposes, we simulated the case where each connection has a rate requirement R_u^* (cf. the model description and Example 1 in Section II) which is constant with time but which varies across connections $u \in \mathcal{U}$ so as to ensure diversity of QoS requirements.

For benchmarking, the first simulated scenario focuses on the static regime where channel conditions do not change throughout the transmission horizon while each user updates his signal covariance matrix following Algorithm 1. To begin with, Fig. 2a depicts the evolution of the user's objective function $\ell(\mathbf{Q};t)$ over time: in tune with Theorem 1, users converge to the minimum of their objective within a few frames, thus optimizing their power/rate trade-off based on their requirements.[5] This is seen clearly in Figs. 2b and 2c where we plot the evolution of the user's total transmit power and the achieved/target rate gap $R(t)/R^*$. Despite the agnostic initialization of Algorithm 1 at very high power levels (representing a pessimistic estimate of channel conditions), the users' transmit power is quickly reduced to the minimum level that can sustain their required rate (corresponding to an achieved/target ratio of 1). This behavior is also clearly seen in Fig. 2d where we plot each user's (average) regret over time.[6] The worst-case upper bound predicted by Theorem 1 (dashed lines) quickly vanishes (at an $\mathcal{O}(1/T)$ decay rate), while the users' actual regret becomes negative within only a few frames, indicating that users are controlling their power optimally with respect to their rate requirements.

Fig. 3 is devoted to fully time-varying systems. Specifically, we consider a wireless network consisting of $U = 4$ wireless 2×2 MIMO connections with mobile transmitters moving at 2km/h and different tolerance levels for their QoS requirements (cf. the model description and Example 1 in Section II). The users' wireless channels are simulated using the extended pedestrian A (EPA) model [25] and the evolution of the aggregate channel gain $\mathrm{tr}[\mathbf{H}\mathbf{H}^\dagger]$ is shown in 3a for reference purposes.

In this time-varying setting, the main challenge for the users is to track the optimum signal covariance profile that balances their transmit power against their achieved throughput (i.e. that minimizes their loss) as this optimum profile evolves over time. To that end, Fig. 3b depicts the evolution of the users' total transmit power under matrix exponential learning (Algorithm 1). As can be seen, the users' transmit power under Algorithm 1 follows closely the evolution of the users' channel gains: users increase power to compensate for poor channel conditions and decrease power in the presence of favorable channel conditions. This is seen further in Fig. 3c where we plot the users' achieved/target rate gap $R(t)/R^*$, averaged over time: the connections that have a softer tolerance for the satisfaction of their QoS requirements (e.g. Connection 1) are very aggressive in reducing transmit power when channel conditions seem to allow it, whereas connections that are less tolerant with respect to their QoS requirements (e.g. Connection 2) are more conservative and transmit at relatively higher powers (resulting in higher rates) as a precaution against deep fading events.

Finally, as in the static channel case, Fig. 3d depicts the users' average regret over time: again, despite the pessimistic high-power initialization of Algorithm 1, the users' regret falls

---

[5]The observed oscillations in some connections have to do with arithmetic issues and can be readily eliminated using a decreasing parameter η.

[6]For simplicity, instead of taking the maximum of (12) over the (infinite) set $\mathcal{X}$, we took the maximum over a sample of 100 covariance matrices in $\mathcal{X}$ (including uniform beamforming profiles with all combinations of antennas active and inactive).
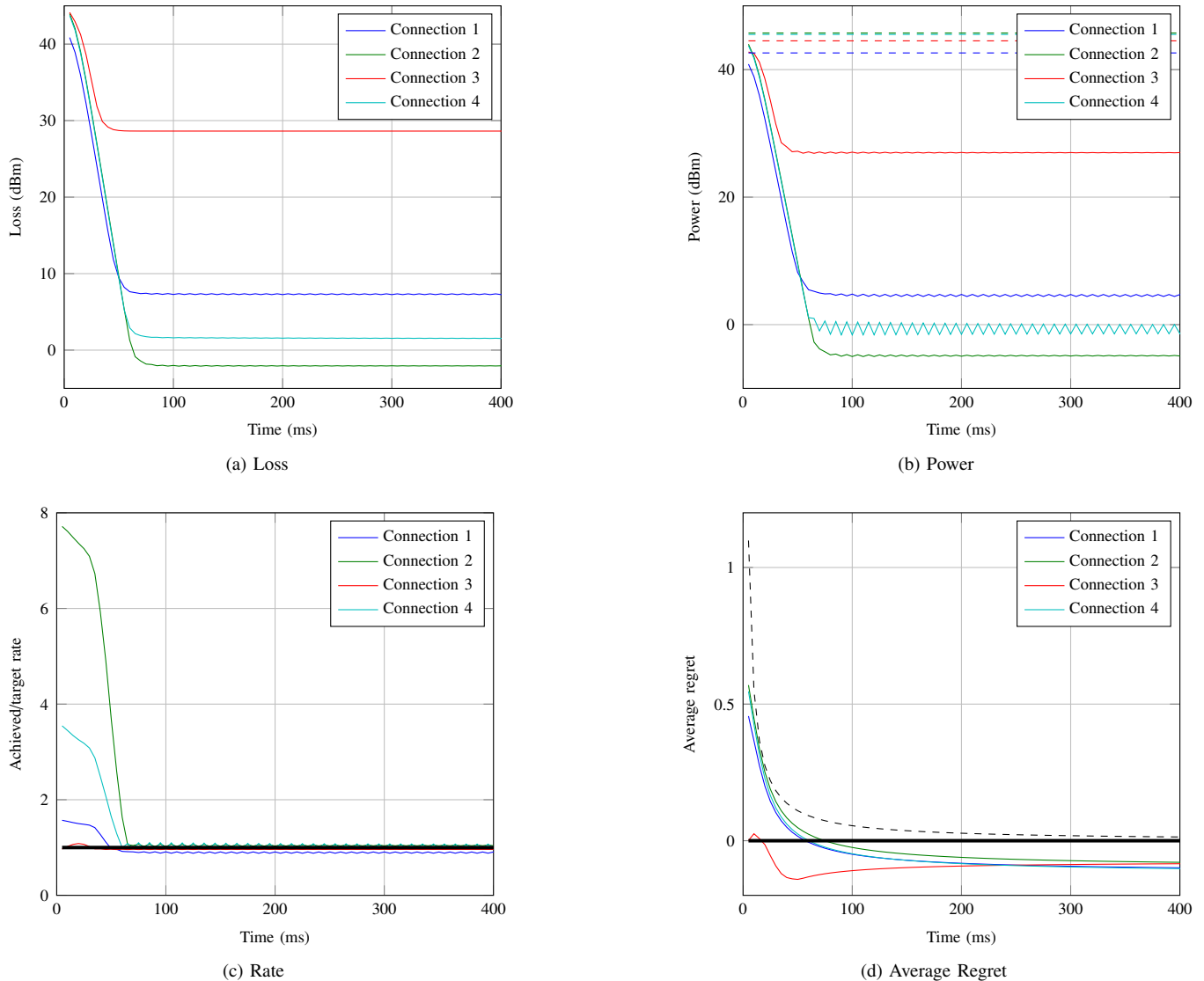
(a) Loss

(b) Power

(c) Rate

(d) Average Regret

Fig. 2. Balancing power against achieved rate in a network consisting of $U = 4$ MIMO connections with static channel conditions. Fig. 2a depicts the evolution of the trade-off objective $\ell(\mathbf{Q}; t)$ under Algorithm 1; in a similar vein, Fig. 2b depicts the users' total transmit power (dashed lines correspond to the users' maximum transmit power), while Fig. 2c shows the achieved/target rate gap R/R^*. Finally, Fig. 2d shows the users' average regret $\mathrm{Reg}(T)/T$ over time: as predicted by Theorem 1, the users' regret quickly becomes negative, indicating the long-term optimality of their transmit policy (given their requirements).

below the no-regret threshold in just a few frames, so users achieve optimality much faster than the $\mathcal{O}(1/T)$ bounds of Theorem 1. The reason for this faster convergence is that the worst-case bounds of Theorem 1 only become relevant in very adverse (or adversarial) channel conditions, occuring for example when users are being jammed by a third party: in standard mobility scenarios (such as the one simulated here), the evolution of the wireless medium is relatively tame from a statistical perspective, so users can learn to track the system much faster than in the adversarial case.

V. Conclusions

In this paper, we examined the trade-off between radiated power and achieved throughput in wireless MIMO systems that evolve dynamically over time as the result of time-varying channel conditions and user QoS requirements. To account for the system's complete lack of stationarity (or any other type of averaging behavior that could allow the use of traditional solution concepts such as Nash/correlated equilibria), we provided a formulation based on online optimization and we derived a matrix exponential learning algorithm that leads to *no regret* – i.e. it is asymptotically optimal in hindsight, irrespective of how the wireless system varies with time. Thanks to the algorithm's no regret property, the system's users are able to track their optimal transmit profile "on the fly", even under randomly changing channel conditions.

Importantly, the proposed algorithm is fully distributed and requires only local CSI that is readily available at each connection in the system. In future extensions of this work, we intend to consider more general MIMO–OFDM systems where

(a) Channel
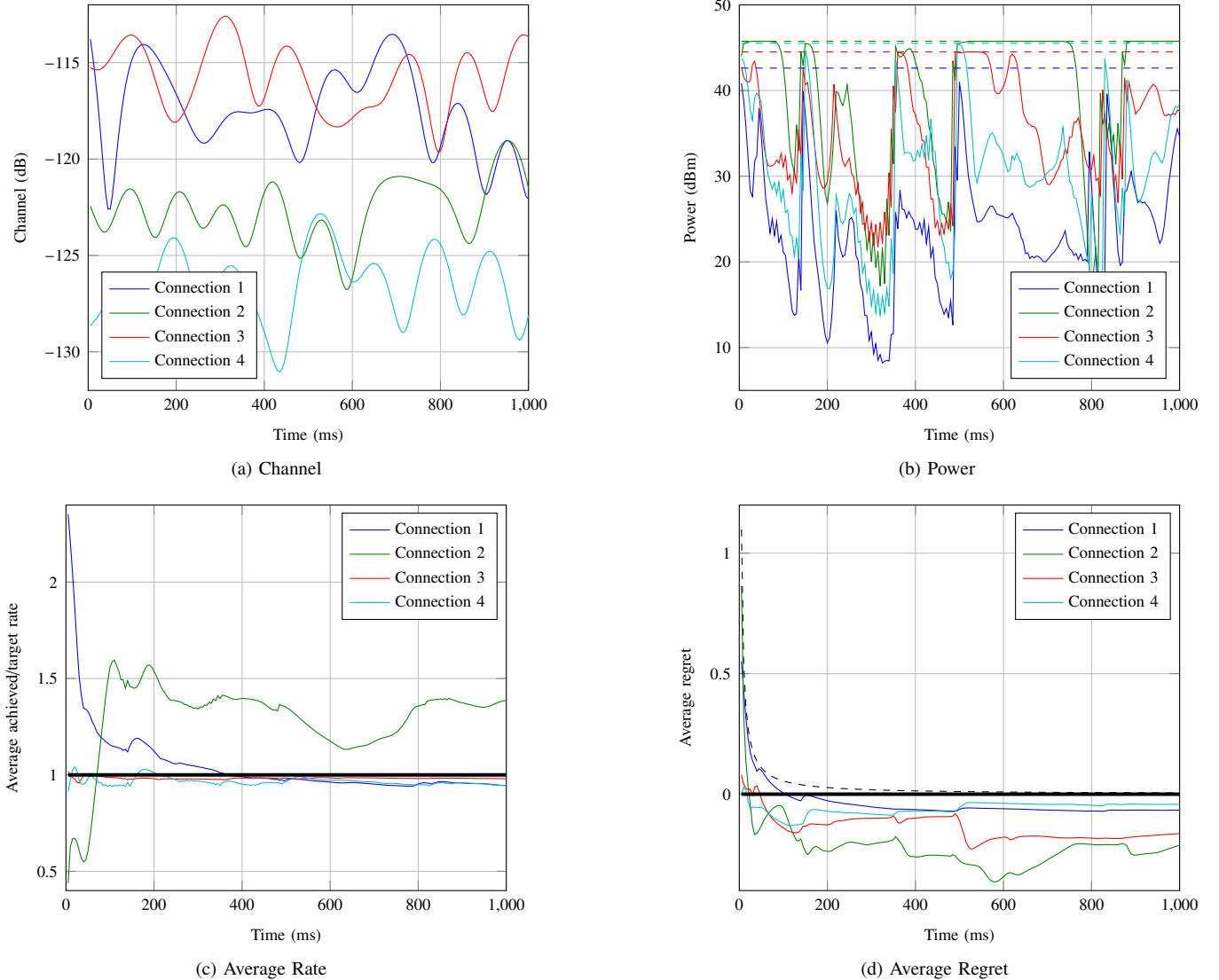

(b) Power

(c) Average Rate

(d) Average Regret

Fig. 3. Balancing power against throughput in a network consisting of $U = 4$ MIMO connections with mobile transmitters and stationary receivers, all moving at 2 km/h (pedestrian speed). For reference purposes, Fig. 3a depicts the evolution of the channel gains $\operatorname{tr}[\mathbf{H}(t)\mathbf{H}^{\dagger}(t)]$ over time. Fig. 3b shows the evolution of the users' total transmit power under Algorithm 1 (the dashed lines represent the users' maximum transmit power), while Fig. 3c shows the achieved/target rate gap $R(t)/R^*$ (averaged over time). Finally, as in the static channel case, Fig. 3d shows the users' average regret $\operatorname{Reg}(T)/T$: as predicted by Theorem 1, the users' regret quickly becomes negative, indicating that their transmit policy is asymptotically optimal in hindsight (given their rate requirements).

connections are established over several subcarriers, and where users only have imperfect CSI at their disposal.

APPENDIX
TECHNICAL PROOFS

Our goal in this appendix is to prove the regret bound (18) for (MXL).

We begin by noting that the loss function $\ell(\mathbf{Q}; t)$ is convex w.r.t. $\mathbf{Q}$, since ϕ is concave and nondecreasing and the Shannon rate function $R(\mathbf{Q}; t)$ is concave in $\mathbf{Q}$ [26]. With this basic convexity result at hand, we obtain:

$$\ell(\mathbf{Q}(t); t) - \ell(\mathbf{Q}^*; t) \leq \operatorname{tr}[(\mathbf{Q}(t) - \mathbf{Q}^*) \cdot \mathbf{V}(t)], \quad (19)$$

where $\mathbf{V}(t) = \nabla_{\mathbf{Q}(t)} \ell(\mathbf{Q}(t); t)$ denotes the gradient of $\ell(\cdot; t)$

evaluated at $\mathbf{Q}(t)$. Accordingly, to establish the no-regret bound (18) for (MXL), it suffices to show that

$$\int_0^T \operatorname{tr}[(\mathbf{Q}(t) - \mathbf{Q}^*) \cdot \mathbf{V}(t)] \, dt \leq \eta^{-1} P \cdot \log(1 + M) \quad (20)$$

for all $\mathbf{Q}^* \in \mathcal{X}$.

Proof of Theorem 1: By (MXL), we readily get:

$$\int_0^T \operatorname{tr}[(\mathbf{Q}(t) - \mathbf{Q}^*) \cdot \mathbf{V}(t)] \, dt = \int_0^T \operatorname{tr}[(\mathbf{Q}^* - \mathbf{Q}(t)) \cdot \dot{\mathbf{Y}}(t)] \, dt$$

$$= \operatorname{tr}[\mathbf{Y}(T) \cdot \mathbf{Q}^*] - \int_0^T \operatorname{tr}[\mathbf{Q}(t)\dot{\mathbf{Y}}(t)] \, dt, \quad (21)$$

where we have used the fact that $\mathbf{Y}(0) = 0$. By the exponential

update step of (MXL), the second term of (21) then becomes:

$$\text{tr}\left[\mathbf{Q}\dot{\mathbf{Y}}\right] = P\frac{\text{tr}\left[\exp(\eta\mathbf{Y})\dot{\mathbf{Y}}\right]}{1 + \text{tr}[\exp(\eta\mathbf{Y})]} = \frac{P}{\eta}\frac{d}{dt}\log\left[1 + \text{tr}[\exp(\eta\mathbf{Y})]\right], \tag{22}$$

and hence, by plugging (22) back into (21) and integrating, we get:

$$\int_0^T \text{tr}\left[(\mathbf{Q}(t) - \mathbf{Q}^*) \cdot \mathbf{V}(t)\right] dt = \text{tr}\left[\mathbf{Y}(T) \cdot \mathbf{Q}^*\right]$$
$$- \frac{P}{\eta}\log\left[1 + \text{tr}[\exp(\eta\mathbf{Y}(T))]\right] + \frac{P}{\eta}\log\left(1 + M\right), \tag{23}$$

where we used again the fact that $\mathbf{Y}(0) = 0$ (implying in turn that $\text{tr}[\exp(\eta\mathbf{Y}(0))] = M$).

To proceed, we will require the inequality:

$$\text{tr}(\mathbf{PX}) \le \log\left(1 + \text{tr}(\exp(\mathbf{X}))\right), \tag{24}$$

valid for all Hermitian $\mathbf{P}, \mathbf{X}$, with $\mathbf{P} \succcurlyeq 0$, $\text{tr}(\mathbf{P}) \le 1$. To that end, let $F(\mathbf{X}) = \log\left(1 + \text{tr}(\exp(\mathbf{X}))\right) - \text{tr}(\mathbf{PX})$, so it suffices to show that $\min_{\mathbf{X}} F(\mathbf{X}) \ge 0$ whenever $\mathbf{P} > 0$ and $\text{tr}(\mathbf{P}) < 1$ (the boundary case $\det(\mathbf{P}) = 0$ or $\text{tr}(\mathbf{P}) = 1$ follows by continuity). Now, since $\text{tr}(\exp(\mathbf{X}))$ is convex in $\mathbf{X}$ and the logarithm is concave and increasing, F will be itself convex, so if it admits a critical point $\mathbf{X}^*$, this point will be a (global) minimizer. By differentiating, we then obtain:

$$\nabla_{\mathbf{X}} F(\mathbf{X}) = \frac{\exp(\mathbf{X})}{1 + \text{tr}(\exp(\mathbf{X}))} - \mathbf{P}. \tag{25}$$

Thus, setting $\nabla_{\mathbf{X}} F(\mathbf{X}) = 0$ and solving for $\mathbf{X}$ yields the (unique) critical point:

$$\mathbf{X}^* = \log\mathbf{P} - \log(1 + t)\mathbf{I}, \tag{26}$$

with $t = \text{tr}(\exp(\mathbf{X}))$. Moreover, setting $p = \text{tr}(\mathbf{P})$ and tracing (25) readily yields:

$$t = p/(1 - p), \tag{27}$$

so the minimum value of F will be:

$$F_{\min} = F(\mathbf{X}^*) = \log(1 + \text{tr}(\exp(\mathbf{X}^*))) - \text{tr}(\mathbf{PX}^*)$$
$$= \log(1 + t) - \text{tr}(\mathbf{P}\log\mathbf{P}) + \log(1 + t)\,\text{tr}(\mathbf{P})$$
$$= -\text{tr}(\mathbf{P}\log\mathbf{P}) - (1 - p)\log(1 - p) \ge 0, \tag{28}$$

where, in the last step, we used the fact that $\mathbf{P} > 0$ and $0 \le \text{tr}(\mathbf{P}) \le 1$.

The above establishes the validity of (24), as claimed. Thus, returning to (23) and setting $\mathbf{P} = \mathbf{Q}^*/P$ (so $\mathbf{P} \succcurlyeq 0$ and $\text{tr}(\mathbf{P}) \le 1$), $\mathbf{X} = \eta\mathbf{Y}(T)$, an immediate application of (24) gives:

$$\int_0^T \text{tr}\left[(\mathbf{Q}(t) - \mathbf{Q}^*) \cdot \mathbf{V}(t)\right] dt \le \frac{P}{\eta}\log\left(1 + M\right), \tag{29}$$

which is simply (20). ∎

REFERENCES

[1] J. Zander, "Performance of optimum transmitter power control in cellular radio systems," *IEEE Trans. Veh. Technol.*, vol. 41, no. 1, pp. 57–62, Feb. 1992.

[2] G. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 641–646, Nov. 1993.

[3] S. Grandhi, R. Vijayan, and D. Goodman, "Distributed power control in cellular radio systems," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 226–228, 1994.

[4] N. Bambos, S. C. Chen, and G. J. Pottie, "Channel access algorithms with active link protection for wireless communication networks with power control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 583–597, Oct. 2000.

[5] T. Holliday, N. Bambos, P. Glynn, and A. Goldsmith, "Distributed power control for time varying wireless networks: Optimality and convergence," in *in Proceedings: Allerton Conference on Communications, Control, and Computing*, 2003.

[6] P. Mertikopoulos, E. V. Belmega, A. L. Moustakas, and S. Lasaulce, "Distributed learning policies for power allocation in multiple access channels," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 96–106, January 2012.

[7] J. Hannan, "Approximation to Bayes risk in repeated play," in *Contributions to the Theory of Games, Volume III*, ser. Annals of Mathematics Studies, M. Dresher, A. W. Tucker, and P. Wolfe, Eds. Princeton, NJ: Princeton University Press, 1957, vol. 39, pp. 97–139.

[8] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[9] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[10] J. Dams, M. Hoefer, and T. Kesselheim, "Convergence time of power-control dynamics," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 11, pp. 2231–2237, Dec. 2012.

[11] I. Stiakogiannakis, P. Mertikopoulos, and C. Touati, "No regrets: Distributed power control under time-varying channels and QoS requirements," in *2014 52st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct. 2014.

[12] B. Latifa, Z. Gao, and S. Liu, "No-regret learning for simultaneous power control and channel allocation in cognitive radio networks," in *Computing, Communications and Applications Conference (ComComAp), 2012*, Jan. 2012, pp. 267–271.

[13] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "Gambling in a rigged casino: The adversarial multi-armed bandit problem," in , *36th Annual Symposium on Foundations of Computer Science, 1995. Proceedings*, Oct. 1995, pp. 322–331.

[14] S. Maghsudi and S. Stanczak, "Joint channel selection and power control in infrastructureless wireless networks: A multi-player multi-armed bandit framework," *IEEE Trans. Veh. Technol.*, vol. PP, no. 99, pp. 1–1, 2014.

[15] P. Mertikopoulos and E. V. Belmega, "Transmit without regrets: online optimization in MIMO–OFDM cognitive radio systems," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, Nov. 2014.

[16] K. Tsuda, G. Rätsch, and M. K. Warmuth, "Matrix exponentiated gradient updates for on-line Bregman projection," *Journal of Machine Learning Research*, vol. 6, pp. 995–1018, 2005.

[17] P. Mertikopoulos, E. V. Belmega, and A. L. Moustakas, "Matrix exponential learning: Distributed optimization in MIMO systems," in *ISIT '12: Proceedings of the 2012 IEEE International Symposium on Information Theory*, 2012, pp. 3028–3032.

[18] S. M. Kakade, S. Shalev-Shwartz, and A. Tewari, "Regularization techniques for learning with matrices," *The Journal of Machine Learning Research*, vol. 13, pp. 1865–1890, 2012.

[19] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[20] H. Bolcskei, D. Gesbert, and A. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 225–234, 2002.

[21] E. Hazan and C. Seshadri, "Efficient learning algorithms for changing environments," in *ICML '09: Proceedings of the 26th International Conference on Machine Learning*, 2009.

[22] N. Cesa-Bianchi, P. Gaillard, G. Lugosi, and G. Stoltz, "Mirror descent meets fixed share (and feels no regret)," in *Advances in Neural Information Processing Systems*, 989-997, Ed., vol. 25, 2012.

[23] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *ICML '03: Proceedings of the 20th International Conference on Machine Learning*, 2003.

[24] J. Kwon and P. Mertikopoulos, "A continuous-time approach to online optimization," 2014, http://arxiv.org/abs/1401.6956.

[25] "User equipment (UE) radio transmission and reception," 3GPP, Technical Specification 36.101 V12.4.0, Jun. 2014. [Online]. Available: http://www.3gpp.org

[26] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.