Online Power Allocation for Opportunistic Radio Access in Dynamic OFDM Networks

Alexandre Marcastel, E Veronica Belmega, Panayotis Mertikopoulos, Inbar Fijalkow

Online Power Allocation for Opportunistic Radio Access in Dynamic OFDM Networks

Alexandre Marcastel*, E. Veronica Belmega*†, Panayotis Mertikopoulos‡† and Inbar Fijalkow*
* ETIS/ENSEA – UCP – CNRS, Cergy-Pontoise, France
† Inria
‡ French National Center for Scientific Research (CNRS), LIG F-38000 Grenoble, France

***Abstract*—User mobility has become a key attribute in the design of optimal resource allocation policies for future wireless networks. This has become increasingly apparent in cognitive radio (CR) systems where the licensed, primary users (PUs) of the network must be protected from harmful interference by the network's opportunistic, secondary users (SUs): here, unpredictability due to mobility requires the implementation of safety net mechanisms that are provably capable of adapting to changes in the users' wireless environment. In this context, we propose a distributed learning algorithm that allows SUs to adjust their power allocation profile (over the available frequency carriers) "on the fly", relying only on strictly causal channel state information. To account for the interference caused to the network's PUs, we incorporate a penalty function in the rate-driven objectives of the SUs, and we show that the proposed scheme matches asymptotically the performance of the best fixed power allocation policy in hindsight. Specifically, in a system with S orthogonal subcarriers and transmission horizon T, this performance gap (known as the algorithm's average *regret*) is bounded from above as $O(T^{-1}\log S)$. We also validate our theoretical analysis with numerical simulations which confirm that the network's SUs rapidly achieve a "no-regret" state under realistic wireless cellular conditions. Moreover, by finetuning the choice of penalty function, the interference induced by the SUs can be kept at a sufficiently low level, thus guaranteeing the PUs' requirements.**

***Index Terms*—Cognitive radio, distributed learning, regret minimization, interference management, OFDM.**

I. Introduction

Design specifications for future and emerging wireless systems target a massive increase in network capacity, fiber-like connection speeds (well into the Gb/s range), and an immersive overall user experience with speed-of-thought connectivity and zero effective latency times. As a result, the telecommunications industry is faced with a formidable challenge: these ambitious design goals necessitate the deployment of new wireless interfaces at a massive scale, but the required infrastructure overhaul is limited by hard economic and technological constraints [1]. In this context, the wireless users' unpredictable behavior, the increasingly flexible character of multi-tier cellular networks, and the increased portability of wireless devices represent major challenges for the design of resource allocation algorithms that are provably capable of adapting to changes in the users' wireless environment [2].

One of the most promising solution candidates for these challenges is that of cognitive radio (CR), a paradigm which allows opportunistic spectrum access by otherwise unlicensed users so as to maximize its utilization [3, 4]. In more detail, cognitive radio systems introduce a hierarchy based on spectrum licensing: on the one hand, the network's primary users (PUs) have leased part of the spectrum and must be protected from harmful interference by opportunistic spectrum access; on the other hand, the network's secondary users (SUs) try to free-ride on unallocated parts of the spectrum without compromising the PUs contractual quality of service guarantees. In particular, SUs are allowed to transmit over a shared set of channels, provided that the interference they induce to the network's PUs is kept below a pre-negotiated threshold.

In this paper, we focus on opportunistic radio access in orthogonal frequency division multiplexing (OFDM) systems where the wireless environment and the users' loads and demands change dynamically over time, in an unpredictable way. Specifically, we study the problem of SUs throughput maximization while keeping the interference caused to the network's PUs under a fixed tolerance set by the PUs. This problem has attracted considerable interest in the literature [5–8], but the vast majority of works on this topic have focused on the case where the users' channels remain static – or, at least, stationary – throughout the transmission horizon.

In more realistic network scenarios however, the users' mobility, their unpredictable behavior (going online and offline in an ad hoc manner), and the complex multi-path fading attributes of the wireless network cause this stringent stationarity assumption to fail. As a result, static solution concepts (such as social optima or Nash equilibria) [5–7] are no longer relevant when the users' wireless environment changes arbitrarily over time. Instead, a suitable solution framework is provided by online optimization and *regret minimization* methods [9, 10] which allow users to adapt to changes in the wireless environment, quickly and efficiently.

To achieve this, we propose a dynamic power allocation policy based on exponential learning [11, 12] that relies only on strictly causal channel state information [7, 13]. To illustrate the performance of the proposed algorithm, we focus on a simple network composed of one PU and several SUs which transmit

This research was supported in part by the Orange Lab Research Chair on IoT within the University of Cergy-Pontoise, by the CNRS funded project REAL.NET-PEPS JCJC 2016, and by ENSEA, Cergy-Pontoise, France.

Panayotis Mertikopoulos was partially supported by the French National Research Agency under grant no. ANR–13–INFR–004-NETLEARN.

simultaneously to a common access point (AP) over several orthogonal frequency bands. In this context, we are able to show that the SUs' average regret vanishes as $O(T^{-1}\log S)$ where S is the number of subcarriers and T is the transmission horizon. As a result, the algorithm is able to reach a no-regret state quickly, even for large numbers of subcarriers.

The closest work to ours is [7] where the authors use a similar learning technique assuming stationary channels and show that SUs converge to a Nash equilibrium; however, this result is no longer relevant in our case because the users' environment evolves *dynamically* over time. This dynamic aspect is present in [13] where the authors use online learning to maximize the SUs' throughput; nonetheless, because interference constraints are absent in [13], the algorithm proposed therein leads users to transmit at full power, thus causing significant interference to the PUs. Instead, the techniques developed herein enable the network's SUs to maximize their throughput while staying below the PUs' interference tolerance level, despite the system's unpredictability.

II. System Model and Problem Formulation

Consider a wireless cognitive radio network with one PU and K SUs, each transmitting to a common receiver via a shared channel over S non-interfering subcarriers. In this multiple access channel (MAC) context, the received signal at the AP in the subcarreir s is given by the familiar baseband model:

$$r_s = \sum_{k=1}^{K} x_{ks}h_{ks} + h_s^{\mathrm{PU}}x_s^{\mathrm{PU}} + w_s, \tag{1}$$

where x_{ks} is the transmitted signal of the k-th SU over subcarrier s, h_{ks} is the associated transfer coefficient between the k-th SU and the AP, x_s^{PU} and h_s^{PU} are the corresponding quantities for the PU, and w_s is the ambient noise in the channel.

For decoding purposes, we assume single user decoding (SUD) at the receiver, meaning that interference from non-designated transmitters is treated as additive (Gaussian) noise. In this case, the Shannon rate of user k at time t will be:

$$R_k(p_k;t) = \sum_{s=1}^{S} \log\left(1 + \frac{g_{ks}(t)p_{ks}}{\sigma_s^2 + \sum_{j\neq k} g_{js}(t)p_{js} + g_s^{PU}p_s^{PU}}\right), \tag{2}$$

where $p_k = (p_{ks})_{s=1}^{S}$ is the transmit power profile of user k, $g_{ks}(t) = |h_{ks}(t)|^2$, $s = 1,\dots,S$ are the associated channel gains at time t, and $\sigma_s^2 = \mathbb{E}[w_s^\dagger w_s]$ is the variance of the noise.

In the context of power-limited users, the users' total transmit power $P_k = \sum_{s=1}^{S} p_{ks}$ will be bounded from above by the maximum transmit power $P_{\max}$ of their wireless devices. Thus, the feasible power region of each user is

$$\mathcal{P}_k = \{p_k \in \mathbb{R}^S : p_{ks} \geq 0 \text{ and } \textstyle\sum_{s=1}^{S} p_{ks} \leq P_{\max}\}. \tag{3}$$

In the absence of other considerations, the unilateral objective of each user would be the maximization of their rate subject to the total power constraint (3) above. However, in a cognitive radio context, the network operator must also safeguard the contractual quality of service (QoS) guarantees that the PU has already paid for typically in the form of a maximum interference tolerance per subcarrier. On that account, the network operator also imposes to each SU the requirement

$$g_{ks}(t)p_{ks} \leq I_{\max}. \tag{4}$$

In contrast to the maximum power constraint of (3), the requirement (4) varies with time (because the SUs' channels themselves vary with time), so it cannot be enforced *a priori*: since there is no way to predict one's channel in advance, it is not possible to devise a policy that always respects this requirement either. Hence, instead of treating (4) as a (dynamic) physical constraint, we incorporate it in the SUs' (dynamic) utility function defined as follows:

$$U_k(p_k;t) = R_k(p;t) - \textstyle\sum_{s=1}^{S} C\left(g_{ks}(t)p_{ks}/I_{\max} - 1\right), \tag{5}$$

where $C(x)$ is a Lipschitz continuous, convex penalty function which is non-decreasing in x.

In the above, the convexity assumption for $C(x)$ essentially acts as an interference control mechanism. Specifically, it implies that the same increase in the incurred interference leads to a higher violation penalty when the network operates in a high-interference stare (as opposed to a mild-interference one). As a result, using a convex penalty scheme drives the network's SUs to transmit at lower powers relative to the PU's QoS requirements. As such our archetypal example will be the piecewise linear cost function:

$$C(x) = \begin{cases} \lambda x & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where λ is a sensitivity parameter that represents the incurred penalty when a SU violates the interference tolerance requirement (4).

In view of all this, we obtain the online problem:

$$\begin{aligned} &\text{maximize} && U_k(p_k;t) \\ &\text{subject to} && p_k \in \mathcal{P}_k \end{aligned} \tag{P}$$

Given that each SU's objective depends *explicitly* on time (via its dependence on the channel gains $g_{ks}(t)$), our aim will be to determine a dynamic power control policy $p_k(t)$ that is as close as possible to maximizing the above objective over time. However, given that it is not possible to predict the channel gains $g_{ks}(t)$ ahead of time, it is not possible to predict the optimal transmit power profile $p_k^*(t)$ which solves (P) in a real-time, online manner. Instead, we will focus on power allocation policies that can be implemented with strictly causal knowledge and which are asymptotically optimal in hindsight, in a sense made precise below.

To make all this precise, fix some horizon T over which the problem (P) is run, and let p_k^* denote the optimum (fixed) power profile over the horizon, i.e. the solution of the time-averaged problem:

$$p_k^* \in \operatorname*{arg\,max}_{p_k\in\mathcal{P}_k} \int_{t=0}^{T} U_k(p_k;t)\,dt, \tag{7}$$

where $\mathcal{P}_k$ is the feasible set of user k defined by the constraints (3). Of course, this solution can only be computed in hindsight

– i.e. assuming that all the parameters of the system (e.g. the channel gains g) are known ahead of the transmission – and will only serve as a theoretical benchmark for our online power allocation policy.

Specifically, to compare the performance of a dynamic power allocation policy $p_k(t)$ to that of the *a posteriori* optimum solution p_k^*, we define a user's (cumulative) *regret* [9, 10] as

$$\mathrm{Reg}_k(T) = \int_{t=0}^{T} U_k(p_k^*; t) - U_k(p_k(t); t)\, dt \tag{8}$$

In other words, a user's regret over the horizon T represents the cumulative performance gap between the proposed policy $p_k(t)$ and the optimum profile p_k^*; in particular, if $\mathrm{Reg}_k(T)$ grows linearly with T, it means that user k is not able to track changes in the system sufficiently fast. With this in mind, we will say that a power control policy leads to *no regret* if

$$\limsup_{T\to\infty} \mathrm{Reg}_k(T)/T \leq 0 \quad \text{for all } k, \tag{9}$$

irrespectively of how the system evolves over time.

If this is the case, it means that there is no fixed power profile yielding a higher utility in the long run; put differently, (9) provides an asymptotic guarantee that ensures that the policy $p(t)$ is at least as good as the mean optimal solution. We will further explore this property in Section IV.

III. Exponential Learning

To devise an online power allocation policy that leads to no regret in (P), our main idea will be as follows: as a first step, we will track the direction of steepest ascent of each user's utility (via its gradient) without taking into account the problem's constraints; subsequently, the resulting trajectory will be mapped back onto the problem's feasible region via an "exponential projection" step.

More precisely, we will consider the exponential learning scheme:

$$\begin{aligned} \dot{y}_{ks} &= v_{ks}, \\ p_{ks} &= P_{\max} \frac{\exp(y_{ks})}{1 + \sum_{s'=1}^{S} \exp(y_{ks'})}, \end{aligned} \tag{XL}$$

where $v_k = \partial_{p_k} U(p_k; t)$ denotes the unilateral (sub)gradient of the k-th user's utility function (for a pseudocode implementation, see Algorithm 1 below). As noted above, the *raison d'être* of the exponentiation step in (XL) is to project the auxiliary variable y_k back to $\mathcal{P}_k$: it is easy to see that $\sum_s p_{ks} \leq P_{\max}$, so the power allocation policy induced by (XL) is a feasible one.

With this in mind, our main theoretical result for (XL) is as follows:

Theorem 1. *The power allocation policy* (XL) *enjoys the regret bound*

$$\mathrm{Reg}_k(T) \leq P_{\max} \log(1 + S). \tag{10}$$

Consequently, the users' average regret $\mathrm{Reg}_k(T)/T$ *vanishes asymptotically as* $O(1/t)$*, i.e.* (XL) *leads to no regret.*

Proof: See Appendix A. ∎

Importantly, the regret bound provided in Theorem 1 is universal and only depends on the "size" of the users' feasible regions (the number of spectral degrees of freedom S and the users' maximum transmit power $P_{\max}$). In particular, the guarantee (10) does not depend on the system's average channel quality, number of users, or other attributes of the system. We thus conclude that (XL) is particularly flexible and can be used "as is" in a fairly wide range of decentralized CR systems: as long as the number of OFDM subcarriers shared by the focal users remains (roughly) constant, the users will converge to a no-regret state in the same rate.

Algorithm 1 Exponential learning (XL).

Parameter: step-size $\delta > 0$.

Initialization: $y_k \leftarrow 0$.

Repeat

allocate powers: $p_{ks} \leftarrow P_{\max} \frac{\exp(y_{ks})}{1+\sum_{s'=1}^{S} \exp(y_{ks'})}$;

get gradient data $v_k = \partial_{p_k} U_k(p_k; t)$;

update scores: $y_k \leftarrow y_k + \delta\, v_k$;

until termination criterion is reached.

IV. Numerical results

To validate our theoretical results we performed extensive numerical simulations of which we exhibit a representative sample below.

Our focus is an uplink cellular network with a fixed AP. Specifically, we consider a wireless system operating over a 10 MHz frequency band centered around the carrier frequency $f_c = 2$ GHz. The cell is a square of side-length equal to 2 km with the AP at its center. The network's PU is randomly positioned inside the cell and we consider $K = 9$ SUs, also placed randomly in the cell, following a Poisson point process. The maximum interference is fixed at $I_{\max} = -83$ dBm and we assume that the SUs' wireless devices have a maximum transmit power of $P_{\max} = 30$ dBm. Furthermore, each SU is assumed to be mobile with a speed chosen arbitrarily between 10 and 130 km/h. Finally, the channels between the wireless users and the AP are generated according to the realistic COST-HATA model for a suburban macro-cellular network [14] with fast- and shadow-fading attributes as in [15].

In Fig. 1, we plot the evolution of the SUs' channel gains and their respective Shannon rates as a function of time. To reduce graphical clutter, we only illustrate this data for three representative SUs at various distances from the AP. Specifically, the distance from the AP of each of the three focal users is $d_2 = 131.8$ m for SU 2, $d_3 = 172.7$ m for SU 3, and $d_7 = 779.6$ m for SU 7; respectively, the SUs' speeds are $v_2 = 50$ km/h, $v_3 = 90$ km/h, and $v_7 = 10$km/h. A first observation is that the SUs' rate is directly correlated to their channel gains; moreover, there are rapid variations in the SUs' throughput that are directly correlated with the variations – and responses – in the SU' power allocation policies.

In Fig. 2, we plot the evolution of the users' transmit powers (in dBm) and the cost for inducing harmful interference to the system's PU. In this case, if the SUs' channel gains are low, the induced interference is also low, so users can transmit at max-

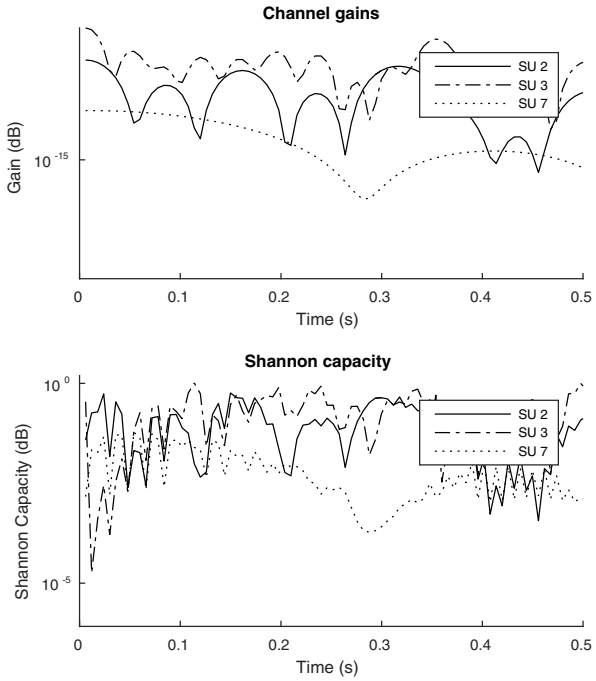

Fig. 1. Evolution of the users' channel gains and Shannon rates as a function of time for SUs $k = 2$, 3, and 7. For each SU, the Shannon rate is directly correlated to the channel gain and depends on the SU's transmit power. The rapid variations of the users' throughput stem from fast changes in the other SU's transmit policies.

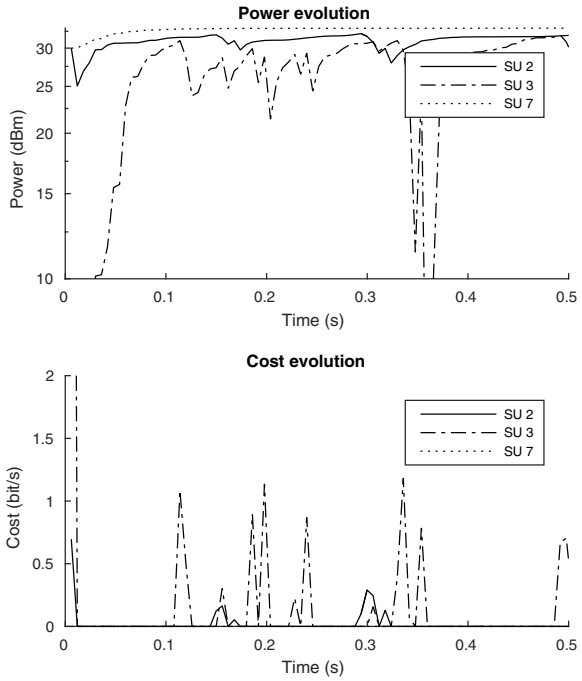

Fig. 2. Evolution of powers and cost functions as function of time for SUs $k = 2$, 3, and 7. When the channel gain is low (e.g. as in the case of SU 7) the SU creates little interference, which implies a zero penalty and transmission at $P_{\max}$. On the contrary, the interference inflicted by SUs $k = 2$ and 3 is often penalized: whenever an SU inflicts harmful interference at time t, the induced penalty at subsequent times is nonzero, so the user's adaptive power allocation policy leads to a reduction in the total transmit power.

imum power – i.e. at $P_{\max} = 30$ dBm (SU 7). On the contrary, when the channel gains become high, the induced interference also increases. As a result, the SUs transmitting at high powers are penalized via the penalty function (6) and decrease their transmit powers as a result thereof. Hence, the penalty function plays a key role in CR interference management.

In Fig. 3, we plot the evolution of the opportunistic users' average regret as a function of time. We see that each SU's regret quickly drops to non-positive values at a rate which depends on the user's individual channels and on the penalty parameter λ – cf. Eq. (6). As a result, the online power allocation policy we propose matches the best fixed transmit profile in hindsight within a few tens of iterations, despite the channels' significant variability over time.

Finally, in Fig. 4, we plot the fraction of times at which the PU's tolerated interference levels are violated – specifically, the fraction of iterations at which at least one SU causes interference above $I_{\max}$ to the PU. As expected, higher values of λ leads to fewer constraint violations. Hence, by combining the exponential learning policy (XL) with the penalty scheme (6), the network operator is able to allow opportunistic spectrum usage while effectively – and efficiently controlling the induced interference and protecting the PU's transmission – and, all this, despite the unpredictable variability of the network's channels over time.

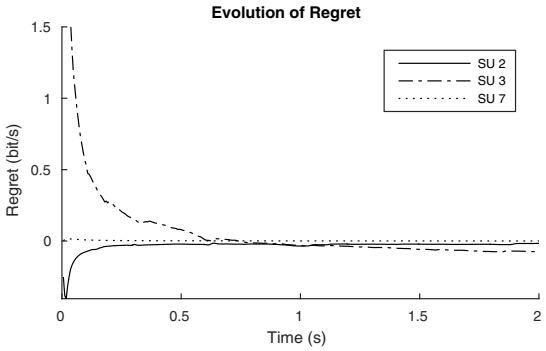

Fig. 3. Evolution of the users' average regret as function of time for SUs $k = 2$, 3, and 7. We see that the SUs' online power allocation policy quickly leads to nonnegative average regret; specifically, (XL) matches the optimal fixed transmit profile in hindsight within a few tens of iterations.

V. Conclusions and perspectives

In this paper, we proposed a distributed power allocation algorithm for the uplink of a time-varying cognitive radio network based on online optimization and exponential learning. Our algorithm allows the network's opportunistic users to achieve an optimal average performance in terms of throughput while ensuring at the same time that the induced interference is kept on average below a maximum, tolerated level. The proposed algorithm is simple, distributed, it relies on strictly causal channel state information, and its gap to the *a posteriori* fixed optimal policy decays as $O(T^{-1} \log S)$ in the number of subcarriers (S) and the transmission horizon (T). All these properties make for a promising power allocation policy in flexible and

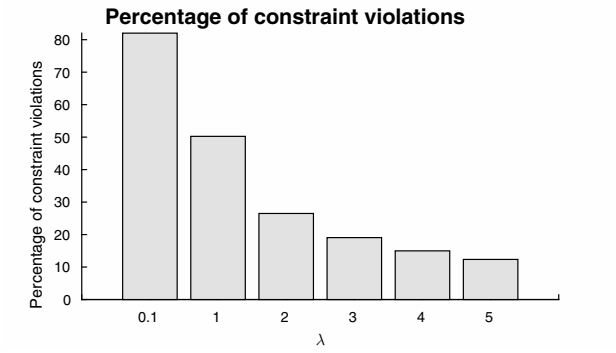

Fig. 4. Fraction of the time when at least one SU violates the maximum interference requirement (4) as function of λ. The higher the penalty parameter λ, the higher the penalty imposed to the SUs when (4); as a result, for high λ, the SUs create less interference beyond the tolerated levels. The network operator can thus manage the SUs' interference to protect the PU simply by tuning λ.

dynamic future wireless communications. Moreover, from a CR viewpoint, we show that the system owner can effectively allow opportunistic access to the spectrum while protecting the primary user's transmissions by tuning a scalar parameter which controls the trade-off between the opportunistic users' rates and their created interference in an adaptive, dynamic way.

Appendix

Proof of Theorem 1: The first step in proving the regret bound (10) is to use the concavity of the users' utility function to write

$$\begin{aligned}\mathrm{Reg}_k(T) &= \int_0^T U_k(p_k^*;t) - U_k(p_k(t);t)\,dt \\ &\le \int_0^T \left\langle v_k(p_k(t);t)\middle| p_k^* - p_k(t)\right\rangle\,dt \\ &= \left\langle y_k(T)\middle| p_k^*\right\rangle - \int_0^T \langle \dot{y}_k(t)|p_k(t)\rangle\,dt, \end{aligned} \tag{11}$$

where we have used the fact that $\dot{y}_k = v_k$ and that $v_k = \partial_{p_k} U_k$ by construction – recall the definition of the policy (XL) in Sec. III.

Note now that the exponentiation in (XL) can be written as

$$\frac{\exp(y_{ks})}{1+\sum_{s'=1}^S \exp(y_k s)} = \frac{\partial}{\partial y_{ks}} \log\left(1+\sum\nolimits_{s'=1}^S \exp(y_{ks'})\right). \tag{12}$$

Thus, letting $f(y_k) = P_{\max}\log\left(1+\sum_{s'=1}^S \exp(y_{ks'})\right)$, the bound (11) becomes:

$$\begin{aligned}\mathrm{Reg}_k(T) &\le \left\langle y_k(T)\middle| p_k^*\right\rangle - \int_0^T \left\langle \dot{y}_k(t)\middle|\nabla_{y_k} f(y_k(t))\right\rangle\,dt \\ &= \left\langle y_k(T)\middle| p_k^*\right\rangle - f(y_k(T)) + f(0), \end{aligned} \tag{13}$$

where we used the fact that (XL) is initialized with $y(0) = 0$.

By Fenchel's inequality [16], we then get

$$f(y) + f^*(p) \ge \langle y|p\rangle\,, \quad \text{for all } p, y \in \mathbb{R}^S, \tag{14}$$

where $f^*(p)$ denotes the convex conjugate of f, viz.

$$f^*(p) = \sup_{y\in\mathbb{R}^S} \langle y|p\rangle - f(y). \tag{15}$$

Thus, substituting in (13), we obtain

$$\mathrm{Reg}_k(T) \le f^*(p_k^*) + f(0) = f^*(p_k^*) + P_{\max}\log(1+S), \tag{16}$$

so we are left to show that $f^*(p) \le 0$ for all $p \in \mathcal{P} \equiv \{p \in \mathbb{R}^S : p_s \le 0 \text{ and } \sum_s p_s \le P_{\max}\}$. To that end, let $x_s = p_s/P_{\max}$; then, it suffices to show that

$$\sum\nolimits_{s=1}^S x_s y_s \le \log\left(1+\sum\nolimits_{s=1}^S \exp(y_s)\right), \tag{17}$$

for all $y \in \mathbb{R}^S$ and for all $x \in \Delta \equiv \{x \in \mathbb{R}^S : x_s \le 0 \text{ and } \sum_s x_s \le 1\}$. However, since the log-sum-exp function is convex in y, Jensen's inequality readily yields:

$$\begin{aligned}\exp\left(\sum\nolimits_{s=1}^S x_s y_s\right) &\le \sum\nolimits_{s=1}^S x_s \exp(y_s) \\ &\le 1 + \sum\nolimits_{s=1}^S x_s \exp(y_s), \end{aligned} \tag{18}$$

and (17) follows by taking logarithms on both sides. We conclude that $f^*(p_k) \le 0$ for all $p_k \in \mathcal{P}_k$, and our claim follows. ∎

References

[1] Qualcomm, "The 1000x data challenge," *Technical Report*, 2014.

[2] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5g be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[3] J. J. Mitola III and G. Q. M. Jr, "Cognitive radio: making software radios more personal," *Personal Communications, IEEE*, vol. 6, no. 4, pp. 13–18, 1999.

[4] A. Goldsmith, S. Jafar, I. Marić, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, 2009.

[5] G. Scutari and D. P. Palomar, "MIMO cognitive radio: A game theoretical approach," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 761–780, February 2010.

[6] R. Masmoudi, E. V. Belmega, I. Fijalkow, and N. Sellami, "Joint scheduling and power allocation in cognitive radio systems," in *Communication Workshop (ICCW), 2015 IEEE International Conference on*. IEEE, 2015, pp. 399–404.

[7] S. D'Oro, P. Mertikopoulos, A. L. Moustakas, and S. Palazzo, "Interference-based pricing for opportunistic multicarrier cognitive radio systems," vol. 14, no. 12, pp. 6536–6549, 2015.

[8] Z. Kenan and T. M. Lok, "Power control for uplink transmission with mobile users," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 5, pp. 2117–2127, 2011.

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

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

[11] P. Mertikopoulos and A. L. Moustakas, "Learning in the presence of noise," in *Game Theory for Networks, 2009. GameNets' 09. International Conference on*. IEEE, 2009, pp. 308–313.

[12] P. Mertikopoulos and W. H. Sandholm, "Learning in games via reinforcement and regularization," *arXiv preprint arXiv:1407.6267*, 2014.

[13] 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, pp. 1987–1999, 2014.

[14] G. F. Pedersen, *COST 231-Digital mobile radio towards future generation systems*. EU, 1999.

[15] G. Calcev, D. Chizhik, B. Göransson, S. Howard, H. Huang, A. Kogiantis, A. F. Molisch, A. L. Moustakas, D. Reed, and H. Xu, "A wideband spatial channel model for system-wide simulations," *Vehicular Technology, IEEE Transactions on*, vol. 56, no. 2, pp. 389–403, 2007.

[16] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1970.