

DISTRIBUTED LEARNING FOR RESOURCE ALLOCATION UNDER UNCERTAINTY

Panayotis Mertikopoulos^{*,‡} E. Veronica Belmega^{‡,‡} Luca Sanguinetti^{§,¶}

^{*} French National Center for Scientific Research (CNRS), LIG F-38000 Grenoble, France

[‡] Inria

[‡] ETIS/ENSEA – UCP – CNRS, Cergy-Pontoise, France

[§] University of Pisa, Dipartimento di Ingegneria dell’Informazione, Pisa, Italy

[¶] Large Networks and System Group, CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France

ABSTRACT

In this paper, we present a distributed matrix exponential learning (MXL) algorithm for a wide range of distributed optimization problems and games that arise in signal processing and data networks. To analyze it, we introduce a novel stability concept that guarantees the existence of a unique equilibrium solution; under this condition, we show that the algorithm converges even in the presence of highly defective feedback that is subject to measurement noise, errors, etc. For illustration purposes, we apply the proposed method to the problem of energy efficiency (EE) maximization in multi-user, multiple-antenna wireless networks with imperfect channel state information (CSI), showing that users quickly achieve a per capita EE gain between 100% and 400%, even under very high uncertainty.

Index Terms— Matrix exponential learning, stochastic optimization, game theory, variational stability, uncertainty.

1. INTRODUCTION

The emergence of massively large heterogeneous networks operating in random, dynamic environments is putting existing system design methodologies under enormous strain and has intensified the need for distributed resource management protocols that remain robust in the presence of randomness and uncertainty. To mention but an example, fifth generation (5G) mobile systems – the wireless backbone of the emerging Internet of things (IoT) paradigm [1] – envision millions of connected devices interacting in randomly-varying environments, typically with very stringent quality of service (QoS) targets that must be met in a reliable, distributed manner [2]. As such, the fusion of game theory, learning and stochastic optimization has been identified as one of the most promising theoretical frameworks for the design of efficient resource allocation policies in large, networked systems [3].

In view of the above, this paper aims to provide a distributed learning algorithm for a broad class of concave games and distributed optimization problems that arise in signal processing and wireless communication networks. Specifically, the proposed learning scheme has been designed with the following operating properties in mind: (i) *Distributedness*: player updates should be based only on local information and measurements; (ii) *Robustness*: feedback and measurements may be subject to random errors and noise; (iii) *Statelessness*: players should not be required to observe the state (or interaction structure) of the system; and (iv) *Flexibility*: players can employ the algorithm in both static and stochastic environments.

To achieve this, we build on the method of *matrix exponential learning* (MXL) that was recently introduced in [4, 5] for throughput maximization in multiple-input and multiple-output (MIMO) networks. The main idea of MXL is that each player tracks the individual gradient of his utility function via an auxiliary “score” matrix, possibly subject to randomness and errors. The players’ actions are then computed via an “exponential mirroring” step that maps these score matrices to the players’ action spaces, and the process repeats.

2. RELATION TO PRIOR WORK

Our work here is related to the online matrix regularization framework of [6] and the exponential weight (EW) algorithm for multi-armed bandit problems [7] (the latter in the case of vector variables on the simplex). MXL schemes have also been proposed for regret minimization in time-varying MIMO systems [8], the aim being to achieve a long-run average throughput that matches the best fixed policy in hindsight. However, the “no-regret” properties derived in these works are neither necessary nor sufficient to ensure convergence to a solution of the underlying game (or optimization problem in the case of a single decision-maker), so the relevant regret minimization literature does not apply to our setting.

Specifically, we show here that: a) MXL can be applied to a broad class of games and stochastic optimization problems (including both matrix and vector variables); b) unlike [4, 5],

This research was supported by the European Research Council under grant no. SG-305123-MORE, the French ANR project NETLEARN (ANR-13-INFR-004), the CNRS project REAL.NET-PEPS JCJC-2016, and by ENSEA, Cergy-Pontoise, France.

the algorithm’s convergence does not require the restrictive structure of a concave potential game [9]; and *c*) MXL remains convergent under very mild assumptions on the underlying stochasticity. Finally, to illustrate the practical benefits of the MXL method, we apply it to the problem of energy efficiency maximization in multi-user MIMO (MU-MIMO) networks with imperfect channel state information (CSI). To the best of our knowledge, this comprises the first distributed solution scheme for general MU-MIMO networks under noisy feedback/CSI.

3. PROBLEM FORMULATION

Consider a finite set of optimizing *players* $\mathcal{N} = \{1, \dots, N\}$, each controlling a positive-semidefinite matrix variable \mathbf{X}_i with the aim of improving their individual well-being. Assuming that this well-being is quantified by a *utility* (or *payoff*) function $u_i(\mathbf{X}_1, \dots, \mathbf{X}_N)$, we obtain the coupled optimization problem

$$\text{for all } i \in \mathcal{N} \quad \begin{cases} \text{maximize} & u_i(\mathbf{X}_1, \dots, \mathbf{X}_N) \\ \text{subject to} & \mathbf{X}_i \in \mathcal{X}_i \end{cases} \quad (1)$$

where \mathcal{X}_i denotes the set of feasible actions of player i (obviously, when $N = 1$, this is a standard semidefinite optimization problem). Specifically, motivated by applications to signal processing and wireless communications, we will focus on feasible action sets of the general form

$$\mathcal{X}_i = \{\mathbf{X}_i \succeq 0 : \|\mathbf{X}_i\| \leq A_i\} \quad (2)$$

where $\|\mathbf{X}_i\| = \sum_{m=1}^M |\text{eig}_m(\mathbf{X}_i)|$ denotes the nuclear matrix (or trace) norm of \mathbf{X}_i , A_i is a positive constant, and the players’ utility functions u_i are assumed individually concave and smooth in \mathbf{X}_i for all $i \in \mathcal{N}$.¹

The coupled multi-agent, multi-objective problem (1) constitutes a *game*, which we denote by \mathcal{G} . Problems of this type are extremely widespread in several areas of signal processing, ranging from computer vision [10, 11] and wireless networks [5, 12–14], to matrix completion and compressed sensing [15], especially in a stochastic setting where: *a*) the objective functions u_i are themselves expectations over an underlying random variable; and/or *b*) the feedback to the optimizers is subject to noise and/or measurement errors.

Setting aside for now any stochasticity issues, we turn to the characterization of the solutions of (1). To this end, the most widely used solution concept is that of a *Nash equilibrium* (NE), defined here as any action profile $\mathbf{X}^* \in \mathcal{X}$ which is *unilaterally stable*, i.e.

$$u_i(\mathbf{X}^*) \geq u_i(\mathbf{X}_i; \mathbf{X}_{-i}^*) \quad \text{for all } \mathbf{X}_i \in \mathcal{X}_i, i \in \mathcal{N}. \quad (3)$$

Thanks to the concavity of each player’s payoff function u_i and the compactness of their action space \mathcal{X}_i , the existence

¹For flexibility, \mathbf{X}_i may also be required to fit a specific block-diagonal form (for instance, a diagonal matrix when working with vector variables); however, for notational simplicity, we do not include this requirement in (2).

of a Nash equilibrium in the case of (1) is guaranteed by the general theory of [16]. As for uniqueness, let

$$\mathbf{V}_i(\mathbf{X}) \equiv \nabla_{\mathbf{X}_i} u_i(\mathbf{X}_i; \mathbf{X}_{-i}) \quad (4)$$

denote the *individual payoff gradient* of player i , and let $\mathbf{V}(\mathbf{X}) \equiv \text{diag}(\mathbf{V}_1(\mathbf{X}), \dots, \mathbf{V}_N(\mathbf{X}))$. Rosen [17] showed that \mathcal{G} admits a unique equilibrium solution when $\mathbf{V}(\mathbf{X})$ satisfies the so-called *diagonal strict concavity* (DSC) condition

$$\text{tr}[(\mathbf{X}' - \mathbf{X})(\mathbf{V}(\mathbf{X}') - \mathbf{V}(\mathbf{X}))] \leq 0 \quad \text{for all } \mathbf{X}, \mathbf{X}' \in \mathcal{X}, \text{ (DSC)}$$

with equality if and only if $\mathbf{X} = \mathbf{X}'$.

More recently, this monotonicity condition was used by Scutari et al. (see [13] and references therein) as the starting point for the convergence analysis of a class of Gauss–Seidel methods for concave games based on variational inequalities [18]. Our approach is similar in scope but relies instead on the following notion of stability:

Definition 1. An action profile $\mathbf{X}^* \in \mathcal{X}$ is called *globally stable* if it satisfies the *variational stability* condition:

$$\text{tr}[(\mathbf{X} - \mathbf{X}^*)\mathbf{V}(\mathbf{X})] \leq 0 \quad \text{for all } \mathbf{X} \in \mathcal{X}. \quad (\text{VS})$$

Obviously, if \mathbf{X}^* satisfies (VS), it is the unique Nash equilibrium of the game; also, (VS) is implied by (DSC) but the converse is not true [19]. In fact, as we show in the next section, the variational stability condition (VS) plays a key role not only in characterizing the structure of the game’s Nash set, but also for determining the convergence properties of the proposed learning scheme.

4. LEARNING UNDER UNCERTAINTY

In this section, we provide a distributed learning scheme that allows players to converge to stable Nash equilibria in a decentralized way under uncertainty information. Intuitively, the main idea of the proposed method is as follows: At each stage $n = 0, 1, \dots$ of the process, each player $i \in \mathcal{N}$ estimates the individual gradient $\mathbf{V}_i(\mathbf{X}(n))$ of his utility function at the current action profile $\mathbf{X}(n)$, possibly subject to measurement errors and noise. Subsequently, every player takes a step along this gradient estimate, and “reflects” this step back to \mathcal{X}_i via an exponential “mirror map”.

More precisely, we will focus on the following *matrix exponential learning* (MXL) scheme (see also Algorithm 1):

$$\begin{aligned} \mathbf{Y}_i(n+1) &= \mathbf{Y}_i(n) + \gamma_n \hat{\mathbf{V}}_i(n), \\ \mathbf{X}_i(n+1) &= A_i \frac{\exp(\mathbf{Y}_i(n+1))}{1 + \|\exp(\mathbf{Y}_i(n+1))\|}, \end{aligned} \quad (\text{MXL})$$

where:

1. $n = 0, 1, \dots$ denotes the stage of the process.
2. the auxiliary matrix variables $\mathbf{Y}_i(n)$ are initialized to an arbitrary (Hermitian) value.

Algorithm 1 Matrix exponential learning (MXL).

Parameter: step-size sequence $\gamma_n \sim 1/n^a$, $a \in (0, 1]$.

Initialization: $n \leftarrow 0$; $\mathbf{Y}_i \leftarrow$ any $M_i \times M_i$ Hermitian matrix.

Repeat

$n \leftarrow n + 1$;

foreach player $i \in \mathcal{N}$ **do**

 play $\mathbf{X}_i \leftarrow A_i \frac{\exp(\mathbf{Y}_i)}{1 + \|\exp(\mathbf{Y}_i)\|}$;

 get gradient feedback $\hat{\mathbf{V}}_i$;

 update auxiliary matrix $\mathbf{Y}_i \leftarrow \mathbf{Y}_i + \gamma_n \hat{\mathbf{V}}_i$;

until termination criterion is reached.

3. $\hat{\mathbf{V}}_i(n)$ is a stochastic estimate of the individual gradient $\mathbf{V}_i(\mathbf{X}(n))$ of player i at stage n (more on this below).
4. γ_n is a decreasing step-size sequence, typically of the form $\gamma_n \sim 1/n^a$ for some $a \in (0, 1]$.

If there were no constraints for the players' actions, $\mathbf{Y}_i(n)$ would define an admissible sequence of play and, *ceteris paribus*, player i would tend to increase his payoff along this sequence. However, this simple ascent scheme does not suffice in our constrained framework, so $\mathbf{Y}_i(n)$ is first exponentiated and subsequently normalized in order to meet the feasibility constraints (2).²

Of course, the outcome of the players' gradient tracking process depends crucially on the quality of the gradient feedback $\hat{\mathbf{V}}_i(n)$ that is available to them. To that end, we do not assume that players can observe each other's actions; instead, we only posit that each player has access to a "black box" feedback mechanism – an *oracle* – that returns an estimate of their individual gradient at a given action profile.³ With this in mind, we will consider the following sources of uncertainty:

- i) The players' gradient observations are subject to noise and/or measurement errors.
- ii) The players' utilities are stochastic expectations of the form $u_i(\mathbf{X}) = \mathbb{E}[\hat{u}_i(\mathbf{X}; \omega)]$ for some random variable ω , and players can only observe the realized gradient of \hat{u}_i .

On account of this, we will focus on the general model:

$$\hat{\mathbf{V}}_i(n) = \mathbf{V}_i(\mathbf{X}(n)) + \mathbf{Z}_i(n), \quad (5)$$

where the noise process $\mathbf{Z}(n)$ satisfies the hypotheses:

(H1) *Zero-mean:*

$$\mathbb{E}[\mathbf{Z}(n) | \mathbf{X}(n)] = \mathbf{0}. \quad (H1)$$

(H2) *Finite mean squared error (MSE):*

$$\mathbb{E}[\|\mathbf{Z}(n)\|_\infty^2 | \mathbf{X}(n)] \leq \sigma_*^2 \quad \text{for some } \sigma_* > 0. \quad (H2)$$

The statistical hypotheses (H1) and (H2) above are fairly mild and allow for a broad range of estimation scenarios. In more

²Note here that $\exp(\mathbf{Y}_i) > 0$ since the gradient matrices \mathbf{V}_i are automatically Hermitian (as derivatives of a real function).

³For ways to construct such oracles in a distributed setting, see [5, 20, 21].

detail, the zero-mean hypothesis (H1) is a minimal requirement for feedback-driven systems, simply positing that there is no *systematic* bias in the players' information. Likewise, (H2) is a bare-bones assumption for the variance of the players' feedback, and it is satisfied by most common error processes – such as uniform, Gaussian, log-normal, and all sub-exponential distributions.

With all this at hand, our main result is as follows:

Theorem 1. *Assume that (MXL) is run with a step-size γ_n such that $\sum_{n=1}^\infty \gamma_n^2 < \sum_{n=1}^\infty \gamma_n = \infty$ and gradient estimates satisfying (H1) and (H2). If \mathbf{X}^* is globally stable, then $\mathbf{X}(n)$ converges to \mathbf{X}^* (a.s.).*

Sketch of proof. The proof is relatively involved so, due to space limitations, we only sketch here the main steps thereof; for a detailed treatment, see [19, 22].

The first step is to consider a deterministic, "mean field" approximation of (MXL) in continuous time, namely

$$\begin{aligned} \dot{\mathbf{Y}}_i &= \mathbf{V}_i(\mathbf{X}(t)), \\ \mathbf{X}(t) &= \frac{A_i \exp(\mathbf{Y}_i(t))}{1 + \|\exp(\mathbf{Y}_i(t))\|}. \end{aligned} \quad (6)$$

If \mathbf{X}^* is stable, it can be shown that the so-called quantum divergence $D_{\text{KL}}(\mathbf{X}^*, \mathbf{X}) = \text{tr}[\mathbf{X}^* (\log \mathbf{X} - \log \mathbf{X}^*)]$ is a strict Lyapunov function for (6) [23, 24], implying that $\mathbf{X}(t)$ converges to \mathbf{X}^* in continuous time. Moreover, under the stated step-size and error variance assumptions, a well-known result from the theory of stochastic approximation [25] shows that $\mathbf{Y}(n)$ is an asymptotic pseudotrajectory (APT) of (6), i.e. the orbits of (6) asymptotically shadow the sequence $\mathbf{Y}(n)$ with arbitrary accuracy over any fixed horizon. By using an exponential concentration inequality for sequences of martingale differences [26], it is possible to estimate the aggregate error between an APT of (6) and an actual solution thereof, allowing us to show that they are both attracted to \mathbf{X}^* (a.s.). \square

5. NUMERICAL RESULTS

In this section, we assess the performance of the MXL algorithm in practical wireless networks by means of numerical simulations. Specifically, driven by current design requirements for 5G mobile networks that target a dramatic decrease in energy-per-bit consumption [14] through the use of multi-antenna transceivers [2], we focus here on the problem of energy efficiency (EE) maximization in MU-MIMO networks.

Our network model consists of a set of N transmitters, each equipped with M transmit antennas and controlling their individual input signal covariance matrix $\mathbf{Q}_i \succeq \mathbf{0}$ subject to the constraint $\text{tr} \mathbf{Q}_i \leq P_{\max}$, where P_{\max} denotes the user's maximum transmit power. Each user's achievable rate is then given by the familiar expression

$$R_i(\mathbf{Q}) = \log \det (\mathbf{W}_{-i} + \mathbf{H}_{ji} \mathbf{Q}_i \mathbf{H}_{ji}^\dagger) - \log \det (\mathbf{W}_{-i}), \quad (7)$$

where \mathbf{H}_{ji} denotes the channel matrix between the j -th transmitter and the i -th receiver, and $\mathbf{W}_{-i} = \mathbf{I} + \sum_{j \neq i} \mathbf{H}_{ji} \mathbf{Q}_j \mathbf{H}_{ji}^\dagger$

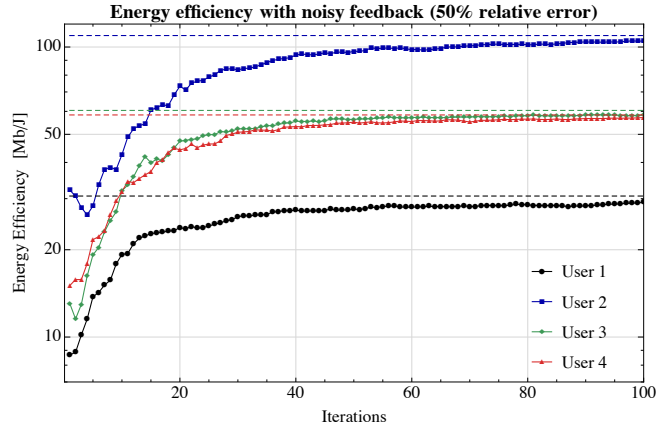
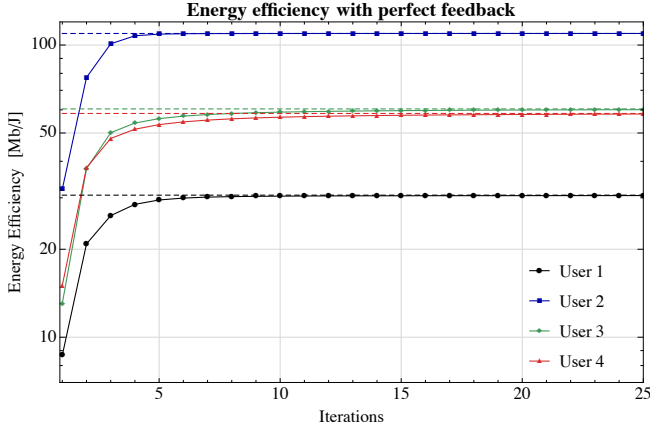


Fig. 1: Performance of MXL in the presence of noise. In both figures, we plot the transmit energy efficiency of wireless users that employ Algorithm 1 in a wireless network with parameters as in Table 1 (to reduce graphical clutter, we only plotted 4 users with diverse channel characteristics). In the absence of noise (left), the system converges to a stable Nash equilibrium state (unmarked dashed lines) within a few iterations. The convergence speed of MXL is slower in the presence of noise (right subfigure) but the algorithm remains convergent under high uncertainty (of the order of $z = 50\%$ of the users’ mean observations).

Table 1: Wireless network simulation parameters

Parameter	Value
Cell size (rectangular)	1 km
Central frequency	2.5 GHz
Total bandwidth	11.2 MHz
Spectral noise density (20 °C)	-174 dBm/Hz
Maximum transmit power	$P_{\max} = 33$ dBm
Non-radiative power	$P_c = 20$ dBm
TX/RX antennas per device	$M = 4, N = 8$

denotes the multi-user interference-plus-noise (MUI) covariance matrix at the i -th receiver. The users’ transmit *energy efficiency* (EE) is thus defined as the ratio between the achievable rate and the total consumed power, i.e.

$$EE_i(\mathbf{Q}) = \frac{R_i(\mathbf{Q})}{P_c + \text{tr}(\mathbf{Q}_i)}, \quad (8)$$

where $P_c > 0$ represents the total power consumed by circuit components at the transmitter [27].

Even though $EE_i(\mathbf{Q})$ is not concave in \mathbf{Q}_i , it can be recast as such via a suitable Charnes-Cooper transformation [28], as discussed in [20]. This leads to a game-theoretic formulation of the general form (1), with randomness and uncertainty entering the process due to the noisy estimation of the users’ MUI covariance matrices, the scarcity of perfect CSI at the transmitter, random measurement errors, etc.

For simulation purposes, we consider a macro-cellular wireless network with parameters as in Table 1. For benchmarking purposes, we first simulate the case where users have perfect CSI measurements at their disposal. In this deterministic regime, the algorithm converges to a stable Nash equilibrium state within a few iterations (for simplicity, we only plotted 4 users with diverse channel characteristics). In turn, this rapid convergence leads to drastic gains in energy efficiency, ranging between $3\times$ and $6\times$ over uniform power allocation schemes.

Subsequently, this simulation cycle was repeated in the presence of observation noise and measurement errors. The intensity of the measurement noise was quantified via the relative error level of the gradient observations $\hat{\mathbf{V}}$, i.e. the standard deviation of $\hat{\mathbf{V}}$ divided by its mean (so a relative error level of $z\%$ means that, on average, the observed matrix $\hat{\mathbf{V}}$ lies within $z\%$ of its true value). We then plotted the users’ energy efficiency over time for a high relative noise level of $z = 50\%$. Fig. 1 shows that the network’s rate of convergence to a Nash equilibrium is negatively impacted by the noise in the users’ measurements; however, MXL remains convergent and the network’s users achieve a per capita gain in energy efficiency between 100% and 400% within a few iterations, despite the noise and uncertainty.

6. CONCLUSIONS

In this paper, we investigated the convergence properties of a distributed matrix exponential learning (MXL) scheme for a general class of concave games with noisy/imperfect feedback. To this end, we introduced a novel stability concept which generalizes Rosen’s diagonal strict concavity condition [17]. Our theoretical analysis reveals that MXL converges to globally stable states from any initialization, and under feedback imperfections of arbitrary magnitude. In view of this, the proposed MXL algorithm exhibits several desirable properties for large-scale resource allocation problems in networks: it is distributed, robust to feedback imperfections, requires only local information on the state of the system, and can be applied in both static and ergodic environments “as is”. To validate our analysis in practical scenarios, we applied the proposed method to the problem of energy efficiency maximization in MU-MIMO interference network. Our numerical results confirm that users quickly reach a stable solution, attaining gains between 100% and 400% in energy efficiency even under very high uncertainty.

REFERENCES

- [1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2347–2376, Fourthquarter 2015.
- [2] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.
- [3] G. Bacci, S. Lasaulce, W. Saad, and L. Sanguinetti, "Game theory for networks: A tutorial on game-theoretic tools for emerging signal processing applications," *IEEE Signal Processing Magazine*, vol. 33, no. 1, pp. 94 – 119, Jan 2016.
- [4] 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.
- [5] P. Mertikopoulos and A. L. Moustakas, "Learning in an uncertain world: MIMO covariance matrix optimization with imperfect feedback," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 5–18, January 2016.
- [6] 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.
- [7] V. G. Vovk, "Aggregating strategies," in *COLT '90: Proceedings of the 3rd Workshop on Computational Learning Theory*, 1990, pp. 371–383.
- [8] 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, November 2014.
- [9] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124 – 143, 1996.
- [10] R. Negrel, D. Picard, and P.-H. Gosselin, "Web-scale image retrieval using compact tensor aggregation of visual descriptors," *IEEE Magazine on MultiMedia*, vol. 20, no. 3, pp. 24–33, 2013.
- [11] M. Law, N. Thome, and M. Cord, "Fantope regularization in metric learning," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1051–1058.
- [12] E. V. Belmega, S. Lasaulce, and M. Debbah, "Power allocation games for mimo multiple access channels with coordination," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 3185–3192, June 2009.
- [13] G. Scutari, F. Facchinei, D. P. Palomar, and J.-S. Pang, "Convex optimization, game theory, and variational inequality theory in multiuser communication systems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 35–49, May 2010.
- [14] Huawei Technologies, "5G: A technology vision," White paper, 2013.
- [15] E. J. Candes, Y. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 199–225, February 2013.
- [16] G. Debreu, "A social equilibrium existence theorem," *Proceedings of the National Academy of Sciences of the USA*, vol. 38, no. 10, pp. 886–893, October 1952.
- [17] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave N -person games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.
- [18] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, ser. Springer Series in Operations Research. Springer, 2003.
- [19] P. Mertikopoulos, E. V. Belmega, R. Negrel, and L. Sanguinetti, "Distributed stochastic optimization via matrix exponential learning," <http://arxiv.org/abs/1606.01190>, 2016.
- [20] P. Mertikopoulos and E. V. Belmega, "Learning to be green: Robust energy efficiency maximization in dynamic MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 743 – 757, April 2016.
- [21] A. D. Flaxman, A. T. Kalai, and H. B. McMahan, "Online convex optimization in the bandit setting: gradient descent without a gradient," in *SODA '05: Proceedings of the 16th annual ACM-SIAM symposium on discrete algorithms*, 2005, pp. 385–394.
- [22] P. Mertikopoulos, "Learning in concave games with imperfect information," <https://arxiv.org/abs/1608.07310>, 2016.
- [23] V. Vedral, "The role of relative entropy in quantum information theory," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 197–234, 2002.
- [24] P. Mertikopoulos and W. H. Sandholm, "Learning in games via reinforcement and regularization," *Mathematics of Operations Research*, 2016, to appear.
- [25] M. Benaïm, "Dynamics of stochastic approximation algorithms," in *Séminaire de Probabilités XXXIII*, ser. Lecture Notes in Mathematics, J. Azéma, M. Émery, M. Ledoux, and M. Yor, Eds. Springer Berlin Heidelberg, 1999, vol. 1709, pp. 1–68.
- [26] V. H. de la Peña, "A general class of exponential inequalities for martingales and ratios," *The Annals of Probability*, vol. 27, no. 1, pp. 537–564, 1999.
- [27] E. Bjornson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3059–3075, June 2015.
- [28] A. Charnes and W. W. Cooper, "Programming with linear fractional functionals," *Naval Research Logistics Quarterly*, vol. 9, pp. 181–196, 1962.