Distributed stochastic optimization via matrix exponential learning

Panayotis Mertikopoulos, Elena Veronica Belmega, Romain Negrel, Luca Sanguinetti

Distributed Stochastic Optimization via Matrix Exponential Learning

Panayotis Mertikopoulos, *Member, IEEE*, E. Veronica Belmega, *Member, IEEE*,
Romain Negrel, and Luca Sanguinetti, *Senior Member, IEEE*

***Abstract*—In this paper, we investigate a distributed learning scheme for a broad class of stochastic optimization problems and games that arise in signal processing and wireless communications. The proposed algorithm relies on the method of matrix exponential learning (MXL) and only requires locally computable gradient observations that are possibly imperfect. To analyze it, we introduce the notion of a stable Nash equilibrium and we show that the algorithm is globally convergent to such equilibria – or locally convergent when an equilibrium is only locally stable. To complement our convergence analysis, we also derive explicit bounds for the algorithm's convergence speed and we test it in realistic multi-carrier/multiple-antenna wireless scenarios where several users seek to maximize their energy efficiency. Our results show that learning allows users to attain a net increase between 100% and 500% in energy efficiency, even under very high uncertainty.**

***Index Terms*—Learning, stochastic optimization, game theory, matrix exponential learning, variational stability, uncertainty.**

I. Introduction

CONSIDER a finite set of optimizing *players* (or *agents*) $\mathcal{K} = \{1, \dots, K\}$, each controlling a matrix variable $\mathbf{X}_k$, and seeking to improve their individual "well-being". Assuming that this "well-being" is quantified by a *utility* (or *payoff*) *function* $u_k(\mathbf{X}_1, \dots, \mathbf{X}_K)$, we obtain the problem

$$\text{for all } k \in \mathcal{K} \quad \begin{cases} \text{maximize} & u_k(\mathbf{X}_1, \dots, \mathbf{X}_K), \\ \text{subject to} & \mathbf{X}_k \in \boldsymbol{\mathcal{X}}_k, \end{cases} \tag{1}$$

where $\boldsymbol{\mathcal{X}}_k$ denotes the set of feasible actions of player k. Specifically, we focus here on feasible action sets of the general form

$$\boldsymbol{\mathcal{X}}_k = \{\mathbf{X}_k \succcurlyeq 0 : \|\mathbf{X}_k\| \leq A_k\} \tag{2}$$

where $\|\mathbf{X}_k\| = \sum_{m=1}^{M} |\mathrm{eig}_m(\mathbf{X}_k)|$ denotes the nuclear norm of $\mathbf{X}_k$, A_k is a positive constant, and the players' utility functions u_k are assumed individually concave and smooth in $\mathbf{X}_k$ for all $k \in \mathcal{K}$.

This research was supported by the European Reseach Council (ERC) under grant no. SG-305123-MORE, the French National Research Agency (ANR) under grants NETLEARN (ANR–13–INFR–004) and ORACLESS (ANR–16–CE33–0004–01), and by ENSEA, Cergy-Pontoise, France.

Part of this work was presented in the 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP 2016) [1].

P. Mertikopoulos is with the French National Center for Scientific Research (CNRS), and the Laboratoire d'Informatique de Grenoble (LIG), F-38000, Grenoble, France.

E. V. Belmega is with ETIS / ENSEA – UCP – CNRS, Cergy-Pontoise, France and Inria.

R. Negrel is with Université Paris-Est, LIGM, Equipe A3SI, ESIEE Paris, Marne-la-Vallée, France. Part of this work was carried out when R. Negrel was with the Université de Normandie, UNICAEN, ENSICAEN, CNRS, GREYC, France.

L. Sanguinetti is with the University of Pisa, Dipartimento di Ingegneria dell'Informazione, Italy, and with the Large Systems and Networks Group (LANEAS), CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France.

The coupled multi-agent, multi-objective problem (1) constitutes a *game*, which we denote by $\mathcal{G}$. As we discuss in the next section, games and optimization problems of this type have been studied extensively in signal processing and wireless communications, especially in a stochastic framework where: *a*) the objective functions u_k are themselves expectations over an underlying random variable (see Ex. II-B below); and/or *b*) the feedback to the optimizers is subject to noise and/or other measurement errors (Ex. II-C).

Accordingly, the main goal of this paper is to provide a learning algorithm that converges to a suitable solution of $\mathcal{G}$ subject to the following desiderata:

(*i*) *Distributedness*: player updates are based on local information and measurements.
(*ii*) *Robustness*: feedback and measurements may be subject to random errors, noise, and delays.
(*iii*) *Statelessness*: players are oblivious to the overall state (or interaction structure) of the system.
(*iv*) *Flexibility*: players can employ the algorithm in both static and ergodic environments.

To achieve this, we build on the method of *matrix exponential learning* (MXL) that was recently introduced in [2] for throughput maximization in multiple-input and multiple-output (MIMO) systems. The main idea of MXL is that each player tracks the individual gradient of their utility function via an auxiliary score matrix and chooses an action by means of an exponential "mirror step" that maps these score matrices to the players' feasible sets. In this general setting, we introduce the notion of *variational stability* (VS), and we show that stable Nash equilibria are locally attracting with high probability, while globally stable equilibria are globally attracting with probability 1. Finally, we derive explicit estimates for the method's convergence rate, which we validate in the context of energy efficiency maximization in multi-antenna systems.

A. Related work

The MXL method studied in this paper has strong ties to matrix regularization [3] and mirror descent methods [4, 5] for (online) convex optimization. In particular, the important special case of real vector variables on the simplex (real diagonal $\mathbf{X}_k$) is closely related to the well-known exponential weight (EW) learning algorithm for multi-armed bandit problems [6]. More recently, MXL schemes were also proposed for single-user regret minimization in online power control [7] and rate maximization problems [8] for dynamic MIMO systems. The goal there was to show that the long-run performance of matrix

exponential learning matches that of the best fixed policy in hindsight, a property known as "no regret" [5, 9]. Nonetheless, in a multi-user, game-theoretic setting, a no-regret strategy may end up assigning positive weight *only* to strictly dominated actions [10]. As a result, regret minimization is neither necessary nor sufficient to ensure convergence to Nash equilibrium.

In [2, 11, 12], a special case of (1) was studied in the context of MIMO rate maximization in the presence of stochasticity, in both single-user [12] and multiple access channels [2, 11]. In both cases, the problem boils down to a (possibly distributed) semidefinite optimization problem [2]. The existence of a single objective function greatly facilitates the analysis; however, many cases of practical interest (such as the examples we discuss in the following section) cannot be modeled as problems of this type, making a potential-based analysis unsuitable.

In this paper, we derive the convergence properties of matrix exponential learning in K-player games of the form (1), and we investigate the algorithm's long-term behavior under feedback errors and uncertainty. Specifically, we consider a broad class of games that satisfy a local monotonicity condition which ensures that Nash equilibria are isolated. In a series of recent papers, Scutari *et al.* used a variant of this condition to establish the convergence of a class of Gauss–Seidel, best-response methods, and successfully applied these algorithms to a wide range of communication problems – for a survey, see [13]. However, under noise and uncertainty, the convergence of best-response methods is often compromised: as an example, in the case of throughput maximization with imperfect feedback, iterative water-filling (a standard best-response scheme) fails to provide any perceptible gains over crude, uniform power allocation policies [2]. Thus, given that randomness, uncertainty and feedback imperfections are ubiquitous in practical systems, we focus throughout on attaining convergence results that are robust to learning impediments of this type.

B. Main contributions and paper outline

The main contribution of this work is the derivation of the convergence properties of matrix exponential learning in games played over bounded regions of positive-semidefinite matrices. To put our theoretical analysis in context, we first discuss three examples from wireless networks and computer vision in Section II. More specifically, we illustrate *i*) a general game-theoretic framework for *contention-based medium access control* (MAC); *ii*) a stochastic optimization formulation for *content-based image retrieval* (a key "Big Data" signal processing problem); and *iii*) the multi-agent problem of *energy efficiency* (EE) maximization in multi-carrier MIMO networks (a critical design feature of green multi-cellular networks).

In Section III, we revisit our core game-theoretic framework, and we introduce the notion of variational stability. Subsequently, in Section IV, we introduce the matrix exponential learning scheme under study, and we detail our assumptions for the uncertainty surrounding the players' objectives and observations. Our main results are presented in Section V: under fairly mild assumptions on the underlying stochasticity, we show that (*i*) the algorithm's only termination states are Nash equilibria; (*ii*) if the game admits a globally (locally) stable equilibrium, then MXL converges globally (locally) to said equilibrium; and (*iii*) on average, MXL converges within ε of a strongly stable equilibrium in $\mathcal{O}(1/\varepsilon^2)$ iterations.

These results greatly extend the recent analysis of [2] for rate maximization in MIMO MAC systems. To further validate our results in MIMO environments, we supplement our analysis with extensive numerical simulations for energy efficiency maximization in multi-carrier MIMO wireless networks in Section VI. To streamline the flow of the paper, most proofs have been relegated to a series of appendices at the end.

Notation: The profile $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_K)$ is identified with the block-diagonal matrix $\operatorname{diag}(\mathbf{X}_k)_{k=1}^K$, and we use the shorthand $(\mathbf{X}_k; \mathbf{X}_{-k})$ to highlight the action $\mathbf{X}_k$ of player k against that of all other players. Also, given $\mathbf{A} \in \mathbb{C}^{d\times d}$, we write $\|\mathbf{A}\| = \sum_{j=1}^d |\operatorname{eig}_j(\mathbf{A})|$ for the nuclear (trace) norm of $\mathbf{A}$ and $\|\mathbf{A}\|_\infty = \max_j |\operatorname{eig}_j(\mathbf{A})|$ for its singular norm.

II. Motivation and Examples

To motivate the general framework of (1), we illustrate below three examples taken from communication networks and computer vision. A reader who is interested in the general theory may skip this section and proceed directly to Sections III–VI.

A. Contention-based medium access

Contention-based medium access control (MAC) aims to provide an efficient means for accessing and sharing a wireless channel in the presence of several interfering wireless users. To model this, consider a set of users indexed by $\mathcal{K} = \{1, \dots, K\}$, each updating their individual channel access probability x_k based on the amount of contention in the network [14]. In practice, users cannot be assumed to know the exact channel access probabilities of other users, so user i infers the level of wireless contention via an aggregate *contention measure* $q_k(x_{-k})$, i.e. a (symmetric) function of the access probability profile $x_{-k} = (x_1, \dots, x_{k-1}, x_{k+1} \dots, x_K)$ of all other users.[1]

With this in mind, the objective of each user is to select their individual channel access probability x_k so as to maximize the benefit derived from acessing the channel more often minus the induced contention $x_k q_k(x_{-k})$ incurred by all other users. This leads to the utility function formulation

$$u_k(x_k; x_{-k}) = U_k(x_k) - x_k q_k(x_{-k}) \tag{3}$$

where U_k is a continuous and nondecreasing function representing the utility of user k when there are no other users in the channel. Thus, in economic terms, u_k simply represents the user's net gain from channel access, discounted by the associated contention cost.

In this way, we obtain the game-theoretic formulation

$$\text{for all } k \in \mathcal{K} \quad \begin{cases} \text{maximize} & u_k(x_k; x_{-k}), \\ \text{subject to} & x_k \in [0, 1] \end{cases} \tag{4}$$

whose solutions can be analyzed through the specification of the utility function $U_k(x_k)$ and the choice of the contention measure q_k. If $U_k(x_k)$ is assumed smooth and strictly concave (see for example [14, 15] and references therein), (4) is a special case of (1) with $M = 1$ and $A_k = 1$.

[1] For instance, a standard contention measure of this type is the conditional collision probability $q_k(\mathbf{x}_{-k}) = 1 - \prod_{\ell\neq k}(1 - x_\ell)$. [14].

B. Metric learning for image similarity search

A key challenge in content-based image retrieval is to design an automatic procedure capable of retrieving documents from a large database based on their similarity to a given request, often distributed over several computing cores in a massively parallel computing grid (or cloud) [16, 17]. To formalize this, an image is typically represented via its *signature*, i.e. a real d-dimensional vector $\mathbf{i} \in \mathbb{R}^d$ that collects and encodes the most distinctive features of said image. Given a database $\mathcal{D}$ of such signatures, each image $\mathbf{i} \in \mathcal{D}$ is further associated with a set $\mathcal{S}_\mathbf{i} \subseteq \mathcal{D}$ of *similar* images and a set $\mathcal{U}_\mathbf{i} \subseteq \mathcal{D}$ of *dissimilar* ones, based on each image's content. Accordingly, the goal is to design a distance metric which is minimized between similar images and is maximized between dissimilar ones.

A widely used distance measure between image signatures is the *Mahalanobis distance*, defined here as

$$d_\mathbf{X}(\mathbf{i},\mathbf{j}) = (\mathbf{i}-\mathbf{j})^\top \mathbf{X}\,(\mathbf{i}-\mathbf{j}), \tag{5}$$

where $\mathbf{X} \succcurlyeq 0$ is a $d \times d$ positive-definite matrix.[2] This so-called "precision matrix" must then be chosen by the optimizer so that $d_\mathbf{X}(\mathbf{i},\mathbf{j}) < d_\mathbf{X}(\mathbf{i},\mathbf{k})$ whenever $\mathbf{i}$ is similar to $\mathbf{j}$ but dissimilar to $\mathbf{k}$. We thus obtain the minimization objective

$$F(\mathbf{X};\mathcal{T}) = \sum_{(\mathbf{i},\mathbf{j},\mathbf{k})\in\mathcal{T}} C(d_\mathbf{X}(\mathbf{i},\mathbf{j}) - d_\mathbf{X}(\mathbf{i},\mathbf{k}) - \varepsilon) + \|\mathbf{X}-\mathbf{I}\|_F^2, \tag{6}$$

where: *a*) $\mathcal{T}$ is the set of all triples $(\mathbf{i},\mathbf{j},\mathbf{k})$ such that $\mathbf{j} \in \mathcal{S}_\mathbf{i}$ and $\mathbf{k} \in \mathcal{U}_\mathbf{i}$; *b*) C is a convex, nondecreasing function that penalizes matrices $\mathbf{X}$ that do not capture similarities and dissimilarities between images; *c*) the parameter $\varepsilon > 0$ reinforces this penalty; and *d*) $\|\cdot\|_F$ denotes the ordinary Frobenius (ℓ^2) norm.[3]

An additional requirement in the above is to employ a low-rank precision matrix $\mathbf{X}$ so as to reduce model complexity and computational costs [16], enable distributed storing and retrieval [20], and better exploit correlations between features that further reduce over-fitting effects [19]. An efficient way to achieve this is to include a trace constraint of the form $\operatorname{tr}\mathbf{X} \le c$ for some $c \ll d$,[4] leading to the feasible region

$$\boldsymbol{\mathcal{X}} = \{\mathbf{X} \in \mathbb{R}^{d\times d} : \mathbf{X} \succcurlyeq 0 \text{ and } \operatorname{tr}\mathbf{X} \le c\}. \tag{7}$$

Thus, combining (6) and (7), we see that metric learning is a special case of the general problem (1) with $K = 1$ optimizers.

Of course, when $\mathcal{D}$ is large, the number of variables involved becomes computationally prohibitive: for instance, (6) may contain up to 10^9–10^{11} terms, even for a modest database of 10^4 images. To circumvent this obstacle, a common approach is to replace $\mathcal{T}$ with a smaller, randomly drawn population sample $\mathcal{W} \subseteq \mathcal{T}$, and then take the average over all such samples. In so doing, we obtain the stochastic optimization problem [21]:

$$\begin{aligned} &\text{minimize} && \mathbb{E}[F(\mathbf{X};\mathcal{W})] \\ &\text{subject to} && \mathbf{X} \succcurlyeq 0, \operatorname{tr}(\mathbf{X}) \le c, \end{aligned} \tag{8}$$

where the expectation is taken over the random samples $\mathcal{W}$.

[2]The baseline case $\mathbf{X} = \mathbf{I}$ corresponds to the Euclidean distance, which is often unsuitable for image discrimination purposes [18].

[3]This last regularization term is included in order to maximize predictive accuracy by reducing the effects of over-fitting to training data [19].

[4]By contrast, an ℓ_0 rank constraint of the form $\operatorname{rank}(\mathbf{X}) \le c$ generically leads to an untractable NP-hard problem formulation.

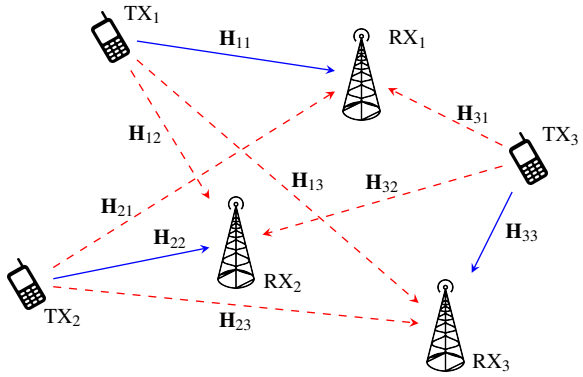


Fig. 1. A multi-user MIMO system with $K = 3$ connections.

The benefit of this formulation is that, at each realization, only a small-size, tractable data sample $\mathcal{W}$ is used for calculations at each computing node. On the down side, the expectation in (8) cannot be calculated, so the optimizer only has access to information on the random, realized gradients $\nabla_\mathbf{X} F(\mathbf{X};\mathcal{W})$. This type of uncertainty is typical of stochastic optimization problems, and the proposed MXL method has been designed precisely with this feedback structure in mind.

C. Energy efficiency maximization

Consider the problem of energy efficiency maximization in multi-user, multiple-carrier MIMO networks – see e.g. [22–24] and references therein. Here, wireless connections are established over transceiver pairs with N_t (resp. N_r) antennas at the transmitter (resp. receiver), and communication takes place over a set of S orthogonal subcarriers. Accordingly, the k-th transmitter is assumed to control their individual input signal covariance matrix $\mathbf{Q}_{ks}$ over each subcarrier $s = 1,\dots,S$, subject to the constraints: *a*) $\mathbf{Q}_{ks} \succcurlyeq 0$ (since each $\mathbf{Q}_{ks}$ is a signal covariance matrix); and *b*) $\operatorname{tr}\mathbf{Q}_k \le P_{\max}$, where $\mathbf{Q}_k = \operatorname{diag}(\mathbf{Q}_{ks})_{s=1}^S$ is the covariance profile of the k-th transmitter, $\operatorname{tr}(\mathbf{Q}_k)$ represents the user's transmit power over all subcarriers, and $P_{\max}$ denotes the user's *maximum* transmit power.

Assuming Gaussian input and single user decoding (SUD) at the receiver, each user's Shannon-achievable throughput is

$$R_k(\mathbf{Q}) = \log\det(\mathbf{W}_{-k} + \mathbf{H}_{kk}\mathbf{Q}_k\mathbf{H}_{kk}^\dagger) - \log\det(\mathbf{W}_{-k}), \tag{9}$$

where $\mathbf{H}_{\ell k}$ denotes the channel matrix profile between the ℓ-th transmitter and the k-th receiver over all subcarriers, $\mathbf{Q} = (\mathbf{Q}_1,\dots,\mathbf{Q}_K)$ denotes the users' transmit profile and $\mathbf{W}_{-k} \equiv \mathbf{W}_{-k}(\mathbf{Q}) = \mathbf{I} + \sum_{\ell\neq k}\mathbf{H}_{\ell k}\mathbf{Q}_\ell\mathbf{H}_{\ell k}^\dagger$ is the multi-user interference-plus-noise (MUI) covariance matrix at receiver k. The users' transmit *energy efficiency* (EE) is then defined as the ratio between their Shannon rate and the total consumed power, i.e.

$$\mathrm{EE}_k(\mathbf{Q}) = \frac{R_k(\mathbf{Q})}{P_c + \operatorname{tr}(\mathbf{Q}_k)}, \tag{10}$$

where $P_c > 0$ represents the total power consumed by circuit components at the transmitter [22].

The energy efficiency function above is not concave w.r.t the covariance matrix $\mathbf{Q}_k$ of user k, but it can be recast as such via a suitable transformation. To this aim, following [24, 25],

consider the adjusted control variables

$$\mathbf{X}_k = \frac{P_c + P_{\max}}{P_{\max}} \frac{\mathbf{Q}_k}{P_c + \operatorname{tr}(\mathbf{Q}_k)}, \tag{11}$$

where the normalization constant $(P_c + P_{\max})/P_{\max}$ implies that $\operatorname{tr}(\mathbf{X}_k) \leq 1$ with equality if and only if $\operatorname{tr}(\mathbf{Q}) = P_{\max}$. The action set of user k is then given by

$$\boldsymbol{\mathcal{X}}_k = \{\operatorname{diag}(\mathbf{X}_{ks})_{s=1}^S : \mathbf{X}_{ks} \succcurlyeq 0 \text{ and } \textstyle\sum_s \operatorname{tr} \mathbf{X}_{ks} \leq 1\}, \tag{12}$$

so, using (11), the energy efficiency expression (10) yields the utility function

$$u_k(\mathbf{X}_k; \mathbf{X}_{-k}) = \frac{P_c + (1 - \operatorname{tr} \mathbf{X}_k) P_{\max}}{P_c(P_c + P_{\max})} \times \log\det\left(\mathbf{I} + \frac{P_c P_{\max}\, \widetilde{\mathbf{H}}_k \mathbf{X}_k \widetilde{\mathbf{H}}_k^\dagger}{P_c + (1 - \operatorname{tr} \mathbf{X}_k) P_{\max}}\right), \tag{13}$$

where $\widetilde{\mathbf{H}}_k = \mathbf{W}_{-k}^{-1/2} \mathbf{H}_{kk}$ is the *effective channel matrix* of user k.

We thus see that multi-user energy efficiency maximization boils down to the general formulation (1) with $A_k = 1$.[5] In this setting, randomness and uncertainty stem from the noisy estimation of the users' MUI covariance matrices (which depend on the other users' behavior), the noise being due to the scarcity of perfect channel state information at the transmitter (CSIT), random measurement errors, etc. To the best of our knowledge, the MXL method discussed in Section IV comprises the first distributed solution scheme for energy efficiency maximization in general multi-user/multi-antenna/multi-carrier networks with local, causal – and possibly imperfect – channel state information (CSI) feedback at the transmitter.

III. Elements from Game Theory

The most widely used solution concept in noncooperative games is that of a *Nash equilibrium* (NE), defined here as an action profile $\mathbf{X}^* \in \boldsymbol{\mathcal{X}}$ which is *unilaterally stable*, i.e.

$$u_k(\mathbf{X}^*) \geq u_k(\mathbf{X}_k; \mathbf{X}_{-k}^*) \quad \text{for all } \mathbf{X}_k \in \boldsymbol{\mathcal{X}}_k,\ k \in \mathcal{K}. \tag{14}$$

In words, $\mathbf{X}^* \in \boldsymbol{\mathcal{X}}$ is a Nash equilibrium when no player can further increase their utility by deviating unilaterally from $\mathbf{X}^*$.

In complete generality, a game need not admit an equilibrium. However, thanks to the concavity of each player's payoff function u_k and the compactness of their action space $\boldsymbol{\mathcal{X}}_k$, existence of a Nash equilibrium is guaranteed by the general theory of [26]. Hence, a natural question that arises is whether such an equilibrium solution is unique or not.

To answer this question, Rosen [27] provided a first-order sufficient condition known as *diagonal strict concavity* (DSC). To state it, define the *individual payoff gradient* of player k as

$$\mathbf{V}_k(\mathbf{X}) \equiv \nabla_{\mathbf{X}_k} u_k(\mathbf{X}_k; \mathbf{X}_{-k}), \tag{15}$$

and let $\mathbf{V}(\mathbf{X}) \equiv \operatorname{diag}(\mathbf{V}_1(\mathbf{X}), \dots, \mathbf{V}_K(\mathbf{X}))$ denote the collective profile of all players' individual gradients. Then, Rosen's monotonicity condition can be stated as:

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

with equality if and only if $\mathbf{X} = \mathbf{X}'$. We then have:

Theorem 1 (Rosen [27])**.** *If a game of the general form* (1) *satisfies* (DSC)*, then it admits a unique Nash equilibrium.*

The above theorem provides a sufficient condition for equilibrium uniqueness, but it does not provide a way for players to compute it – especially in a decentralized setting with no information exchange between players. More recently, (DSC) 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 the theory of variational inequalities [28]. Our approach is similar in scope but relies instead on the following stability notion:

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

$$\operatorname{tr}[(\mathbf{X} - \mathbf{X}^*)\, \mathbf{V}(\mathbf{X})] \leq 0 \ \text{ for all } \mathbf{X} \text{ sufficiently close to } \mathbf{X}^*. \tag{VS}$$

In particular, if (VS) holds for all $\mathbf{X} \in \boldsymbol{\mathcal{X}}$, we say that $\mathbf{X}^*$ is *globally stable*.

Mathematically, (VS) is implied by (DSC);[6] the converse however does not hold, even when (VS) holds globally. Moreover, stability plays a key role in the characterization of Nash equilibria, as shown in the following proposition:

Proposition 1. *If $\mathbf{X}^* \in \boldsymbol{\mathcal{X}}$ is stable, then it is an isolated Nash equilibrium; specifically, if $\mathbf{X}^*$ is globally stable, it is the game's unique Nash equilibrium.*

Proof: Suppose $\mathbf{X}^*$ is stable, pick some $\mathbf{X}_k$ close to $\mathbf{X}_k^*$, and let $\mathbf{X} = (\mathbf{X}_k; \mathbf{X}_{-k}^*)$. Then, by (VS), we get $\operatorname{tr}[(\mathbf{X}_k - \mathbf{X}_k^*)\, \mathbf{V}_k(\mathbf{X}_k; \mathbf{X}_{-k}^*)] < 0$, implying that u_k is decreasing along the ray $\mathbf{X}_k^* + t(\mathbf{X}_k - \mathbf{X}_k^*)$. Since this covers all rays starting at $\mathbf{X}_k^*$, we conclude that $\mathbf{X}^*$ is the game's unique equilibrium in the neighborhood of $\mathbf{X}^*$ where (VS) holds. ∎

In addition to characterizing the structure of the game's Nash set, variational stability also determines the convergence properties of the learning scheme we develop in the next section. More precisely, as we show in Section V, local stability implies that a Nash equilibrium is locally attracting, while global stability implies that it is globally so.

On that account, it is crucial to have a verifiable criterion for Nash equilibrium stability. We accomplish this by appealing to a second-order condition similar to the second derivative test in calculus. To state it, define the Hessian of a game as follows:

Definition 2. The *Hessian* of a game is the (symmetric) matrix $\mathbf{D}(\mathbf{X}) = (\mathbf{D}_{k\ell}(\mathbf{X}))_{k,\ell \in \mathcal{K}}$ with blocks

$$\mathbf{D}_{k\ell}(\mathbf{X}) = \tfrac{1}{2} \nabla_{\mathbf{X}_k} \nabla_{\mathbf{X}_\ell} u_\ell(\mathbf{X}) + \tfrac{1}{2} [\nabla_{\mathbf{X}_\ell} \nabla_{\mathbf{X}_k} u_k(\mathbf{X})]^\dagger. \tag{16}$$

The terminology "Hessian" reflects the fact that when (1) is a single-agent optimization problem ($K = 1$), $\mathbf{D}(\mathbf{X})$ is simply the Hessian of the optimizer's objective. Thus, just as negative-definiteness of the Hessian of a function guarantees (strong) concavity and the existence of a unique maximizer, we have:

[5] Strictly speaking, the block-diagonal constraint in (12) does not appear in (2), but it is satisfied automatically by the learning method presented in Sec. IV.

[6] Simply note that Nash equilibria are solutions of the variational inequality $\operatorname{tr}[(\mathbf{X} - \mathbf{X}^*)\mathbf{V}(\mathbf{X}^*)] \leq 0$ [13]. Then, (VS) follows by setting $\mathbf{X}^* = \mathbf{X}'$ in (DSC).

Proposition 2. *If* $\mathbf{D}(\mathbf{X}) \prec 0$ *for all* $\mathbf{X} \in \boldsymbol{\mathcal{X}}$, $\mathcal{G}$ *admits a unique, globally stable Nash equilibrium. More generally, if* $\mathbf{X}^*$ *is a Nash equilibrium and* $\mathbf{D}(\mathbf{X}^*) \prec 0$, $\mathbf{X}^*$ *is stable and isolated.*

Proof: For the first statement, assume that $\mathbf{D}(\mathbf{X}) \prec 0$ for all $\mathbf{X} \in \boldsymbol{\mathcal{X}}$ and note that $\mathbf{D}(\mathbf{X})$ is simply the symmetrized G-matrix of the game in the sense of Rosen [27, p. 524]. Theorem 6 in [27] states that, if this G-matrix is negative-definite, the game satisfies (DSC), so it admits a unique Nash equilibrium $\mathbf{X}^*$ by Theorem 1. Since (DSC) further implies (VS) for all $\mathbf{X} \in \boldsymbol{\mathcal{X}}$, our claim follows. For our second assertion, if $\mathbf{X}^*$ is a Nash equilibrium and $\mathbf{D}(\mathbf{X}^*) \prec 0$, we have $\mathbf{D}(\mathbf{X}) \prec 0$ for all $\mathbf{X}$ in some convex product neighborhood $\mathcal{U}$ of $\mathbf{X}^*$ (by continuity). Then, by restricting $\mathcal{G}$ to $\mathcal{U}$ and reasoning as in the global case above, we conclude that $\mathbf{X}^*$ is the game's unique equilibrium in $\mathcal{U}$, i.e. $\mathbf{X}^*$ is an isolated Nash equilibrium. ∎

Remark 1. In the special class of potential games, the players' payoff functions are aligned along the game's potential [29], whose maximizers are Nash equilibria. In this context, (DSC) boils down to strict concavity of the potential function, which in turn implies the existence of a unique Nash equilibrium. Likewise, Proposition 2 similarly reduces to the second order condition of concave potential functions that guarantees uniqueness of the solution (strictly negative definite Hessian matrix). In light of this, Nash equilibria of concave potential games are automatically stable in the sense of Definition 1.

IV. Learning under Uncertainty

The goal of this section is to describe a learning process that leads to Nash equilibrium in a decentralized environment with imperfect feedback and no information exchange between players. The main idea of the proposed method is as follows: First, at every stage $n = 1, 2\dots$ of the process, each player $k \in \mathcal{K}$ takes a step along the individual gradient of their utility function (possibly subject to nose). The output is then "mirrored" back to $\mathcal{K}$ via a suitable matrix exponentiation map, and the process repeats.

More precisely, consider the *matrix exponential learning* (MXL) scheme

$$\begin{aligned} \mathbf{X}_k(n) &= A_k \frac{\exp(\mathbf{Y}_k(n-1))}{1 + \|\exp(\mathbf{Y}_k(n-1))\|}, \\ \mathbf{Y}_k(n) &= \mathbf{Y}_k(n-1) + \gamma_n \hat{\mathbf{V}}_k(n), \end{aligned} \tag{MXL}$$

where:

1) $n = 1, 2, \dots$ denotes the stage of the process.
2) the auxiliary matrix variables $\mathbf{Y}_k$ are initialized to some arbitrary (Hermitian) value $\mathbf{Y}_k(0)$ for all $k = 1, \dots, K$.
3) $\hat{\mathbf{V}}_k(n)$ is a stochastic estimate of the individual gradient $\mathbf{V}_k(\mathbf{X}(n))$ of player k at stage n (more on this below).
4) γ_n is a decreasing step-size sequence, typically of the form $\gamma_n \propto 1/n^\beta$.

Setting aside for a moment the precise nature of the stochastic estimates $\hat{\mathbf{V}}_k(n)$, we note that the update of the auxiliary matrix variable $\mathbf{Y}_k(n)$ in (MXL) acts as a "steepest ascent" step along the estimated direction of each player's individual payoff gradient $\mathbf{V}_k(\mathbf{X}(n))$. As such, if there were no constraints for the players' actions, $\mathbf{Y}_k(n)$ would define an admissible sequence of play and, *ceteris paribus*, each player would tend to unilaterally increase their payoff along this sequence. However, this simple ascent scheme does not suffice in our constrained framework, so $\mathbf{Y}_k(n)$ is exponentiated and normalized in order to meet the feasibility constraints (2).[7]

Algorithm 1 Matrix exponential learning (MXL).

Parameter: step-size sequence $\gamma_n \sim 1/n^\beta$, $\beta \in (0, 1]$.
Initialization: $n \leftarrow 0$; $\mathbf{Y}_k \leftarrow$ any $M_k \times M_k$ Hermitian matrix.
Repeat
  $n \leftarrow n + 1$;
  foreach player $k \in \mathcal{K}$ **do**
    play $\mathbf{X}_k \leftarrow A_k \dfrac{\exp(\mathbf{Y}_k)}{1 + \|\exp(\mathbf{Y}_k)\|}$;
    get gradient feedback $\hat{\mathbf{V}}_k$;
    update auxiliary matrix $\mathbf{Y}_k \leftarrow \mathbf{Y}_k + \gamma_n \hat{\mathbf{V}}_k$;
until termination criterion is reached.

Of course, the outcome of the players' gradient tracking process depends crucially on the quality of the available feedback $\hat{\mathbf{V}}_k(n)$. 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 (cf. Ex. II-C).
ii) The players' utility functions are themselves stochastic expectations of the form $u_k(\mathbf{X}) = \mathbb{E}[\hat{u}_k(\mathbf{X}; \omega)]$ for some random variable ω, and the players can only observe the (stochastic) gradient of $\hat{u}_k$ (cf. Ex. II-B).
iii) Any combination of the above.

In view of all this, we will focus on the general model:

$$\hat{\mathbf{V}}_k(n) = \mathbf{V}_k(\mathbf{X}(n)) + \mathbf{Z}_k(n), \tag{17}$$

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

(H1) *Zero-mean:*

$$\mathbb{E}[\mathbf{Z}(n) \,|\, \mathcal{F}_{n-1}] = 0. \tag{H1}$$

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

$$\mathbb{E}[\|\mathbf{Z}(n)\|_\infty^2 \,|\, \mathcal{F}_{n-1}] \leq \sigma_*^2 \quad \text{for some } \sigma_* > 0. \tag{H2}$$

In the above, $\mathcal{F}_n$ denotes the history (natural stochastic basis) of $\mathbf{Z}(n)$ up to stage n. As such, the statistical hypotheses (H1) and (H2) above are fairly mild and allow for a broad range of estimation scenarios (in particular, we will *not* be assuming that the errors are i.i.d. or bounded). In more 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 all Gaussian, log-normal, uniform and sub-exponential distributions.

In other words, Hypotheses (H1) and (H2) simply mean that the players' individual gradient estimates $\hat{\mathbf{V}}_k$ are *unbiased and*

[7] Recall here that $\mathbf{V}_k$ is Hermitian as the derivative of a real function with respect to a Hermitian matrix variable. Moreover, the normalization step in (MXL) subsequently ensures that the resulting matrix has $\|\mathbf{X}_k\| \leq A_k$, so the sequence $\mathbf{X}_k(n)$ induced by (MXL) meets the problem's feasibility constraints.

bounded in mean square, i.e.

$$\mathbb{E}[\hat{\mathbf{V}}_k(n) \,\big|\, \mathcal{F}_{n-1}] = \mathbf{V}_k(\mathbf{X}(n)), \tag{18a}$$

$$\mathbb{E}[\|\hat{\mathbf{V}}_k(n)\|_\infty^2 \,\big|\, \mathcal{F}_{n-1}] \leq V_k^2 \quad \text{for some } V_k > 0. \tag{18b}$$

For generality, all of our results will be stated under (H1) and (H2) – or, equivalently, (18). For a more refined analysis under tighter assumptions for the noise, see Section V-C below.

V. Convergence Analysis and Implementation

By construction, the recursion (MXL) with gradient feedback satisfying Hypotheses (H1) and (H2) enjoys the following desirable properties:

(P1) *Distributedness:* players only require individual gradient information as feedback.
(P2) *Robustness:* the players' feedback could be stochastic, imperfect, or otherwise perturbed by random noise.
(P3) *Statelessness:* players do not need to know the state of the system.
(P4) *Reinforcement:* players tend to increase their individual utilities.

The above shows that (MXL) is a promising candidate for learning in games and distributed optimization problems of the general form (1). Accordingly, our aim in this section is to examine the long-term convergence properties of (MXL).

A. Convergence properties

We begin by showing that the algorithm's termination states are Nash equilibria with probability 1:

Theorem 2. *Assume that Algorithm 1 is run with a decreasing step-size sequence γ_n such that $\sum_{n=1}^\infty \gamma_n^2 < \sum_{n=1}^\infty \gamma_n = \infty$ and gradient observations satisfying (H1) and (H2). If it exists, $\lim_{n\to\infty} \mathbf{X}(n)$ is a Nash equilibrium of the game (a.s.).*

The proof of Theorem 2 is presented in detail in Appendix B and is essentially by contradiction. To provide some intuition, if the limit of (MXL) is not a Nash equilibrium, at least one player of the game must be dissatisfied, thus experiencing a repelling drift – in the limit and on average. Owing to the algorithm's exponentiation step, this "repulsion" can be quantified via the so-called *von Neumann* (or *quantum*) *entropy* [30]; by using the law of large numbers, it is then possible to estimate the impact of the noise and show that this entropy diverges to infinity, thus obtaining the desired contradiction.

In words, Theorem 2 shows that if (MXL) converges, it converges to a Nash equilibrium; however, it does not provide any guarantee that the algorithm converges in the first place. A sufficient condition for convergence based on equilibrium stability is given below:

Theorem 3. *Assume that Algorithm 1 is run with a step-size sequence γ_n such that $\sum_{n=1}^\infty \gamma_n^2 < \sum_{n=1}^\infty \gamma_n = \infty$ and gradient observations satisfying (H1) and (H2). If $\mathbf{X}^*$ is globally stable, then $\mathbf{X}(n)$ converges to $\mathbf{X}^*$ (a.s.).*

The proof of Theorem 3 is given in Appendix C. In short, it comprises the following steps: First, we consider a deterministic, continuous-time variant of (MXL) and we show that globally stable equilibria are global attractors of said system. To this end, we introduce a matrix version of the so-called *Fenchel coupling* [31, 32] and we show that this coupling plays the role of a Lyapunov function in continuous time. Subsequently, we derive the evolution of the discrete-time stochastic system (MXL) by using the method of stochastic approximation [33] and the theory of concentration inequalities [34] to control the gap between continuous and discrete time.

From a practical point of view, an immediate corollary of Theorem 3 is the following second-derivative test:

Corollary 1. *If the game's Hessian matrix $\mathbf{D}(\mathbf{X})$ is negative-definite for all $\mathbf{X} \in \boldsymbol{\mathcal{X}}$, Algorithm 1 converges to the game's (necessarily) unique Nash equilibrium (a.s.).*

The above results show that (MXL) converges to stable equilibria under very mild assumptions on the underlying stochasticity (zero-mean errors and finite conditional variance). Building on this, our next result shows that local stability implies local convergence with arbitrarily high probability:

Theorem 4. *Assume that Algorithm 1 is run with feedback satisfying (H1) and (H2), and a small enough step-size sequence γ_n with $\sum_{n=1}^\infty \gamma_n^2 < \sum_{n=1}^\infty \gamma_n = \infty$. If $\mathbf{X}^*$ is (locally) stable, then it is locally attracting with arbitrarily high probability; specifically, for every $\varepsilon > 0$, there exists a neighborhood U_ε of $\mathbf{X}^*$ such that*

$$\mathbb{P}(\mathbf{X}(n) \to \mathbf{X}^* \,|\, \mathbf{X}(1) \in U_\varepsilon) \geq 1 - \varepsilon. \tag{19}$$

Theorem 4 is proven in Appendix D. The key difference with the proof of Theorem 3 is that, since we only assume the existence of a *locally stable* equilibrium $\mathbf{X}^*$, the drift of (MXL) does not point towards $\mathbf{X}^*$ globally, so $\mathbf{X}(n)$ may exit the basin of attraction of $\mathbf{X}^*$ in the presence of high uncertainty. However, by invoking (H2) and Doob's maximal inequality for martingale difference sequences [35], it can be shown that this happens with arbitrarily small probability, leading to the probabilistic convergence result (19).

Furthermore, as in the case of globally stable equilibria, we also have the following easy-to-check convergence criterion:

Corollary 2. *Let $\mathbf{X}^*$ be a Nash equilibrium of $\mathcal{G}$ such that $\mathbf{D}(\mathbf{X}^*) \prec 0$. Then (MXL) converges locally to $\mathbf{X}^*$ with arbitrarily high probability.*

B. Rate of convergence

Theorems 2–4 give a fairly complete picture of the *qualitative* convergence properties of (MXL): the algorithm's only possible end-states are Nash equilibria, and the existence of a globally (resp. locally) stable Nash equilibrium implies its global (resp. local) convergence. On the other hand, these results do not address the *quantitative* aspects of the algorithm's long-term behavior (for instance, its convergence speed); in what follows, we study precisely this question.

We begin by introducing the so-called *quantum Kullback–Leibler divergence* (or *von Neumann relative entropy*) [30] defined here as:

$$D_{\mathrm{KL}}(\mathbf{X}^*, \mathbf{X}) = \operatorname{tr}[\mathbf{X}^*(\log \mathbf{X}^* - \log \mathbf{X})]. \tag{20}$$

By Klein's inequality [30], $D_{\mathrm{KL}}(\mathbf{X}^*, \mathbf{X}) \geq 0$ with equality if and only if $\mathbf{X} = \mathbf{X}^*$, so $D_{\mathrm{KL}}(\mathbf{X}^*, \mathbf{X})$ represents a (convex) measure of the distance between $\mathbf{X}$ and $\mathbf{X}^*$. With this in mind, we introduce below the following quantitative measure of stability:

Definition 3. Given $B > 0$, $\mathbf{X}^*$ is called *B-strongly stable* if

$$\mathrm{tr}[(\mathbf{X} - \mathbf{X}^*)\,\mathbf{V}(\mathbf{X})] \leq -B\,D_{\mathrm{KL}}(\mathbf{X}^*, \mathbf{X}) \quad \text{for all } \mathbf{X} \in \boldsymbol{\mathcal{X}}. \tag{21}$$

Under this refinement of equilibrium stability,[8] we obtain the following quantitative result:

Theorem 5. *Assume that Algorithm 1 is run with the step-size sequence $\gamma_n = \gamma/n$ and gradient observations satisfying (H1) and (H2). If $\mathbf{X}^*$ is B-strongly stable and $1 < B\gamma \leq 2$, we have*

$$\mathbb{E}[D_{\mathrm{KL}}(\mathbf{X}^*, \mathbf{X}(n))] \tag{22}$$

for all $n \geq 2$. In particular:

$$\mathbb{E}[\|\mathbf{X}(n) - \mathbf{X}^*\|] = \mathcal{O}(n^{-1/2}). \tag{23}$$

The explicit convergence rate provided by Theorem 5 (proven in Appendix E) can be finetuned further by choosing γ so as to minimize the constant $\gamma^2 \max\{2, 1/(B\gamma - 1)\}$ that appears in (22). Carrying this out, the step-size sequence $\gamma_n = 3/(2Bn)$ leads to the optimized convergence rate

$$\mathbb{E}[D_{\mathrm{KL}}(\mathbf{X}^*, \mathbf{X}(n))] \leq \frac{9V^2}{2B^2 n}. \tag{24}$$

It is also possible to estimate the algorithm's convergence rate for equilibrium states that are only locally strongly stable. However, given that the algorithm's convergence is probabilistic in this case, the resulting bounds are also probabilistic and, hence, more complicated to present.

Remark 2. Algorithm 1 could be run with $\gamma > 2/B$ and a bound similar to (22) would still hold for $n \geq \lceil B\gamma \rceil$. However, since the optimized step-size $\gamma_n = 3/(2Bn)$ that achieves (24) is already covered by the condition $B\gamma \leq 2$, we do not present this more general case here.

C. Practical implementation

We close this section with a discussion of certain issues pertaining to the practical implementation of Algorithm 1:

a) On the step-size sequence γ_n: Using a decreasing step-size γ_n in (MXL) may appear counter-intuitive because it implies that new information enters the algorithm with decreasing weights. As evidenced by Theorem 5, the reason for this is that a constant step-size might cause the process to overshoot and lead to an oscillatory behavior around the algorithm's end-state. In the noiseless, deterministic regime, these oscillations can be dampened by using forward-backward splitting or accelerated descent methods [4]. However, in the presence of noise, the use of a decreasing step-size is essential in order to dissipate measurement noise and other stochastic effects, explaining why it is not possible to improve on (22) by using a constant step.

That being said, the "$\ell^2 - \ell^1$" summability condition $\sum_n \gamma_n^2 < \sum_n \gamma_n = \infty$ required by Theorems 2–4 is closely linked to the finite mean squared error hypothesis (H2). In practice however, the feedback noise process $\mathbf{Z}(n)$ often satisfies tighter moment bounds (e.g. Gaussian/exponential processes have finite moments of all orders), so this summability condition can be relaxed further in order to allow for more aggressive policies of the form $\gamma_n \propto 1/n^\beta$ for some $\beta \leq 1/2$.

To that end, consider the q-th moment condition

$$\mathbb{E}[\|\mathbf{Z}(n)\|_\infty^q \,|\, \mathcal{F}_{n-1}] \leq \sigma_*^q \quad \text{for some } \sigma_* > 0. \tag{H2$'$}$$

Under this refinement of (H2), our analysis in Appendices B–D shows that Theorems 2–4 continue to hold under the milder step-size summability condition $\sum_n \gamma_n^{1+q/2} < \sum_n \gamma_n = \infty$. Specifically, as long as $\mathbf{Z}(n)$ has finite moments up to some order $q > 2(1-\beta)/\beta$, it is possible to use a step-size policy of the form $\gamma_n \propto 1/n^\beta$ for *any* $\beta \in (0, 1]$. We use this observation in Section VI where we carry out an extensive numerical analysis of Algorithm 1 with step-sizes of the form $\gamma_n \propto 1/\sqrt{n}$.

Finally, we also note here that it is possible for each player to choose their own step-size policy, provided that the summability requirements discussed above hold for each player separately (see Proposition C.2 in Appendix C). This property of MXL becomes particularly important in conjunction with the possibility of asynchronous implementation described below.

b) Distributed implementation: An implicit assumption in (MXL) is that user updates – albeit local – are concurrent. This can be achieved via a global update timer that synchronizes the players' updates; however, in a fully decentralized setting, even this degree of coordination may be challenging to attain. In light of this, we discuss below an asynchronous variant of Algorithm 1 where each player's updates follow an individual, independent schedule.

Specifically, assume that each player $k \in \mathcal{K}$ has an individual timer that triggers an `UpdateEvent`, i.e. a request for gradient feedback and, subsequently, an update of $\mathbf{X}_k$. Of course, the players' gradient estimates $\hat{\mathbf{V}}_k$ will then be subject to delays and asynchronicities (in addition to noise), so the update structure of Algorithm 1 must be modified appropriately. To that end, let $\mathcal{K}_n \subseteq \mathcal{K}$ be the set of players that update their actions at the n-th overall `UpdateEvent` (typically $|\mathcal{K}_n| = 1$ if players update at random times), and let $d_k(n)$ be the corresponding number of epochs that have elapsed since the last update of player k. We then obtain the following asynchronous variant of (MXL):

$$\begin{aligned} \mathbf{X}_k(n) &= A_k \frac{\exp(\mathbf{Y}_k(n-1))}{1 + \|\exp(\mathbf{Y}_k(n-1))\|}, \\ \mathbf{Y}_k(n) &= \mathbf{Y}_k(n-1) + \gamma_{n_k}\, \mathbb{1}\{k \in \mathcal{K}_n\} \cdot \hat{\mathbf{V}}_k(n), \end{aligned} \tag{25}$$

where n_k denotes the number of updates performed by player k up to epoch n, and the (asynchronous) estimate $\hat{\mathbf{V}}_k(n)$ satisfies

$$\mathbb{E}[\hat{\mathbf{V}}_k(n) \,|\, \mathcal{F}_{n-1}] = \mathbf{V}_k(\mathbf{X}_1(n - d_1(n)), \dots, \mathbf{X}_K(n - d_K(n))). \tag{26}$$

To extend our convergence analysis to this setting, a possible way would be to exploit the asynchronous stochastic approximation analysis of [36, 37]. However, this analysis would take us too far afield, so we relegate it to future work.

c) Computational complexity and numerical stability: From the point of view of computational complexity, the bottleneck of each iteration of Algorithm 1 is the matrix exponentiation step $\mathbf{Y}_k \mapsto \exp(\mathbf{Y}_k)$. Since matrix exponentiation has the

[8] Since $D_{\mathrm{KL}}(\mathbf{X}^*, \mathbf{X}) \geq 0$ with equality if and only if $\mathbf{X} = \mathbf{X}^*$, it follows that strongly stable states are necessarily stable.

TABLE I
WIRELESS NETWORK SIMULATION PARAMETERS

Parameter	Value
Cell size (rectangular)	1 km
User density	500 users/km^2
Time frame duration	5 ms
Wireless propagation model	COST Hata
Central frequency	2.5 GHz
Total bandwidth	11.2 MHz
OFDM subcarriers	1024
Subcarrier spacing	11 kHz
Spectral noise density (20 °C)	−174 dBm/Hz
Maximum transmit power	$P_{\max} = 33$ dBm
Non-radiative power consumption	$P_c = 20$ dBm
Tx/Rx antennas per link	4/8

same complexity as matrix multiplication, this step has polynomial complexity with a low degree on the input dimension of $\mathbf{X}_k$. In particular, each exponentiation requires $\mathcal{O}(M_k^\omega)$ floating point operations, where the complexity exponent can be as low as $\omega = 2.373$ if the players employ fast Coppersmith–Winograd matrix multiplication methods [38].

Finally, regarding the numerical stability of Algorithm 1, the only possible source of arithmetic errors is the exponentiation/normalization step. Indeed, if the eigenvalues of $\mathbf{Y}_k$ are large, this step could incur an overflow where both the numerator and the denominator evaluate to machine infinity. This potential instability can be fixed as follows: If y_k denotes the largest eigenvalue of $\mathbf{Y}_k$ and we let $\mathbf{Y}'_k = \mathbf{Y}_k - y_k\mathbf{I}$, the algorithm's exponentiation step can be rewritten as:

$$\mathbf{X}_k \leftarrow \frac{\exp(\mathbf{Y}'_k)}{\exp(-y_k) + \|\exp(\mathbf{Y}'_k)\|}. \tag{27}$$

Thanks to this shift, the elements of the numerator are now bounded from above by 1 (because the largest diagonal element of $\mathbf{Y}'_k$ is $e^{y_k - y_k} = 1$), so there is no danger of encountering a numerical indeterminacy of the type `Inf/Inf`. Thus, to avoid computer arithmetic issues, we employ the stable expression (27) in all numerical implementations of Algorithm 1.

d) Updates and feasibility: As we discussed earlier, the exponentiation/normalization step of Algorithm 1 ensures that the players' action variables $\mathbf{X}_k(n)$ satisfy the game's feasibility constraints (2) at each stage n. In the noiseless case ($\mathbf{Z} = 0$), feasibility is automatic because, by construction, $\mathbf{V}_k$ (and, hence, $\mathbf{Y}_k$) is Hermitian. That said, in the presence of noise, there is no reason to assume that the estimates $\hat{\mathbf{V}}_k$ are Hermitian, so the updates $\mathbf{X}_k$ may also fail to be feasible. To rectify this, we tacitly assume that each player corrects such errors by replacing $\hat{\mathbf{V}}_k$ with $(\hat{\mathbf{V}}_k + \hat{\mathbf{V}}_k^\dagger)/2$ in (MXL). Since this error-correcting operation is linear in its input, Hypotheses (H1) and (H2) continue to apply, and our analysis holds verbatim.

VI. Numerical Results

In this section, we assess the performance of MXL via numerical simulations. For concreteness, we focus on the case of energy efficiency maximization in practical multi-user MIMO networks (cf. Section II-C), but our conclusions apply to a wide range of parameters and scenarios. To the best of our knowledge, this comprises the first distributed solution scheme for general multi-user/multi-antenna/multi-carrier networks under imperfect feedback/CSI and mobility considerations.

Our basic network setup consists of a macro-cellular OFDMA wireless network with access points deployed on a rectangular grid with cell size 1 km (for a quick overview of simulation parameters, see Table I). Signal transmission and reception occurs over a 10 MHz band divided into 1024 subcarriers around a central frequency of $f_c = 2.5$ GHz. We further assume a frame-based time-division duplexing (TDD) scheme with frame duration $T_f = 5$ ms: transmission takes place during the uplink phase while the network's access points process the received signal and provide feedback during the downlink phase. Finally, signal propagation is modeled after the widely used COST Hata model with spectral noise density equal to −174 dBm/Hz at 20 °C.

The network is populated by wireless transmitters (users) following a homogeneous Poisson point process with intensity $\rho = 500$ users/km^2. Each wireless transmitter is further assumed to have 4 transmit antennas, a maximum transmit power of $P_{\max} = 40$ dBm and circuit (non-radiative) power consumption of $P_c = 20$ dBm. In each cell, orthogonal frequency division multiplexing (OFDM) subcarriers are allocated to wireless users randomly so that different users are assigned to disjoint carrier sets. We then focus on a set of $K = 25$ users, each located at a different cell and sharing $S = 8$ common subcarriers. Finally, at the receiver end, we consider 8 receive antennas per connection and a receiver noise figure of 7 dB.

To assess the performance and robustness of the MXL algorithm, we first focus on a scenario with stationary users and static channel conditions. Specifically, in Fig. 2, each user runs Algorithm 1 with a variable step-size $\gamma_n \sim n^{-1/2}$ and initial transmit power $P_0 = P_{\max}/2 = 26$ dBm (allocated uniformly across different antennas and subcarriers), and we plot the users' transmit energy efficiency over time. For benchmarking purposes, we first assume that users have perfect CSI measurements at their disposal. In this deterministic regime, the algorithm converges to a stable equilibrium 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× and 6× over uniform power allocation schemes.

Subsequently, this simulation cycle was repeated in the presence of observation noise and errors. The intensity of the feedback 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' transmit energy efficiency over time for noise levels $z = 25\%$, 50%, and 100% (corresponding to moderate, high, and extremely high uncertainty respectively). Fig. 2 shows that the network's rate of convergence to a Nash equilibrium is negatively impacted by the magnitude of the noise; remarkably however, MXL retains its convergence properties and the network's users achieve a 100% per capita gain in energy efficiency within a few tens of iterations, even under extremely high uncertainty (of the order of $z = 100\%$).

To examine the algorithm's performance in a fully dynamic network environment, Fig. 3 focuses on mobile users with

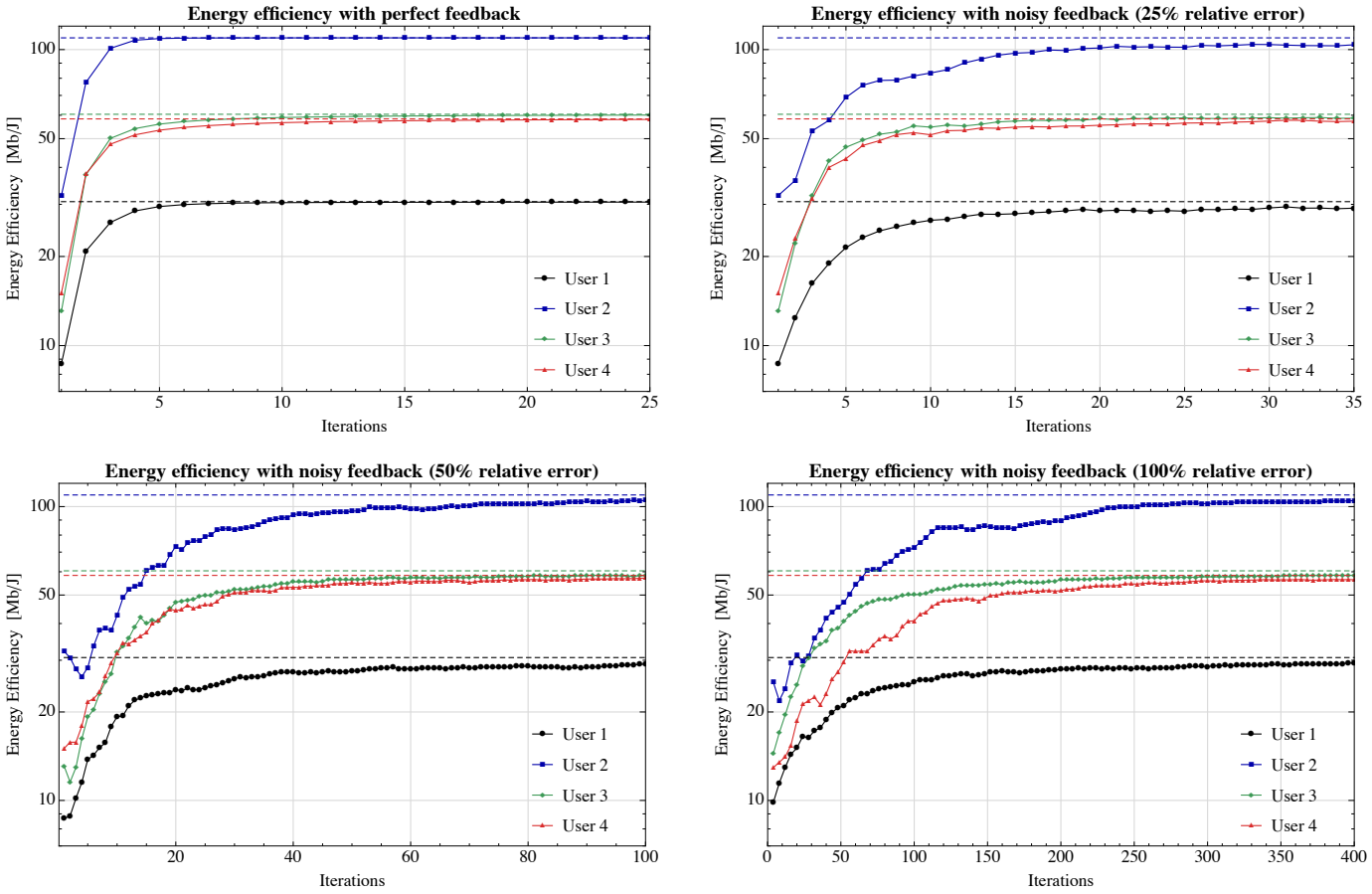


Fig. 2. Performance of MXL in the presence of noise. In all figures, we plot the transmit energy efficiency of wireless users that employ Algorithm 1 in a wireless network with parameters as described in the main text (to reduce graphical clutter, we only plotted 4 users with diverse channel characteristics). In the absence of noise (upper 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 but the algorithm remains convergent under very high degrees of uncertainty (up to relative error levels of 100%; bottom right).

channels that vary with time due to (Rayleigh) fading, path loss fluctuations, etc. To simulate this scenario, we used the standard extended typical urban (ETU) model for the users' environment and the extended pedestrian A (EPA) and extended vehicular A (EVA) models to emulate pedestrian (3–5 km/h) and vehicular movement (30–130 km/h) respectively [39]. In Fig. 3(a), we plotted the channel gains ($\mathrm{tr}[\mathbf{H}\mathbf{H}^\dagger]$) of 4 users with diverse mobility and distance characteristics (two pedestrian and two vehicular users, one closer and one farther away from their intended receiver). As can be seen, the users' channels exhibit significant fluctuations (in the range of a few dB) over different time scales, so the Nash equilibrium set of the energy efficiency game described in Section II-C will evolve itself over time. Nevertheless, despite the channels' variability, Fig. 3(b) shows that MXL adapts to this highly volatile network environment very quickly, allowing users to track the game's instantaneous equilibrium with remarkable accuracy. For comparison, we also plotted the users' achieved energy efficiency under a uniform power allocation policy, which is known to be optimal under isotropic fading conditions [40]. Because urban environments are not homogeneous and/or isotropic (even on average), uniform power allocation fails to adapt and performs consistently worse than MXL (achieving an energy efficiency ratio between 2× and 6× lower than that of MXL).

Finally, in Fig. 4 we examine the gap between Nash equilibrium solutions and socially optimum states that maximize the system's overall energy efficiency. To simplify the analysis, we focus on the case where all users transmit to a common receiver who employs a centralized successive interference cancellation scheme to decode the users' messages.[9] In this setting, the system's Shannon sum-rate is given by [41]:

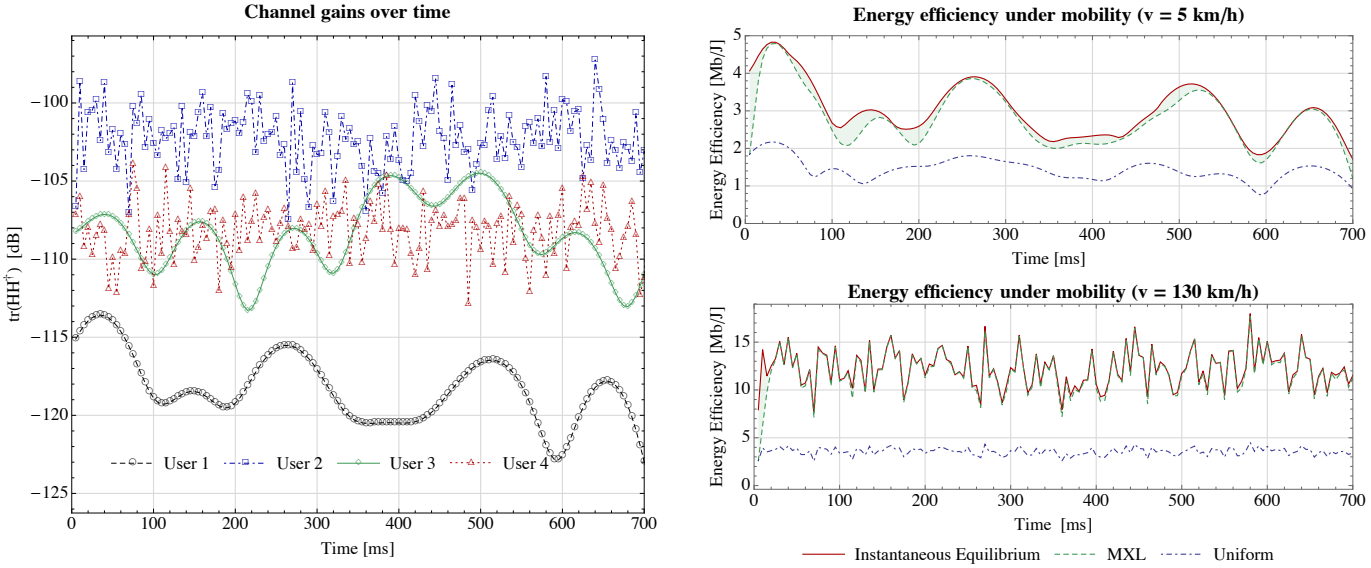
$$R_{\mathrm{sys}}(\mathbf{Q}_1,\ldots,\mathbf{Q}_K) = \log\det\left(\mathbf{I} + \textstyle\sum_k \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^\dagger\right), \qquad (28)$$

leading to the system-wide energy efficiency expression

$$\mathrm{EE}_{\mathrm{sys}}(\mathbf{Q}_1,\ldots,\mathbf{Q}_K) = \frac{R_{\mathrm{sys}}(\mathbf{Q}_1,\ldots,\mathbf{Q}_K)}{\sum_k (P_c + \mathrm{tr}(\mathbf{Q}_k))}. \qquad (29)$$

To maximize (29), it is possible to employ a centralized MXL scheme with the same update structure as Algorithm 1, but geared instead to the objective (29). In brief, the basic elements are as follows: First, the variable change of Section II can be used to define a new set of variables $\mathbf{Z} = \mathbf{Z}(\mathbf{Q})$ that map the problem (29) to an equivalent concave problem with objective function $u_{\mathrm{sys}}(\mathbf{Z})$ and feasible region $\mathcal{Z}$, as in (13). The centralized matrix exponential learning (CMXL) algorithm

[9] This (Gaussian) MAC model is a special case of the interference channel described in Section II. In general interference channels, the sum-rate R_{sys} is non-concave, so the determination of socially optimum states becomes an NP-hard problem whose solution is beyond the scope of the current paper.



(a) Channel gain evolution for different user velocities

(b) Equilibrium tracking under mobility

Fig. 3. Performance of MXL in a dynamic network environment with mobile users moving at $v \in \{3, 5, 30, 130\}$ km/h. The users' achieved energy efficiency tracks the system's (evolving) equilibrium remarkably well, even under rapidly changing channel conditions.

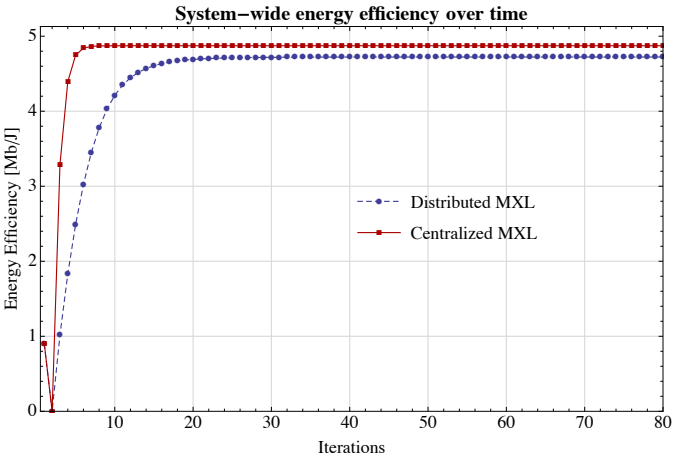


Fig. 4. Unilateral versus system-wide energy efficiency maximization.

is then defined by replacing the gradient step in (MXL) with $\nabla_{\mathbf{Z}} u_{\text{sys}}(\mathbf{Z})$, and then defining a modified exponential mapping as in Proposition A.1.

In Fig. 4, we assess the performance of Algorithm 1 with respect to the system-wide objective (29) by plotting the system's energy efficiency for $K = 5$ users that employ the MXL and CMXL methods described above. Even when users follow the non-cooperative MXL algorithm, they reduce power significantly so the system's energy efficiency ends up being within 5% of its maximum value – a surprisingly small gap between socially and unilaterally efficient states.

VII. Conclusions and Perspectives

In this paper, we examined a distributed matrix exponential learning algorithm for stochastic semidefinite optimization problems and games that arise in key areas of signal processing and wireless communications (ranging from image-based similarity search to MIMO systems). The main idea of the proposed method is to track the players' individual payoff gradients in a dual, unconstrained space, and then map this process back to the players' action spaces via an "exponential mirror" step. Thanks to the aggregation of the players' payoff gradients, the algorithm is capable of operating under uncertainty and feedback noise, two impediments that can have a detrimental effect on more aggressive best-response methods.

To analyze the proposed algorithm, we introduced the notion of a stable Nash equilibrium, and we showed that the algorithm is globally convergent to such equilibria – or locally convergent when an equilibrium is only locally stable. Our convergence analysis also revealed that, on average, the algorithm converges to an ε-neighborhood of a Nash equilibrium (in terms of the Kullback–Leibler distance) within $\mathcal{O}(1/\varepsilon^2)$ iterations. To validate our theoretical analysis, we also tested the algorithm's performance in realistic multi-carrier/multiple-antenna wireless scenarios where several users seek to maximize their energy efficiency: in this setting, users quickly reach a Nash equilibrium and attain gains between 100% and 500% in energy efficiency, even under very high uncertainty.

The above results are particularly promising and suggest that our analysis applies to an even wider setting than the game-theoretic framework (1) – for instance, games with convex action sets that are not necessarily of the form (2). Another natural question that arises is whether it is possible to run the proposed MXL without *any* gradient information. We intend to explore these directions at depth in future work.

Appendix A
The Exponentiation Step

In this appendix, our goal will be to establish certain properties of the exponential map of Algorithm 1 that are crucial in the stationarity and convergence analysis of the next appendices. For simplicity, we only treat the case $A_k = 1$; the general case follows by a trivial rescaling so we do not present it.

With a fair degree of hindsight, we begin by introducing the modified von Neumann entropy [30]:

$$h(\mathbf{X}) = \operatorname{tr}[\mathbf{X}\log\mathbf{X}] + (1-\operatorname{tr}(\mathbf{X}))\log(1-\operatorname{tr}(\mathbf{X})). \tag{A.1}$$

The convex conjugate of h over the spectrahedron $\mathcal{D} = \{\mathbf{X} \in \mathbb{H}^M_+ : \operatorname{tr}(\mathbf{X}) \leq 1\}$ is then defined as:

$$h^*(\mathbf{Y}) = \max\{\operatorname{tr}[\mathbf{Y}\mathbf{X}] - h(\mathbf{X}) : \mathbf{X} \in \mathcal{D}\}, \tag{A.2}$$

with $\mathbf{Y} \in \mathbb{H}^M$. As it turns out, the exponentiation step of the XL algorithm is simply the (matrix) derivative of h^*:

Proposition A.1. *With notation as above, we have:*

$$h^*(\mathbf{Y}) = \log(1+\operatorname{tr}(\exp(\mathbf{Y}))), \tag{A.3}$$

and

$$\nabla h^*(\mathbf{Y}) = \mathbf{G}(\mathbf{Y}) \equiv \frac{\exp(\mathbf{Y})}{1+\operatorname{tr}[\exp(\mathbf{Y})]}. \tag{A.4}$$

Proof: Since the von Neumann entropy is strictly convex [30] and becomes infinitely steep at the boundary of $\boldsymbol{\mathcal{X}}$, it follows that the maximization problem (A.2) admits a unique solution $\mathbf{X} \in \mathcal{D}^\circ$. Hence, by the first-order Karush–Kuhn–Tucker (KKT) conditions for the problem (A.2), we get:

$$\mathbf{Y} - \log\mathbf{X} + \log(1-\operatorname{tr}(\mathbf{X}))\mathbf{I} = 0, \tag{A.5}$$

where we used the fact that $\nabla h(\mathbf{X}) = \mathbf{I} + \log\mathbf{X} - \log(1-\operatorname{tr}(\mathbf{X}))\mathbf{I} - \mathbf{I} = \log\mathbf{X} - \log(1-\operatorname{tr}(\mathbf{X}))\mathbf{I}$. Exponentiating (A.5) then yields

$$\mathbf{X} = (1-\operatorname{tr}(\mathbf{X}))\exp(\mathbf{Y}), \tag{A.6}$$

and, after taking traces on both sides, we obtain $\operatorname{tr}(\mathbf{X}) = \operatorname{tr}(\exp(\mathbf{Y}))/[1+\operatorname{tr}(\exp(\mathbf{Y}))]$. Combining the above then yields

$$\mathbf{X} = \frac{\exp(\mathbf{Y})}{1+\operatorname{tr}[\exp(\mathbf{Y})]}, \tag{A.7}$$

and our claim follows by substituting in (A.2). ∎

In addition to the above, the von Neumann entropy also provides a "congruence" measure between the primal variables $\mathbf{X}$ and the auxiliary "dual" variables $\mathbf{Y}$. Specifically, following [31], we introduce here the *Fenchel coupling*:

$$F(\mathbf{X},\mathbf{Y}) = h(\mathbf{X}) + h^*(\mathbf{Y}) - \operatorname{tr}[\mathbf{Y}\mathbf{X}] = D_{\mathrm{KL}}(\mathbf{X}, \mathbf{G}(\mathbf{Y})). \tag{A.8}$$

By Fenchel's inequality, we have $F(\mathbf{X},\mathbf{Y}) \geq 0$ with equality if and only if $\mathbf{Y} = \nabla h(\mathbf{X})$ – or, equivalently, iff $\mathbf{X} = \mathbf{G}(\mathbf{Y})$. More importantly for our purposes, we also have the following approximation lemma:

Proposition A.2. *For all $\mathbf{X} \in \mathcal{D}$ and for all $\mathbf{Y}, \mathbf{Z} \in \mathbb{H}^M$, we have*

$$F(\mathbf{X}, \mathbf{Y}+\mathbf{Z}) \leq F(\mathbf{X},\mathbf{Y}) + \operatorname{tr}[\mathbf{Z}(\mathbf{G}(\mathbf{Y}) - \mathbf{X})] + \|\mathbf{Z}\|^2_\infty. \tag{A.9}$$

Proof: By the definition of the Fenchel coupling, we get:

$$\begin{aligned} F(\mathbf{X}, \mathbf{Y}+\mathbf{Z}) &= h(\mathbf{X}) + h^*(\mathbf{Y}+\mathbf{Z}) - \operatorname{tr}[(\mathbf{Y}+\mathbf{Z})\mathbf{X}] \\ &\leq h(\mathbf{X}) + h^*(\mathbf{Y}) + \operatorname{tr}[\mathbf{Z}\,\mathbf{G}(\mathbf{Y})] + \|\mathbf{Z}\|^2_\infty \\ &\quad - \operatorname{tr}[\mathbf{Y}\mathbf{X}] - \operatorname{tr}[\mathbf{Z}\mathbf{X}] \\ &= F(\mathbf{X},\mathbf{Y}) + \operatorname{tr}[\mathbf{Z}(\mathbf{G}(\mathbf{Y}) - \mathbf{X})] + \|\mathbf{Z}\|^2_\infty, \end{aligned} \tag{A.10}$$

where the expansion of h^* in the second line follows from the fact that the von Neumann entropy is $1/2$-strongly convex (from the duality of strong convexity and strong smoothness, and the fact that h^* is 2-strongly smooth) [3]. ∎

Appendix B
Stationarity Analysis

We begin with the proof of Theorem 2 regarding the possible termination states of Algorithm 1:

Proof of Theorem 2: Let $\mathbf{V}^* = \mathbf{V}(\mathbf{X}^*)$ and assume that $\mathbf{X}^*$ is not a Nash equilibrium. By Eq. (14), this implies that $\operatorname{tr}[(\mathbf{X}'_k - \mathbf{X}^*_k)\mathbf{V}^*_k] > 0$ for some player $k \in \mathcal{K}$ and some $\mathbf{X}' \in \boldsymbol{\mathcal{X}}_k$. Therefore, by continuity, there exists some $a > 0$ such that

$$\operatorname{tr}[(\mathbf{X}'_k - \mathbf{X}_k)\,\mathbf{V}''_k] \geq a > 0, \tag{B.1}$$

for all $\mathbf{X}$ in a small enough neighborhood U of $\mathbf{X}^*$ in $\boldsymbol{\mathcal{X}}$ and for all $\mathbf{V}''_k$ sufficiently close to $\mathbf{V}^*_k$.

Since $\mathbf{X}(n) \to \mathbf{X}^*$ as $n \to \infty$, we may assume that $\mathbf{X}(n) \in U$ for all n.[10] The recursion (MXL) then yields:

$$\mathbf{Y}(n) = \mathbf{Y}(0) + \tau_n \bar{\mathbf{V}}(n), \tag{B.2}$$

where we have set $\tau_n = \sum_{j=1}^n \gamma_j$ and

$$\bar{\mathbf{V}}(n) = \frac{1}{\tau_n}\sum_{j=1}^n \gamma_j \hat{\mathbf{V}}(j) = \frac{1}{\tau_n}\sum_{j=1}^n \gamma_j \mathbf{V}(\mathbf{X}(j)) + \frac{1}{\tau_n}\sum_{j=1}^n \gamma_j \mathbf{Z}(j) \tag{B.3}$$

denotes the γ-weighted time average of the received gradient estimates $\hat{\mathbf{V}}(n)$. By the strong law of large numbers for martingale differences [35, Theorem 2.18], we get $\lim_{n\to\infty} n^{-1}\sum_{j=1}^n \mathbf{Z}(j) = 0$ (a.s.), so the last term of (B.3) also converges to zero (a.s.) by Hardy's summability criterion [42, Theorem 14] applied to the weight sequence $w_{j,n} = \gamma_j/\tau_n$.[11] Thus, given that $\mathbf{X}(n) \in U$ for all n, we conclude that $\bar{\mathbf{V}}(n) \to \mathbf{V}^*$ as $n \to \infty$.

Now, with notation as in Appendix A, let $h_k(\mathbf{X}_k) = \operatorname{tr}[\mathbf{X}_k\log\mathbf{X}_k] + (1-\operatorname{tr}(\mathbf{X}_k))\log(1-\operatorname{tr}(\mathbf{X}_k))$. Since $\mathbf{X}_k(n+1) = \nabla h^*_k(\mathbf{Y}_k(n))$ by Prop. A.1, we will also have $\nabla h_k(\mathbf{X}_k(n+1)) = \mathbf{Y}_k(0) + \tau_n\bar{\mathbf{V}}_k(n)$ by the general theory of convex conjugation. In turn, this implies that

$$h_k(\mathbf{X}'_k) - h_k(\mathbf{X}_k(n+1)) \geq \operatorname{tr}[(\mathbf{Y}_k(0) + \tau_n\bar{\mathbf{V}}_k(n))\,(\mathbf{X}'_k - \mathbf{X}_k(n+1))], \tag{B.4}$$

by the convexity of h_k. However, since $\lim_{n\to\infty}\bar{\mathbf{V}}_k(n) = \mathbf{V}^*_k$ and $\lim_{n\to\infty}\tau_n = \infty$, Eq. (B.1) yields $h_k(\mathbf{X}'_k) - h_k(\mathbf{X}_k(n+1)) \gtrsim a\tau_n$, so $h_k(\mathbf{X}'_k) - h_k(\mathbf{X}_k(n+1)) \to \infty$ as $n \to \infty$, a contradiction. ∎

Appendix C
Global Convergence

We begin our analysis with an auxiliary result for (MXL) in continuous time. Specifically, consider the dynamics

$$\begin{aligned} \dot{\mathbf{Y}} &= \mathbf{V}(\mathbf{X}), \\ \mathbf{X} &= \mathbf{G}(\mathbf{Y}), \end{aligned} \tag{MXL$_c$}$$

obtained by taking the continuous-time limit of (MXL). Our first auxiliary result is that globally stable states are globally attracting under (MXL$_c$):

[10] Note here that the more general sumability requirement $\sum_{n=1}^\infty \gamma_n^{1+q/2} < \sum_{n=1}^\infty \gamma_n = \infty$ is not affected if we start the sequence at some finite $n_0 > 0$.

[11] If every player is using their individual step-size sequence $\gamma_{k,n}$, the sums in (B.3) can be decomposed into player-by-player components, and the law of large numbers and Hardy's criterion can be applied to each player separately to yield the required result.

Proposition C.1. *Let $\mathbf{X}^*$ be a globally stable Nash equilibrium and let $\mathbf{X}(t)$ be a solution of* (MXL$_c$)*. Then,* $\lim_{t\to\infty}\mathbf{X}(t)=\mathbf{X}^*$.

Proof: Let $H(t)=F(\mathbf{X}^*,\mathbf{Y}(t))$. Then

$$\dot H=\operatorname{tr}[\dot{\mathbf{Y}}\,\nabla h^*(\mathbf{Y})]-\operatorname{tr}[\dot{\mathbf{Y}}\,\mathbf{X}^*]=\operatorname{tr}[(\mathbf{X}-\mathbf{X}^*)\,\mathbf{V}(\mathbf{X})],\tag{C.1}$$

i.e. $\dot H\le 0$ with equality if and only if $\mathbf{X}=\mathbf{X}^*$ (recall that $\mathbf{X}^*$ is assumed globally stable). This implies that $H(t)$ is nonincreasing, and hence converges to some $c\ge 0$ as $t\to\infty$. Hence, by compactness, there exists some $\hat{\mathbf{X}}\in\mathcal{X}$ and a sequence $t_n\uparrow\infty$ such that $\mathbf{X}(t_n)\to\hat{\mathbf{X}}$ as $n\to\infty$.

Assume now that $\hat{\mathbf{X}}\neq\mathbf{X}^*$, so there exists some $a>0$ and a neighborhood U of $\hat{\mathbf{X}}$ such that $\operatorname{tr}[(\mathbf{X}-\mathbf{X}^*)\,\mathbf{V}(\mathbf{X}^*)]\le -a$ for all $\mathbf{X}\in U$. Since $\|\dot{\mathbf{X}}\|$ is bounded from above (recall that $\mathbf{G}$ is Lipschitz), there exists some $\delta>0$ such that $\mathbf{X}(t)\in U$ for all $t\in[t_n,t_n+\delta]$ and all $n\ge 0$. In that case however, (C.1) yields:

$$\begin{aligned}\lim_{t\to\infty}H(t)&\le H(0)+\sum_{n=1}^{\infty}\int_{t_n}^{t_n+\delta}\operatorname{tr}[(\mathbf{X}(t)-\mathbf{X}^*)\,\mathbf{V}(\mathbf{X}(t))]\,dt\\&\le H(0)-\sum_{n=1}^{\infty}a\delta=-\infty,\end{aligned}\tag{C.2}$$

a contradiction. This shows that $\mathbf{X}^*$ is the only potential ω-limit point of $\mathbf{X}(t)$; since $\mathbf{X}(t)$ admits at least one ω-limit, it follows that $\mathbf{X}(t)\to\mathbf{X}^*$, as claimed. ∎

With this auxiliary result at hand, we are finally in a position to prove our global convergence result:

Proof of Theorem 3: We first note that the recursion (MXL) can be written in the more succinct form

$$\mathbf{Y}(n)=\mathbf{Y}(n-1)+\gamma_n\left[\mathbf{V}(\mathbf{G}(\mathbf{Y}(n-1)))+\mathbf{Z}(n)\right].\tag{C.3}$$

Since $\mathbf{V}(\mathbf{X})$ is differentiable for almost all $\mathbf{X}\in\mathcal{X}$ by Alexandrov's theorem, Propositions 4.1 and 4.2 in [33] show that $\mathbf{X}(n)$ is an *asymptotic pseudotrajectory* (APT) of (MXL$_c$), i.e. the sequence of play generated by (MXL) are asymptotically close to solution segments of (MXL$_c$) of arbitrary length – for a precise statement, see [33, Sec. 3].[12]

Assume now that $\mathbf{X}(n)$ remains at a minimal positive distance from $\mathbf{X}^*$. Since $\mathbf{X}^*$ is globally stable, we will have $\operatorname{tr}[(\mathbf{X}(n)-\mathbf{X}^*)\,\mathbf{V}(\mathbf{X}(n))]\le -a$ for some $a>0$ and for all n. Furthermore, if we let $D_n=F(\mathbf{X},\mathbf{Y}(n-1))$, Proposition A.2 yields:

$$\begin{aligned}D_{n+1}&=F(\mathbf{X}^*,\mathbf{Y}(n-1)+\gamma_n\hat{\mathbf{V}}(n))\\&\le D_n+\gamma_n v_n+\gamma_n\xi_n+\gamma_n^2\|\hat{\mathbf{V}}(n)\|_\infty^2,\end{aligned}\tag{C.4}$$

where we have set $v_n=\operatorname{tr}[(\mathbf{X}(n)-\mathbf{X}^*)\,\mathbf{V}(\mathbf{X}(n))]$ and $\xi_n=\operatorname{tr}[(\mathbf{X}(n)-\mathbf{X}^*)\,\mathbf{Z}(n)]$. Hence, telescoping (C.4) yields:

$$D_{n+1}\le D_1-\tau_n\left(a-\textstyle\sum_{j=1}^n w_{j,n}\,\xi_j\right)+\sum_{j=1}^n\gamma_j^2\|\hat{\mathbf{V}}(j)\|_\infty^2,\tag{C.5}$$

where $\tau_n=\sum_{j=1}^n\gamma_j$ and $w_{j,n}=\gamma_j/\tau_n$. By the strong law of large numbers for martingale differences [35, Theorem 2.18], we have $n^{-1}\sum_{j=1}^n\xi_j\to 0$ (a.s.); hence, given that $\gamma_{n+1}/\gamma_n\le 1$, Hardy's summability criterion [42, Thm. 14] applied to the sequence $w_{j,n}=\gamma_j/\tau_n$ yields $\sum_{j=1}^n w_{j,n}\xi_j\to 0$ (a.s.). Finally, since $\mathbb{E}[\sum_{j=1}^n\gamma_j^2\|\hat{\mathbf{V}}(j)\|_\infty^2]\le V^2\sum_{j=1}^\infty\gamma_j^2<\infty$, Doob's martingale convergence theorem [35, Theorem 2.5] shows that $\sum_{j=1}^n\gamma_j^2\|\hat{\mathbf{V}}(j)\|_\infty^2$ is finite (a.s.).[13]

Since $\tau_n\to\infty$ by assumption, the above implies that the RHS of (C.5) tends to $-\infty$ (a.s.); this contradicts the fact that $D_n\ge 0$, so we conclude that $\mathbf{X}(n)$ visits a compact neighborhood of $\mathbf{X}^*$ infinitely often (viz. there exists a sequence $n_j\uparrow\infty$ such that $\mathbf{X}(n_j)$ lies in said neighborhood). Since $\mathbf{X}^*$ attracts any initial condition $\mathbf{G}(\mathbf{Y}(0))$ under the continuous-time dynamics (MXL$_c$), Theorem 6.10 in [33] shows that $\mathbf{X}(n)$ converges to $\mathbf{X}^*$ (a.s.), as claimed. ∎

We close this section by showing that $\mathbf{Y}(n)$ remains an APT of (MXL$_c$) even if each player employs their individual step-size policy $\gamma_{k,n}$:

Proposition C.2. *Suppose* (MXL) *is run with player-specific step-size policies $\gamma_{k,n}$ such that $\sum_n\gamma_{k,n}^2<\sum_n\gamma_{k,n}=\infty$ for all $k\in K$. Then, $\mathbf{Y}(n)$ is an APT of* (MXL$_c$)*.*

Proof: Since the blocks $\mathbf{Z}_k$ and $\mathbf{Z}_{k'}$ are disjoint for $k\neq k'$, we obtain $\|\sum_k\gamma_{k,n}\mathbf{Z}_k(n)\|^2=\sum_k\|\gamma_{k,n}\mathbf{Z}_k(n)\|^2$ in the product norm on $\mathcal{X}$. Thus, Burkholder's inequality [35, Chap. 2] yields

$$\begin{aligned}&\mathbb{E}\left[\sup_{j<m<n}\left\|\textstyle\sum_{\ell=j}^m\sum_{k\in K}\gamma_{k,\ell}^2\mathbf{Z}_k(\ell)\right\|\right]\\&\qquad\le C\sum_{k\in K}\mathbb{E}\left[\textstyle\sum_{\ell=j}^n\gamma_{k,\ell}^2\|\mathbf{Z}_{k,\ell}\|_\infty^2\right],\end{aligned}\tag{C.6}$$

for some universal constant C. To proceed, let $\Delta(t,T)=\sup_{0\le h\le T}\left\|\int_t^{t+h}\bar{\mathbf{Z}}(s)\,ds\right\|_\infty$ where $\bar{\mathbf{Z}}$ is the linear interpolation of the discrete-time process $\mathbf{Z}(n)$ at each epoch n [33, p. 12]. Then, by applying's Hölder's inequality as in [33, p. 15], (C.6) readily gives $\mathbb{E}[\Delta(t,T)^2]\le C\sum_{k\in K}\int_t^{t+T}\bar\gamma_k^2(s)\,ds$, where, in similar notation, $\bar\gamma_k(t)$ denotes the linear interpolation of the discrete-time step-size sequence $\gamma_{k,n}$. We thus get

$$\begin{aligned}\sum_m\mathbb{E}[\Delta(mT,T)^2]&\le C_T\sum_{k\in K}\int_0^\infty\bar\gamma_k^2(s)\,ds\\&=\mathcal{O}\left(\sum_{k\in K}\sum_{n=0}^\infty\gamma_{k,n}^2\right)<\infty,\end{aligned}\tag{C.7}$$

so $\lim_{m\to\infty}\Delta(mT,T)\to 0$ by the Borel–Cantelli lemma. Arguing as in the proof of [33, Prop. A.1], it then follows that $\mathbf{Y}(n)$ is an APT of (MXL$_c$), as claimed. ∎

Remark 3. Applying Hölder's inequality [33, p. 15] to the derivation of (C.7) above, Proposition C.2 can also be extended to the case where (H2′) holds and the players' step-sizes satisfy $\sum_{n=1}^\infty\gamma_{k,n}^{1+q/2}<\infty$ for all $k\in K$. The rest of the argument remains essentially identical so, for concision, we omit it.

Appendix D
Local Convergence

Proof of Theorem 4: By the definition of local stability, there exists a neighborhood U of $\mathbf{X}^*$ in $\mathcal{X}$ such that (VS) holds for all $\mathbf{X}\in U$. Assume now that $m>0$ is taken sufficiently small so that $\mathbf{G}(\mathbf{Y})\in U$ whenever $F(\mathbf{X}^*,\mathbf{Y})<3m$ (the existence of such a positive m follows from the fact that $\mathbf{G}(\mathbf{Y})\to\mathbf{X}^*$ if $F(\mathbf{X}^*,\mathbf{Y})\to 0$). By Eq. (C.1), it then follows that the set $U_{3m}=$

[12] In particular, [33, Prop. 4.2] shows that, under (H2′), $\mathbf{Y}(n)$ is an APT of (MXL$_c$) for any step-size sequence γ_n such that $\sum_n\gamma_n^{1+q/2}<\sum_n\gamma_n=\infty$.

[13] By Hölder's inequality [33, p. 15], the same conclusion holds under (H2′) and the summability condition $\sum_n\gamma_n^{1+q/2}=\infty$.

$\{\mathbf{Y} : F(\mathbf{X}^*, \mathbf{Y}) \leq 3m\}$ is invariant under (MXL$_c$). Hence, by shadowing the proof of Theorem 3, it suffices to show that there exists an open set $U_\varepsilon \subseteq U_{3m}$ such that $\mathbb{P}(\mathbf{Y}(n) \in U_{3m} \text{ for all } n) \geq 1 - \varepsilon$ whenever $\mathbf{Y}(0) \in U_\varepsilon$.

To that end, let $D_n = F(\mathbf{X}^*, \mathbf{Y}(n-1))$ and assume that $D_1 \leq m$. Then, (C.4) yields

$$D_n \leq m + \sum_{j=1}^n \gamma_j v_j + \sum_{j=1}^n \gamma_j \xi_j + \sum_{j=1}^n \gamma_j^2 \|\hat{\mathbf{V}}(j)\|_\infty^2, \tag{D.1}$$

where $v_j = \operatorname{tr}[(\mathbf{X}(j) - \mathbf{X}^*)\mathbf{V}(j)]$ and $\xi_j = \operatorname{tr}[(\mathbf{X}(j) - \mathbf{X}^*)\mathbf{Z}(j)]$. We first claim that $\sup_n \sum_{j=1}^n \gamma_j \xi_j \leq m$ for all n with probability at least $1 - \varepsilon/2$ if γ_j is chosen appropriately. Indeed, letting $S_n = \sum_{j=1}^n \gamma_j \xi_j$, Doob's inequality [35, Theorem 2.1] gives

$$\mathbb{P}(\sup_{1 \leq j \leq n} |S_j| \leq m) \leq \frac{\mathbb{E}[|S_n|^2]}{m^2} \leq \frac{\sigma_*^2 \|\boldsymbol{\mathcal{X}}\|^2 \sum_{j=1}^n \gamma_j^2}{m^2}, \tag{D.2}$$

where we used the fact that $\mathbb{E}[\xi_j \xi_\ell] = 0$ if $j \neq \ell$. Hence, letting $\Gamma_2 = \sum_{j=1}^\infty \gamma_j^2$, it follows that $\mathbb{P}(\sup_n S_n \leq \varepsilon) \leq \Gamma_2 \sigma_*^2 \|\boldsymbol{\mathcal{X}}\|^2 / m^2 \leq \varepsilon/2$ if Γ_2 is sufficiently small. Moreover, letting $R_n = \sum_{j=1}^n \gamma_j^2 \|\hat{\mathbf{V}}(j)\|_\infty^2$ and working as above, Doob's inequality again shows that $\mathbb{P}(\sup_n R_n \geq m) \leq m^{-1} \lim_{n\to\infty} \mathbb{E}[R_n] \leq m^{-1} \Gamma_2 V^2 \leq \frac{\varepsilon}{2}$, if γ_n is taken small enough.[14]

Combining all of the above, we get $D_n \leq 3m + \sum_{j=1}^n \gamma_j v_j$ with probability at least $1 - \varepsilon$. Since $\mathbf{G}(U_{3m}) \subseteq U$ by construction, we have $v_j \leq 0$ for all $j = 1, \dots, n$, and we conclude that $D_n \leq 3m$ with probability at least $1 - \varepsilon$. This implies that $\mathbb{P}(\mathbf{Y}(n) \in U_{3m} \text{ for all } n) \geq 1 - \varepsilon$, as claimed. ∎

Appendix E
Rates of Convergence

In this last appendix, our goal is to derive the convergence rate of matrix exponential learning:

Proof of Theorem 5: Let $\bar{D}_n = \mathbb{E}[F(\mathbf{X}^*, \mathbf{Y}(n-1))] = \mathbb{E}[D_{\mathrm{KL}}(\mathbf{X}^*, \mathbf{X}(n))]$. Then, taking expectations in (C.4) yields

$$\begin{aligned} \bar{D}_{n+1} &\leq \bar{D}_n + \gamma_n \mathbb{E}[\operatorname{tr}[(\mathbf{X}(n) - \mathbf{X}^*)\mathbf{V}(n)]] + \gamma_n^2 \mathbb{E}[\|\hat{\mathbf{V}}(n)\|_\infty^2] \\ &\leq (1 - \gamma_n B)\bar{D}_n + \gamma_n^2 V^2, \end{aligned} \tag{E.1}$$

where, in the second line, we used (18b) and the assumption that $\mathbf{X}^*$ is B-strongly stable. With this in mind, we claim that $\bar{D}_n \leq A/n$ for all $n \geq 2$, where $A = \max\{2, 1/(B\gamma - 1)\}\gamma^2 V^2$. Proceeding by induction, note that (E.1) yields $\bar{D}_2 \leq (1 - B\gamma)\bar{D}_1 + \gamma^2 V^2 < \gamma^2 V^2 \leq A/2$, so (22) holds for $n = 2$. Assume now that $\bar{D}_n \leq A/n$ for some $n \geq 2$, so it suffices to show that $\bar{D}_{n+1} \leq A/(n+1)$ as well. Clearly, with $1 - B/\gamma_n = 1 - B\gamma/n \geq 0$ for all $n \geq 2$ (by assumption), we only need to show that

$$\left(1 - \frac{B\gamma}{n}\right)\frac{A}{n} + \frac{\gamma^2 V^2}{n^2} \leq \frac{A}{n+1}. \tag{E.2}$$

Rearranging this last equation, we get $nA \leq (n+1)(AB\gamma - \gamma^2 V^2)$. However, by definition, we have $A \geq \gamma^2 V^2/(B\gamma - 1)$ or, equivalently, $AB\gamma - \gamma^2 V^2 \geq A$, and our claim follows.

Finally, for the bound (23), recall that h is $(1/2)$-strongly convex over $\boldsymbol{\mathcal{X}}$ [3], so $D_{\mathrm{KL}}(\mathbf{X}^*, \mathbf{X}) \geq \frac{1}{4}\|\mathbf{X}^* - \mathbf{X}\|^2$ by the general theory on Bregman divergences [5, p. 148]. ∎

[14] As in the proof of Theorem 3, Hölder's inequality [33, Eq. 14] shows that the same holds under (H2′) and the summability condition $\sum_n \gamma_n^{1+q/2} = \infty$.

References

[1] P. Mertikopoulos, E. V. Belmega, and L. Sanguinetti, "Distributed learning for resource allocation under uncertainty," in *GLOBALSIP '16: Proceedings of the 2016 IEEE Global Conference on Signal and Information Processing*, 2016.

[2] 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.

[3] 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.

[4] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, ser. Applied Optimization. Kluwer Academic Publishers, 2004, no. 87.

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

[6] V. G. Vovk, "Aggregating strategies," in *COLT '90: Proceedings of the 3rd Workshop on Computational Learning Theory*, 1990, pp. 371–383.

[7] I. Stiakogiannakis, P. Mertikopoulos, and C. Touati, "Adaptive power allocation and control in time-varying multi-carrier MIMO networks," working paper, http://arxiv.org/abs/1503.02155.

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

[10] Y. Viossat and A. Zapechelnyuk, "No-regret dynamics and fictitious play," *Journal of Economic Theory*, vol. 148, no. 2, pp. 825–842, March 2013.

[11] 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.

[12] H. Yu and M. J. Neely, "Dynamic power allocation in MIMO fading systems without channel distribution information," http://arxiv.org/abs/1512.08419, 2015.

[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] L. Chen, S. H. Low, and J. C. Doyle, "Random access game and medium access control design," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, pp. 1303 – 1316, Aug 2010.

[15] T. Cui, L. Chen, and S. H. Low, "A game-theoretic framework for medium access control," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1116–1127, Sept. 2008.

[16] 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.

[17] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 12, pp. 1349–1380, 2000.

[18] A. Bellet, A. Habrard, and M. Sebban, "A survey on metric learning for feature vectors and structured data," *arXiv preprint arXiv:1306.6709*, 2013.

[19] 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.

[20] D. Lim, G. Lanckriet, and B. McFee, "Robust structural metric learning," in *Proc. of the 30th Intl. Conf. on Machine Learning*, 2013, pp. 615–623.

[21] L. Bottou, "Stochastic gradient descent tricks," in *Neural Networks: Tricks of the Trade*. Springer, 2012, pp. 421–436.

[22] 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.

[23] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah, "Energy-efficient power control: A look at 5G wireless technologies," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1668–1683, April 2016.

[24] 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.

[25] A. Charnes and W. W. Cooper, "Programming with linear fractional functionals," *Naval Research Logistics Quarterly*, vol. 9, pp. 181–196, 1962.

[26] 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.

[27] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave *N*-person games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.

[28] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, ser. Springer Series in Operations Research. Springer, 2003.

[29] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124 – 143, 1996.

[30] V. Vedral, "The role of relative entropy in quantum information theory," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 197–234, 2002.

[31] P. Mertikopoulos and W. H. Sandholm, "Learning in games via reinforcement and regularization," *Mathematics of Operations Research*, vol. 41, no. 4, pp. 1297–1324, November 2016.

[32] P. Mertikopoulos, "Learning in concave games with imperfect information," https://arxiv.org/abs/1608.07310, 2016.

[33] 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.

[34] 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.

[35] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application*, ser. Probability and Mathematical Statistics. New York: Academic Press, 1980.

[36] P. Coucheney, B. Gaujal, and P. Mertikopoulos, "Penalty-regulated dynamics and robust learning procedures in games," *Mathematics of Operations Research*, vol. 40, no. 3, pp. 611–633, August 2015.

[37] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press and Hindustan Book Agency, 2008.

[38] A. M. Davie and A. J. Stothers, "Improved bound for complexity of matrix multiplication," *Proceedings of the Royal Society of Edinburgh, Section A: Mathematics*, vol. 143, no. 2, pp. 351–369, 4 2013.

[39] 3GPP, "User equipment (UE) radio transmission and reception," White paper, Jun. 2014.

[40] D. P. Palomar, J. M. Cioffi, and M. Lagunas, "Uniform power allocation in MIMO channels: a game-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, p. 1707, July 2003.

[41] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications and Related Technologies*, vol. 10, no. 6, pp. 585–596, 1999.

[42] G. H. Hardy, *Divergent Series*. Oxford University Press, 1949.

Panayotis Mertikopoulos (M' 11) received the Ptychion degree in physics (summa cum laude) from the University of Athens in 2003, the M.Sc. and M.Phil. degrees in mathematics from Brown University in 2005 and 2006 respectively (both summa cum laude), and the Ph.D. degree in physics from the University of Athens in 2010. During 2010–2011, he was a post-doctoral researcher at the Economics Department of École Polytechnique, Paris, France. Since 2011, he has been a CNRS Researcher at the Laboratoire d'Informatique de Grenoble, Grenoble, France. P. Mertikopoulos was an Embeirikeion Foundation Fellow between 2003 and 2006, and received the best paper award in NetGCoop '12. He was a track co-chair of IFORS 2014, TPC co-chair of WiOpt 2014, and general co-chair of AlgoGT 2013. His main research interests lie in learning, optimization, game theory, and their applications to wireless networks and signal processing.

E. Veronica Belmega (S'08–M'10) received the M.Sc. (engineer diploma) degree from the University Politehnica of Bucharest, Bucharest, Romania, in 2007, and the M.Sc. and Ph.D. degrees both from the University Paris-Sud 11, Orsay, France, in 2007 and 2010, respectively. From 2010 to 2011, she was a Postdoctoral Researcher in a joint project between Princeton University, N.J., USA and the Alcatel-Lucent Chair on Flexible Radio in Supélec, France. She is currently an Assistant Professor with ETIS/ENSEA–Université de Cergy-Pontoise–CNRS, Cergy-Pontoise, France, and Inria. She was one of the ten recipients of the L'Oréal–UNESCO–French Academy of Science Fellowship: "For young women doctoral candidates in science" in 2009.

Romain Negrel received the M.Sc. and Ph.D. degrees both from the University of Cergy-Pontoise, Cergy-Pontoise, France, in 2011 and 2014, respectively. From 2015 to 2016, he was a Postdoctoral Researcher with the GREYC Lab and the University of Normandy, Caen, France. He is currently an Assistant Professor with LIGM/ESIEE Paris-Université Paris-Est, Marne-la-Vallée, France.

Luca Sanguinetti (SM'15) received the Laurea Telecommunications Engineer degree (cum laude) and the Ph.D. degree in information engineering from the University of Pisa, Italy, in 2002 and 2005, respectively. Since 2005 he has been with the Dipartimento di Ingegneria dell'Informazione of the University of Pisa. In 2004, he was a visiting Ph.D. student at the German Aerospace Center (DLR), Oberpfaffenhofen, Germany. During the period June 2007 – June 2008, he was a postdoctoral associate in the Dept. Electrical Engineering at Princeton. During the period June 2010 - Sept. 2010, he was selected for a research assistantship at the Technische Universität Munchen. From July 2013 to July 2015, he was with the Alcatel–Lucent Chair on Flexible Radio, Supélec, Gif-sur-Yvette, France. He is currently an Assistant Professor at the Dipartimento di Ingegneria dell'Informazione of the University of Pisa.

L. Sanguinetti is serving as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE SIGNAL PROCESSING LETTERS. He served as Lead Guest Editor of IEEE JOURNAL ON SELECTED AREAS OF COMMUNICATIONS Special Issue on "Game Theory for Networks" and as an Associate Editor for IEEE JOURNAL ON SELECTED AREAS OF COMMUNICATIONS (series on Green Communications and Networking). His expertise and general interests span the areas of communications and signal processing, game theory and random matrix theory for wireless communications. He was the co-recipient of two best paper awards: *IEEE Wireless Commun. and Networking Conference (WCNC) 2013* and *IEEE Wireless Commun. and Networking Conference (WCNC) 2014*. He was also the recipient of the FP7 Marie Curie IEF 2013 "Dense deployments for green cellular networks".