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 and/or obsolete. 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. We also derive an explicit linear bound for the algorithm's convergence speed, which remains valid under measurement errors and uncertainty of arbitrarily high variance. To validate our theoretical analysis, we test the algorithm 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.

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, Grenoble, France. R. Negrel is with GREYC, CNRS UMR 6072, ENSICAEN Université de Caen Basse-Normandie, France. L. Sanguinetti is with the University of Pisa, Dipartimento di Ingegneria dell'Informazione, Italy, and also with the Large Systems and Networks Group (LANEAS), CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France.

arXiv:1606.01190v1 [cs.IT] 3 Jun 2016

I. Introduction

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

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

where $\mathcal{X}_i$ denotes the set of feasible actions of player i. Specifically, we will focus on feasible action sets of the general form

$$\mathcal{X}_i = \{\mathbf{X}_i \succcurlyeq 0 : \|\mathbf{X}_i\| \le A_i\} \tag{2}$$

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

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 are extremely widespread in signal processing, wireless communications and information theory, especially in a stochastic framework where: *a*) the objective functions u_i 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 measurement errors (Ex. II-C). Accordingly, our main goal will be 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 by the authors of [1] in the context of throughput maximization in multiple-input and multiple-output (MIMO) systems. In a nutshell, the main idea of the proposed method is that each player tracks the individual gradient of his utility function via an auxiliary score matrix, possibly subject to randomness and/or feedback imperfections. The players' actions are then computed via an "exponential projection" step that maps these score matrices to the players' action spaces, and the process repeats. Thus, building on the preliminary results of [1] for MIMO throughput maximization, we show here that *a*) MXL can

be applied to a much broader class of (stochastic) optimization problems and games of the general form (1); *b*) the algorithm's convergence does not require the (restrictive) structure of a potential game [2]; *c*) MXL converges to equilibrium (locally or globally) under very mild assumptions for the noise and uncertainty surrounding the optimizers' decisions; and *d*) we derive explicit estimates for the method's rate of convergence.

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 $\mathbf{X}_i$ with constant trace) 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 the context of online power control [7] and throughput maximization [8] for dynamic MIMO systems. The goal there was to show that, in the presence of temporal variabilities, 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, (external) regret minimization is neither necessary nor sufficient to ensure convergence to a Nash equilibrium (or other game-theoretic solution concept).

In [1, 11, 12], a special case of (1) was studied in the context of MIMO throughput maximization in the presence of stochasticity, in both single-user [12] and multiple access channels [1, 11]. In both cases, the problem boils down to a semidefinite optimization problem, possibly distributed over several agents [1]. 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, so a potential-based analysis is often unsuitable for practical applications.

In this paper, we derive the convergence properties of matrix exponential learning in concave *N*-player games, and we investigate the algorithm's long-term behavior under feedback errors and uncertainty. Specifically, building on earlier work by Rosen [13], we consider a broad class of games that satisfy a local *variational stability* condition which ensures that Nash equilibria are locally isolated. In a series of recent papers, Scutari et al. used a stronger, global 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 panoramic survey, see [14] and references therein. However, in the presence of 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) [15] fails to provide any perceptible gains over crude, uniform power allocation policies [1]. 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 for games played over bounded regions of positive-semidefinite matrices in the presence of feedback noise and randomness. 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 transmit *energy efficiency* (EE) maximization in multi-carrier MIMO networks (a critical design feature of emerging green multi-cellular networks).

In Section III, we revisit our core game-theoretic framework, and we discuss the notion of a *stable* Nash equilibrium. Subsequently, in Section IV, we introduce our matrix exponential learning scheme, and we detail our assumptions for the stochasticity affecting the players' objectives and observations. Our main results are then presented in Section V and can be summarized as follows: Under fairly mild assumptions on the underlying stochasticity (zero-mean feedback errors with finite conditional variance), we show that (*i*) the algorithm's only termination states are Nash equilibria; (*ii*) if the game admits a globally (locally) stable equilibrium, then the algorithm converges globally (locally) to said equilibrium; and (*iii*) on average, the algorithm converges to an ε-neighborhood of an extreme (resp. interior) strongly stable equilibrium within $\mathcal{O}(1/\varepsilon)$ iterations (resp. $\mathcal{O}(1/\varepsilon^2)$).

The above results greatly extend and generalize the recent analysis of [1] for throughput maximization in MIMO MAC systems. To further validate our results in MIMO environments, our theoretical analysis is supplemented with extensive numerical simulations for energy efficiency maximization in multi-carrier MIMO wireless networks in Section VI. To streamline the flow of the paper, technical proofs have been relegated to a series of appendices at the end.

Notation: In what follows, the profile $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$ is identified with the block-diagonal matrix $\mathrm{diag}(\mathbf{X}_i)_{i=1}^N$, and we use the game-theoretic shorthand $(\mathbf{X}_i; \mathbf{X}_{-i})$ when we need to focus on the action $\mathbf{X}_i$ of

player i against that of all other players. Also, given $\mathbf{A} \in \mathbb{C}^{d \times d}$, we write $\|\mathbf{A}\| = \sum_{j=1}^{d} |\text{eig}_j(\mathbf{A})|$ for the nuclear (trace) norm of $\mathbf{A}$ and $\|\mathbf{A}\|_\infty = \max_j |\text{eig}_j(\mathbf{A})|$ for its (dual) 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 (and reflecting the subjective tastes and interests of the authors). A reader who is interested in the general theory may skip this section and proceed directly to Sections III–V.

The first example below defines a game-theoretic model for the interactions between wireless users in networks with contention-based medium access. The second example revolves around metric learning for similarity-based image search and showcases the range of problems where the proposed method applies. Finally, the third example focuses on energy efficiency (EE) maximization in MIMO multi-user networks, a key aspect of future and emerging wireless networks: to the best of our knowledge, MXL provides the first distributed and provably convergent algorithm for energy efficiency maximization in with noisy feedback and imperfect channel state information.

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 competing wireless users that interfere with each other. To model this, consider a set of wireless users indexed by $\mathcal{N} = \{1, \ldots, N\}$, each updating their individual channel access probability x_i in response to the amount of contention in the network [16]. In practice, wireless 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* $q_i(\mathbf{x}_{-i})$ which is determined as a (symmetric) function of the access probability profile $\mathbf{x}_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1} \ldots, x_N)$ of all other users.[1]

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

$$u_i(x_i; \mathbf{x}_{-i}) = U_i(x_i) - x_i q_i(\mathbf{x}_{-i}) \tag{3}$$

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

---

[1]For instance, a standard contention measure of this type is the conditional collision probability $q_i(\mathbf{x}_{-i}) = 1 - \prod_{j \neq i}(1 - x_j)$. [16].

We thus obtain the multi-agent, multi-objective formulation

$$
\text{for all } i \in \mathcal{N} \quad
\begin{cases}
\text{maximize} & u_i(x_i; \mathbf{x}_{-i}), \\
\text{subject to} & x_i \in [0, 1].
\end{cases}
\tag{4}
$$

whose solutions can be analyzed through the specification of the utility function $U_i(x_i)$ and the choice of the contention measure q_i. If $U_i(x_i)$ is assumed to be continuously differentiable and strictly concave (see for example [16, 17] and references therein), (4) is a special case of (1) with $M = 1$ and $A_i = 1$. We further note the unilateral optimization problems above are interconnected because of the dependence of the contention measure $q_i(\mathbf{x}_{-i})$ on the access probabilities of all users; as a result, the theory of noncooperative games theory arises naturally as the most suitable framework to analyze and solve (4) [17].

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) [18, 19]. 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 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 $\mathbf{k}$. With this in mind, we obtain the minimization objective

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

---

[2] The baseline case $\mathbf{X} = \mathbf{I}$ corresponds to the Euclidean distance, which is often unsuitable for image discrimination purposes [20].

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 [18], enable distributed storing and retrieval [22], and better exploit correlations between features that further reduce over-fitting effects [21]. A computationally efficient way to achieve this is to include a trace constraint of the form $\operatorname{tr}\mathbf{X} \leq c$ for some $c \ll d$,[4] leading to the feasible region

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

Combining (6) and (7), we see that metric learning in image retrieval is a special case of the general problem (1) with $N = 1$ optimizers. However, when $\mathcal{D}$ is large, the number of variables involved is computationally prohibitive: for instance, (6) may contain up to 10^9–10^{11} terms for a modest database with 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 [23]:

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

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

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 flip side however, 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. [24–26] and references therein. Here, wireless connections are established over transceiver pairs with M (resp. N) antennas at the transmitter (resp. receiver), and communication takes place over a set of

---

[3]This last regularization term is included in order to maximize predictive accuracy by reducing the effects of over-fitting to training data [21].

[4]By contrast, an ℓ_0 rank constraint of the form $\operatorname{rank}(\mathbf{X}) \leq c$ generically leads to an untractable NP-hard problem formulation.
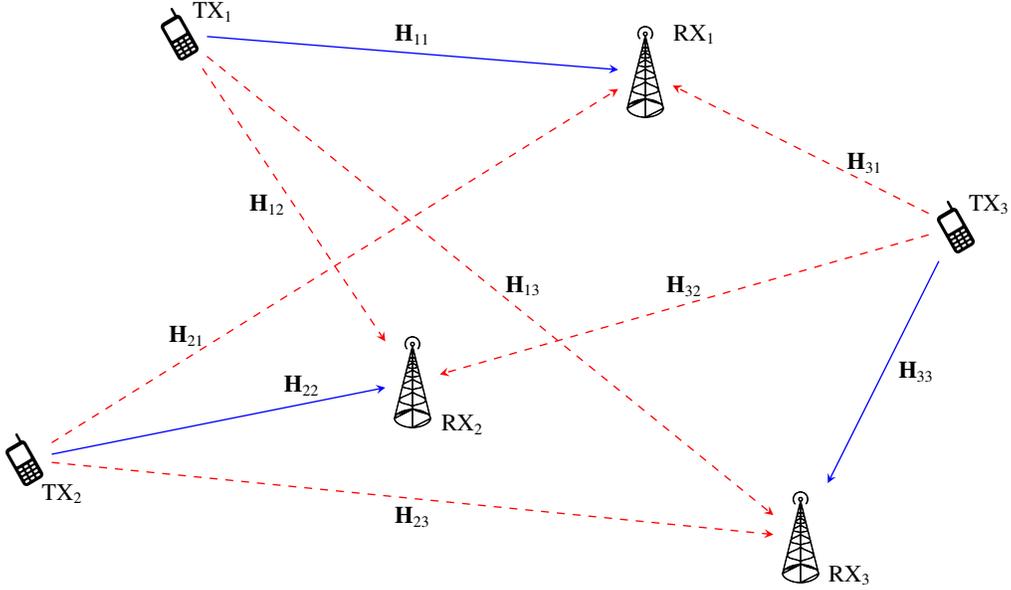
Fig. 1. A multi-user MIMO system with $N = 3$ connections.

S orthogonal subcarriers. Accordingly, the i-th transmitter is assumed to control his individual input signal covariance matrix $\mathbf{Q}_{is}$ over each subcarrier $s = 1, \ldots, S$, subject to the constraints: a) $\mathbf{Q}_{is} \succcurlyeq 0$ (since each $\mathbf{Q}_{is}$ is a signal covariance matrix); and b) $\operatorname{tr} \mathbf{Q}_i \leq P_{\max}$, where $\mathbf{Q}_i = \operatorname{diag}(\mathbf{Q}_{is})_{s=1}^S$ is the covariance profile of the i-th transmitter, $\operatorname{tr}(\mathbf{Q}_i)$ 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 given by the well-known expression

$$R_i(\mathbf{Q}) = \log \det \left(\mathbf{W}_{-i} + \mathbf{H}_{ii} \mathbf{Q}_i \mathbf{H}_{ii}^\dagger \right) - \log \det \left(\mathbf{W}_{-i} \right), \tag{9}$$

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

$$\operatorname{EE}_i(\mathbf{Q}) = \frac{R_i(\mathbf{Q})}{P_c + \operatorname{tr}(\mathbf{Q}_i)}, \tag{10}$$

where $P_c > 0$ represents the total power consumed by circuit components at the transmitter [24].

The energy efficiency function above is not concave w.r.t the covariance matrix $\mathbf{Q}_i$ of user i, but it can be recast as such via a suitable Charnes-Cooper transformation [27]. To this aim, following [26], consider the

adjusted control variables

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

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

$$\mathcal{X}_i = \{\text{diag}(\mathbf{X}_{is})_{s=1}^S : \mathbf{X}_{is} \succcurlyeq 0 \text{ and } \sum_s \text{tr} \, \mathbf{X}_{is} \leq 1\}, \tag{12}$$

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

$$u_i(\mathbf{X}_i; \mathbf{X}_{-i}) = \frac{P_c + (1 - \text{tr} \, \mathbf{X}_i) \, P_{\max}}{P_c(P_c + P_{\max})} \log \det \left(\mathbf{I} + \frac{P_c P_{\max} \, \widetilde{\mathbf{H}}_i \mathbf{X}_i \widetilde{\mathbf{H}}_i^{\dagger}}{P_c + (1 - \text{tr} \, \mathbf{X}_i) P_{\max}} \right), \tag{13}$$

where $\widetilde{\mathbf{H}}_i = \mathbf{W}_{-i}^{-1/2} \mathbf{H}_{ii}$ denotes the *effective channel matrix* of user i.

We thus see that the resulting energy efficiency maximization problem is of the general form (1) with $A_i = 1$.[5] In this context, 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 as any action profile $\mathbf{X}^* \in \mathcal{X}$ which is *unilaterally stable* in the sense that

$$u_i(\mathbf{X}^*) \geq u_i(\mathbf{X}_i; \mathbf{X}_{-i}^*) \quad \text{for all } \mathbf{X}_i \in \mathcal{X}_i, \, i \in \mathcal{N}. \tag{14}$$

In other words, $\mathbf{X}^* \in \mathcal{X}$ is a Nash equilibrium when no single player can further increase his individual utility assuming that the actions of all other players remain unchanged.

In complete generality, a game need not admit an equilibrium. However, thanks to the concavity of each player's payoff function u_i and the compactness of their action space $\mathcal{X}_i$, the existence of a Nash equilibrium in the case of (1) is guaranteed by the general theory of [28]. Hence, a natural question that arises is whether such an equilibrium solution is unique or not.

---

[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.

To answer this question, Rosen [13] provided a first-order sufficient condition for general N-person concave games known as *diagonal strict concavity* (DSC). To state it, define the *individual payoff gradient* of player i as

$$\mathbf{V}_i(\mathbf{X}) \equiv \nabla_{\mathbf{X}_i} u_i(\mathbf{X}_i; \mathbf{X}_{-i}), \tag{15}$$

and let $\mathbf{V}(\mathbf{X}) \equiv \mathrm{diag}(\mathbf{V}_1(\mathbf{X}), \dots, \mathbf{V}_N(\mathbf{X}))$ denote the collective profile of all players' individual gradients. With this definition at hand, Rosen's condition can be stated as:

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

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

Theorem 1 (Rosen [13])**.** *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 and/or imperfect feedback. More recently, (DSC) was used by Scutari et al. (see [14] and references therein) as the starting point for the convergence analysis of a class of Gauss–Seidel methods for concave games based on variational inequalities [29]. Our approach is similar in scope but relies instead on the following notion of stability:

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

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

In particular, if (VS) holds for all $\mathbf{X} \in \mathcal{X}$, 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. Still, as is the case with DSC, stability plays a key role in the characterization of Nash equilibria:

Proposition 1. *If $\mathbf{X}^* \in \mathcal{X}$ 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 $\mathbf{X}_i$ close to $\mathbf{X}_i^*$, and let $\mathbf{X} = (\mathbf{X}_i; \mathbf{X}_{-i}^*)$. Then, by (VS), we get $\mathrm{tr}[(\mathbf{X}_i - \mathbf{X}_i^*)\,\mathbf{V}_i(\mathbf{X}_i; \mathbf{X}_{-i}^*)] < 0$, implying that u_i is decreasing along the ray $\mathbf{X}_i^* + t(\mathbf{X}_i - \mathbf{X}_i^*)$. Since this covers

---

[6]Simply note that Nash equilibria are solutions of the variational inequality $\mathrm{tr}[(\mathbf{X} - \mathbf{X}^*)\mathbf{V}(\mathbf{X}^*)] \leq 0$ [14]. Then, (VS) follows by setting $\mathbf{X}^* = \mathbf{X}'$ in (DSC).

all rays starting at $\mathbf{X}_i^*$, we conclude that $\mathbf{X}^*$ is the game's unique equilibrium in the neighborhood of $\mathbf{X}^*$ where (VS) holds. Our assertion then follows immediately. ∎

The variational stability condition (VS) is important not only for characterizing the structure of the game's Nash set, but also for determining the convergence properties of the proposed learning scheme. 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. Therefore, it is of paramount importance to have a verifiable criterion for the stability of a Nash equilibrium. This can be accomplished by appealing to a second-order condition (similar to the second derivative test in calculus).

Specifically, define the *Hessian* of a game as follows:

Definition 2. The *Hessian matrix* of a game is the block matrix $\mathbf{D}(\mathbf{X}) = (\mathbf{D}_{ij}(\mathbf{X}))_{i,j\in\mathcal{N}}$ with blocks

$$\mathbf{D}_{ij}(\mathbf{X}) = \tfrac{1}{2}\nabla_{\mathbf{X}_i}\nabla_{\mathbf{X}_j}u_j(\mathbf{X}) + \tfrac{1}{2}[\nabla_{\mathbf{X}_j}\nabla_{\mathbf{X}_i}u_i(\mathbf{X})]^\dagger. \tag{16}$$

The terminology "Hessian" reflects the fact that, in the single-player case where (1) is a single-agent optimization problem, $\mathbf{D}(\mathbf{X})$ 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:

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

Proof: The first statement of the proposition can be proved as follows. Assume first that $\mathbf{D}(\mathbf{X}) \prec 0$ for all $\mathbf{X} \in \mathcal{X}$. Then, from [13, Theorem 6] it follows that that the game satisfies (DSC) and thus admits a unique Nash equilibrium $\mathbf{X}^*$; since (DSC) implies (VS) for all $\mathbf{X} \in \mathcal{X}$, our claim is immediate. The proof for the second statement is as follows. If $\mathbf{X}^*$ is a Nash equilibrium and $\mathbf{D}(\mathbf{X}^*) \prec 0$, we will have $\mathbf{D}(\mathbf{X}) \prec 0$ for all $\mathbf{X}$ in some convex neighborhood $\mathcal{U}$ of $\mathbf{X}^*$ (by continuity). Then, by applying [13, Theorem 6] to $\mathcal{U}$ and reasoning as in the global case above, it follows that $\mathbf{X}^*$ is the game's unique equilibrium in $\mathcal{U}$, as claimed. ∎

Remark 1. In the special class of potential games, the players' payoff functions are aligned along a common goal, the game's potential [2]; as a result, the maximizers of the game's potential 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. Similarly, the sufficient condition of Proposition 2 reduces to the second order condition of concave potential functions that guarantees uniqueness of the solution (strictly negative

---

Algorithm 1 Matrix exponential learning (MXL).

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

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

Repeat
> $n \leftarrow n + 1$;
>
> **foreach** player $i \in \mathcal{N}$ **do**
>> play $\mathbf{X}_i \leftarrow A_i \dfrac{\exp(\mathbf{Y}_i)}{1 + \|\exp(\mathbf{Y}_i)\|}$;
>> get gradient feedback $\hat{\mathbf{V}}_i$;
>>
>> update auxiliary matrix $\mathbf{Y}_i \leftarrow \mathbf{Y}_i + \gamma_n \hat{\mathbf{V}}_i$;
>
> **until** termination criterion is reached.

---

definite Hessian matrix).

IV. LEARNING UNDER UNCERTAINTY

The goal of this section is to provide a learning algorithm that allows players to converge to a Nash equilibrium in a decentralized environment with no information exchange between players and/or imperfect feedback. To that end, the main idea of the proposed learning scheme is as follows: First, at every stage $n = 0, 1, \ldots$ of the process, each player $i \in \mathcal{N}$ tracks the individual gradient of his utility function via an auxiliary "score" matrix $\mathbf{Y}_i(n)$, possibly subject to feedback/stochasticity-induced errors. Subsequently, the players' actions $\mathbf{X}_i(n)$ are computed via an "exponential projection" step that maps the gradient tracking matrix $\mathbf{Y}_i(n)$ back to the player's action space $\mathcal{X}_i$, and the process repeats.

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

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

where:

1) $n = 0, 1, \ldots$ denotes the stage of the process.

2) the auxiliary matrix variables $\mathbf{Y}_i(n)$ are initialized to an arbitrary (Hermitian) value.

3) $\hat{\mathbf{V}}_i(n)$ is a stochastic estimate of the individual gradient $\mathbf{V}_i(\mathbf{X}(n))$ of player i at stage n (more on this below).

4) γ_n is a decreasing step-size sequence, typically of the form $\gamma_n \sim 1/n^a$ for some $a \in (0, 1]$.

Setting aside for a moment the precise nature of the stochastic estimates $\hat{\mathbf{V}}_i(n)$, we note that the update of the auxiliary matrix variable $\mathbf{Y}_i(n)$ in (MXL) acts as a "steepest ascent" step along the estimated direction of each player's individual payoff gradient $\mathbf{V}_i(\mathbf{X}(n))$.[7] As such, if there were no constraints for the players' actions, $\mathbf{Y}_i(n)$ would define an admissible sequence of play and, *ceteris paribus*, player i would tend to increase his payoff along this sequence. However, this simple ascent scheme does not suffice in our constrained framework, so $\mathbf{Y}_i(n)$ is first exponentiated and subsequently normalized in order to meet the feasibility constraints (2).[8]

Of course, the outcome of the players' gradient tracking process depends crucially on the quality of the gradient feedback $\hat{\mathbf{V}}_i(n)$ that is available to them. 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 $u_i(\mathbf{X}) = \mathbb{E}[\hat{u}_i(\mathbf{X}; \omega)]$ for some random variable ω, and the players can only observe the (stochastic) gradient of $\hat{u}_i$ (cf. Ex. II-B).

iii) Any combination of the above.

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

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

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

(H1) *Zero-mean:*

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

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

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

The statistical hypotheses (H1) and (H2) above are fairly mild from and allow for a broad range of estimation scenarios.[9] 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, Hypothesis (H2) is a bare-bones assumption for the variance of the players' feedback, and it is satisfied by most common

---

[7]The step-size sequence γ_n further finetunes this process and its role is discussed in detail later.

[8]Recall here that $\mathbf{V}_i$ is Hermitian as the derivative of a real function with respect to a Hermitian matrix variable [30]. Moreover, the normalization step in (MXL) subsequently ensures that the resulting matrix has $\|\mathbf{X}_i\| \leq A_i$, so the sequence $\mathbf{X}_i(n)$ induced by (MXL) meets the problem's feasibility constraints.

[9]In particular, we will *not* be assuming that the errors are i.i.d. or bounded.

error processes – such as Gaussian, log-normal, uniform and all sub-Gaussian distributions. In other words, Hypotheses (H1) and (H2) simply mean that the players' individual gradient estimates $\hat{\mathbf{V}}_i$ are *unbiased and bounded in mean square*, i.e.

$$\mathbb{E}[\hat{\mathbf{V}}_i(n) \mid \mathbf{X}(n)] = \mathbf{V}_i(\mathbf{X}(n)), \tag{18a}$$

$$\mathbb{E}[\|\hat{\mathbf{V}}_i(n)\|_\infty^2 \mid \mathbf{X}(n)] \le V_i^2 \quad \text{for some } V_i > 0. \tag{18b}$$

Even though (H2) will be our main error control assumption, it will also be useful to consider the more refined hypothesis:

(H2′) *Subexponential moment growth:* for all $p \in \mathbb{N}$,

$$\mathbb{E}[\|\mathbf{Z}(n)\|_\infty^p \mid \mathbf{X}(n)] \le \frac{p!}{2}\sigma_*^p \quad \text{for some } \sigma_* > 0. \tag{H2′}$$

Clearly, (H2) is implied by (H2′) but the latter may fail in certain heavy-tailed error distributions (such as Pareto-tailed ones). From a practical point of view, this is not particularly restrictive as (H2′) continues to apply to a wide range of practical scenarios (including all Gaussian and sub-Gaussian error distributions), and is considerably more general than the "finite errors" assumptions studied in the literature [5].

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 will be to examine the long-term convergence properties of (MXL) and its practical implementation attributes.

A. Convergence

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

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 $\mathbf{X}(n)$ converges, it does so to 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* [31]; then, by using the law of large numbers, it is possible to control the impact of the noise and show that this entropy diverges to infinity, thus obtaining the desired contradiction.

Stated concisely, Theorem 2 shows that if (MXL) converges, it converges to a Nash equilibrium. However, the theorem 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 do this, we introduce a matrix version of the so-called *Fenchel coupling* [32] and we show that it 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 aggregate error between continuous and discrete time.

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

Corollary 1. *If the game's Hessian matrix $\mathbf{D}(\mathbf{X})$ is negative-definite for all $\mathbf{X} \in \mathcal{X}$, 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). The following theorem shows that, under the slightly sharper assumption (H2′), local stability implies local convergence with arbitrarily

high probability:

Theorem 4. *Assume that Algorithm 1 is run with gradient observations satisfying* (H1) *and* (H2′)*, and small enough step-sizes γ_n such that $\sum_{n=1}^{\infty} \gamma_n^{1+q} < \sum_{n=1}^{\infty} \gamma_n = \infty$ for some $q \in (0, 1)$. 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*

$$\mathbb{P}(\mathbf{X}(n) \to \mathbf{X}^* \mid \mathbf{X}(0) \in U_\varepsilon) \geq 1 - \varepsilon. \tag{19}$$

Theorem 4 is proven in Appendix D, building on the (global) proof of Theorem 3. The key difference with Theorem 3 is that, since we only assume the existence of a *locally stable* equilibrium $\mathbf{X}^*$, the drift of (MXL) has no reason to point towards $\mathbf{X}^*$ globally. As a result, $\mathbf{X}(n)$ may exit the basin of $\mathbf{X}^*$ in the presence of high uncertainty. However, by invoking (H2′) and the Borel–Cantelli lemma, it is possible to show that this happens with arbitrarily small probability, leading to the probabilistic convergence result (19).

As in the case of globally stable equilibria, Theorem 4 leads to the following easy-to-check condition for local convergence:

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

B. Rate of convergence

Combining Theorems 2–4 gives a fairly complete picture of the *qualitative* convergence properties of the exponential learning algorithm (MXL): the only possible end-states of (MXL) are Nash equilibria, and the existence of a globally (resp. locally) stable Nash equilibrium implies the algorithm's global (resp. local) convergence to said equilibrium. On the other hand, these theorems do not address the *quantitative* aspects of the algorithm's long-term behavior, such as its rate of convergence. In what follows, we study precisely this question.

We begin by introducing the so-called *quantum Kullback–Leibler divergence* (or *von Neumann relative entropy*) to measure distances on $\mathcal{X}$ [31]. Specifically, the quantum divergence between $\mathbf{X}$ and $\mathbf{X}^*$ is defined as

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

By Klein's inequality [31], $D_{\mathrm{KL}}(\mathbf{X}^*, \mathbf{X}) \geq 0$ 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

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

Under this refinement of equilibrium stability,[10] 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 $\gamma > B^{-1}$, we have*

$$\mathbb{E}[D_{\text{KL}}(\mathbf{X}^*, \mathbf{X}(n))] \leq \frac{\gamma^2 V^2}{B\gamma - 1}\frac{1}{n}. \tag{22}$$

In particular, if $\mathbf{X}^*$ lies in the interior of $\mathcal{X}$, then

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

Otherwise, if $\mathbf{X}^*$ is an extreme point of $\mathcal{X}$, we have

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

The explicit bounds of Theorem 5 (proven in Appendix E) have several interesting consequences:

i) The rate of convergence to interior versus extreme states of $\mathcal{X}$ is substantially different because the von Neumann divergence grows quadratically when $\mathbf{X}^*$ is interior and linearly when $\mathbf{X}^*$ is extreme. Intuitively, the reason for this is that, in the case of interior states, the algorithm has to slow down considerably near $\mathbf{X}^*$ in order for random oscillations around $\mathbf{X}^*$ to dissipate. On the other hand, in the case of extreme points, the algorithm has a constant drift towards the boundary of $\mathcal{X}$ and there are no possibilities of overshooting (precisely because the equilibrium state in question is an extreme point of $\mathcal{X}$). As a result, random oscillations become irrelevant near extreme points, thus allowing for faster convergence in that case.

ii) Thanks to the explicit expression (22), we see that if the stability constant B can be estimated ahead of time, we can attain the convergence rate

$$\mathbb{E}[D_{\text{KL}}(\mathbf{X}^*, \mathbf{X}(n))] \leq \frac{4V^2}{B^2 n}, \tag{24}$$

achieved for the optimized step-size sequence $\gamma_n = 2/(Bn)$. 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 that case, the resulting bounds are also probabilistic and, hence, more complicated to present.

---

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

C. Practical implementation

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

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

b) 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 oscillatory behavior around the algorithm's end-state. In the deterministic regime, these oscillations can be dampened by using forward-backward splitting or accelerated descent methods [35]. 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.

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 $\mathbf{Y}_i \mapsto \exp(\mathbf{Y}_i)$. Since matrix exponentiation has the same complexity as matrix multiplication, this step has polynomial complexity with a low degree on the input dimension of $\mathbf{X}_i$ [36]. In particular, each exponentiation requires $\mathcal{O}(M_i^\omega)$ 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 [37].

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 $\mathbf{Y}_i$ 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_i denotes the largest eigenvalue of $\mathbf{Y}_i$ and we let $\mathbf{Y}_i' = \mathbf{Y}_i - y_i\mathbf{I}$, the algorithm's exponentiation step can be rewritten as:

$$\mathbf{X}_i \leftarrow \frac{\exp(\mathbf{Y}_i')}{\exp(-y_i) + \|\exp(\mathbf{Y}_i')\|}. \tag{25}$$

Thanks to this shift, the elements of the numerator are now bounded from above by 1 (because the largest diagonal element of $\mathbf{Y}_i'$ is $e^{y_i - y_i} = 1$), 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 (25) in all numerical implementations of Algorithm 1.

   d) Asynchronous implementation: An implicit assumption in (MXL) is that user updates – albeit local – are concurrent. This can often 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. Thus, to overcome this synchronicity limitation, we discuss below an asynchronous variant of Algorithm 1 where each player updates his action based on an individual, independent schedule.

   Specifically, assume that each player $i \in \mathcal{N}$ has an individual timer that triggers an update event, i.e. a request for gradient feedback and, subsequently, an update of $\mathbf{X}_i$. Of course, the players' gradient estimates $\hat{\mathbf{V}}_i$ 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 $\mathcal{N}_n \subseteq \mathcal{N}$ be the set of players that update their actions at the n-th overall update event (typically $|\mathcal{N}_n| = 1$ if players update at random times), and let $d_i(n)$ be the corresponding number of epochs that have elapsed since the last update of player i. We then obtain the following asynchronous variant of (MXL):

$$
\begin{aligned}
\mathbf{Y}_i(n + 1) &= \mathbf{Y}_i(n) + \gamma_{n_i} \, \mathbb{1}\{i \in \mathcal{N}_n\} \cdot \hat{\mathbf{V}}_i(n), \\
\mathbf{X}_i(n + 1) &= A_i \frac{\exp(\mathbf{Y}_i(n + 1))}{1 + \|\exp(\mathbf{Y}_i(n + 1))\|},
\end{aligned}
\tag{26}
$$

where n_i denotes the number of updates performed by player i up to epoch n while the (asynchronous) estimate $\hat{\mathbf{V}}_i(n)$ satisfies

$$
\mathbb{E}[\hat{\mathbf{V}}_i(n)] = \mathbf{V}_i(\mathbf{X}_1(n - d_1(n)), \dots, \mathbf{X}_N(n - d_N(n))).
\tag{27}
$$

Assuming that the delays $d_i(n)$ are bounded and the players' update rates are strictly positive (in the sense that $\liminf_{n \to \infty} n_i/n > 0$), it is possible to employ the analysis of [38] to show that Theorems 2–4 remain true under the asynchronous variant (26). However, a detailed analysis would take us too far afield, so we relegate it to future work.

VI. Numerical Results

   In this section, we assess the performance of the MXL algorithm via numerical simulations. For concreteness, we focus on the case of transmit 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.

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	$P_c = 20$ dBm
Transmit antennas per device	$M = 4$
Receive antennas per link	$N = 8$

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 [39, 40] 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 $M = 4$ 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 $N = 25$ users, each located at a different cell of a 5×5 square cell cluster and sharing $S = 8$ common subcarriers. Finally, at the receiver end, we consider $N = 8$ receive antennas per connection and a receiver
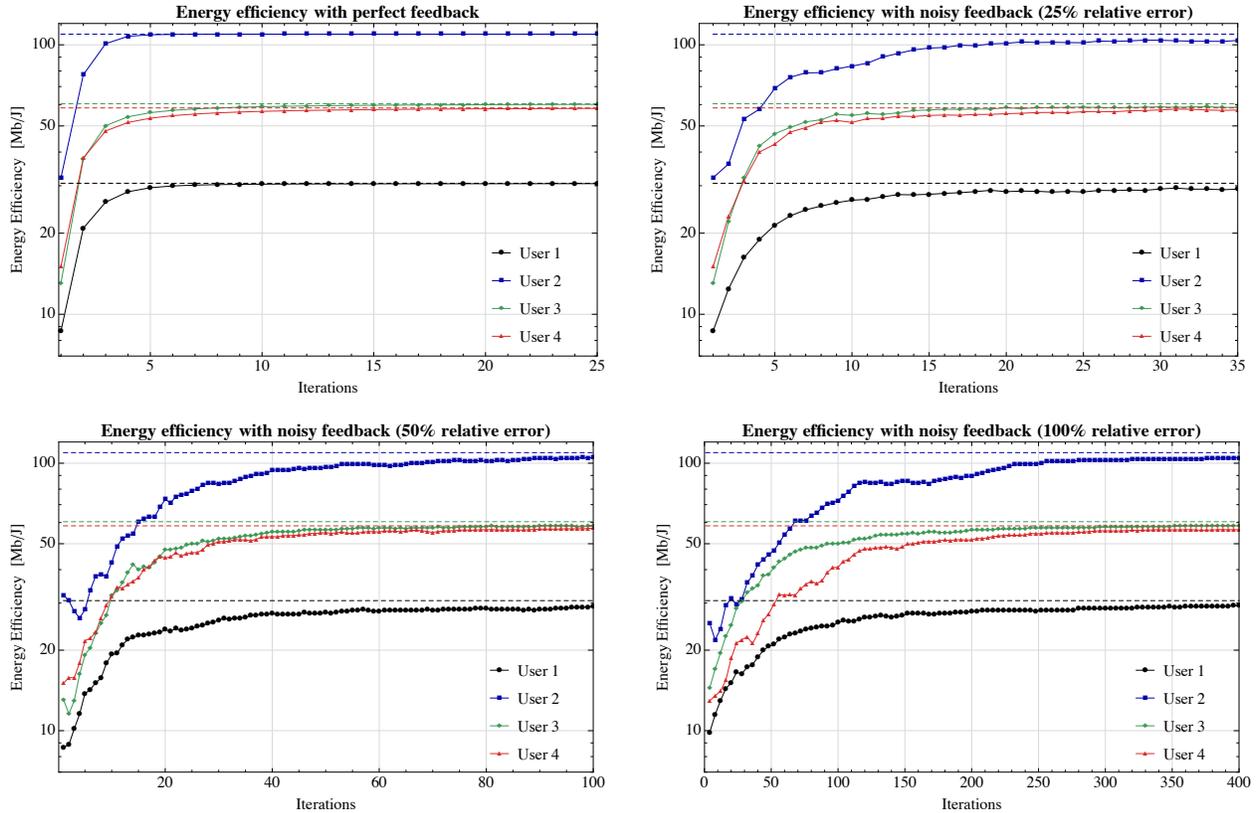
Fig. 2. Performance of MXL in the presence of noise. In all figures, we plot the transmit energy efficiency of wireless users that employ Algorithm 1 in a wireless network with parameters as described in the main text (to reduce graphical clutter, we only plotted 4 users with diverse channel characteristics). 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).

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, Fig. 2(a) assumes that users have perfect CSI measurements at their disposal. In this deterministic regime, the algorithm converges to a stable Nash equilibrium state within a few iterations (for simplicity, we only plotted 4 users with diverse channel characteristics). In turn, this rapid convergence leads to drastic gains in energy efficiency, ranging between 3× and 6× over uniform power allocation schemes.

(a) Channel gain evolution for different user velocities

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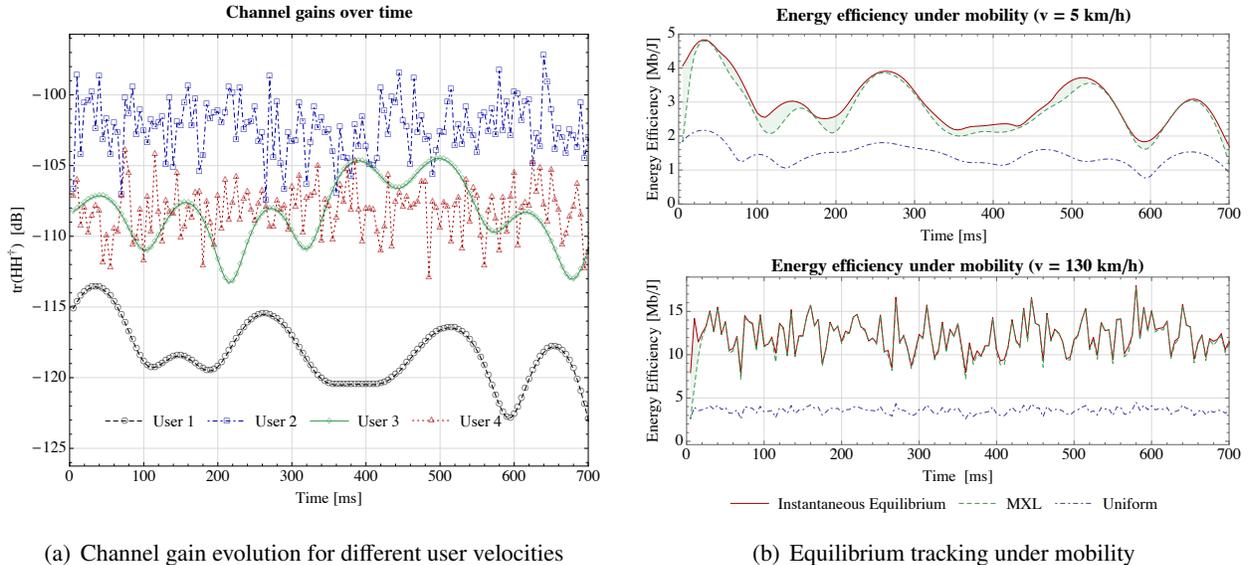
Fig. 3. Performance of MXL in a dynamic network environment with mobile users moving at $v \in \{3, 5, 30, 130\}$ km/h. The users' achieved energy efficiency tracks the system's (evolving) equilibrium remarkably well, even under rapidly changing channel conditions.

Subsequently, the simulation cycle above was repeated in the presence of observation noise and measurement errors. The intensity of the measurement noise was quantified via the relative error level of the gradient observations $\hat{\mathbf{V}}$, i.e. the standard deviation of $\hat{\mathbf{V}}$ divided by its mean (so a relative error level of $z\%$ means that, on average, the observed matrix $\hat{\mathbf{V}}$ lies within $z\%$ of its true value). We then plotted the users' 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 $z = 100\%$).

Finally, to assess the algorithm's performance in a fully dynamic network environment, Fig. 3 focuses on mobile users with 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 [41].

In Fig. 3(a), we plotted the channel gains ($\mathrm{tr}[\mathbf{HH}^{\dagger}]$) 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 [42]. Because urban environments are not homogeneous and/or isotropic (even on average), uniform power allocation fails to adapt to the changing wireless landscape and performs consistently worse than MXL (achieving an energy efficiency ratio between 2× and 6× lower than that of MXL).

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 and wireless medium access control). 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 projection" 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 $\mathcal{O}(1/\varepsilon)$ 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 $A_i = 1$; 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 [31]:

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

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

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

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:*

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

and

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

Proof: Since the von Neumann entropy is strictly convex [31] and becomes infinitely steep at the boundary of $\mathcal{X}$, 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:

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

Solving for $\mathbf{X}$ then yields

$$\mathbf{X} = \frac{\exp(\mathbf{Y})}{1 + \text{tr}[\exp(\mathbf{Y})]}, \tag{A.6}$$

and our claim follows by substituting the above 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 [32], we introduce here the *Fenchel coupling*:

$$F(\mathbf{X}, \mathbf{Y}) = h(\mathbf{X}) + h^*(\mathbf{Y}) - \text{tr}[\mathbf{YX}] = D_{\text{KL}}(\mathbf{X}, \mathbf{G}(\mathbf{Y})). \tag{A.7}$$

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*

$$F(\mathbf{X}, \mathbf{Y} + \mathbf{Z}) \leq F(\mathbf{X}, \mathbf{Y}) + \text{tr}[\mathbf{Z}(\mathbf{G}(\mathbf{Y}) - \mathbf{X})] + \|\mathbf{Z}\|_\infty^2. \tag{A.8}$$

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

$$F(\mathbf{X}, \mathbf{Y} + \mathbf{Z}) = h(\mathbf{X}) + h^*(\mathbf{Y} + \mathbf{Z}) - \text{tr}[(\mathbf{Y} + \mathbf{Z})\mathbf{X}]$$

$$\leq h(\mathbf{X}) + h^*(\mathbf{Y}) + \text{tr}[\mathbf{Z}\ \mathbf{G}(\mathbf{Y})] + \|\mathbf{Z}\|_\infty^2$$

$$- \text{tr}[\mathbf{YX}] - \text{tr}[\mathbf{ZX}]$$

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

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 prove 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 $\text{tr}[(\mathbf{X}_i' - \mathbf{X}_i^*)\mathbf{V}_i^*] > 0$ for some player $i \in \mathcal{N}$ and some $\mathbf{X}' \in \mathcal{X}_i$. Therefore, by continuity, there exists some $a > 0$ such that

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

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

Since $\mathbf{X}(n) \to \mathbf{X}^*$ as $n \to \infty$, we may assume that $\mathbf{X}(n) \in U$ for all n (note that the step-size condition $\sum_{n=1}^\infty \gamma_n^2 < \sum_{n=1}^\infty \gamma_n = \infty$ is not affected if we start the sequence at some finite $n_0 > 0$). The recursion (MXL) then yields:

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

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

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

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 [43, 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 [44, Theorem 14] applied to the weight sequence $w_{j,n} = \gamma_j/t_n$. 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 $h_i(\mathbf{X}_i) = \operatorname{tr}[\mathbf{X}_i \log \mathbf{X}_i] + (1 - \operatorname{tr}(\mathbf{X}_i)) \log(1 - \operatorname{tr}(\mathbf{X}_i))$. Since $\mathbf{X}_i(n) = \nabla h_i^*(\mathbf{Y}_i(n))$ by Proposition A.1, we will also have $\nabla h_i(\mathbf{X}_i(n)) = \mathbf{Y}_i(0) + t_n \bar{\mathbf{V}}_i(n)$ by the general theory of convex conjugation. In turn, this implies that

$$h_i(\mathbf{X}_i') - h_i(\mathbf{X}_i(n)) \geq \operatorname{tr}[(\mathbf{Y}_i(0) + t_n \bar{\mathbf{V}}_i(n))(\mathbf{X}_i' - \mathbf{X}_i(n))], \tag{B.4}$$

by the convexity of h_i. However, since $\lim_{n\to\infty} \bar{\mathbf{V}}_k(n) = \mathbf{V}_i^*$ and $\lim_{n\to\infty} t_n = \infty$, Eq. (B.1) yields

$$h_i(\mathbf{X}_i') - h_i(\mathbf{X}_i(n)) \gtrsim at_n, \tag{B.5}$$

so $h_i(\mathbf{X}_i') - h_i(\mathbf{X}_i(n)) \to \infty$ as $n \to \infty$, a contradiction. $\blacksquare$

Appendix C

Global Convergence

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

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

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

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 $H(t) = F(\mathbf{X}^*, \mathbf{Y}(t))$. A simple differentiation then yields

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

i.e. $\dot{H} \leq 0$ 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 $c \geq 0$ as $t \to \infty$. Hence, by compactness, there exists some $\hat{\mathbf{X}} \in \mathcal{X}$ and a sequence $t_n \nearrow \infty$ such that $\mathbf{X}(t_n) \to \hat{\mathbf{X}}$ as $n \to \infty$.

Assume now ad absurdum that $\hat{\mathbf{X}} \neq \mathbf{X}^*$, so there exists some $a > 0$ and a neighborhood U of $\hat{\mathbf{X}}$ such that $\operatorname{tr}[(\mathbf{X} - \mathbf{X}^*) \mathbf{V}(\mathbf{X}^*)] \leq -a$ for all $\mathbf{X} \in U$; furthermore, 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 for all $n \geq 0$. In that case however, (C.1) yields:

$$\lim_{t \to \infty} H(t) \leq H(0) + \sum_{n=1}^{\infty} \int_{t_n}^{t_n+\delta} \text{tr}[(\mathbf{X}(t) - \mathbf{X}^*) \mathbf{V}(\mathbf{X}(t))] \, dt$$

$$\leq H(0) - \sum_{n=1}^{\infty} a\delta = -\infty, \tag{C.2}$$

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:

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

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 $X(n)$ is an *asymptotic pseudotrajectory* of (MXL_c), i.e. the iterates of (MXL) are asymptotically close to solution segments of (MXL_c) of arbitrary length – for a precise statement, see [33, Sec. 3].

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

$$D_{n+1} = F(\mathbf{X}^*, \mathbf{Y}(n) + \gamma_n \hat{\mathbf{V}}(n))$$

$$\leq D_n + \gamma_n v_n + \gamma_n \xi_n + \gamma_n^2 \|\hat{\mathbf{V}}(n)\|_\infty^2, \tag{C.4}$$

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

$$D_{n+1} \leq D_0 - t_n \left(a - \sum_{j=1}^n w_{j,n} \xi_j\right) + \sum_{j=1}^n \gamma_j^2 \|\hat{\mathbf{V}}(j)\|_\infty^2, \tag{C.5}$$

where $t_n = \sum_{j=1}^n \gamma_j$ and $w_{j,n} = \gamma_j/t_n$. By the strong law of large numbers for martingale differences [43, Theorem 2.18], we have $n^{-1} \sum_{j=1}^n \xi_j \to 0$ (a.s.); hence, given that $\gamma_{n+1}/\gamma_n \leq 1$, Hardy's summability criterion [44, Thm. 14] applied to the sequence $w_{j,n} = \gamma_j/t_n$ yields $\sum_{j=1}^n w_{j,n} \xi_j \to 0$ (a.s.). Finally, since γ_n is square-summable and $\gamma_n \mathbf{Z}(n)$ is a martingale difference with finite variance, Theorem 6 in [45] shows that $\sum_{n=1}^{\infty} \gamma_n^2 \|\hat{\mathbf{V}}(n)\|_\infty^2 < \infty$ (a.s.).

Since $t_n \to \infty$ by assumption, the above implies that the RHS of (C.5) tends to $-\infty$ (a.s.); this contradicts the fact that $D_n \geq 0$, so we conclude that $\mathbf{X}(n)$ visits a compact neighborhood of $\mathbf{X}^*$ infinitely often (viz. there exists a sequence $n_k \nearrow \infty$ such that $\mathbf{X}(n_k)$ lies in said neighborhood). Since $\mathbf{X}^*$ attracts any initial

condition $\mathbf{X}(0) = \mathbf{G}(\mathbf{Y}(0))$ under the continuous-time dynamics (MXL$_c$), Theorem 6.10 in [33] then shows that $\mathbf{X}(n)$ converges to $\mathbf{X}^*$ (a.s.), as claimed. ∎

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 $F(\mathbf{X}^*, \mathbf{Y}) < 4m$ (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 $U_{4m} = \{\mathbf{Y} : F(\mathbf{X}^*, \mathbf{Y}) \leq 4m\}$ is invariant under (MXL$_c$). Hence, by shadowing the proof of Theorem 3, it suffices to show that there exists an open set $U_\varepsilon \subseteq U_{4m}$ such that $\mathbb{P}(\mathbf{Y}(n) \in U_{4m} \text{ for all } n) \geq 1 - \varepsilon$ whenever $\mathbf{Y}(0) \in U_\varepsilon$.

To that end, let $D_n = F(\mathbf{X}^*, \mathbf{Y}(n))$ and assume that $F(\mathbf{X}^*, \mathbf{Y}(0)) \leq m$. Then, (C.4) yields

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

$$\leq m + \sum_{j=1}^{n} \gamma_j v_j + \sum_{j=1}^{n} \gamma_j \xi_j + \sum_{j=1}^{n} \gamma_j^2 \left(L^2 + K \|\mathbf{Z}(j)\|_\infty^2 \right), \tag{D.1}$$

where $v_j = \mathrm{tr}[(\mathbf{X}(j) - \mathbf{X}^*)\mathbf{V}(j)]$, $\xi_j = \mathrm{tr}[(\mathbf{X}(j) - \mathbf{X}^*)\mathbf{Z}(j)]$, K is a sufficiently large positive constant, and $L = \sup_{\mathbf{X} \in \mathcal{X}} \|\mathbf{V}(\mathbf{X})\|_\infty$. By the subexponential moment growth condition (H2′) and the martingale concentration inequalities of [34, Theorem 1.2A], it follows that $\mathbb{P}(\sum_{j=1}^{n} \gamma_j \xi_j \geq m) \leq \exp\left(-\frac{m^2}{2\sigma_*^2(\sum_{j=1}^{n} \gamma_j^2 + \gamma_1 m/\sigma_*)}\right) \leq \varepsilon$ if γ is chosen sufficiently small (specifically, so that $\gamma_1 m/\sigma_* + \sum_{j=1}^{\infty} \gamma_j^2 \leq m^2 |\log \varepsilon|^{-1}$). Likewise, (H2′) implies that the tails of $\mathbf{Z}$ are subexponential, so there exists some constant $A > 0$ such that $\mathbb{P}(\|\mathbf{Z}(j)\|_\infty \geq m/\gamma_j) \leq A \exp(-m/\gamma_j)$. By the summability assumption for γ_j and the Borel-Cantelli lemma, it follows that $\mathbb{P}(\|\mathbf{Z}(j)\|_\infty \geq m/\gamma_j \text{ for infinitely many } j) = 0$, so the last term of (D.1) is bounded from above by m (a.s.) if γ is chosen small enough. Finally, by choosing γ sufficiently small so that $\sum_{j=1}^{\infty} \gamma_j^2 \leq m/L^2$, the third term of (D.1) is also bounded by m.

Combining all of the above, we obtain $D_n \leq 4m + \sum_{j=1}^{n} \gamma_j v_j$ with probability at least $1 - \varepsilon$. Since $\mathbf{G}(U_{4m}) \subseteq U$ by construction, we have $v_j \leq 0$ for all $j = 1, \ldots, n$, and we conclude that $D_n \leq 4m$ with probability at least $1 - \varepsilon$. This implies that $\mathbb{P}(\mathbf{Y}(n) \in U_{4m} \text{ for all } n) \leq 1 - \varepsilon$, 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 $\bar{D}_n = \mathbb{E}[F(\mathbf{X}^*, \mathbf{Y}(n))]$. Then, taking expectations in (C.4) yields

$$\bar{D}_{n+1} \le \bar{D}_n + \gamma_n \, \mathbb{E}[\mathrm{tr}[(\mathbf{X}(n) - \mathbf{X}^*) \, \mathbf{V}(n)]] + \gamma_n^2 \, \mathbb{E}[\|\hat{\mathbf{V}}(n)\|_\infty^2]$$

$$\le (1 - \gamma_n B)\bar{D}_n + \gamma_n^2 V^2, \tag{E.1}$$

where, in the second line, we used (18b) and the assumption that $\mathbf{X}^*$ is B-strongly stable. With this in mind, assume inductively that $\bar{D}_n \le A/n$ for some $A > 0$; it will then suffice to show that $\bar{D}_{n+1} \le A/(n + 1)$ and that we can choose $A = \gamma^2 V^2/(B\gamma - 1)$. Therefore, substituting in (E.1), it suffices to show that

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

Rearranging this last equation, we get

$$nA \le (n + 1)(AB\gamma - \gamma^2 V^2), \tag{E.3}$$

which shows that it suffices to pick $A = AB\gamma - \gamma^2 V^2$. Solving for A then yields $A = \gamma^2 V^2/(B\gamma - 1)$, as claimed.

Finally, the bound (23b) follows by noting that $D_{\mathrm{KL}}(\mathbf{X}^*, \mathbf{X}) = \Omega(\|\mathbf{X}^* - \mathbf{X}\|)$ if $\mathbf{X}^*$ is an extreme point of $\mathcal{X}$; likewise, the bound (23a) follows by noting that $D_{\mathrm{KL}}(\mathbf{X}^*, \mathbf{X}) = \Omega(\|\mathbf{X}^* - \mathbf{X}\|^2)$ if $\mathbf{X}^*$ is interior, and subsequently applying Jensen's inequality. ∎

REFERENCES

[1] 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.

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

[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, "Primal-dual subgradient methods for convex problems," *Mathematical Programming*, vol. 120, no. 1, pp. 221–259, 2009.

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

[14] 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.

[15] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, 2004.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

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

[21] 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.

[22] 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.

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

[24] 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.

[25] 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.

[26] 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.

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

[28] 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.

[29] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, ser. Springer Series

in Operations Research.   Springer, 2003.

[30] J. Dattorro, *Convex Optimization & Euclidean Distance Geometry*.   Palo Alto, CA, USA: Meboo Publishing, 2005.

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

[32] P. Mertikopoulos and W. H. Sandholm, "Learning in games via reinforcement and regularization," *Mathematics of Operations Research*, 2016, to appear.

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

[36] C. Moler and C. van Loan, "Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later," *SIAM Review*, vol. 45, no. 1, pp. 3–49, 2003.

[37] 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.

[38] 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.

[39] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Veh. Technol.*, vol. 29, no. 3, pp. 317–325, August 1980.

[40] COST Action 231, "Digital mobile radio towards future generation systems," European Commission, final report, 1999.

[41] 3GPP, "User equipment (UE) radio transmission and reception," White paper, Jun. 2014.

[42] 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.

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

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

[45] Y. S. Chow, "Convergence of sums of squares of martingale differences," *The Annals of Mathematical Statistics*, vol. 39, no. 1, 1968.