Distributed stochastic optimization via matrix exponential learning
Panayotis Mertikopoulos, Elena Belmega, Romain Negrel, Luca Sanguinetti

To cite this version:
Panayotis Mertikopoulos, Elena Belmega, Romain Negrel, Luca Sanguinetti. Distributed stochastic optimization via matrix exponential learning. IEEE Transactions on Signal Processing, Institute of Electrical and Electronics Engineers, 2017, 65 (9), pp.2277-2290. 10.1109/tsp.2017.2656847. hal-01382285

HAL Id: hal-01382285
https://hal.archives-ouvertes.fr/hal-01382285
Submitted on 23 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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) \( \mathcal{K} = \{1, \ldots, K\} \), each controlling a matrix variable \( X_k \), and seeking to improve their individual “well-being”. Assuming that this “well-being” is quantified by a utility (or payoff) function \( u_k(X_1, \ldots, X_K) \), we obtain the problem

\[
\text{maximize } u_k(X_1, \ldots, X_K), \quad \text{subject to } X_k \in \mathcal{X}_k,
\]

(1)

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

\[
\mathcal{X}_k = \{ X_k \geq 0 : \| X_k \| \leq A_k \}
\]

(2)

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

This research was supported by the European Research 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 ASI, 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), CentralaSupélec, 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 desidera:

(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 \( X_k \)) 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. 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 $\varepsilon$ of a strongly stable equilibrium in $O(1/\varepsilon^2)$ 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 $X = (X_1, \ldots, X_K)$ is identified with the block-diagonal matrix $\text{diag}(X_k)_{k=1}^K$, and we use the shorthand $(X_k; X_{\neq k})$ to highlight the action $X_k$ of player $k$ against that of all other players. Also, given $A \in \mathbb{C}^{d\times d}$, we write $||A|| = \sum_{j=1}^d |\text{eig}_j(A)|$ for the nuclear (trace) norm of $A$ and $||A||_\infty = \max_j |\text{eig}_j(A)|$ 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 $\mathcal{K} = \{1, \ldots, K\}$, 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 $q_k(x_{\neq k})$, i.e. a (symmetric) function of the access probability profile $x_{\neq k} = (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_K)$ of all other users.\footnote{For instance, a standard contention measure of this type is the conditional collision probability $q_k(x_{\neq k}) = 1 - \prod_{\ell \neq k}(1 - x_\ell)$. [14].}

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 accessing the channel more often minus the collision probability incurred by all other users. This leads to the utility function formulation

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

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

$$\begin{align*}
\text{for all } k \in \mathcal{K}, \quad & \max_{x_k \in [0, 1]} \left\{ \begin{array}{ll}
& u_k(x_k; x_{\neq k}), \\
& \text{subject to } x_k \in [0, 1]
\end{array} \right\}, \quad \text{subject to } x_k \in [0, 1] \\
& \text{subject to } x_k \in [0, 1]
\end{align*} \quad (4)$$

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 $A_k = 1$.\footnote{For instance, a standard contention measure of this type is the conditional collision probability $q_k(x_{\neq k}) = 1 - \prod_{\ell \neq k}(1 - x_\ell)$. [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 $i \in \mathbb{R}^d$ that collects and encodes the most distinctive features of said image. Given a database $D$ of such signatures, each image $i \in D$ is further associated with a set $S_i \subseteq D$ of similar images and a set $U_i \subseteq D$ 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

$$d_X(i, j) = (i - j)^\top X (i - j),$$

where $X \succ 0$ is a $d \times d$ positive-definite matrix. This so-called “precision matrix” must then be chosen by the optimizer so that

$$d_X(i, j) < d_X(i, k)$$

whenever $i$ is similar to $j$ but dissimilar to $k$.

We thus obtain the minimization objective

$$F(X; \mathcal{T}) = \sum_{i, j, k \in \mathcal{T}} C(d_X(i, j) - d_X(i, k) - \epsilon) + \|X - I\|_F^2,$$

where: $\mathcal{T}$ is the set of all triples $(i, j, k)$ such that $j \in S_i$ and $k \in U_i$; $C$ is a convex, nondecreasing function that penalizes matrices $X$ that do not capture similarities and dissimilarities between images; $\epsilon$ the parameter $\epsilon > 0$ reinforces this penalty; and $d \|X\|_F$ denotes the ordinary Frobenius ($\ell^2$) norm.

An additional requirement in the above is to employ a low-rank precision matrix $X$ 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 $\text{tr}(X) \leq c$ for some $c \ll d^4$, leading to the feasible region

$$\mathcal{X} = \{X \in \mathbb{R}^{d \times d} : X \succ 0 \text{ and tr}(X) \leq c\}.$$ (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 $D$ is large, the number of variables involved becomes computationally prohibitive: for instance, (6) may contain up to $10^6$–$10^{11}$ 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]:

$$\text{minimize } \mathbb{E}[F(X; \mathcal{W})]$$

subject to $X \succ 0$, tr$(X) \leq c$, (8)

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

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

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

4By contrast, an $\ell_2$ rank constraint of the form $\text{rank}(X) \leq c$ generically leads to an untractable NP-hard problem formulation.

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 $\nabla_X F(X; \mathcal{W})$. This type of uncertainty is typical of stochastic optimization problems, and the proposed MVL 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 $N_i$ (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 $Q_k$, over each subcarrier $s = 1, \ldots, S$, subject to the constraints: a) $Q_{ks} \succ 0$ (since each $Q_{ks}$ is a signal covariance matrix); and b) $\text{tr}(Q_k) \leq P_{\text{max}}$, where $Q_k = \text{diag}(Q_{ks})_{s=1}^S$ is the covariance profile of the $k$-th transmitter, $\text{tr}(Q_k)$ represents the user’s transmit power over all subcarriers, and $P_{\text{max}}$ 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

$$R_k(Q) = \log \det(W_{-k} + H_{kk} Q_k H_{kk}^\dagger) - \log \det(W_{-k}),$$

where $H_{kk}$ denotes the channel matrix profile between the $k$-th transmitter and the $k$-th receiver over all subcarriers, $Q = (Q_1, \ldots, Q_K)$ denotes the users’ transmit profile and $W_{-k} = W_{-k}(Q) = I + \sum_{s \neq k} H_{ts} Q_k H_{tk}^\dagger$ 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.

$$\text{EE}_k(Q) = \frac{R_k(Q)}{P_c + \text{tr}(Q_k)},$$

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 $Q_k$ of user $k$, but it can be recast as such via a suitable transformation. To this aim, following [24, 25],

Fig. 1. A multi-user MIMO system with $K = 3$ connections.
consider the adjusted control variables

\[ X_k = \frac{P_c + P_{\text{max}}}{P_{\text{max}}} \frac{Q_k}{P_{\text{max}}} \left( P_c + \text{tr}(Q_k) \right)^{-1} \]  

(11)

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

\[ \mathcal{X}_k = \{ \text{diag}(X_{ki})_{i=1}^K : X_{ki} \succeq 0 \text{ and } \sum_i \text{tr}(X_{ki}) \leq 1 \} \]  

(12)

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

\[ u_k(X_k; X_{-k}) = \frac{P_c + (1 - \text{tr}(X_k)) P_{\text{max}}}{P_c(P_c + P_{\text{max}})} \times \log \det \left( I + \frac{P_c P_{\text{max}} \bar{H}_k X_k \bar{H}_k^\dagger}{P_c + (1 - \text{tr}(X_k)) P_{\text{max}}} \right) \]  

(13)

where \(\bar{H}_k = W^{-1/2}_k H_{kk}\) 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 \(X^* \in \mathcal{X}\) which is unilaterally stable, i.e.

\[ u_k(X^*; X_{-k}) \geq u_k(X_k; X_{-k}) \quad \text{for all } X_k \in \mathcal{X}_k, k \in K. \]  

(14)

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

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 \(\mathcal{X}_k\), 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

\[ V_k(X) \equiv \nabla_{X_k} u_k(X_k; X_{-k}) \]  

(15)

and let \(V(X) \equiv \text{diag}(V_1(X), \ldots, V_K(X))\) denote the collective profile of all players’ individual gradients. Then, Rosen’s monotonicity condition can be stated as:

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

\[ \text{(DSC)} \]

with equality if and only if \(X = X'\). 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 \(X^* \in \mathcal{X}\) is called *stable* if it satisfies the variational stability condition:

\[ \text{tr}[(X - X^*)(V(X))] \leq 0 \quad \text{for all } X \text{ sufficiently close to } X^*. \]  

(VS)

In particular, if (VS) holds for all \(X \in \mathcal{X}\), we say that \(X^*\) 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 \(X^* \in \mathcal{X}\) is stable, then it is an isolated Nash equilibrium; specifically, if \(X^*\) is globally stable, it is the game’s unique Nash equilibrium.*

**Proof:** Suppose \(X^*\) is stable, pick some \(X_k\) close to \(X_k^*\), and let \(X = (X_k; X_{-k})\). Then, by (VS), we get \(\text{tr}[(X_k - X_k^*)V(X_k; X_{-k})] < 0\), implying that \(u_k\) is decreasing along the ray \(X_k^* + t(X_k - X_k^*)\). Since this covers all rays starting at \(X_k^*\), we conclude that \(X^*\) is the game’s unique equilibrium in the neighborhood of \(X^*\) 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 \(D(X) = (D_{kl}(X))_{k,l \in K}\) with blocks

\[ D_{kl}(X) = \frac{1}{2} \nabla_{X_k} \nabla_{X_k} u_k(X) + \frac{1}{2} [\nabla_{X_k} \nabla_{X_k} u_k(X)]^\top. \]  

(16)

The terminology “Hessian” reflects the fact that when (1) is a single-agent optimization problem \((K = 1)\), \(D(X)\) 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:

\(^6\)Simply note that Nash equilibria are solutions of the variational inequality \(\text{tr}[(X - X^*)(V(X))] \leq 0\) [13]. Then, (VS) follows by setting \(X^* = X^*\) in (DSC).
Proposition 2. If \( D(X) < 0 \) for all \( X \in \mathcal{X} \), \( G \) admits a unique, globally stable Nash equilibrium. More generally, if \( X^* \) is a Nash equilibrium and \( D(X^*) < 0 \), \( X^* \) is stable and isolated.

Proof: For the first statement, assume that \( D(X) < 0 \) for all \( X \in \mathcal{X} \) and note that \( D(X) \) 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 \( X^* \) by Theorem 1. Since (DSC) further implies (VS) for all \( X \in \mathcal{X} \), our claim follows. For our second assertion, if \( X^* \) is a Nash equilibrium and \( D(X^*) < 0 \), we have \( D(X) < 0 \) for all \( X \) in some convex product neighborhood \( \mathcal{U} \) of \( X^* \) (by continuity). Then, by restricting \( G \) to \( \mathcal{U} \) and reasoning as in the global case above, we conclude that \( X^* \) is the game’s unique equilibrium in \( \mathcal{U} \), i.e. \( X^* \) 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 \( n = 1, 2, \ldots \) of the process, each player \( k \in K \) takes a step along the individual gradient of their utility function (possibly subject to noise). The output is then “mirrored” back to \( K \) via a suitable matrix exponentiation map, and the process repeats.

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

\[
X_k(n) = A_k \frac{\exp(Y_k(n-1))}{1 + \|\exp(Y_k(n-1))\|^\gamma},
\]

\[
Y_k(n) = Y_k(n-1) + \gamma(n) \hat{V}_k(n),
\]

(Algorithm 1

\[\text{Matrix exponential learning (MXL).} \]

Parameter: step-size sequence \( \gamma(n) \sim 1/n^2, \beta \in (0, 1] \).

Initialization: \( n \leftarrow 0; \ Y_k \leftarrow \text{any } M_k \times M_k \text{ Hermitian matrix.} \)

Repeat

\[
\begin{aligned}
& n \leftarrow n + 1; \\
& \text{foreach player } k \in K \text{ do} \\
& \quad \text{play } X_k \leftarrow A_k \frac{\exp(Y_k)}{1 + \|\exp(Y_k)\|}; \\
& \quad \text{get gradient feedback } \hat{V}_k; \\
& \quad \text{update auxiliary matrix } Y_k \leftarrow Y_k + \gamma(n) \hat{V}_k;
\end{aligned}
\]

until termination criterion is reached.

Algorithm 1) The players’ utility functions are themselves stochastic expectations of the form \( u_k(X) = E[u_k(X; \omega)] \) for some random variable \( \omega \), and the players can only observe the (stochastic) gradient of \( u_k \) (cf. Ex. II-B).

iii) Any combination of the above.

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

\[\hat{V}_k(n) = V_k(X(n)) + Z_k(n), \quad (17)\]

where the stochastic noise process \( Z(n) \) satisfies the hypotheses:

\[E[Z(n) \mid F_{n-1}] = 0. \quad (H1)\]

\[E[\|Z(n)\|^2 \mid F_{n-1}] \leq \sigma^2 \quad \text{for some } \sigma > 0. \quad (H2)\]

In the above, \( F_n \) denotes the history (natural stochastic basis) of \( Z(n) \) 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 \( \hat{V}_k \) are unbiased and...
bounded in mean square, i.e.
\[
\begin{align}
E[\hat{V}_k(n) | \mathcal{F}_{n-1}] &= V_k(X(n)), \tag{18a} \\
E[|\hat{V}_k(n)|^2]_{\mathcal{F}_{n-1}} &\leq V_k^2 \quad \text{for some } V_k > 0, \tag{18b}
\end{align}
\]
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 \( \gamma_n \) 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, \( \lim_{n \to \infty} X(n) \) 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 \( \gamma_n \) such that \( \sum_{n=1}^{\infty} \gamma_n^2 < \sum_{n=1}^{\infty} \gamma_n = \infty \) and gradient observations satisfying (H1) and (H2). If \( X^* \) is globally stable, then \( X(n) \) converges to \( X^* \) (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 \( D(X) \) is negative-definite for all \( X \in \mathcal{X} \), 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). 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 \( \gamma_n \) with \( \sum_{n=1}^{\infty} \gamma_n^2 < \sum_{n=1}^{\infty} \gamma_n = \infty \). If \( X^* \) is (locally) stable, then it is locally attracting with arbitrarily high probability: specifically, for every \( \varepsilon > 0 \), there exists a neighborhood \( U_\varepsilon \) of \( X^* \) such that
\[
P(X(n) \to X^* \mid X(1) \in U_\varepsilon) \geq 1 - \varepsilon. \tag{19}
\]

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 \( X^* \), the drift of (MXL) does not point towards \( X^* \) globally, so \( X(n) \) may exit the basin of attraction of \( X^* \) 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 \( X^* \) be a Nash equilibrium of \( \mathcal{G} \) such that \( D(X^*) \) < 0. Then (MXL) converges locally to \( X^* \) 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:
\[
D_{\text{KL}}(X^*,X) = \text{tr}[X^*(\log X^* - \log X)]. \tag{20}
\]
By Klein’s inequality [30], \( D_{KL}(X^*, X) \geq 0 \) with equality if and only if \( X = X^* \), so \( D_{KL}(X^*, X) \) represents a (convex) measure of the distance between \( X \) and \( X^* \). With this in mind, we introduce below the following quantitative measure of stability:

**Definition 3.** Given \( B > 0, X^* \) is called \( B \)-strongly stable if

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

Under this refinement of equilibrium stability, 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 \( X^* \) is \( B \)-strongly stable and \( 1 < B y \leq 2 \), we have

\[
E[D_{KL}(X^*, X(n))] \leq 9\gamma^2 / 2B^2n, \tag{24}
\]

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 \( n \geq [B y] \). However, since the optimized step-size \( \gamma_n = 3/(2Bn) \) that achieves (24) is already covered by the condition \( B y \leq 2 \), 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 \( \gamma_n \):** Using a decreasing step-size \( \gamma_n \) 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 lead to an oscillatory behavior around the algorithm’s end-state. a constant step-size might cause the process to overshoot and

Remark. That being said, the \( \ell^2 - \ell^1 \) summability condition \( \sum \gamma_n \leq \infty \) required by Theorems 2–4 is closely linked to the finite mean squared error hypothesis (H2). In practice however, the feedback noise process \( Z(n) \) 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 \( \beta \leq 1/2 \).

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 \( k \in K \) has an individual timer that triggers an UpdateEvent, i.e. a request for gradient feedback and, subsequently, an update of \( X_k \). Of course, the players’ gradient estimates \( \hat{V}_k \) 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 \( K_n = [K]_n \) be the set of players that update their actions at the \( n \)-th overall UpdateEvent (typically \( |K_n| = 1 \) 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):

\[
X_k(n) = A_k \exp(Y_k(n - 1)) \tag{25}
\]

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

\[
E[\hat{V}_k(n) | F_{n-1}] = V_k(X_1(n - d_1(n)), \ldots, X_K(n - d_K(n))). \tag{26}
\]

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 \( Y_k \mapsto \exp(Y_k) \). Since matrix exponentiation has the
same complexity as matrix multiplication, this step has polynomial complexity with a low degree on the input dimension of $X_k$. In particular, each exponentiation requires $O(M_k^{2.373})$ 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 $Y_k$ are large, this step could incur an overflow wherever 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 $Y_k$ and we let $Y'_k = Y_k - y_k I$, the algorithm’s exponentiation step can be rewritten as:

$$X_k \leftarrow \frac{\exp(Y'_k)}{\exp(-y_k) + \|\exp(Y'_k)\|}. \quad (27)$$

Thanks to this shift, the elements of the numerator are now bounded from above by 1 (because the largest diagonal element of $Y'_k$ is $e^{y_k-y_k} = 1$), so there is no danger of encountering a numerical indeterminacy of the type $\text{Inf}/\text{Inf}$. 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 $X_k(n)$ satisfy the game’s feasibility constraints (2) at each stage $n$. In the noiseless case ($Z = 0$), feasibility is automatic because, by construction, $V_k$ (and, hence, $Y_k$) is Hermitian. That said, in the presence of noise, there is no reason to assume that the estimates $V_k$ are Hermitian, so the updates $X_k$ may also fail to be feasible. To rectify this, we tacitly assume that each player corrects such errors by replacing $V_k$ with $(V_k + V_k^\dagger)/2$ 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 1). 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 $20^\circ C$.

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 $P_{\text{max}} = 40$ 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 $P_0 = P_{\text{max}}/2 = 26$ 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 3x and 6x 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 $\tilde{\nabla}$, i.e. the standard deviation of $\tilde{\nabla}$ divided by its mean (so a relative error level of $\%$ means that, on average, the observed matrix $\tilde{\nabla}$ lies within $\%$ of its true value). We then plotted the users’ transmit energy efficiency over time for noise levels $\% = 25\%$, $50\%$, 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 $\% = 100\%$).

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

| 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 ($20^\circ C$) | $-174$ dBm/Hz |
| Maximum transmit power | $P_{\text{max}} = 40$ dBm |
| Non-radiative power consumption | $P_c = 20$ dBm |
| Tx/Rx antennas per link | 4/8 |
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 not homogeneous and 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 (trHHH) 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. In this setting, the system’s Shannon sum-rate is given by [41]:

\[
R_{\text{sys}}(Q_1, \ldots, Q_K) = \log \det \left( I + \sum_k H_k Q_k H_k^H \right),
\]

leading to the system-wide energy efficiency expression

\[
\text{EE}_{\text{sys}}(Q_1, \ldots, Q_K) = \frac{R_{\text{sys}}(Q_1, \ldots, Q_K)}{\sum_k (P_k + \text{tr}(Q_k))}.
\]

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 \( Z = Z(Q) \) that map the problem (29) to an equivalent concave problem with objective function \( u_{\text{sys}}(Z) \) 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 \( R_{\text{sys}} \) 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.
Equilibrium tracking under mobility

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 $\epsilon$-neighborhood of a Nash equilibrium (in terms of the Kullback–Leibler distance) within $O(1/\epsilon^2)$ 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.

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 $\epsilon$-neighborhood of a Nash equilibrium (in terms of the Kullback–Leibler distance) within $O(1/\epsilon^2)$ 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 EXPONENTIAL MAPPING 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]:

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

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

$$h^*(Y) = \max \{\text{tr}(XY) - h(X) : X \in D\}, \quad (A.2)$$

with $Y \in H^M$. 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:

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

and

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

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

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

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

$$X = (1 - \text{tr}(X))\exp(Y), \quad (A.6)$$

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

$$X = \frac{\exp(Y)}{1 + \text{tr}(\exp(Y))}, \quad (A.7)$$

and our claim follows by substituting in (A.2). \n
In addition to the above, the von Neumann entropy also provides a “congruence” measure between the primal variables $X$ and the auxiliary “dual” variables $Y$. Specifically, following [31], we introduce here the *Fenchel coupling*:

$$F(X, Y) = h(X) + h^*(Y) - \text{tr}(XY) = D_{\text{KL}}(X, G(Y)). \quad (A.8)$$

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

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

$$F(X, Y + Z) \leq F(X, Y) + \text{tr}(Z(G(Y)) - X) + \|Z\|_{\text{loc}}^2. \quad (A.9)$$

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

$$F(X, Y + Z) = h(X) + h^*(Y + Z) - \text{tr}(Y + Z)X \leq h(X) + h^*(Y) + \text{tr}(Z(G(Y) - X) + \|Z\|_{\text{loc}}^2$$

$$\leq \text{tr}(YX) - \text{tr}(ZX) \leq F(X, Y) + \text{tr}(Z(G(Y) - X)) + \|Z\|_{\text{loc}}^2. \quad (A.10)$$

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 $V^* = V(X^*)$ and assume that $X^*$ is not a Nash equilibrium. By Eq. (14), this implies that $\text{tr}([X_k^* - X_j^*]V_k^*) > 0$ for some players $k, j \in K$ and some $X^* \in \mathcal{X}_k$. Therefore, by continuity, there exists some $a > 0$ such that

$$\text{tr}([X_k^* - X_j^*]V_k^*) \geq a > 0, \quad (B.1)$$

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

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

$$Y(n) = Y(0) + \tau_n \tilde{V}(n), \quad (B.2)$$

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

$$\tilde{V}(n) = \frac{1}{\tau_n} \sum_{j=1}^{n} \gamma_j \tilde{V}(j) = \frac{1}{\tau_n} \sum_{j=1}^{n} \gamma_j V(X(j)) + \frac{1}{\tau_n} \sum_{j=1}^{n} \gamma_j Z(j). \quad (B.3)$$

(B.3) denotes the $\gamma$-weighted time average of the received gradient estimates $\tilde{V}(n)$. By the strong law of large numbers for martingale differences [35, Theorem 2.18], we get $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{n} \gamma_j Z(j) = 0 \text{ (a.s.)}$, 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 $w_{j,n} = \gamma_j/\tau_n$.\footnote{\textsuperscript{11}} Thus, given that $X(n) \in U$ for all $n$, we conclude that $\tilde{V}(n) \rightarrow V^*$ as $n \rightarrow \infty$.

Now, with notation as in Appendix A, let $h_k(X_k) = \text{tr}[X_k \log X_k] + (1 - \text{tr}(X_k)) \log(1 - \text{tr}(X_k))$. Since $X_k(n + 1) = \nabla h_k'(Y_k(n))$ by Prop. A.1, we will also have $\nabla h_k(X_k(n + 1)) = Y_k(n) + \tau_n \tilde{V}(n)$ by the general theory of convex conjugation. In turn, this implies that

$$h_k(X_k(n)) - h_k(X_k(n + 1)) \geq \text{tr}([Y_k(n) + \tau_n \tilde{V}(n)] - X_k(n + 1)), \quad (B.4)$$

by the convexity of $h_k$. However, since $\lim_{n \rightarrow \infty} \tilde{V}(n) = V^*$ and $\lim_{n \rightarrow \infty} \tau_n = \infty$, Eq. (B.1) yields $h_k(X_k(n)) - h_k(X_k(n + 1)) \geq a \tau_n$, so $h_k(X_k(n)) - h_k(X_k(n + 1)) \rightarrow \infty$ as $n \rightarrow \infty$, a contradiction. \hfill \blacksquare

**Appendix C**

**Global Convergence**

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

$$\dot{Y} = V(X), \quad (MXL_c)$$

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

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

\textsuperscript{11}If every player is using their individual step-size sequence $\gamma(n)$, 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 $X^*$ be a globally stable Nash equilibrium and let $X(t)$ be a solution of \((\text{MXL}_{\alpha})\). Then, $\lim_{t \to \infty} X(t) = X^*$. 

Proof: Let $H(t) = F(X^*, Y(t))$. Then
\[ H = \text{tr}[Y \nabla h^*(Y)] - \text{tr}[X \nabla^2 f(X) V(X)], \tag{C.1} \]
i.e. $H \leq 0$ with equality if and only if $X = X^*$ (recall that $X^*$ 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 $\tilde{X} \in \mathcal{X}$ and a sequence $t_n \uparrow \infty$ such that $X(t_n) \to \tilde{X}$ as $n \to \infty$.

Assume now that $\tilde{X} \neq X^*$, so there exists some $a > 0$ and a neighborhood $U$ of $\tilde{X}$ such that $\text{tr}[(X - X^*) V(X')] \leq -a$ for all $X \in U$. Since $\|X\|$ is bounded from above (recall that $G$ is Lipschitz), there exists some $\delta > 0$ such that $X(t) \in U$ for all $t \in [t_n, t_n + \delta]$ and all $n \geq 0$. In that case however, (C.1) yields:
\[ \lim_{t \to \infty} H(t) = H(0) - \sum_{n=1}^{\infty} \int_{t_n}^{t_n+\delta} \text{tr}[(X(t) - X^*) V(X(t))] dt \leq H(0) - a \delta = -\infty, \tag{C.2} \]
a contradiction. This shows that $X^*$ is the only possible $\omega$-limit point of $X(t)$; since $X(t)$ admits at least one $\omega$-limit, it follows that $X(t) \to X^*$, as claimed.

With this auxiliary result at hand, we finally are 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
\[ Y(n) = Y(n-1) + \gamma_n [V(G(Y(n-1))) + Z(n)]. \tag{C.3} \]
Since $V(X)$ is differentiable for almost all $X \in \mathcal{X}$ by Alexandrov’s theorem, Propositions 4.1 and 4.2 in [33] show that $X(n)$ is an asymptotic pseudotrajectory (APT) of (MXL$_{\alpha}$), i.e. the sequence of play generated by (MXL) are asymptotically close to solution segments of (MXL$_{\alpha}$) of arbitrary length for a precise statement, see [33, Sec. 3].

Assume now that $X(n)$ remains at a minimal positive distance from $X^*$. Since $X^*$ is globally stable, we will have $\text{tr}[\{X(n) - X^*\} V(X(n))] \leq -a$ for some $a > 0$ and for all $n$. Furthermore, if we let $D_n = F(X, Y(n-1))$, Proposition A.2 yields:
\[ D_{n+1} = F(X^*, Y(n-1) + \gamma_n [V(G(Y(n-1))) + Z(n)] \]
\[ \leq D_n + \gamma_n \nu + \gamma_n^2 \|V(n)\|_{\infty}, \tag{C.4} \]
where we have set $\nu = \text{tr}[(X(n) - X^*) V(X(n))]$ and $\xi_n = \text{tr}[\{X(n) - X^*\} Z(n)]$. Hence, telescoping (C.4) yields:
\[ D_n \leq D_1 - \tau_n \left( \gamma - \sum_{j=1}^{\infty} w_{jn} \xi_j + \sum_{j=1}^{\infty} \gamma_j^2 \|V(j)\|_{\infty} \right) \tag{C.5} \]
where $\tau_n =\sum_{j=1}^{\infty} \gamma_j$ and $w_{jn} = \gamma_j/\tau_n$. By the strong law of large numbers for martingale differences [35, Theorem 2.18], we have $\sum_{j=1}^{\infty} \xi_j = 0$ (a.s.); hence, given that $\gamma_{n+1}/\gamma_n \leq 1$, Hardy’s summability criterion [42, Thm. 14] applied to the sequence $w_{jn} = \gamma_j/\tau_n$ yields $\sum_{j=1}^{\infty} w_{jn} \xi_j \to 0$ (a.s.). Finally, since $E[\sum_{j=1}^{\infty} \gamma_j^2 \|V(j)\|_{\infty}^2] \leq V \sum_{j=1}^{\infty} \gamma_j^2 < \infty$, Doob’s martingale convergence theorem [35, Theorem 2.5] shows that $\sum_{j=1}^{\infty} \gamma_j^2 \|V(j)\|_{\infty}^2$ is finite (a.s.).

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 $D_n \geq 0$, so we conclude that $X(n)$ visits a compact neighborhood of $X^*$ infinitely often (viz. there exists a sequence $n_j \uparrow \infty$ such that $X(n_j)$ lies in said neighborhood). Since $X^*$ attracts any initial condition $G(Y(0))$ under the continuous-time dynamics (MXL$_{\alpha}$), Theorem 6.10 in [33] shows that $X(n)$ converges to $X^*$ (a.s.), as claimed.

We close this section by showing that $Y(n)$ remains an APT of (MXL$_{\alpha}$) 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 $\sum_{n} \gamma_{k,n}^2 \gamma_{k,n} = \infty$ for all $k \in K$. Then, $Y(n)$ is an APT of (MXL$_{\alpha}$).

Proof: Since the blocks $Z_k$ and $Z_k^*$ are disjoint for $k \neq k'$, we obtain $\|\sum_{k} \gamma_{k,n} Z_k(n)\| = \sum_{k} \|\gamma_{k,n} Z_k(n)\|_2$ in the product norm on $\mathcal{X}$. Thus, Burkholder’s inequality [35, Chap. 2] yields
\[ \mathbb{E} \left[ \sup_{j=0}^{m} \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \gamma_{k,n}^2 [Z_k(n)]_2 \right] \leq C \sum_{k=0}^{\infty} \mathbb{E} \left[ \sum_{j=0}^{\infty} \gamma_{k,n}^2 [Z_k(t)]_2 \right], \tag{C.6} \]
for some universal constant $C$. To proceed, let $\Delta(t, T) = \sup_{0 \leq s \leq T} \|\tilde{Z}(t, s)\|_{\infty}$, where $\tilde{Z}$ is the linear interpolation of the discrete-time process $Z(t)$ at each epoch $n$ [33, p. 12]. Then, by applying Hölder’s inequality as in [33, p. 15], (C.6) readily gives $\mathbb{E} \left[ \Delta(t, T)^2 \right] \leq C \sum_{k=0}^{\infty} \mathbb{E} \left[ \sum_{j=0}^{\infty} \gamma_{k,n}^2 \gamma_{k,n} \right] ds$, where, in similar notation, $\tilde{Z}(t)$ denotes the linear interpolation of the discrete-time step-size sequence $\gamma_{k,n}$. We thus get
\[ \sum_{m=0}^{\infty} \mathbb{E} \left[ \Delta(mT, T)^2 \right] \leq C \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \gamma_{k,n}^2 \gamma_{k,n} \ln \frac{1}{\epsilon} \phi_{k,n}^2, \tag{C.7} \]
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 $Y(n)$ is an APT of (MXL$_{\alpha}$), 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 $\sum_{n=1}^{\infty} \gamma_{k,n}^2 \gamma_{k,n} < \infty$ 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 $X^*$ in $\mathcal{X}$ such that $V(Y)$ holds for all $X \in U$. Assume now that $m > 0$ is taken sufficiently small so that $G(Y) \in U$ whenever $F(X^*, Y) < 3m$ (the existence of such a positive $m$ follows from the fact that $G(Y) \to X^*$ if $F(X^*, Y) \to 0$). By Eq. (C.1), it then follows that the set $U_{3m} = 13$In particular, [33, Prop. 4.2] shows that, under (H2'), $Y(n)$ is an APT of (MXL$_{\alpha}$) for any step-size sequence $\gamma_n$ such that $\sum_{n} \gamma_n^{1+q/2} < \sum_{n} \gamma_n = \infty$.
\( Y : F(X^*, Y) \leq 3m \) is invariant under (M XL ). Hence, by shadowing the proof of Theorem 3, it suffices to show that there exists an open set \( U_e \subseteq U_{3m} \) such that \( P(Y(n) \in U_{3m} \text{ for all } n) \geq 1 - e \) whenever \( Y(0) \in U_e \).

To that end, let \( D_n = F(X^*, Y(n-1)) \) and assume that \( D_1 \leq m \). Then, (C.4) yields
\[
D_n \leq m + \sum_{j=1}^{n} \gamma_j \xi_j + \sum_{j=1}^{n} \gamma_j \norm{\tilde{V}(j)}_{\infty}^2 .
\]
where, in the second line, we used (18b) and the assumption that \( \gamma_j \xi_j \leq m \) for all \( n \) with probability at least \( 1 - e/2 \) if \( \gamma_j \) is chosen appropriately. Indeed, letting \( S_n = \sum_{j=1}^{n} \gamma_j \xi_j \), Doob's inequality \([35, \text{Theorem 2.1}]\) gives
\[
P(\sup_{1 \leq t \leq n} |S_t| \leq m) \leq \frac{\mathbb{E}[|S_n|^2]}{m^2} \leq \frac{\sigma^2 \norm{X}^2}{m^2} \sum_{j=1}^{n} \gamma_j^2 .
\]

where we used the fact that \( \mathbb{E}[\xi_j \xi_t] = 0 \) if \( j \neq t \). Hence, letting \( \Gamma_2 = \sum_{j=1}^{\infty} \gamma_j^2 \), it follows that \( P(\sup_{1 \leq t \leq n} |S_t| \leq \epsilon) \geq \Gamma_2 \sigma^2 \norm{X}^2 / m^2 \geq \epsilon / 2 \) if \( \Gamma_2 \) is sufficiently small. Moreover, letting \( R_n = \sum_{j=1}^{n} \gamma_j \norm{\tilde{V}(j)}_{\infty}^2 \), and working as above, Doob's inequality again shows that \( P(\sup_{1 \leq t \leq n} R_t \geq m) \leq m^{-1} \lim_{n \to \infty} \mathbb{E}[R_n] \leq m^{-1} \Gamma_2 V^2 \leq \frac{\epsilon}{2} \), if \( \Gamma_2 \) is taken small enough.

Combining all of the above, we get \( D_n \leq 3m + \sum_{j=1}^{n} \gamma_j \xi_j \) with probability at least \( 1 - e \). Since \( \mathbb{G}(U_{3m}) \subseteq U \) by construction, we have \( u_j \leq 0 \) for all \( j = 1, \ldots, n \), and we conclude that \( D_n \leq 3m \) with probability at least \( 1 - e \). This implies that \( P(Y(n) \in U_{3m} \text{ for all } n) \geq 1 - e \), 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 \( \bar{D}_n = \mathbb{E}[F(X^*, Y(n-1))] = \mathbb{E}[D_{KL}(X^*, Y(n))] \). Then, taking expectations in (C.4) yields
\[
\bar{D}_{n+1} \leq \bar{D}_n + \gamma_n \mathbb{E}[\text{tr}(h(X(n)-X^*) V(n))] + \gamma_n^2 \mathbb{E}[\norm{\tilde{V}(n)}_{\infty}^2].
\]

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

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

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

[24] A. Charnes and W. W. Cooper, “Programming with linear fractional functions,” *Naval Research Logistics Quarterly*, vol. 9, pp. 181–196, 1962.

[25] 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.

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

[27] E. V. 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.

[28] P. Coucheney, B. Gaujal, and P. Mertikopoulos, “Penalty-regulated dynamic and regularized,” *Mathematics of Operations Research*, vol. 41, no. 4, pp. 1297–1324, November 2016.

[29] P. Mertikopoulos and W. H. Sandholm, “Learning in games via reinforcement and regularization,” *Mathematics of Operations Research*, vol. 40, no. 3, pp. 611–633, August 2015.

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

[31] 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.

[32] A. Charnes and W. W. Cooper, “Programming with linear fractional functions,” *Naval Research Logistics Quarterly*, vol. 9, pp. 181–196, 1962.

[33] 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.

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

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

[36] D. Monderer and L. S. Shapley, “Potential games,” *Econometrica*, vol. 61, no. 2, pp. 283–292, 1993.

[37] M. Benaïm, “Dynamics of stochastic approximation algorithms,” in *Séminaire de Probabilités XXXIII*, ser. Lecture Notes in Mathematics, vol. 1781, Springer, 2001.

[38] V. Vedral, “The role of relative entropy in quantum information theory,” *Reviews of Modern Physics*, vol. 74, no. 1, pp. 1–41, 2002.

[39] D. Monderer and L. S. Shapley, “Potential games,” *Econometrica*, vol. 61, no. 2, pp. 283–292, 1993.

[40] M. Benaïm, “Dynamics of stochastic approximation algorithms,” in *Séminaire de Probabilités XXXIII*, ser. Lecture Notes in Mathematics, vol. 1781, Springer, 2001.

[41] V. Vedral, “The role of relative entropy in quantum information theory,” *Reviews of Modern Physics*, vol. 74, no. 1, pp. 1–41, 2002.

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