Distributed Optimization in Multi-User MIMO Systems with Imperfect and Delayed Information

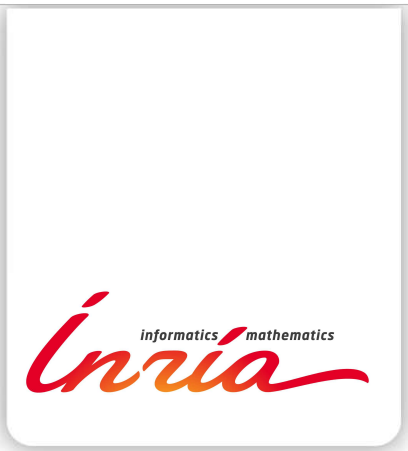
Pierre Coucheney, Bruno Gaujal, Panayotis Mertikopoulos

Distributed Optimization in Multi-User MIMO Systems with Imperfect and Delayed Information

Pierre Coucheney, Bruno Gaujal, Panayotis Mertikopoulos

RESEARCH REPORT
N° 8426
Dec. 2013
Project-Team Mescal

ISSN 0249-6399 ISRN INRIA/RR--8426--FR+ENG



Distributed Optimization in Multi-User MIMO Systems with Imperfect and Delayed Information

Pierre Coucheney*, Bruno Gaujal†, Panayotis Mertikopoulos ‡

Project-Team Mescal

Research Report n° 8426 — Dec. 2013 — 19 pages

Abstract: Starting from an entropy-driven reinforcement learning scheme for multi-agent environments, we develop a distributed algorithm for robust spectrum management in Gaussian multiple-input, multiple-output (MIMO) uplink channels. In continuous time, our approach to optimizing the transmitters' signal distribution relies on the method of matrix exponential learning, adjusted by an entropy-driven barrier term which generates a distributed, convergent algorithm in discrete time. As opposed to traditional water-filling methods, the algorithm's convergence speed can be controlled by tuning the users' learning rate; accordingly, entropy-driven learning algorithms in MIMO systems converge arbitrarily close to the optimum signal covariance profile within a few iterations (even for large numbers of users and/or antennas per user), and this convergence remains robust even in the presence of imperfect (or delayed) measurements and asynchronous user updates.

Key-words: Distributed optimization; matrix exponential learning; multiple access channels; MIMO; stochastic approximation.

* PRiSM, University of Versailles, 45 avenue des Etats-Unis, 78035 Versailles, France
† Inria and University of Grenoble Alpes, 38330 Grenoble, France
‡ French National Center for Scientific Research (CNRS) and University of Grenoble Alpes, 38330 Grenoble, France

RESEARCH CENTRE GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Optimisation distribuée dans les systèmes MIMO avec information imparfaite et retardée

Résumé : Dans cet article, nous développons un algorithme distribué pour l'optimisation spectrale d'un canal Gaussien MIMO (antennes entrées et sorties multiples), à partir d'un schéma d'apprentissage multi-agents reposant sur l'entropie. En temps continu, notre approche pour optimiser la distribution du signal des transmetteurs utilise une technique d'apprentissage par matrices exponentielles, ajustée par une barrière entropique qui permet de concevoir un algorithme en temps discret de calcul de la configuation optimale. Contrairement au cas des méthodes classiques par remplissage, la vitesse de convergence peut être controlée en réglant le taux d'apprentissage des transmetteurs. Cela permet à notre algorithme de converger vers la configuration de matrice de covariance optimale en quelques itérations (même avec un grand nombre d'antennes et d'utilisateurs). Cette convergence est robuste aux imperfections et aux retards sur les mesures du signal et aux désynchronisations entre utilisateurs.

Mots-clés : Optimisation distribuée; apprentissage exponentiel, MIMO, canaux à accès multiples; approximation stochastique

Contents

1 Introduction

The seminal prediction that the use of multiple antennas in radio signal transmission and reception can lead to substantial performance gains [1, 2] has made multiple-input and multiple-output (MIMO) technologies an integral component of most state-of-the-art wireless communication protocols, ranging from 3G LTE, 4G and HSPA+, to 802.11n WiFi and WiMax (to name but a few). However, given the decentralized nature of some of these protocols (e.g. the latter two), it is not clear how users may benefit from the use of multiple antenna technologies at a network level. Even worse, when centrally controlled protocols (such as the former ones) are deployed at massively large scales (e.g. in densely populated urban environments), their complexity is such that the advantages of using multiple antennas only become apparent if distributed optimization methods are employed to manage the network's spectrum.

In this often unregulated context, the radio spectrum is shared by all users, so the intended receiver of a signal has to cope with unwarranted interference from a large number of transmitters (a factor which severely limits the capacity of the wireless system in question). On that account, and given that the theoretical performance limits of MIMO systems still elude us (even in basic network models such as the interference channel), a widespread approach is to use the mutual information for Gaussian input and noise as a performance metric, and to optimize (the covariance of) the input signal distribution of each transmitter in the presence of interference from all other users.

In this paper, we focus on uplink MIMO systems consisting of several non-cooperative (and mutually independent) Gaussian transmitters who upload data to a receiver (conceivably representing a set of collocated terminals or even a set of non-collocated receivers connected over a high-speed backbone network). This vector Gaussian multiple access channel (MAC) has attracted significant interest in the wireless literature [2–4], and it is well-known that attaining its capacity boils down to solving a nonlinear optimization problem over a set of positive-semidefinite matrices representing the users' feasible input covariance matrices (i.e. the spread of their symbol distributions over their antennas).

Traditionally, this semidefinite problem is solved by water-filling techniques [3], properly adapted to multi-user environments [4–6]. Unfortunately however, the convergence speed of iterative water-filling methods decreases linearly with the number of users in the system (making such methods unsuitable for

large networks), whereas the convergence of faster, simultaneous water-filling methods [6] is conditional on certain "mild-interference" conditions which fail to hold even in the simple special case of parallel multiple access channels (PMACs) [7].

To overcome these limitations, the authors of [8] proposed an alternative approach based on the method of matrix exponential learning whose convergence speed scales well with the number of users in the system. However, just like water-filling, this method relies on perfect channel state information at the transmitter (CSIT) and accurate measurements at the receiver, an assumption which breaks down in rapidly evolving, unregulated networks. Consequently, a major challenge arises when this information can only be estimated in an imperfect manner, or when only delayed (and, hence, potentially obsolete) measurements are available: for instance, the analysis of [9] can be used to show that in the presence of stochasticity, exponential learning may converge with positive probability to a globally suboptimal signal covariance profile, even in the very simple case of a single user transmitting over two parallel channels. Moreover, in the absence of a centralized scheduler, enforcing simultaneous user updates is all but impossible, so it is not clear if methods that rely on synchronous decision-taking can be implemented in decentralized environments.

To address these issues, we introduce an adjusted variant of the matrix exponential learning method of [8] which penalizes zero eigenvalues in the transmission covariance profile. The resulting method of adjusted exponential learning actually applies to a wide class of nonlinear semidefinite programming problems, but we chose to focus here on the MIMO MAC case for simplicity and concreteness. In this context, the dynamics' entropy-driven adjustment term is controlled by a temperature parameter which allows the system to converge to a nonsingular covariance profile that is arbitrarily close to an optimum one. With the limit of this learning process being non-singular, the powerful stochastic approximation techniques of [10] then allow us to show that the algorithm converges very fast (in practice, within a few iterations), even for large numbers of users and/or antennas per user, and even in the presence of large estimation errors or asynchronous updates.

2 System Model

As we mentioned in the introduction, we will model our uplink system as a vector Gaussian multiple access channel consisting of a finite set of transmitters $k \in \mathcal{K} \equiv \{1, \dots, K\}$, each equipped with m_k antennas, and each transmitting simultaneously to a base receiver with m_0 antennas. More precisely, this system will be represented by the familiar baseband model

$$\mathbf{y} = \sum\nolimits_k \mathbf{H}_k \mathbf{x}_k + \mathbf{z}, \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^{m_0}$ denotes the aggregate signal reaching the receiver, $\mathbf{x}_k \in \mathbb{C}^{m_k}$ is the message transmitted by user $k \in \mathcal{K}$, $\mathbf{H}_k \in \mathbb{C}^{m_0 \times m_k}$ is the associated $m_0 \times m_k$ (complex) channel matrix, and $\mathbf{z} \in \mathbb{C}^{m_0}$ is the noise in the channel, including thermal, atmospheric and other peripheral interference effects, and assumed to be a (zero-mean) circularly symmetric complex Gaussian random vector with non-singular covariance (taken equal to $\mathbf{I}$ after a change of basis).

The signal model (1) is quite general, so, depending on the structure of the channel matrices $\mathbf{H}_k$, it applies to several telecommunications systems, ranging from digital subscriber line (DSL) uplink networks with Tœplitz circulant $\mathbf{H}_k$, to code division multiple access (CDMA) and/or frequency division multiple access (FDMA) radio networks [11]. For the sake of concreteness however, we will stick here with the interpretation of (1) as an ad hoc multi-user MIMO multiple access channel with $\mathbf{H}_k$ representing the channel of each link, assumed to remain static and fixed for the duration of the transmission.

With this in mind, the average transmit power of user k will be

$$P_k = \mathbb{E}\left[\|\mathbf{x}_k\|^2\right] = \mathrm{tr}(\mathbf{Q}_k), \tag{2}$$

where the expectation is taken over the codebook of user k and $\mathbf{Q}_k$ denotes the corresponding signal covariance matrix:

$$\mathbf{Q}_k = \mathbb{E}\left[\mathbf{x}_k \mathbf{x}_k^\dagger\right]. \tag{3}$$

Then, assuming successive interference cancellation (SIC) at the receiver, the maximum information transmission rate will be achieved for random Gaussian codes and will be given by

$$\Phi(\mathbf{Q}) = \log\det\left(\mathbf{I} + \textstyle\sum_k \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^\dagger\right), \tag{4}$$

where $\mathbf{Q} = (\mathbf{Q}_1, \ldots, \mathbf{Q}_k) = \bigoplus_k \mathbf{Q}_k$ denotes the block diagonal (direct) sum of the individual matrices $\mathbf{Q}_k$ [2]. In this way, we obtain the classical sum rate maximization problem [4, 5]:

$$\begin{aligned} &\text{maximize} && \Phi(\mathbf{Q}), \\ &\text{subject to} && \mathbf{Q}_k \in \mathcal{X}_k \quad (k = 1, \ldots, K), \end{aligned} \tag{P}$$

where $\mathcal{X}_k = \{\mathbf{Q}_k \in \mathbb{C}^{m_k \times m_k} : \mathbf{Q}_k \succcurlyeq 0, \operatorname{tr}(\mathbf{Q}_k) = P_k\}$ is the set of feasible signal covariance matrices of user k that satisfy the power constraint $\operatorname{tr}(\mathbf{Q}_k) = P_k$.[1]

Needless to say, solving (P) leads to a globally optimum transmit spectrum which maximizes the aggregate information rate decodable by the receiver. On the other hand, if sophisticated interference cancellation techniques are not available (for instance, if the complexity of calculating the users' decoding order is prohibitive), each transmitter's signal must be decoded by treating interference by other users as noise. In this single user decoding (SUD) regime, the users' unilateral objective is to maximize their *individual* achievable rates

$$u_k(\mathbf{Q}) = \log\det\left(\mathbf{W}_k + \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^\dagger\right) - \log\det\left(\mathbf{W}_k\right), \tag{5}$$

where

$$\mathbf{W}_k = \mathbf{I} + \textstyle\sum_{\ell \neq k} \mathbf{H}_\ell \mathbf{Q}_\ell \mathbf{H}_\ell^\dagger \tag{6}$$

is the multi-user interference-plus-noise matrix of user k. From a game-theoretic perspective, we will thus say that the system is at *Nash equilibrium* when no user can improve his individual achievable rate u_k by unilaterally changing his signal covariance matrix $\mathbf{Q}_k$, i.e. when

$$u_k(\mathbf{Q}) \geq u_k(\mathbf{Q}_k'; \mathbf{Q}_{-k}) \quad \text{for all } k \in \mathcal{K} \text{ and } \mathbf{Q}_k' \in \mathcal{X}_k, \tag{7}$$

where $(\mathbf{Q}_k'; \mathbf{Q}_{-k}) \equiv (\mathbf{Q}_1, \ldots, \mathbf{Q}_k', \ldots, \mathbf{Q}_K)$ is the standard game-theoretic shorthand notation for unilateral deviations.

A striking feature of this model is that users are individually aligned with the global objective function Φ, so the solutions of (P) coincide with the Nash equilibria of (5). Indeed, as was shown in [12], Φ is a *potential* for the game defined by the utility functions (5) in the sense that the individual rates u_k satisfy the property $u_k(\mathbf{Q}_k; \mathbf{Q}_{-k}) - u_k(\mathbf{Q}_k'; \mathbf{Q}_{-k}) = \Phi(\mathbf{Q}_k; \mathbf{Q}_{-k}) - \Phi(\mathbf{Q}_k'; \mathbf{Q}_{-k})$ [13]. As a result, with Φ concave, *the game's Nash equilibria will be precisely the solutions of* (P).

3 The dynamics of adjusted matrix exponential learning

In the iterative water-filling approach of [4], it is assumed that transmitters (viewed as decision-makers) have perfect knowledge of their channel matrices $\mathbf{H}_k$ and of the user-specific interference-plus-noise matrices (6). The latter can be computed from the aggregate signal-plus-noise covariance matrix

$$\mathbf{W} = \mathbf{I} + \textstyle\sum_\ell \mathbf{H}_\ell \mathbf{Q}_\ell \mathbf{H}_\ell^\dagger \tag{8}$$

[1] More general power constraints of the form $\operatorname{tr}(\mathbf{Q}_k) \leq P_k$ or spectral mask constraints and null shaping constraints limiting transmission in certain bands could also be considered, but we will not be treating them in this paper for reasons of simplicity and space constraints.

by subtracting $\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^\dagger$, so these information requirements boil down to perfect channel state information at the transmitter (CSIT) and accurate measurements of $\mathbf{W}$ at the receiver (who can then broadcast this information to the network's users via e.g. a dedicated radio channel).

Under the same information assumptions, the authors of [8] considered the matrix exponential learning dynamics:

$$\begin{aligned}\dot{\mathbf{Y}}_k &= \mathbf{V}_k,\\ \mathbf{Q}_k &= P_k\frac{\exp(\mathbf{Y}_k)}{\mathrm{tr}[\exp(\mathbf{Y}_k)]},\end{aligned} \tag{XL}$$

where $\mathbf{Y}_k$ is an auxiliary Hermitian "scoring" matrix and

$$\mathbf{V}_k = \nabla_{\mathbf{Q}_k^*}\Phi = \mathbf{H}_k^\dagger\mathbf{W}^{-1}\mathbf{H}_k \tag{9}$$

denotes the (conjugate) gradient of Φ w.r.t. $\mathbf{Q}_k$.[2] These dynamics were shown in [8] to always converge to a solution of (P), so the question which arises is whether this convergence is preserved if the "perfect CSIT" assumption is relaxed.

To obtain a discretization of (XL) in a stochastic environment with imperfect state information, consider the recursive scheme

$$\mathbf{Y}_k(n+1) = \mathbf{Y}_k(n) + \gamma_n\hat{\mathbf{V}}_k, \tag{10}$$

where γ_n is a variable step size, $\hat{\mathbf{V}}_k$ is a stochastically perturbed estimate of the true derivative matrix $\mathbf{V}_k$ of the sum rate function Φ, and $\mathbf{Q}_k$ is obtained by exponentiating $\mathbf{Y}_k$ as in (XL). If γ_n vanishes moderately fast (i.e. $\sum_n\gamma_n = +\infty$ but $\sum_n\gamma_n^2 < \infty$), and more importantly, if there is no systematic error in the measurements for $\mathbf{V}_k$ (i.e. if $\mathbb{E}[\hat{\mathbf{V}}_k] = \mathbf{V}_k$), one could hope to apply the stochastic approximation techniques of [10, 14] to show that the discretized *stochastic* version of (XL) converges to an optimum signal covariance profile. This, however, is not the case: in PMAC networks where all the matrices $\mathbf{H}_k$ and $\mathbf{Q}_k$ are diagonal in a common basis, the exponential learning dynamics (XL) reduce to the well-known *replicator dynamics* of evolutionary game theory [15], i.e.

$$\dot{q}_{k\alpha} = q_{k\alpha}\left(V_{k\alpha} - \textstyle\sum_{\beta=0}^{m_k} q_{k\beta}V_{k\beta}\right), \tag{RD}$$

where $q_{k\alpha}$ is the α-th diagonal element of $\mathbf{Q}_k$ and $V_{k\alpha} = \frac{\partial\Phi}{\partial q_{k\alpha}}$ denotes the α-th element of $\mathbf{V}_k$ (which is also diagonal now) [16]. Thus, given that the discrete replicator dynamics with linear costs are known to converge with positive probability to a global minimum of the objective function in stochastic environments [9], the Euler-type discretization (10) of (XL) may lead to similarly unwarranted behavior.

From a mathematical viewpoint, the problem in discretizing (XL) is that the scoring matrices $\mathbf{Y}_k$ may fail to remain bounded for all time, so the tracking techniques of stochastic approximation do not apply – this also explains why the discretization of (XL) in a stochastic environment might yield a different limit than the perfect information, deterministic counterpart of [8].

Motivated by the analysis of [17] for finite games, we will instead consider the adjusted matrix exponential learning scheme:

$$\begin{aligned}\dot{\mathbf{Y}}_k &= \mathbf{V}_k - \tau\mathbf{Y}_k,\\ \mathbf{Q}_k &= P_k\frac{\exp(\mathbf{Y}_k)}{\mathrm{tr}[\exp(\mathbf{Y}_k)]},\end{aligned} \tag{XL$_\tau$}$$

where $\tau > 0$ is a temperature-like parameter whose role will be illustrated below. On the one hand, the term $-\tau\mathbf{Y}_k$, (XL$_\tau$) readily yields $\mathbf{Y}_k(t) = \int_0^t e^{-\tau(t-s)}\mathbf{V}_k(\mathbf{Q}(s))\,ds$, and with $\mathbf{V}_k$ bounded over $\mathcal{X} \equiv \prod_k\mathcal{X}_k$ (Lemma 7), the auxiliary matrices $\mathbf{Y}_k$ will remain themselves bounded for all time, so the theory of stochastic approximation applies. On the other hand, the convergence properties of (XL) are not the same

[2] Namely, if $Q_{\alpha\beta} = X_{\alpha\beta} + iY_{\alpha\beta}$, the components of $\nabla_{\mathbf{Q}^*}\Phi$ will be $\frac{\partial\Phi}{\partial X_{\alpha\beta}} + i\frac{\partial\Phi}{\partial Y_{\alpha\beta}}$.

as those of (XL_τ), so it is not clear if (XL_τ) will end up solving (P). Our strategy will thus be to establish the convergence properties of (XL_τ) in continuous time, and to then employ the stochastic approximation tools of [10] to derive a robust solution algorithm for (P) based on (XL_τ).

We begin with the former task, leaving the latter for the next section. Our main result for the adjusted dynamics (XL_τ) is that they converge to an approximate solution of (P), and that the error of this approximation tends to 0 as $\tau \to 0$:

Theorem 1. *Let $\mathbf{Q}(t)$ be a solution orbit of (XL_τ), and let $h(\mathbf{Q}) = \operatorname{tr}(\mathbf{Q}\log\mathbf{Q})$ denote the (negative) von Neumann entropy of $\mathbf{Q}$. Then the orbit $\mathbf{Q}(t)$ remains in $\mathcal{X}$ for all $t \geq 0$ and it converges to the (unique) maximum point of the free energy function $F(\mathbf{Q}) = \Phi(\mathbf{Q}) - \tau h(\mathbf{Q})$. In particular, $\mathbf{Q}(t)$ converges within $\varepsilon(\tau)$ of a solution of (P), and the approximation error $\varepsilon(\tau)$ becomes vanishingly small as $\tau \to 0$.*

Remark. As a direct consequence of the temperature adjustment term $-\tau\mathbf{Y}_k$, the dynamics (XL_τ) do not maximize the original sum rate objective Φ but the "perturbed" objective $F = \Phi - \tau h$ instead. If we interpret the problem's true objective function (i.e. the users' sum rate) as the internal energy of a thermodynamical system, then Theorem 1 simply states that we end up maximizing the "free" energy of the system which is available for useful, mechanical work [18].[3] At zero temperature, all the energy of the system is free for mechanical work, so we recover our original objective.

We will prove Theorem 1 via two intermediate results of independent interest (both proven in Appendix A). The first one concerns the evolution of the users' covariance matrices:

Proposition 2. *Let $\mathbf{Q}(t)$ be a solution orbit of the adjusted dynamics (XL_τ). Then, $\mathbf{Q}(t)$ is a solution of the temperature-adjusted entropy-driven dynamics:*

$$\dot{\mathbf{Q}} = \int_0^1 \mathbf{Q}^{1-s}\mathbf{V}_\tau\mathbf{Q}^s \, ds - \operatorname{tr}(\mathbf{Q}\mathbf{V}_\tau)/\operatorname{tr}(\mathbf{Q})\,\mathbf{Q} \tag{ED_τ}$$

where $\mathbf{V}_\tau = \mathbf{V} - \tau\log\mathbf{Q}$ and $\mathbf{V} = (\mathbf{V}_1, \dots, \mathbf{V}_K) = \bigoplus_k \mathbf{V}_k$ is the block diagonal sum of the gradient matrices (9).

The evolution equation (ED_τ) owes its name to the fact that the entropy-adjusted objective $F(\mathbf{Q}) = \Phi(\mathbf{Q}) - \tau h(\mathbf{Q})$ is a strict Lyapunov function for (ED_τ)/(XL_τ); more precisely:

Proposition 3. *Let $F(\mathbf{Q}) = \Phi(\mathbf{Q}) - \tau\operatorname{tr}(\mathbf{Q}\log\mathbf{Q})$ denote the free energy of $\mathbf{Q}$, and let $\mathbf{Q}(t)$ be a solution orbit of the temperature-adjusted dynamics (XL_τ). The gradient of F with respect to $\mathbf{Q}$ is $\nabla_{\mathbf{Q}^*}\Phi = \mathbf{V}_\tau - \tau\mathbf{I} = \mathbf{V} - \tau(\mathbf{I} + \log\mathbf{Q})$, and $\frac{d}{dt}F(\mathbf{Q}(t)) \geq 0$ with equality if and only if $\mathbf{V}_\tau \propto \mathbf{I}$.*

Thanks to these two results, we then obtain:

Proof of Theorem 1. As we noted before, the dynamics (XL_τ) readily yield $\mathbf{Y}_k(t) = \int_0^t e^{-\tau(t-s)}\mathbf{V}_k(\mathbf{Q}(s))ds$, so $\mathbf{Y}_k(t)$ will be Hermitian (because $\mathbf{V}_k$ is), and hence $\mathbf{Q}_k(t)$ will be positive-definite with trace equal to P_k; as a result, $\mathbf{Q}(t)$ will remain in the feasible set $\mathcal{X}$ for all $t \geq 0$. Furthermore, by the same integral expression, we deduce that $\mathbf{Y}$ will be bounded in norm by M/τ where $M = \sup\{|\mathbf{V}(\mathbf{Q})| : \mathbf{Q} \in \mathcal{X}\} < +\infty$ (see also Lemma 7). As a result, $\mathbf{Q}(t)$ will be contained in a compact set $K \subseteq \mathcal{X}$ which is well-separated from the boundary $\operatorname{bd}(\mathcal{X})$ of $\mathcal{X}$, i.e. all the eigenvalues of $\mathbf{Q}(t)$ will remain a bounded distance away from zero and every ω-limit of $\mathbf{Q}(t)$ will belong to the interior of $\mathcal{X}$. It then follows from Proposition 3 and the general theory of Lyapunov functions that $\mathbf{Q}(t)$ will converge to a set of points satisfying the stationarity condition $\mathbf{V}_\tau = \mathbf{V} - \tau\mathbf{Y} \propto \mathbf{I}$.

Since the von Neumann entropy $h(\mathbf{Q}) = \operatorname{tr}(\mathbf{Q}\log\mathbf{Q})$ is strongly convex and becomes infinitely steep at the boundary of $\mathcal{X}$, the perturbed objective (free energy) $F = \Phi - \tau h$ will admit a unique maximum point

[3] In statistical mechanics, the Helmholtz free energy is actually defined as $-F$, owing to the fact that h and Φ are minus their physical counterparts (of entropy and potential energy respectively).

over $\mathcal{X}$, and this point will be interior [19]. By deriving the Karush–Kuhn–Tucker (KKT) conditions for F, it is then easy to show that $\mathbf{V}_\tau(\mathbf{Q}) \propto \mathbf{I}$ if and only if $\mathbf{Q}$ is the (unique) maximizer of F, thus establishing the convergence part of our claim. The fact that the unique maximum of F will be within $\varepsilon(\tau)$ of a solution of (P) with $\varepsilon(\tau) \to 0$ as $\tau \to 0$ is then a consequence of Berge's maximum theorem [20]. □

4 Discretization and stochastic approximation

4.1 Simultaneous updates

To use the adjusted exponential learning dynamics (XL_τ) as a distributed solution method for the sum rate optimization problem (P), we will first consider the Euler discretization

$$\begin{aligned} \mathbf{Y}_k(n+1) &= \mathbf{Y}_k(n) + \gamma_n[\mathbf{V}_k(n) - \tau\mathbf{Y}_k(n)], \\ \mathbf{Q}_k(n+1) &= P_k \frac{\exp(\mathbf{Y}_k(n+1))}{\mathrm{tr}[\exp(\mathbf{Y}_k(n+1)]}, \end{aligned} \tag{11}$$

where the gradient matrix $\mathbf{V}_k(n)$ is given by (9) and γ_n is a variable step size which tends to 0 as $n \to \infty$. Since the domain of $\mathbf{Y}_k$ is unconstrained and the global attractor of the continuous-time dynamics (XL_τ) is a compact, connected set, the general theory of deterministic approximation implies that the Euler discretization (11) of (XL_τ) will converge to the global attractor of (XL_τ). Therefore, assuming that the gradient matrices $\mathbf{V}_k$ can be calculated in an error-free manner (which, in turn, boils down to assuming perfect CSIT and perfect knowledge of $\mathbf{W}$ at the receiver), Theorem 1 shows that the discrete-time scheme (11) will converge to an approximate solution of (P), and the error $\varepsilon(\tau)$ of this approximation will tend to zero as $\tau \to 0$.

Let us now consider the more realistic case of a stochastic environment with imperfect state information. In particular, we will consider two types of random perturbations in (11): (*i*) the transfer matrices $\mathbf{H}_k$ can only be measured at the transmitter end up to some random observational error; and (*ii*) the receiver can only estimate the covariance $\mathbf{W}$ of the aggregate received signal $\mathbf{y}$ via random sampling – assumed to occur in between the update cycles of the discrete-time process (11).

Even though these two randomness sources are independent of one another, the gradient matrices $\mathbf{V}_k$ depend nonlinearly on $\mathbf{H}_k$ and $\mathbf{W}$, so care must be taken to construct an unbiased estimator of $\mathbf{V}_k$ from noisy estimates of $\mathbf{H}_k$ and $\mathbf{W}$. Fortunately, there exist well-known unbiased estimators for the precision (inverse covariance) matrix $\mathbf{P} = \mathbb{E}[\mathbf{y}\mathbf{y}^\dagger]^{-1}$ of a multivariate Gaussian random variable [21], so, in the case of Gaussian input and noise, imperfect knowledge of $\mathbf{H}_k$ and $\mathbf{W}$ may be treated as a zero-mean random observation error on $\mathbf{V}_k$ (see Appendix B for more details).

In view of the above, assume that at the n-th iteration of (11), the network's users observe a perturbed version $\hat{\mathbf{V}}_k$ of their true gradient matrices $\mathbf{V}_k$, i.e.

$$\hat{\mathbf{V}}_k(n) = \mathbf{V}_k(n) + \mathbf{\Xi}_k(n). \tag{12}$$

for some random error process $\mathbf{\Xi}_k$. We then obtain the following stochastic approximation of the entropic dynamics (XL_τ):

$$\begin{aligned} \mathbf{Y}_k(n+1) &= \mathbf{Y}_k(n) + \gamma_n[\hat{\mathbf{V}}_k(n) - \tau\mathbf{Y}_k(n)], \\ \mathbf{Q}_k(n+1) &= P_k \frac{\exp(\mathbf{Y}_k(n+1))}{\mathrm{tr}[\exp(\mathbf{Y}_k(n+1))]}, \end{aligned} \tag{SA_τ}$$

with the assumptions:

(A1) The observational errors $\mathbf{\Xi}_k(n)$ of (12) are i.i.d. bounded (a.s.) random processes with $\mathbb{E}[\mathbf{\Xi}_k(n)] = 0$ for all n.

(A2) The variable step γ_n satisfies the "$\ell^2 - \ell^1$" summability condition $\sum_n \gamma_n = \infty$ and $\sum_n \gamma_n^2 < \infty$ (a typical choice being $\gamma_n = 1/n^\alpha$ with $1/2 < \alpha \leq 1$).

As a first (and technically crucial) step in the convergence analysis of (SA_τ), we show that the process $\mathbf{Y}_k(n)$ generated from (SA_τ) is *stable*, i.e. its iterates remain bounded (a.s.):

Lemma 4. *If $\gamma_n < 1/\tau$ for all sufficiently large n, the stochastic approximation process $\mathbf{Y}_k$ of (SA_τ) is bounded almost surely.*

Proof of Lemma 4. With $\mathbf{\Xi}_k$ and $\mathbf{V}_k$ bounded almost surely (by assumption and by Lemma 7 respectively), the unbiased estimator $\hat{\mathbf{V}}_k = \mathbf{V}_k + \mathbf{\Xi}_k$ will also be bounded in norm (a.s.), say by some $M > 0$. Additionally, since the steps γ_n are themselves eventually bounded by $1/\tau$, we will also have $0 < 1 - \gamma_n\tau \leq 1$ for all n greater than some $n_0 \in \mathbb{N}$. Then, for $n \geq n_0$, the definition (SA_τ) of $\mathbf{Y}_k(n)$ and the bound $|\hat{\mathbf{V}}_k(n)| < M$ readily yield $|\mathbf{Y}_k(n+1)| \leq (1 - \tau\gamma_n)|\mathbf{Y}_k(n)| + \gamma_n M$. We thus obtain the following cases:

- If $\tau|\mathbf{Y}_k(n)| \geq M$, then $|\mathbf{Y}_k(n+1)| \leq |\mathbf{Y}_k(n)| + \gamma_n(M - \tau|\mathbf{Y}_k(n)|) \leq |\mathbf{Y}_k(n)|$, so $\mathbf{Y}_k$ decreases in norm.
- On the other hand, if $\tau|\mathbf{Y}_k(n)| \leq M$, we will have $|\mathbf{Y}_k(n+1)| \leq (1 - \gamma_n\tau)M/\tau + \gamma_n M = M/\tau$.

It follows that $|\mathbf{Y}_k(n+1)|$ will either decrease or be uniformly bounded by M/τ, and our claim follows by induction. □

Remark. In the unadjusted regime $\tau = 0$ of [8], the proof above no longer holds. This is actually one of the main reasons for the introduction of a temperature-controlled perturbation to exponential learning: the analysis of [17] shows that stochastic approximations of (XL) may well be unstable for $\tau = 0$.

By the stability of (SA_τ) we then obtain:

Theorem 5. *Under the assumptions* (A1) *and* (A2)*, the discrete-time learning scheme* (SA_τ) *with noisy measurements converges almost surely to the (unique) maximum point of the perturbed sum rate function $F(\mathbf{Q}) = \Phi(\mathbf{Q}) - \tau h(\mathbf{Q})$. In particular, $\mathbf{Q}(n)$ converges to within $\varepsilon(\tau)$ of an optimum covariance profile of* (P)*, and the error $\varepsilon(\tau)$ vanishes as $\tau \to 0$.*

Proof of Theorem 5. By Lemma 4 and assumption A1, both $\mathbf{Y}_k(n)$ and the error processes $\mathbf{\Xi}_k(n)$ will be bounded almost surely (so, in particular, $\mathbb{E}[|\mathbf{V}_k(n)|^2]$ will also be bounded. Moreover, by Proposition (3), the dynamics (XL_τ) admit a strict Lyapunov function and a unique rest point (which obviously has measure zero in the set $\mathcal{X}$ of feasible covariance matrices). The theorem then follows from the general theory of stochastic approximation – see e.g. Theorem 5.7 in [14]. □

4.2 Asynchronous updates

Even though the discrete-time process (SA_τ) with imperfect state information converges arbitrarily close to an optimum signal covariance profile, it is not clear how it can be implemented in the absence of a centralized scheduler that could synchronize the users' update schedule. We will thus consider here a fully decentralized setting where each user updates his signal covariance matrix $\mathbf{Q}_k$ based on an individual timer – and, hence, independently of other users. In this case, the estimates for the signal-plus-noise precision matrix $\mathbf{W}^{-1}$ that are periodically calculated and broadcasted by the receiver (based on his own timer) might themselves suffer from delays induced by this lack of synchronization, so our formulation will need to account for delayed as well as asynchronous updates.

To account for all that, let us index the transmitters' updates by $n \in \mathbb{N}$ so that at each update period n, only a subset $\mathcal{K}_n$ of the K transmitters change their covariance matrices $\mathbf{Q}_k$. To model each player's individual timer, let $N_k(n)$ denote the number of updates that have been performed by user k up to epoch n,

and let $d_k(n)$ denote the number of steps elapsed between the update and its measurement by the receiver. The discretization (11) of (XL_τ) with asynchronous and delayed updates then becomes:

$$\begin{aligned} \mathbf{Y}_k(n+1) &= \mathbf{Y}_k(n) + \gamma_{N_k(n)}\, \mathbb{1}(k \in \mathcal{K}_n) \cdot [\hat{\mathbf{V}}_k(n) - \tau \mathbf{Y}_k(n)], \\ \mathbf{Q}_k(n+1) &= P_k \frac{\exp(\mathbf{Y}_k(n+1))}{\mathrm{tr}[\exp(\mathbf{Y}_k(n+1))]}, \end{aligned} \tag{a-SA$_\tau$}$$

where

$$\hat{\mathbf{V}}_k(n) = \mathbf{H}_k^\dagger (\mathbf{I} + \sum\nolimits_\ell \mathbf{H}_\ell \mathbf{Q}_\ell(n - d_\ell(n)) \mathbf{H}_\ell^\dagger)^{-1} \mathbf{H}_k. \tag{13}$$

By definition, $\mathbf{Y}_k(n)$ and $\mathbf{Q}_k(n)$ are only modified at step $n+1$ if $k \in \mathcal{K}_n$, so every user user $k \in \mathcal{K}$ only needs to keep track of his individual timer in (a-SA$_\tau$).

The discretization (a-SA$_\tau$) is an asynchronous stochastic approximation in the sense of [10, Chap. 7], so it will converge (a.s.) to the same limit as the synchronous version (SA$_\tau$) under some mild assumptions. More precisely, we have:

Theorem 6. *Assume that the set-valued process $\mathcal{K}_n$ which specifies the set of users updating their covariance matrix at step n is a homogeneous recurrent Markov chain (i.e. all users' update rates are finite and nonzero) and that the delay functions $d_k(n)$ are bounded (a.s.). Then, the conclusions of Theorem 5 still hold for the asynchronous scheme (a-SA$_\tau$) with delayed measurements and step-size $\gamma_n = 1/n$.*

Proof of Theorem 6. Following Theorems 2 and 3 in [10], the recursion (a-SA$_\tau$) may be viewed as a stochastic approximation of the rate-adjusted dynamics

$$\dot{\mathbf{Y}}_k = \eta_k \left[\mathbf{V}_k - \tau \mathbf{Y}_k\right], \tag{14}$$

where $\eta_k = \lim_n N_k(n)/n > 0$ is the update rate of user k (the existence and positivity of this limit follow from the ergodicity of the process $\mathcal{K}_n$). This multiplicative factor does not alter the rest points of the original dynamics (XL_τ), so, given that the free energy $F = \Phi - \tau h$ remains a strict Lyapunov function for (14), the proof of Theorem 5 still applies. □

5 Algorithms and numerical results

5.1 Algorithmic implementation

If all users share a common update timer, the stochastic approximation (SA$_\tau$) yields the following synchronized algorithm:

Algorithm 1 Adjusted matrix exponential learning (AXL)

$n \leftarrow 0$;

foreach transmitter $k \in \mathcal{K}$ **do**

  initialize the Hermitian score matrix $\mathbf{Y}_k \in \mathbb{H}^{m_k}$.

Repeat

  $n \leftarrow n+1$;

  Receiver measures and broadcasts the matrix $\hat{\mathbf{P}} = \hat{\mathbf{W}}^{-1}$;

  foreach transmitter $k \in \mathcal{K}$ **do**

    Measure channel matrix $\hat{\mathbf{H}}_k$;

    Update score matrix $\mathbf{Y}_k \leftarrow \mathbf{Y}_k + \gamma_n \left[\hat{\mathbf{H}}_k \hat{\mathbf{W}}^{-1} \hat{\mathbf{H}}_k^\dagger - \tau \mathbf{Y}_k\right]$;

    Update covariance matrix $\mathbf{Q}_k \leftarrow P_k \dfrac{\exp(\mathbf{Y}_k)}{\mathrm{tr}[\exp(\mathbf{Y}_k)]}$.

Remark 1. To ensure the algorithm's convergence, the variable step γ_n should satisfy the $\ell^2-\ell^1$ summability assumption (A2). However, as we shall see in the simulations section below, γ_n can actually be taken constant (and quite large) without compromising the algorithm's convergence (all the while speeding it up considerably).

Remark 2. From the point of view of distributed implementation, the adjusted exponential learning (AXL) algorithm has the following desirable properties:

(P1) It is *distributed*: users only update their individual variables using the same information as in distributed water-filling (namely the broadcast of $\mathbf{W}^{-1}$).

(P2) It is *stateless*: users do not need to know the state of the system (they might as well be oblivious to each other).

(P3) It is *reinforcing*: users tend to increase their individual transmission rates u_k.

(P4) It is *stable*: the matrix exponentials can be calculated in a numerically stable and efficient manner [22].

The synchronization of the users' updates can be obtained via the broadcast of the precision matrix $\mathbf{W}^{-1}$ which can double up as a beacon, triggering an update event at the transmitters. Thanks to Theorem 5, the AXL algorithm will then converge arbitrarily close to an optimum covariance profile, even in the presence of measurement errors in the estimation of $\mathbf{W}$ and $\mathbf{H}_k$ (see Appendix B for details on how to adjust the measurement process in that case).

On the other hand, update synchronization is not desirable in fully decentralized settings (for instance, if users come and go in the system, each user should start with a fresh step size γ_0 instead of relying on a globally diminishing one). In such settings, one can employ a fully distributed variant of AXL by simply basing the algorithm on the asynchronous discretization (a-SA$_\tau$) instead of (SA$_\tau$).

To wit, assume that each transmitter is equipped with an individual timer $t_k(n)$, $n \in \mathbb{N}$, whose ticks indicate the update events of user k (i.e. the transmitter updates its covariance matrix at each update event regardless of the global state). In this way, we obtain the following asynchronous variant of AXL:

Algorithm 2 Asynchronous adjusted exp. learning (a-AXL)

$n \leftarrow 0$;

Initialize the Hermitian score matrix $\mathbf{Y}$.

Repeat

  At each UpdateEvent

    $n \leftarrow n + 1$;

    Measure channel matrix $\hat{\mathbf{H}}$;

    Update score matrix $\mathbf{Y} \leftarrow \mathbf{Y} + \gamma_n[\hat{\mathbf{H}}\hat{\mathbf{W}}^{-1}\hat{\mathbf{H}}^\dagger - \tau\mathbf{Y}]$ using latest broadcasted value of $\hat{\mathbf{W}}^{-1}$;

    Update covariance matrix $\mathbf{Q} \leftarrow P\dfrac{\exp(\mathbf{Y})}{\mathrm{tr}[\exp(\mathbf{Y})]}$.

This asynchronous adjusted exponential learning (a-AXL) algorithm is run independently by each transmitter – though, of course, if all transmitters share a common timer, a-AXL reduces to the synchronous AXL variant above. Then, provided that all individual timers t_k have positive finite rate (i.e. $\lim t_k(n)/n$ exists and is finite),[4] the update sequence generated by a-AXL agrees with the assumptions of Theorem 6, so the algorithm will converge arbitrarily close to an optimum covariance profile.

[4] More precisely, the set-valued process $\mathcal{K}_n$ used in (a-SA$_\tau$) to indicate the set of transmitters updating their covariance matrices at

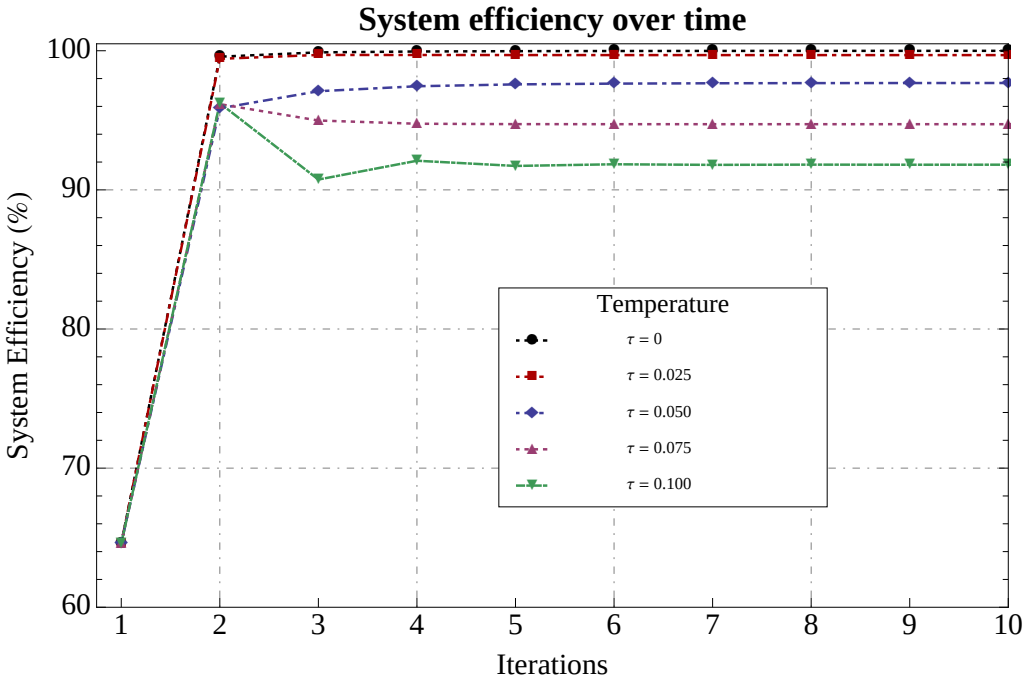


Figure 1: The effect of the temperature parameter τ of the adjusted exponential learning (AXL).

5.2 Numerical results

To assess the performance of the adjusted matrix exponential algorithm AXL, we simulated in Fig. 1 a multi-user uplink MIMO system consisting of a wireless base receiver with 5 antennas and $K = 25$ transmitters, each with a random number m_k of transmit antennas picked uniformly between 2 and 6. Each user's channel matrix $\mathbf{H}_k$ was drawn from a complex Gaussian distribution at the outset of the transmission (but remained static once picked), and we then ran the AXL algorithm with a large constant step size for different values of the temperature parameter τ. The performance of the algorithm over time was assessed by plotting the normalized efficiency ratio

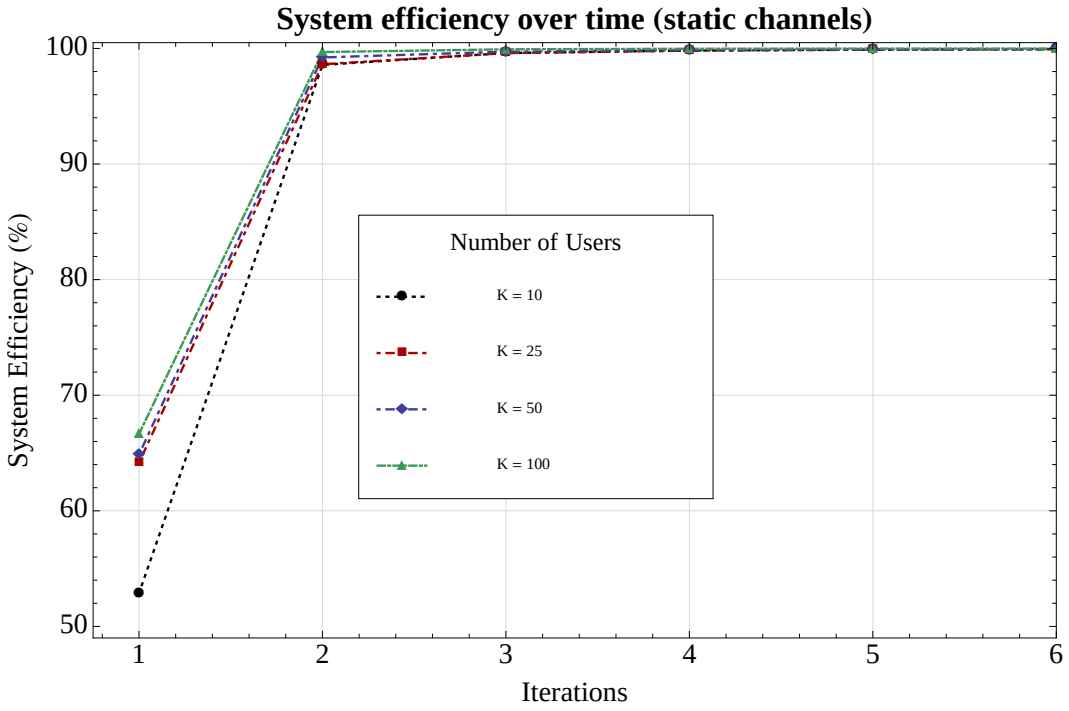
$$\mathrm{eff}(n) = \frac{\Phi_n - \Phi_{\min}}{\Phi_{\max} - \Phi_{\min}}, \tag{15}$$

where Φ_n denotes the users' sum rate at the n-th iteration of the algorithm, and $\Phi_{\max}$ (resp. $\Phi_{\min}$) is the maximum (resp. minimum) value of Φ over the system's set $\mathcal{X}$ of feasible covariance matrices.[5] In tune with Theorem 1, AXL converges within a few iterations (effectively, within a single iteration for low τ), but the end value of the users' sum rate deteriorates when the temperature gets higher.

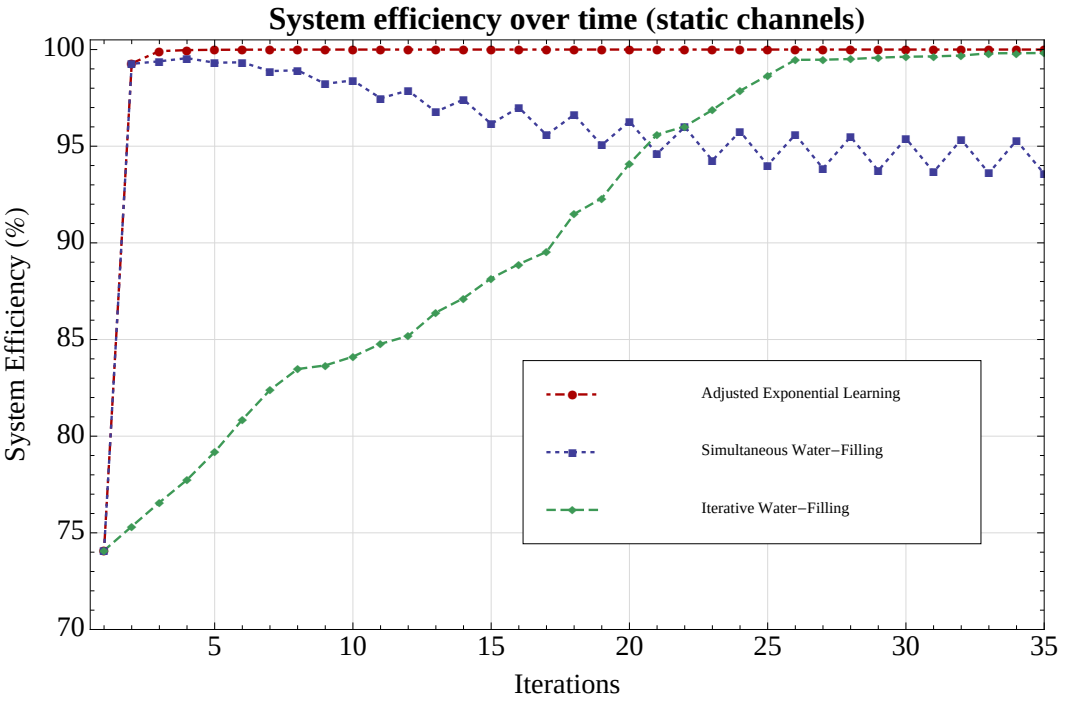
In Fig. 2, we fix the algorithm's temperature parameter to a low level ($\tau = 10^{-3}$) that ensures effective convergence to the system's sum capacity, and we investigate the algorithm's convergence speed as a

the n-th update event may be obtained from the individual timers $t_k(n)$ as follows: first, let $\mathcal{K}(t) = \{k \in \mathcal{K} : t_k(n) = t \text{ for some } n \in \mathbb{N}\}$ denote the set of players updating at time t and let $n(t) = \mathrm{card}\{s \leq t : \mathcal{K}(s) \neq \varnothing\}$ be the total number of update epochs up to time t. We will then have $\mathcal{K}_n = \mathcal{K}(\inf\{t : n(t) = n\})$ and $N_k(n) = \sum_{r=1}^{n} \mathbb{1}(k \in \mathcal{K}_r)$, so the limit $\lim t_k(n)/n$ exists and is finite if and only if the limit $\lim N_k(n)/n$ exists and is positive.

[5]The reason for using this efficiency measure instead of the user's sum rate Φ directly, was to eliminate any scaling artifacts arising e.g. from Φ taking values in a very narrow band close to its maximum value.

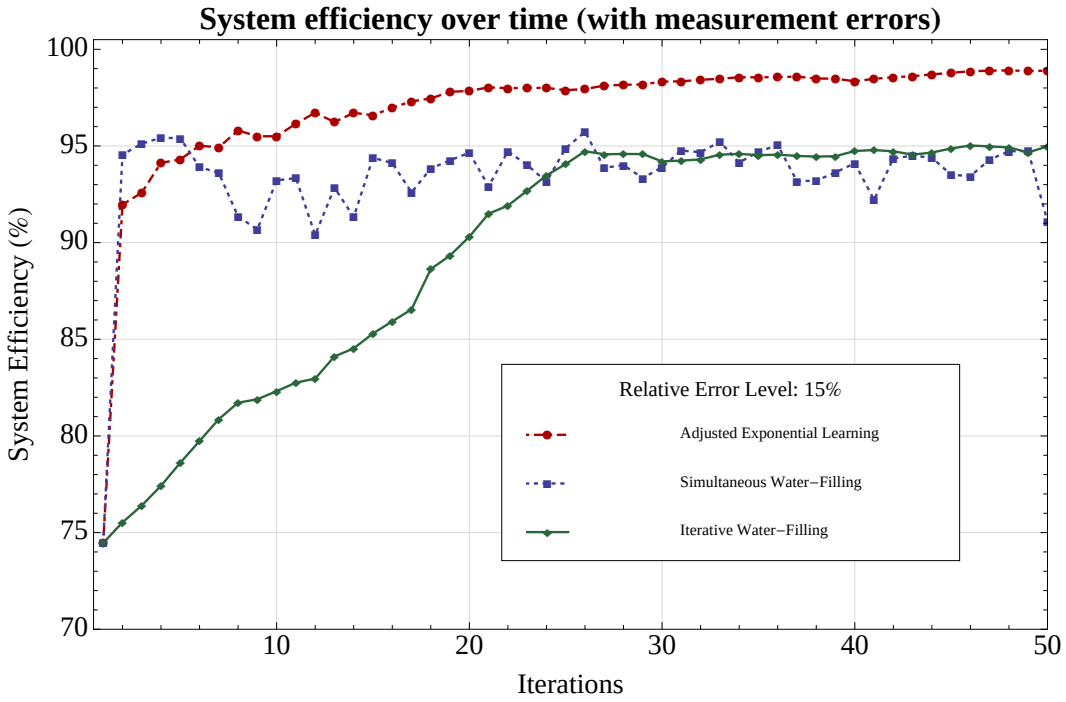


(a) The performance of AXL for different numbers of users.

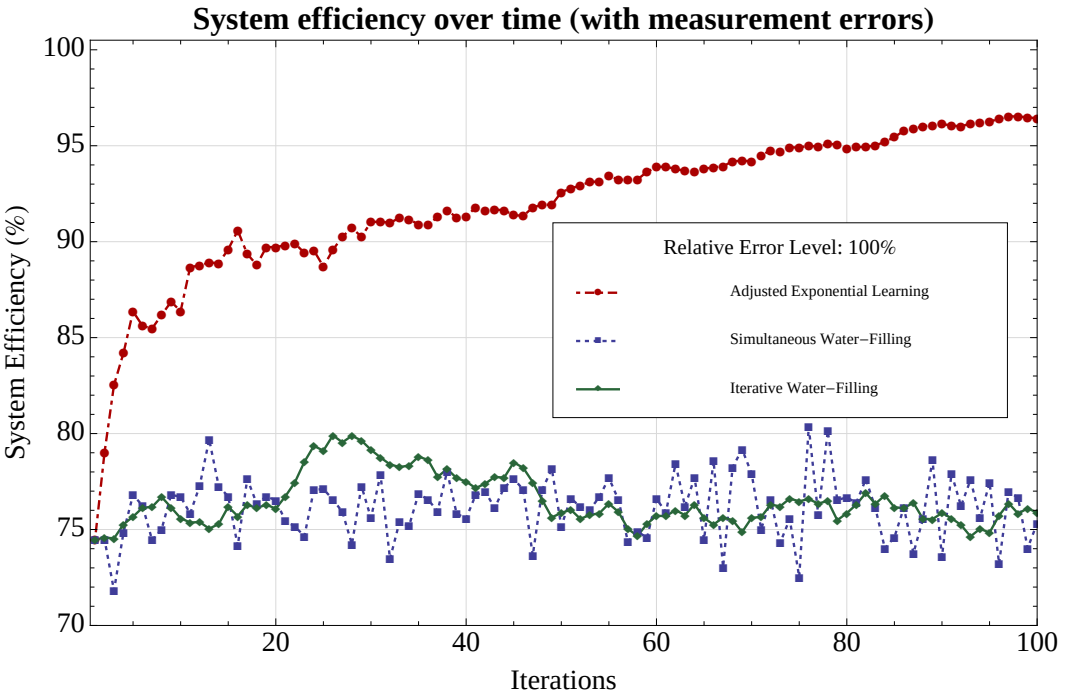


(b) Adjusted exponential learning vs. water-filling for $K = 25$ users.

Figure 2: The convergence speed of adjusted exponential learning as a function of the number of transmitters compared to water-filling techniques.



(a) Learning with an average relative error level of 15%.



(b) Learning with an average relative error level of 100%.

Figure 3: The robustness of entropy-driven learning in the presence of measurement errors: in contrast to water-filling methods, the entropy-driven learning attains the channel's sum capacity, even in the presence of very high measurement errors.

function of the number of transmitters, using existing water-filling methods as a benchmark. Specifically, in Fig. 2(a), we ran the AXL algorithm for a multi-user uplink MIMO system with $K = 10$, 25, 50 and 100 users using a large, constant step size; as a result of this parameter tuning, AXL achieves the system's sum capacity within one or two iterations, even for large numbers of users. Importantly, as can be seen in Fig. 2(b), this represents a marked improvement over water-filling methods, even in moderately-sized systems with $K = 25$ users: on the one hand, iterative water-filling (IWF) [4] is significantly slower than AXL (it requires $\mathcal{O}(K)$ iterations to achieve the same performance level as the first iteration of AXL), whereas simultaneous water-filling (SWF) [6] fails to converge altogether as a result of the users' optimum transmit directions turning out to be anti-aligned in the simulated channel realization.

The robustness of AXL is investigated in Fig. 3 where we simulate an uplink MIMO system consisting of $K = 25$ transmitters with imperfect channel state information (CSI) and noisy measurements at the receiver. For simplicity, we modeled these errors as additive i.i.d. zero-mean Gaussian perturbations to the matrices $\mathbf{V}_k = \mathbf{H}_k\mathbf{W}^{-1}\mathbf{H}_k^\dagger$ that are used in the update step of AXL, and the strength of these perturbations was controlled by the ratio of the errors' standard deviation to the matrix norm of $\mathbf{V}_k$ (so a relative error level of $\eta = 100\%$ means that the measurement error has the same magnitude as the measured variable). We then plotted the efficiency of AXL over time for average error levels of $\eta = 15\%$ and $\eta = 100\%$, and we ran the iterative and simultaneous water-filling algorithms with the same relative error levels for comparison. As can be seen in Fig. 3, the performance of water-filling methods remains acceptable at low error levels (attaining 90–95% of the system's sum capacity), but when the measurement noise gets higher, water-filling offers no perceptible advantage over the users' initial choice of covariance matrices; by contrast, AXL retains its convergence even for relative error levels as high as 100% – though, of course, its convergence speed is negatively affected.

Finally, to account for changing channel conditions, we also plotted the performance of AXL for non-static channels following the well-known Jakes model of Rayleigh fading [23]. More precisely, in Fig. 4, we consider a MIMO uplink system with 3 receive antennas, $K = 10$ users with 2 antennas each, transmitting at a frequency of $f = 2\,\text{GHz}$ and with average pedestrian velocities of $v = 5\,\text{km/h}$ (corresponding to a channel coherence time of 108 ms). We then ran the AXL algorithm with an update period of $\delta = 3\,\text{ms}$, and we plotted the achieved sum rate $\Phi(t)$ at time t versus the maximum attainable sum rate $\Phi_{\max}(t)$ given the channel matrices $\mathbf{H}_k(t)$ at time t, and versus the "uniform" sum rate that users could achieve by spreading their power uniformly over their antennas. As a result of its high convergence speed, AXL tracks the system's sum capacity remarkably well, despite the changing channel conditions; moreover, the sum rate difference between the learned transmit covariance profile and the uniform one shows that this tracking is not an artifact of the system's sum capacity falling within a narrow band of what could be attained by spreading power uniformly over antennas.

6 Conclusions

In this paper, we introduced a class of distributed algorithms for robust spectrum management in multi-user MIMO systems based on a temperature-adjusted variant of matrix exponential learning. By penalizing zero eigenvalues in the transmitters' covariance matrices, this entropy-driven adjustment generates a discrete-time algorithm which tracks the continuous-time dynamics of adjusted exponential learning and converges arbitrarily close to the system's optimum transmit spectrum. In contrast to traditional water-filling methods, the algorithm's convergence speed can be controlled by tuning the users' learning rate, so entropy-driven learning converges within a few iterations, even for large numbers of users and/or antennas per user. Finally, thanks to this adjustment term, the algorithm remains robust in the presence of stochastic perturbations: it converges even when the transmitters only have imperfect (or delayed) CSI at their disposal, or even if they update in a fully asynchronous manner and independently of one another.

The optimization method of adjusted exponential learning method actually applies to a wide range of

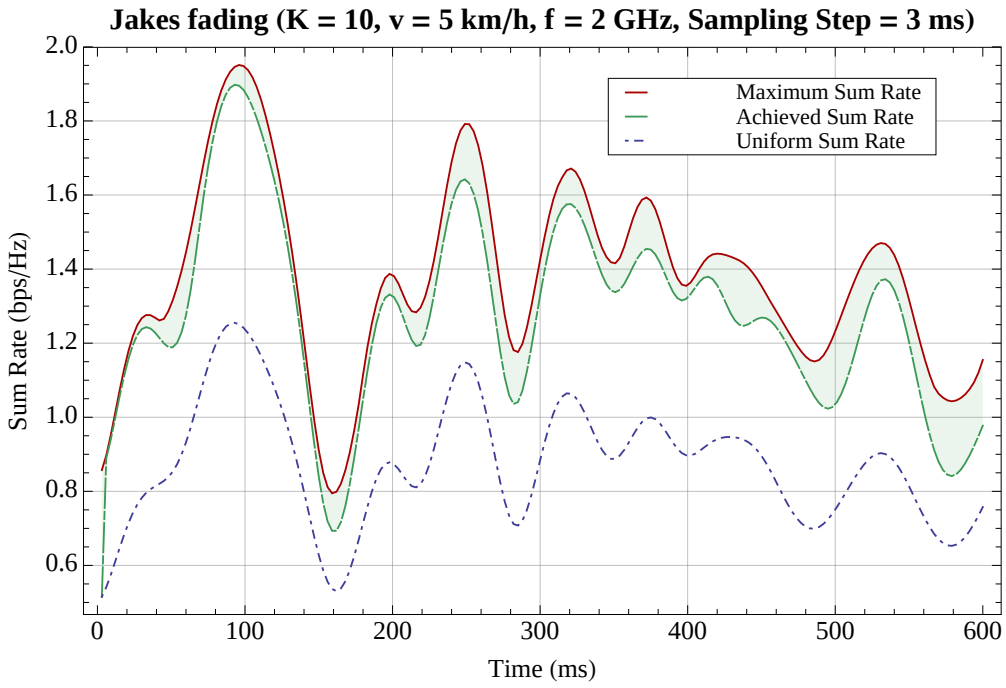


Figure 4: The performance of entropy-driven learning under changing channel conditions (following the Jakes model for Rayleigh fading with parameters indicated in the figure caption).

semidefinite problems; we focused here on the MIMO MAC for simplicity, but in the future, we aim to extend the method to more general channel models (such as the interference channel).

A Appendix: Convergence of adjusted exponential learning

We begin this appendix with a technical lemma that establishes the boundedness of the gradient matrices $\mathbf{V}_k$:

Lemma 7. *Let* $\mathbf{V}_k = \mathbf{H}_k^\dagger \mathbf{W}^{-1} \mathbf{H}_k$ *be the (conjugate) gradient of* Φ *as in* (9)*. Then,* $\mathbf{V}_k$ *is bounded in norm over* $\mathcal{X}$*.*

Proof. Since $\mathcal{X}$ is compact, the eigenvalues of $\mathbf{W} = \mathbf{I} + \sum_\ell \mathbf{H}_\ell \mathbf{Q}_\ell \mathbf{H}_\ell^\dagger$ are bounded above; moreover, with $\mathbf{H}_\ell \mathbf{Q}_\ell \mathbf{H}_\ell^\dagger \succcurlyeq 0$ for all $\ell \in \mathcal{K}$, the eigenvalues of $\mathbf{W}$ will also be bounded below by 1. It thus follows that the L^2 spectral norm $\|\mathbf{W}^{-1}\|_2$ (and, hence, any norm) of $\mathbf{W}^{-1}$ will be bounded over $\mathcal{X}$, so the same will hold for $\mathbf{V}_k = \mathbf{H}_k^\dagger \mathbf{W}^{-1} \mathbf{H}_k$ as well. □

Proof of Proposition 2. For simplicity, we will suppress the user index k and we will assume that $\operatorname{tr}(\mathbf{Q}) = 1$. Letting $\mathcal{Z} = \operatorname{tr}[\exp(\mathbf{Y})]$, we then get:

$$\begin{aligned}
\dot{\mathbf{Q}} &= \frac{1}{\mathcal{Z}} \frac{d}{dt} \exp(\mathbf{Y}) - \frac{\exp(\mathbf{Y})}{\mathcal{Z}^2} \frac{d\mathcal{Z}}{dt} \\
&= \mathcal{Z}^{-1} \int_0^1 e^{(1-s)\mathbf{Y}} \dot{\mathbf{Y}} e^{s\mathbf{Y}} \, ds - \mathcal{Z}^{-2} e^{\mathbf{Y}} \operatorname{tr}\left[\dot{\mathbf{Y}} e^{\mathbf{Y}}\right] \\
&= \int_0^1 \mathbf{Q}^{1-s} \mathbf{V}_\tau \mathbf{Q}^s \, ds - \operatorname{tr}(\mathbf{Q}\mathbf{V}_\tau) \, \mathbf{Q},
\end{aligned} \tag{16}$$

where the second line is an application of Fréchet's derivative formula for exponentials [24] and the last one follows by recalling that $\mathbf{Q} = \exp(\mathbf{Y})/\operatorname{tr}[\exp(\mathbf{Y})]$, so $\dot{\mathbf{Y}} = \mathbf{V} - \tau\mathbf{Y} = \mathbf{V} - \tau(\log\mathbf{Q} + \mathfrak{Z}\mathbf{I}) = \mathbf{V}_\tau - \tau\mathfrak{Z}\mathbf{I}$. □

Proof of Proposition 3. Suppressing the user index k for simplicity, and assuming without loss of generality that $\operatorname{tr}(\mathbf{Q}) = 1$, we will first show that the gradient of F with respect to $\mathbf{Q}$ is $\mathbf{V} - \tau(\mathbf{I} + \log\mathbf{Q})$. By the definition of $\mathbf{V}$, this boils down to showing that $\nabla_{\mathbf{Q}^*} h = \mathbf{I} + \log\mathbf{Q}$, so, in turn, it suffices to show that $\frac{d}{dt}\big|_{t=0} h(\mathbf{Q} + t\mathbf{Z}) = \operatorname{tr}[\mathbf{Z}(\mathbf{I} + \log\mathbf{Q})]$ for every Hermitian matrix $\mathbf{Z}$. Accordingly, letting $q_\alpha(t)$ denote the eigenvalues of $\mathbf{Q} + t\mathbf{Z}$, we will have $h(\mathbf{Q} + t\mathbf{Z}) = \sum_\alpha q_\alpha(t)\log q_\alpha(t)$ and hence $\frac{dh}{dt} = \sum_\alpha \dot{q}_\alpha \log q_\alpha + \sum_\alpha \dot{q}_\alpha$. However, if $\mathbf{u}_\alpha$ is the eigenvector corresponding to q_α, the relation $(\mathbf{Q} + t\mathbf{Z})\mathbf{u}_\alpha = q_\alpha\mathbf{u}_\alpha$ gives

$$\mathbf{Z}\mathbf{u}_\alpha + (\mathbf{Q} + t\mathbf{Z})\dot{\mathbf{u}}_\alpha = \dot{q}_\alpha\mathbf{u}_\alpha + q_\alpha\dot{\mathbf{u}}_\alpha, \tag{17}$$

so, after multiplying from the left by $\mathbf{u}_\alpha^\dagger$, we get:

$$\dot{q}_\alpha = \mathbf{u}_\alpha^\dagger\mathbf{Z}\mathbf{u}_\alpha + \mathbf{u}_\alpha^\dagger(\mathbf{Q} + t\mathbf{Z})\dot{\mathbf{u}}_\alpha - q_\alpha\mathbf{u}_\alpha^\dagger\dot{\mathbf{u}}_\alpha = \mathbf{u}_\alpha^\dagger\mathbf{Z}\mathbf{u}_\alpha. \tag{18}$$

By plugging this back to dh/dt, we then obtain $\frac{d}{dt}\big|_{t=0} h(\mathbf{Q}+t\mathbf{Z}) = \sum_\alpha(1+\log q_\alpha)\mathbf{u}_\alpha^\dagger\mathbf{Z}\mathbf{u}_\alpha = \operatorname{tr}[\mathbf{Z}(\mathbf{I}+\log\mathbf{Q})]$, as claimed.

For the second part of the proposition, our calculation above gives $\dot{F} = \operatorname{tr}[\dot{\mathbf{Q}}(\mathbf{V}_\tau - \tau\mathbf{I})] = \operatorname{tr}[\dot{\mathbf{Q}}\mathbf{V}_\tau]$ (recall that $\operatorname{tr}(\mathbf{Q}) = 1$, so $\operatorname{tr}(\dot{\mathbf{Q}}) = 0$). Invoking Proposition 2, we then obtain:

$$\dot{F} = \int_0^1 \operatorname{tr}\left[\mathbf{Q}^{1-s}\mathbf{V}_\tau\mathbf{Q}^s\mathbf{V}_\tau\right] ds - \operatorname{tr}(\mathbf{Q}\mathbf{V}_\tau)^2, \tag{19}$$

so our assertion follows from the Jensen-like inequality for matrices proved in Lemma 8 below. □

Lemma 8. *Let* $\mathbf{V}$ *be Hermitian and let* $\mathbf{Q} > 0$ *have* $\operatorname{tr}(\mathbf{Q}) = 1$. *Then,* $\operatorname{tr}(\mathbf{Q}^{1-s}\mathbf{V}\mathbf{Q}^s\mathbf{V}) \geq \operatorname{tr}(\mathbf{Q}\mathbf{V})^2$, *with equality iff* $\mathbf{V} \propto \mathbf{I}$.

Proof. Let $a = (1-s)/2$, $b = s/2$, and set $\mathbf{A} = \mathbf{Q}^{1/2}$, $\mathbf{B} = \mathbf{Q}^a\mathbf{V}\mathbf{Q}^b$. Then, the Cauchy–Schwarz inequality for matrices gives $\operatorname{tr}(\mathbf{A}\mathbf{A}^\dagger)\operatorname{tr}(\mathbf{B}\mathbf{B}^\dagger) \geq |\operatorname{tr}(\mathbf{A}\mathbf{B}^\dagger)|^2$ with equality iff $\mathbf{A} \propto \mathbf{B}$. On the other hand, we will have $\operatorname{tr}(\mathbf{A}\mathbf{A}^\dagger) = \operatorname{tr}\mathbf{Q} = 1$ and $\operatorname{tr}(\mathbf{B}\mathbf{B}^\dagger) = \operatorname{tr}[\mathbf{Q}^a\mathbf{V}\mathbf{Q}^b\mathbf{Q}^b\mathbf{V}\mathbf{Q}^a] = \operatorname{tr}[\mathbf{Q}^{1-s}\mathbf{V}\mathbf{Q}^s\mathbf{V}]$, leading to:

$$\begin{aligned} 1 \cdot \operatorname{tr}[\mathbf{Q}^{1-s}\mathbf{V}\mathbf{Q}^s\mathbf{V}] &\geq |\operatorname{tr}[\mathbf{Q}^{1/2}\mathbf{Q}^{s/2}\mathbf{V}\mathbf{Q}^{(1-s)/2}]|^2 \\ &= |\operatorname{tr}(\mathbf{Q}\mathbf{V})|^2 = \operatorname{tr}(\mathbf{Q}\mathbf{V})^2, \end{aligned} \tag{20}$$

where the last equality follows from the fact that $\operatorname{tr}(\mathbf{Q}\mathbf{V})$ is real. This inequality holds as an equality if and only if $\mathbf{Q}^{1/2} \propto \mathbf{Q}^a\mathbf{V}\mathbf{Q}^b$, so with $a + b = 1/2$, this last condition is equivalent to $\mathbf{V} \propto \mathbf{I}$, as claimed. □

B Unbiased estimators for $\mathbf{W}^{-1}$ and $\mathbf{V}_k$

In this appendix, we present an unbiased procedure with which the receiver and the transmitters may estimate the physical quantities involved in the the discretization (11) of (XL_τ), based at each step on direct signal measurements that may be subject to observational (but not systematic) errors.

We first consider the random perturbations induced on the measurement of $\mathbf{W}$ by signal sampling at the receiver end. To wit, recall that $\mathbf{W}$ is just the covariance matrix of the aggregate received signal $\mathbf{y} \in \mathbb{C}^{m_0}$: indeed, $\mathbb{E}[\mathbf{y}\mathbf{y}^\dagger] = \mathbb{E}\left[\mathbf{z}\mathbf{z}^\dagger\right] + \sum_k \mathbf{H}_k\,\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^\dagger\right]\mathbf{H}_k^\dagger = \mathbf{I} + \sum_k \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^\dagger = \mathbf{W}$, so $\mathbf{P} \equiv \mathbf{W}^{-1}$ is simply the precision (inverse covariance) matrix of a multivariate Gaussian random variable.

An unbiased estimate for the covariance $\mathbf{W}$ of $\mathbf{y}$ may be obtained from a systematically unbiased sample $\mathbf{y}_1, \dots, \mathbf{y}_M$ of $\mathbf{y}$ by means of the classical estimator $\hat{\mathbf{W}} = \frac{1}{M}\sum_{j=1}^M \mathbf{y}_j\mathbf{y}_j^\dagger$.[6] On the other hand, given

[6] Since the expected value $\mathbb{E}[\mathbf{y}] = 0$ of $\mathbf{y}$ need not be estimated itself from the data sample, we do not need to include the $M/(M-1)$ bias correction factor in the estimate of $\mathbf{W}$.

that $\hat{\mathbf{W}}^{-1}$ is a biased estimator of $\mathbf{W}^{-1}$ (and hence introduces a systematic error to the measurement process) [21], we cannot use this classical covariance estimate for $\mathbf{W}^{-1}$. Instead, following [21], an unbiased estimate of the precision matrix $\mathbf{P} = \mathbf{W}^{-1}$ of $\mathbf{y}$ will be given by the corrected expression:

$$\hat{\mathbf{P}} = \frac{M - m_0 - 1}{M} \hat{\mathbf{W}}^{-1}, \tag{21}$$

where $\hat{\mathbf{W}} = \frac{1}{M} \sum_{j=1}^{M} \mathbf{y}_j \mathbf{y}_j^\dagger$ as before. Thus, to obtain $\mathbf{W}^{-1}$, the receiver only needs to take $M > m_0 + 1$ periodic measurements of $\mathbf{y}$ between iteration cycles, and then broadcast the unbiased estimate $\hat{\mathbf{P}}$ of $\mathbf{W}^{-1}$ to the network's users.

Similarly, in the absence of perfect channel state information at the transmitter, the users will need to obtain an unbiased estimate of the unilateral gradient matrices $\mathbf{V}_k = \mathbf{H}_k^\dagger \mathbf{W}^{-1} \mathbf{H}_k$ from the broadcasted value of $\mathbf{W}$ and using imperfect measurements of their channel matrices $\mathbf{H}_k$. However, an added complication here is that the estimated matrix $\hat{\mathbf{V}}_k$ must be itself Hermitian – otherwise, $\mathbf{Q}_k$ need not be positive-definite and the algorithm might fail to be well-posed.

To accommodate this requirement, an unbiased Hermitian estimate for $\mathbf{V}_k$ may be obtained from a sample $\mathbf{H}_{k,1}, \ldots, \mathbf{H}_{k,M}$ ($M > 1$) of $\mathbf{H}_k$ which is subject to zero-mean observational errors (assumed independent across users) via the expression:

$$\hat{\mathbf{V}}_k = \frac{1}{2(M-1)} \sum_{j=1}^{M-1} \left(\mathbf{H}_{k,j}^\dagger \hat{\mathbf{P}} \mathbf{H}_{k,j+1} + \mathbf{H}_{k,j+1}^\dagger \hat{\mathbf{P}} \mathbf{H}_{k,j} \right), \tag{22}$$

where $\hat{\mathbf{P}}$ is the unbiased estimate (21) of $\mathbf{W}^{-1}$. Indeed, if the sample measurements $\mathbf{H}_{k,j}$ are independent realizations of some random variable $\hat{\mathbf{H}}_k$ with $\mathbb{E}[\hat{\mathbf{H}}_k] = \mathbf{H}_k$, we will have:

$$\begin{aligned} \mathbb{E}[\hat{\mathbf{V}}_k] &= \frac{1}{2(M-1)} \sum_{j=1}^{M-1} \mathbb{E}\left[\mathbf{H}_{k,j}^\dagger \hat{\mathbf{P}} \mathbf{H}_{k,j+1} + \mathbf{H}_{k,j+1}^\dagger \hat{\mathbf{P}} \mathbf{H}_{k,j} \right] \\ &= \frac{1}{2(M-1)} \sum_{j=1}^{M-1} 2\, \mathbb{E}[\hat{\mathbf{H}}_k^\dagger]\, \mathbb{E}[\hat{\mathbf{P}}]\, \mathbb{E}[\hat{\mathbf{H}}_k] = \mathbf{H}_k^\dagger \mathbf{W}^{-1} \mathbf{H}_k, \end{aligned}$$

where we have used the independence of the samples to decorrelate the expectations in the second equality, and relied on the unbiasedness of $\hat{\mathbf{P}}$ and $\hat{\mathbf{H}}_k$ for the last one. Thus, with $\mathbb{E}[\hat{\mathbf{V}}_k] = \mathbf{V}_k$, our construction of an unbiased estimator for $\mathbf{V}_k$ is complete.

References

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

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

[3] R. S. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: capacity region and multiuser water-filling," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 773–785, May 1993.

[4] 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.

[5] G. Scutari, D. P. Palomar, and S. Barbarossa, "Competitive design of multiuser MIMO systems based on game theory: a unified view," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1089–1103, September 2008.

[6] G. Scutari, "Simultaneous iterative water-filling for Gaussian frequency-selective interference channels," in *ISIT '06: Proceedings of the 2006 International Symposium on Information Theory*, 2006.

[7] P. Mertikopoulos, E. V. Belmega, A. L. Moustakas, and S. Lasaulce, "Dynamic power allocation games in parallel multiple access channels," in *ValueTools '11: Proceedings of the 5th International Conference on Performance Evaluation Methodologies and Tools*, 2011.

[8] 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.

[9] D. Lamberton, G. Pagès, and P. Tarrès, "When can the two-armed bandit algorithm be trusted?" *The Annals of Applied Probability*, vol. 14, no. 3, pp. 1424–1454, 2004.

[10] V. S. Borkar, *Stochastic approximation*. Cambridge University Press and Hindustan Book Agency, 2008.

[11] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology*. NJ, USA: Prentice Hall, 1999.

[12] E. V. Belmega, S. Lasaulce, M. Debbah, and A. Hjørungnes, "Distributed power allocation policies in MIMO channels," in *EUSIPCO '10: Proceedings of the 2010 European Signal Processing Conference*, 2010.

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

[14] M. Benaïm, "Dynamics of stochastic approximation algorithms," *Séminaire de probabilités de Strasbourg*, vol. 33, 1999.

[15] P. D. Taylor and L. B. Jonker, "Evolutionary stable strategies and game dynamics," *Mathematical Biosciences*, vol. 40, no. 1-2, pp. 145–156, 1978.

[16] P. Mertikopoulos, E. V. Belmega, A. L. Moustakas, and S. Lasaulce, "Distributed learning policies for power allication in multiple access channels," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 96–106, January 2012.

[17] P. Coucheney, B. Gaujal, and P. Mertikopoulos, "Entropy-driven dynamics and robust learning procedures in games," Inria, Tech. Rep., 2013, http://arxiv.org/abs/1303.2270.

[18] L. D. Landau and E. M. Lifshitz, "Statistical physics," in *Course of Theoretical Physics*. Oxford: Pergamon Press, 1976, vol. 5.

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

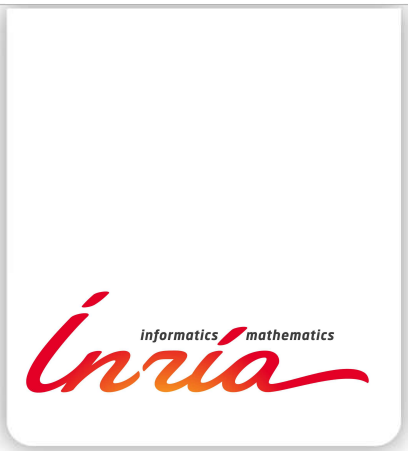
[20] C. Berge, *Topological Spaces*. New York: Dover, 1997.

[21] T. Anderson, *An introduction to multivariate statistical analysis*, 3rd ed. Wiley-Interscience, 2003.

[22] C. Moler and C. F. 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.

[23] G. Calcev, D. Chizhik, B. Göransson, S. Howard, H. Huang, A. Kogiantis, A. F. Molisch, A. L. Moustakas, D. Reed, and H. Xu, "A wideband spatial channel model for system-wide simulations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, p. 389, March 2007.

[24] N. J. Higham, *Functions of Matrices: Theory and Computation*. Philadelphia, PA: SIAM, 2008.



RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399