Entropy-Driven Optimization Dynamics for Gaussian Vector Multiple Access Channels

Panayotis Mertikopoulos
Centre National de la Recherche Scientifique (CNRS)
University of Grenoble (LIG laboratory)
E-mail: panayotis.mertikopoulos@imag.fr

Aris L. Moustakas
National & Kapodistrian University of Athens
Department of Physics
E-mail: arislm@phys.uoa.gr

***Abstract*—We develop a distributed optimization method for finding optimum input signal covariance matrices in Gaussian vector multiple access channels (solving also an equivalent game-theoretic formulation of the problem). Since ordinary gradient ascent violates the problem's semidefiniteness constraints, we introduce an entropic barrier term whose Hessian allows us to write a gradient-like flow which behaves well with respect to the problem's constraints, and which allows users to achieve the channel's capacity. The algorithm's convergence speed can be tuned by adjusting the underlying entropy function (and thus changing the spectral geometry of the cone of semidefinite matrices), so, in practice, users are able to achieve capacity within a few iterations, even for large numbers of users and/or antennas per user.**

***Index Terms*—Distributed optimization; entropy functions; Hessian gradient flows; multiple access channel; MIMO.**

I. Introduction

Owing to the seminal prediction [1, 2] that the use of multiple antennas in transmitting and receiving radio signals can lead to substantial performance gains, multiple-input and multiple-output (MIMO) technologies have become 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 that some of these protocols are decentralized (e.g. the latter two), it is not clear how users may benefit from the use of multiple antenna technologies at a network level; worse yet, even 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 in order to bring about the advantages of using multiple antennas, distributed optimization methods become a necessity.

Accordingly, given that the theoretical limits of MIMO systems still elude us even in basic network models (such as the interference channel), it is useful to start instead with the mutual information for Gaussian input and noise, and to optimize the input signal distribution of each transmitter in the presence of interference from all other users. One such Gaussian channel model which has attracted significant interest in the wireless literature is the MIMO multiple access channel (MAC) which consists of a single receiver (or a suitably interconnected set thereof) who decodes the simultaneous signals of several (non-cooperative and mutually independent) transmitters.

From a theoretical point of view, attaining the capacity of this channel means solving a nonlinear optimization problem over a set of positive-definite matrices representing the users' feasible input covariance matrices [2]. However, due to the implicit nature of the problem's semidefiniteness constraints, standard gradient ascent and/or interior point methods do not readily apply, so the problem is usually solved by water-filling techniques [3], properly adapted to multi-user environments [4, 5]. Unfortunately however, iterative and/or sequential water-filling algorithms converge very slowly when the number of users is large, whereas the convergence of faster, simultaneous water-filling methods [6] depends on the channels satisfying certain "mild-interference" conditions that fail to hold even in the simpler parallel multiple access channel (PMAC) setting (itself a reduced version of the full MIMO MAC problem) [7].

In this paper, we take a different approach from water-filling and develop a class of distributed dynamical optimization schemes based on the Hessian–Riemannian approach of [8, 9]. Specifically, we endow the cone of semidefinite matrices with an entropic barrier function, and we use the Hessian of this function to derive a matrix-valued gradient-like system for the problem at hand. Importantly, in the special case of the quantum von Neumann entropy, our entropy-driven method reduces to the matrix exponential learning algorithm of [10, 11] and thus serves to explain the convergence properties of said algorithm from a more systematic point of view.

Of course, different entropy functionals (such as the von Neumann entropy of quantum statistical mechanics or the matrix equivalent of the Tsallis entropy) yield different dynamics, but it turns out that all such optimization procedures converge to a solution of our signal optimization problem – albeit at different rates.[1] In fact, by suitably tuning the underlying entropy function, we show that users are able to achieve the channel's sum capacity within a few iterations (even for large numbers of users and/or antennas per user); as such, even in the presence of rapidly changing channel conditions (due to e.g. fading), users will still able to track the optimum transmit spectrum in a small fraction of the channel's coherence time.

This work was funded by the research program CROWN, through the Operational Program Education and Lifelong Learning 2007-2013 of NSRF, co-financed by EU and Greek national funds.

[1] In fact, the resulting optimization methods apply to a much wider class of semidefinite problems; we only focus here on vector Gaussian multiple access channels for concreteness and simplicity.

Paper Outline: After presenting our system model in Section II, we develop the mathematical groundwork of our paper in Section III. Our main theoretical results are then presented in a self-consistent manner in Section IV, while Section V is devoted to numerical simulations.

II. System Model

Our system model comprises a vector Gaussian multiple access channel with a finite set of wireless users $k \in \mathcal{K} \equiv \{1, \dots, K\}$, each equipped with m_k antennas, and each transmitting simultaneously to a wireless base receiver with n 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 $\mathbf{y} \in \mathbb{C}^n$ denotes the aggregate signal reaching the receiver, $\mathbf{x}_k \in \mathbb{C}^{m_k}$ is the message transmitted by user $k \in \mathcal{K}$, $\mathbf{H}_k \in \mathbb{C}^{n \times m_k}$ is the associated $n \times m_k$ (complex) channel matrix, and $\mathbf{z} \in \mathbb{C}^n$ is the noise in the channel, 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).

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

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

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

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

Thus, assuming successive interference cancellation (SIC) at the receiver, it is well-known that for a given profile $\mathbf{Q} = (\mathbf{Q}_1, \dots, \mathbf{Q}_K)$ of covariance matrices, the maximum information transmission rate will be achieved for Gaussian codebooks and will be given by the well-known expression [2]:

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

As such, given that users can only transmit with finite power $\operatorname{tr}(\mathbf{Q}_k) = P_k$,we are led to the sum rate maximization problem [4, 5]:

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

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 matrices of user k.

Clearly, the solution points of (SRP) correspond to a centralized, 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, due to decentralization or a very large number of users), then, by means of single user decoding (SUD) techniques, the users' objective would be instead 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

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

is the multi-user interference-plus-noise matrix of user k. In this way, from a game-theoretic perspective, the system will be at *Nash equilibrium* (also known as a *selfish optimum*) when no user can improve his individual achievable rate u_k by unilaterally changing his signal covariance matrix $\mathbf{Q}_k$. More specifically, we will say that $\mathbf{Q}$ is a *Nash profile* 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 we have employed the standard shorthand for unilateral deviations: $(\mathbf{Q}_k'; \mathbf{Q}_{-k}) \equiv (\mathbf{Q}_1, \dots, \mathbf{Q}_k', \dots, \mathbf{Q}_K)$.

A striking feature of this model is that users are individually aligned with the global objective function Φ, so the solutions of (SRP) coincide with the Nash equilibria of (5). Indeed, as was shown in [12], Φ is a *potential* [13] for the game defined by (5) in the sense that the individual rates u_k satisfy the property $u_k(\mathbf{Q}_k; \mathbf{Q}_{-k}) - u_k(\mathbf{Q}_k'; \mathbf{Q}_{-k}) = \Phi(\mathbf{Q}_k; \mathbf{Q}_{-k}) - \Phi(\mathbf{Q}_k'; \mathbf{Q}_{-k})$. As a result, given that Φ is concave, *the game's Nash equilibria* (7) *will be* precisely *the maximum points of* (SRP).

III. Matrix Entropy Functions and Hessian Flows

A natural (but ultimately doomed) approach to solving the maximization problem (SRP) would be to ascend the matrix differential $d\Phi \equiv \mathbf{V} = (\mathbf{V}_1, \dots, \mathbf{V}_K)$ of Φ w.r.t. $\mathbf{Q}$, where

$$\mathbf{V}_k = \frac{\partial \Phi}{\partial \mathbf{Q}_k^*} = \mathbf{H}_k^\dagger \mathbf{W}^{-1} \mathbf{H}_k, \tag{8}$$

and $\mathbf{W} = \mathbf{I} + \sum_{\ell \in \mathcal{K}} \mathbf{H}_\ell \mathbf{Q}_\ell \mathbf{H}_\ell^\dagger$ is the aggregate signal-plus-noise covariance matrix at the receiver end. The benefit of this approach is that Φ increases along the trajectories of the gradient system $\dot{\mathbf{Q}} = \mathbf{V}$, but its crippling disadvantage is that following the direction of steepest ascent in this way will generically violate the feasibility constraints of (SRP). As a result, this section will be devoted to developing a dynamical scheme which behaves well with respect to the semidefiniteness constraints $\mathbf{Q}_k \succcurlyeq 0$ of (SRP).

Given that $\mathcal{X} = \prod_k \mathcal{X}_k$ is a convex set, (SRP) can be viewed more simply as a constrained concave maximization problem in some real space $\mathbb{C}^N$. Thus, if $\mathbf{x} = \operatorname{vec}(\mathbf{Q}) \in \mathbb{C}^N$ is the vectorization of $\mathbf{Q}$, the main question that arises is what to use in place of the gradient system $\dot{\mathbf{x}} = \mathbf{v} \equiv \operatorname{vec}(\mathbf{V})$. To that end, we will consider the alternative gradient system

$$\dot{\mathbf{x}} = \mathbf{\Omega} \mathbf{v} \tag{9}$$

where $\mathbf{\Omega}$ is a suitably chosen positive-definite matrix $\mathbf{\Omega}$ which vanishes at the boundary of $\mathcal{X}$ (so that solution orbits of (9) remain in $\mathcal{X}$ for all time while still ascending Φ).

In the case where the feasible region is just the unit simplex Δ of $\mathbb{C}^N$, the analysis of [8, 9] shows that a particularly well-suited candidate for $\mathbf{\Omega}$ is the Hessian of a *negative entropy function* on Δ, viz. a strictly convex function $h\colon \Delta \to \mathbb{R}$ such that $|dh| \to +\infty$ near $\operatorname{bd}(\Delta)$. To wit, in the case of the negative Shannon entropy $h(x) = \sum_j x_j \log x_j$, the calculations of [9] (see also [14]) show that the system $\dot{\mathbf{x}} = \operatorname{Hess}(h)^{-1} \cdot \mathbf{v}$ is equivalent to the *replicator equation* of evolutionary biology

$$\dot{x}_j = x_j \left(\tfrac{\partial \phi}{\partial x_j} - \sum\nolimits_\ell x_\ell \tfrac{\partial \phi}{\partial x_\ell}\right), \tag{10}$$

which is known to converge to the maximum of ϕ whenever $\phi\colon \Delta \to \mathbb{R}$ is concave [9].

In our matrix setting, the definition of (negatives of) entropy functions on the cone $\mathcal{C}$ of positive-semidefinite $m \times m$ matrices may be extended as follows:

Definition 1. Let $\theta\colon [0, +\infty) \to \mathbb{R} \cup \{+\infty\}$ be a smooth function such that *i.* $\theta(x) < +\infty$ for all $x > 0$, *ii.* $\lim_{x\to 0^+} \theta'(x) = -\infty$, and *iii.* $\theta''(x) > 0$ for all x. Furthermore, let $\mathbf{Q} \in \mathcal{C}$, and set $\theta(\mathbf{Q}) \equiv \mathbf{U}\mathbf{\Theta}\mathbf{U}^\dagger$ where $\mathbf{\Theta} = \operatorname{diag}(\theta(q_1), \dots, \theta(q_m))$, $q_1, \dots, q_m$ are the eigenvalues of $\mathbf{Q}$ and $\mathbf{U}$ is the corresponding matrix of eigenvectors. Then, we define the *matrix entropy function associated to* θ as

$$h(\mathbf{Q}) = \operatorname{tr}(\theta(\mathbf{Q})) = \textstyle\sum_j \theta(q_j). \quad \mathbf{Q} \in \mathcal{C}, \tag{11}$$

In this case, θ will be called the *kernel* of h.

Notational Convention. In statistical physics and information theory, the entropy is a *concave* function, so Definition 1 describes *minus* entropies. However, carrying around the adjective negative would be needlessly frustrating, so in the rest of the paper, "entropy" will be taken to mean "minus entropy".

As an example of the above construction, note that the *von Neumann quantum entropy* $h_1(\mathbf{Q}) = \operatorname{tr}(\mathbf{Q} \log \mathbf{Q})$ corresponds simply to the entropic kernel $\theta_1(x) = x \log x$, so if $\mathbf{Q}$ is diagonal, we fall back to the familiar expression $\sum_j q_j \log q_j$. Similarly, if we let $\theta_r(x) = (1-r)^{-1}(x - x^r)$ for $0 < r < 1$, then we readily obtain the *Tsallis quantum entropy* $h_r(\mathbf{Q}) = (1-r)^{-1} \operatorname{tr}(\mathbf{Q} - \mathbf{Q}^r)$ which tends to the ordinary Shannon entropy as $r \to 1^-$.

With this in mind, the Hessian gradient flow system which corresponds to h will be

$$\dot{\mathbf{Q}} = \operatorname{Hess}(h)^{-1} \cdot \mathbf{V} \tag{HGF}$$

where $\mathbf{V} = \frac{\partial \Phi}{\partial \mathbf{Q}^*}$ is the matrix differential of Φ as before, and $\operatorname{Hess}(h)$ denotes the Hessian *tensor* of h. Importantly, the "$\cdot$" sign in (HGF) does *not* denote matrix multiplication, but the *application* of $\operatorname{Hess}(h)$ to $\mathbf{V}$: to view the operator $\operatorname{Hess}(h)$ as a (positive-definite) matrix $\mathbf{Q}$ and $\mathbf{V}$ would have to be vectorized, in which case we fall back to (9).

That said, since h is a function of matrices, it is all but impossible to obtain an analytic expression (vectorized or not) for its Hessian operator $\operatorname{Hess}(h)$ via ordinary matrix calculus, so we will have to take a radically different approach in order to be able to use (HGF). To that end, we will thus make use of the *Bregman divergence associated to h* defined as:

$$D_h(\mathbf{P} \,\|\, \mathbf{Q}) \equiv h(\mathbf{P}) - h(\mathbf{Q}) - \operatorname{tr}(dh \cdot (\mathbf{P} - \mathbf{Q})), \quad \mathbf{P}, \mathbf{Q} \in \mathcal{C}, \tag{12}$$

where $dh = \frac{\partial h}{\partial \mathbf{Q}^*}$ is the matrix differential of h. From the general theory of Bregman functions [15], it is known that D_h defined as above is convex in $\mathbf{Q}$ and positive everywhere except at $\mathbf{P}$ where it vanishes (in the case of the von Neumann entropy, this result is known as Klein's inequality [16]). For our purposes however, the most important property of D_h is that it characterizes Hessian flows as follows:

Proposition 1. *Let* $h(\mathbf{Q}) = \operatorname{tr}(\theta(\mathbf{Q}))$ *be a matrix entropy function with kernel* θ *and let* $\operatorname{Hess}(h)$ *be its Hessian operator. Then,* $\mathbf{Q}(t)$ *is a solution trajectory of the Hessian gradient flow system* (HGF) *if and only if*

$$\frac{d}{dt} D_h(\mathbf{P} \,\|\, \mathbf{Q}(t)) = \operatorname{tr}(\mathbf{V}(\mathbf{P} - \mathbf{Q})) \quad \textit{for all } \mathbf{P} \in \mathcal{C}. \tag{13}$$

Sketch of proof: By differentiating the Bregman divergence with respect to time, we get

$$\tfrac{d}{dt} D_h(\mathbf{P} \,\|\, \mathbf{Q}(t)) = \operatorname{tr}(\dot{\mathbf{Q}} \cdot \operatorname{Hess}(h) \cdot (\mathbf{P} - \mathbf{Q})), \tag{14}$$

so the direct implication follows immediately. To obtain the converse, note that if (14) holds, then it holds for every Hermitian matrix $\mathbf{P} - \mathbf{Q}$, so we must have $\dot{\mathbf{Q}} \cdot \operatorname{Hess}(h) = \mathbf{V}$. ■

The *raison d'être* of this proposition is that if we manage to design a dynamical system which satisfies (13), then this system will automatically be a Hessian gradient flow, and will thus converge to a solution of Φ by the general theory of [8, 9]. Albeit roundabout (and not particularly simpler-sounding), this approach will enable us to obtain a useable form for (HGF) so our first step will be to take advantage of the trace in (12) in order to obtain an explicit expression for D_h:

Lemma 1. *Let* $h(\mathbf{Q}) = \operatorname{tr}(\theta(\mathbf{Q}))$ *be as in* (11)*. Then, the Bregman divergence of h will be:*

$$D_h(\mathbf{P} \,\|\, \mathbf{Q}) = \operatorname{tr}\left[\theta(\mathbf{P}) - \theta(\mathbf{Q})\right] - \operatorname{tr}\left[\theta'(\mathbf{Q})(\mathbf{P} - \mathbf{Q})\right]. \tag{15}$$

Sketch of proof: Let $\mathbf{Z} = \mathbf{P} - \mathbf{Q}$ and set $f(\xi) = h(\mathbf{Q} + \xi\mathbf{Z})$. Then, if $q_j(\xi)$ is an eigenvalue of $\mathbf{Q} + \xi\mathbf{Z}$, we will have $\dot{q}_j = P_{jj} - q_j$ where $P_{jj} = \mathbf{u}_j^\dagger \mathbf{P} \mathbf{u}_j$ is the (j, j)-th element of $\mathbf{P}$ expressed in the eigenbasis of $\mathbf{Q}$. Then, a differentiation of $f(\xi)$ readily yields $f'(0) = \sum_j \dot{q}_j(0)\theta'(q_j(0)) = \operatorname{tr}(\theta'(\mathbf{Q})\mathbf{P}) - \operatorname{tr}(\theta'(\mathbf{Q})\mathbf{Q})$ and our claim follows by noting that $\operatorname{tr}(dh \cdot (\mathbf{P} - \mathbf{Q})) = f'(0)$. ■

Remark. As a corollary of this lemma, note that in the case of the von Neumann entropy we immediately obtain the well-known expression for the quantum Kullback-Leibler divergence $D_{\mathrm{KL}}(\mathbf{P} \,\|\, \mathbf{Q}) = \operatorname{tr}(\mathbf{P} \log \mathbf{P}) - \operatorname{tr}(\mathbf{P} \log \mathbf{Q})$ [16].

In light of the above, our strategy to determine the Hessian operator $\operatorname{Hess}(h)^{-1}$ of h will be as follows: letting $\dot{\mathbf{Q}} = \mathbf{V}^\sharp$ where $\mathbf{V}^\sharp$ is a matrix to be determined, we will apply (13) in order to get a relation between $\mathbf{V}^\sharp$ and $\mathbf{V}$; then, solving this relation for $\mathbf{V}^\sharp$, we will derive the required operation of $\operatorname{Hess}(h)^{-1}$ on $\mathbf{V}$. Unfortunately, this is still impossible to do in an arbitrary basis, so we will work in the eigenbasis of $\mathbf{Q}$ and proceed to derive the evolution of the eigen-decomposition of $\mathbf{Q}$ when $\dot{\mathbf{Q}} = \mathbf{V}^\sharp$:

Proposition 2. *Let* $\dot{\mathbf{Q}} = \operatorname{Hess}(h)^{-1} \cdot \mathbf{V} \equiv \mathbf{V}^\sharp$ *and let* $\{q_j, \mathbf{u}_j\}$ *be a smooth eigen-decomposition of* $\mathbf{Q}(t)$*. If* $V^\sharp_{jk} = \mathbf{u}_j^\dagger \mathbf{V}^\sharp \mathbf{u}_k$ *denotes the* (j,k)*-th component of* $\mathbf{V}^\sharp$ *in the eigenbasis of* $\mathbf{Q}$*, then:*

$$\dot{q}_j = V^\sharp_{jj}, \tag{16a}$$

$$\dot{\mathbf{u}}_k = \textstyle\sum_{j \neq k} \frac{V^\sharp_{jk}}{q_k - q_j} \mathbf{u}_j. \tag{16b}$$

Sketch of proof: Simply note that the equation $\dot{\mathbf{Q}} = \mathbf{V}^\sharp$ implies the identity $\dot{q}_k \delta_{jk} = (q_j - q_k)\mathbf{u}_j^\dagger \cdot \dot{\mathbf{u}}_k + V^\sharp_{jk}$. ■

With this proposition at hand, we may finally calculate $\mathbf{V}^\sharp$:

Proposition 3. *Let* $\mathbf{V}^\sharp = \mathrm{Hess}(h)^{-1}\cdot\mathbf{V}$ *and let* $\{q_j, \mathbf{u}_j\}$ *be a smooth eigen-decomposition of* $\mathbf{Q}\in\mathcal{C}$. *If* $V^\sharp_{jk} = \mathbf{u}_j^\dagger\mathbf{V}^\sharp\mathbf{u}_k$, *then:*

$$V^\sharp_{jk} = \begin{cases} \frac{1}{\theta''(q_j)}V_{jj} & \text{if } j = k,\\ \frac{q_j-q_k}{\theta'(q_j)-\theta'(q_k)}V_{jk} & \text{otherwise.}\end{cases} \tag{17}$$

Sketch of proof: Using Lemma 1 to differentiate $H(t) \equiv D_h(\mathbf{P}\,\|\,\mathbf{Q}(t))$, we have:

$$\dot{H}(t) = \tfrac{d}{dt}\Big[-\textstyle\sum_j\theta(q_j) - \mathrm{tr}(\mathbf{P}\cdot\theta'(\mathbf{Q})) + \sum_j q_j\theta'(q_j)\Big]. \tag{18}$$

Of these terms, the first and the third may be calculated immediately using Proposition 2: $\frac{d}{dt}\sum_j\theta(q_j) = \mathrm{tr}(\mathbf{\Theta}'\mathbf{V}^\sharp)$ and $\frac{d}{dt}\sum_j q_j\theta'(q_j) = \mathrm{tr}(\mathbf{\Theta}'\mathbf{V}^\sharp) + \mathrm{tr}(\mathbf{Q}\mathbf{\Theta}''\mathbf{V}^\sharp)$, where $\mathbf{\Theta}' = \mathrm{diag}(\theta'(q_1),\dots,\theta'(q_m))$ and $\mathbf{\Theta}''$ is defined analogously. For the second term now, since $\mathbf{P}$ and $\mathbf{Q}$ do not necessarily commute, write $\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$ and $\dot{\mathbf{U}} = \mathbf{U}\mathbf{A}$ where $\mathbf{A}$ is the skew-Hermitian matrix defined by (16b). Then, by expanding $\frac{d}{dt}\mathrm{tr}(\mathbf{P}\cdot\theta'(\mathbf{Q})) = \mathrm{tr}(\mathbf{PUA\Theta'U}^\dagger) - \mathrm{tr}(\mathbf{PU\Theta'AU}^\dagger) + \mathrm{tr}(\mathbf{PU}\dot{\mathbf{\Theta}}'\mathbf{U}^\dagger)$, Proposition 2 and some algebra yield $\frac{d}{dt}\mathrm{tr}(\mathbf{P}\cdot\theta'(\mathbf{Q})) = \sum_{j,k}P_{kj}\frac{\theta'(q_k)-\theta'(q_j)}{q_k-q_j}V^\sharp_{jk}$ where $P_{kj} = \mathbf{u}_k^\dagger\mathbf{P}\mathbf{u}_j$ and we are using the continuity convention that $\frac{\theta'(q_k)-\theta'(q_j)}{q_k-q_j} = \theta''(q_k)$ when $q_j = q_k$. Regrouping these last terms then yields $\dot{H}(t) = \sum_{j,k}(q_j\delta_{jk} - P_{kj})\frac{\theta'(q_k)-\theta'(q_j)}{q_k-q_j}V^\sharp_{jk}$, and since $\dot{H} = \mathrm{tr}((\mathbf{Q}-\mathbf{P})\mathbf{V})$ for all $\mathbf{P}$ by Prop. 1, our claim follows. ■

IV. Convergence Analysis

Thanks to the calculations of the previous section, we are now in a position to state our main result:

Theorem 1. *Let* $\theta\colon [0,+\infty)\to\mathbb{R}\cup\{+\infty\}$ *be an entropic kernel (cf. Def. 1), and let* $\mathbf{Q} = (\mathbf{Q}_1,\dots,\mathbf{Q}_K)\in\mathcal{X}$ *denote the users' signal covariance matrices. Furthermore, let* $\{q_{k\alpha},\mathbf{u}_{k\alpha}\}_{\alpha=1}^{m_k}$ *be the eigen-decomposition of* $\mathbf{Q}_k$, *and let* $V^k_{\alpha\beta} = \mathbf{u}_{k\alpha}^\dagger\cdot\mathbf{V}\cdot\mathbf{u}_{k\beta}$ *be the components of the matrix* $\mathbf{V} = d\Phi$ *of* (8) *in this basis. Then:*

1) *The entropic eigen-dynamics*

$$\begin{aligned}\dot{q}_{k\alpha} &= \tfrac{1}{\theta''(q_{k\alpha})}\Big(V^k_{\alpha\alpha} - \theta''_{k,h}(q)\textstyle\sum_{\beta=1}^{m_k}V^k_{\beta\beta}/\theta''(q_{k\beta})\Big),\\ \dot{\mathbf{u}}_{k\alpha} &= \textstyle\sum_{\beta\neq\alpha}\frac{V^k_{\alpha\beta}}{\theta'(q_{k\beta})-\theta'(q_{k\alpha})}\mathbf{u}_{k\beta},\end{aligned} \tag{ED$_\theta$}$$

where $\theta''_{k,h}(q) \equiv \Big(\sum_{\beta=1}^{m_k}1/\theta''(q_{k\beta})\Big)^{-1}$, *stay in* $\mathcal{X}$ *for all* $t\geq 0$.

2) $\Phi(\mathbf{Q}(t))$ *is non-decreasing along the solutions of* (ED$_\theta$).

3) *If* $\mathbf{Q}_k(0)\succ 0$, *then* (ED$_\theta$) *converges to a solution of* (SRP).

Sketch of proof: By Proposition 3, we see that the dynamics (ED$_\theta$) are simply the Hessian flow dynamics (HGF) minus the projection of the gradient $\mathbf{V}$ of Φ on the tangent space to $\mathcal{X}$ (where we have the additional trace constraint $\mathrm{tr}(\mathbf{Q}_k) = \sum_{\alpha=1}^{m_k}q_{k\alpha} = P_k$); as a result, $\mathcal{X}$ remains invariant under (ED$_\theta$). The second and third parts of the theorem then follow from this observation (viz. that (ED$_\theta$) is a Hessian gradient flow on $\mathcal{X}$) and Theorem 4.7 in [9]. ■

Remark 1. As an important special case of Theorem 1, note that if we take the Shannon entropy kernel $\theta(x) = x\log x$, then (ED$_\theta$) becomes:

$$\begin{aligned}\dot{q}_{k\alpha} &= q_{k\alpha}\Big(V^k_{\alpha\alpha} - P_k^{-1}\textstyle\sum_{\beta=1}^{m_k}q_{k\beta}V^k_{\beta\beta}\Big),\\ \dot{\mathbf{u}}_{k\alpha} &= \textstyle\sum_{\beta\neq\alpha}\frac{V^k_{\alpha\beta}}{\log q_{k\beta}-\log q_{k\alpha}}\mathbf{u}_{k\beta},\end{aligned} \tag{XD}$$

which is precisely the eigen-dynamics of the matrix exponential learning scheme that was introduced in [10, 11]. In this way, we see that the convergence properties of exponential learning are not owed to any special properties of the exponential mapping, but to the fact that (XD) is a Hessian gradient flow system.

Remark 2. Theorem 1 shows that the entropic dynamics (ED$_\theta$) solve (SRP) for any admissible choice of entropy, but there is no mention of the actual *rate* of convergence. As we shall see in the next section, the rate of convergence can be tuned by adjusting the underlying entropy function: for instance, for the Shannon entropy, if $\mathbf{Q}(t)\to\mathbf{P}$, then $\|\mathbf{Q}(t)-\mathbf{P}\| = \mathcal{O}(e^{-ct})$ for some $c>0$.

Now, based on the entropy-driven dynamics (ED$_\theta$) of Theorem 1, the algorithm that we will use to solve (SRP) will be:

Algorithm 1 Entropy-driven Gradient Ascent (EGA)

Require: Entropy kernel $\theta(x)$, initial transmit directions $\mathbf{u}_{k\alpha}\in\mathbb{C}^{m_k}$ and eigenvalues $p_{k\alpha}>0$, $\sum_{\alpha=1}^{m_k}p_{k\alpha} = P_k$.
$t\leftarrow 0$;
repeat
  $t\leftarrow t+1$;
  for all $k\in\mathcal{K}$ **do**
    $\{\mathbf{u}_{k\alpha}\}\leftarrow\Big\{\mathbf{u}_{k\alpha} + \delta(t)\sum_{\beta\neq\alpha}\frac{V^k_{\alpha\beta}}{\theta'(q_{k\beta})-\theta'(q_{k\alpha})}\mathbf{u}_{k\beta}\Big\}$;
    $\{q_{k\alpha}\}\leftarrow\Big\{q_{k\alpha} + \frac{\delta(t)}{\theta''(q_{k\alpha})}\Big(V^k_{\alpha\alpha} - \theta''_{k,h}(q)\sum_{\beta=1}^{m_k}V^k_{\beta\beta}/\theta''(q_{k\beta})\Big)\Big\}$;
  end for
until required accuracy is reached or transmission ends.

In terms of information, EGA requires exactly the same amount of information as distributed water-filling – viz. knowledge of the channel matrices $\mathbf{H}_k$ and the aggregate signal-plus-noise covariance matrix $\mathbf{W}$ of (8). Furthermore, EGA *does not* require any diagonalization steps: since it updates eigenvalues and eigenvectors directly, the users' signal matrices are already diagonalized. Finally, since the update steps of EGA no longer contain a linear search for the water level, its computational overhead is also considerably lighter than that of water-filling.

Now, as far as EGA's convergence is concerned, we have:

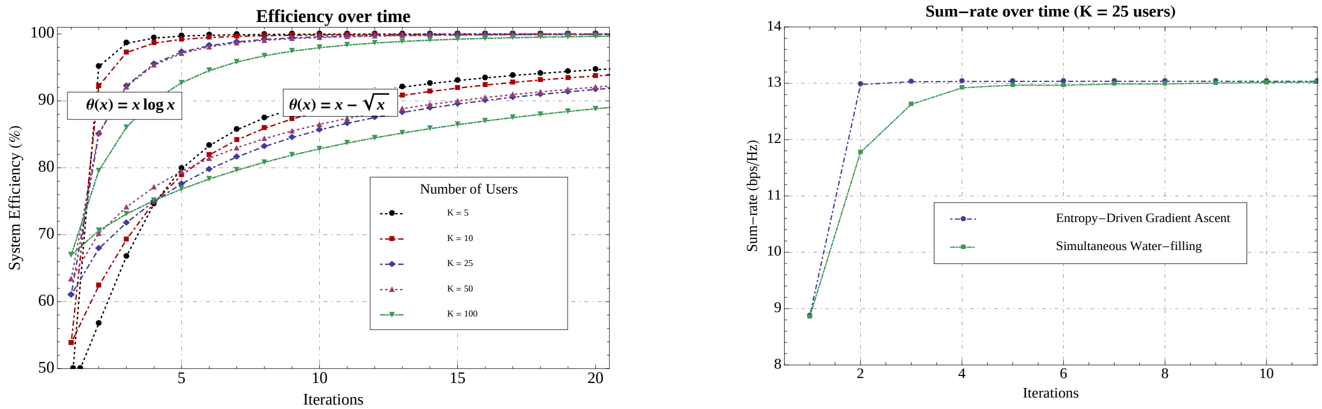
Theorem 2. *For any step sequence* $\delta(t)$ *with* $\sum_t\delta(t) = \infty$ *and small enough* $\sum_t\delta(t)^2$, *EGA converges to a solution of (SRP).*

Sketch of proof: If $\delta(t)$ is small enough, then EGA remains in $\mathcal{X}$ for all iterations; thus, as EGA represents an Euler discretization of (ED$_\theta$) with decreasing step size, our claim follows trivially from standard results in the theory of stochastic approximation [17]. ■

Remark. Even though Theorem 2 applies to decreasing step sizes (e.g. $\delta(t) = 1/t$), numerical experiments show that we can take a constant step size to speed up the algorithm without jeopardizing convergence (see Fig. 1).

V. Numerical Simulations

To assess the convergence properties of EGA, we simulated in Fig. 1 a MIMO MAC system consisting of $K = 5, 10, 25, 50$ and 100 transmitters, each with a random number of antennas



(a) The effect of entropy on the rate of convergence of EGA.

(b) Convergence of Shannon-driven EGA and of simultaneous water-filling.

Fig. 1. Convergence of different variants of the EGA algorithm and a comparison to simultaneous water-filling.

(between 2 and 10), a receiver with $n = 5$ antennas, and random channel matrices $\mathbf{H}_k$. We then ran EGA with a constant step and tracked the normalized efficiency ratio

$$\mathrm{eff}(t) = \frac{\Phi(t) - \Phi_{\min}}{\Phi_{\max} - \Phi_{\min}}, \tag{19}$$

where $\Phi_{\min}$ and $\Phi_{\max}$ are the minimum and maximum values of Φ over $\mathcal{X}$ respectively, and $\Phi(t)$ is the sum rate at time t.

In Fig. 1(a), we see that the speed of EGA depends crucially on the choice of entropy function: the Shannon variant ($\theta(x) = x \log x$) converges within a few iterations (even for $K = 100$ users), whereas the variant corresponding to the Tsallis-$\frac{1}{2}$ entropy $h_{1/2}(x) = 2\,\mathrm{tr}(\mathbf{Q} - \mathbf{Q}^{1/2})$ converges at a significantly slower rate. Finally, for benchmark purposes, we also plotted in Fig. 1(b) the evolution of the system's sum-rate for EGA and simultaneous water-filling (SWF) [6]: remarkably, by simply choosing a larger step size, we see that the Boltzmann variant of EGA outperforms even the SWF algorithm and essentially achieves the channel's sum capacity in a single iteration.

VI. Conclusions and Future Directions

In this paper, we introduced a class of dynamic distributed algorithms for solving semidefinite concave maximization problems in the context of multi-user MIMO networks. The method relies on endowing the space of semidefinite matrices with a matrix entropy function which acts as a barrier that enforces the problem's positivity constraints; the Hessian flow induced by this entropy then allows the network's users to achieve the system's sum capacity in but a few iterations, even for large numbers of users. The choice of entropy actually determines the algorithm's convergence speed, so this is a very promising method to develop fast convergence algorithms.

Finally, we note that this method can be extended to MIMO optimization problems with other type of constraints (e.g. spectral masks), and its rapid convergence ensures robustness even in the presence of rapidly changing channel conditions.

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] J. Bolte and M. Teboulle, "Barrier operators and associated gradient-like dynamical systems for constrained minimization problems," *SIAM Journal on Control and Optimization*, vol. 42, no. 4, pp. 1266–1292, 2003.

[9] F. Alvarez, J. Bolte, and O. Brahic, "Hessian Riemannian gradient flows in convex programming," *SIAM Journal on Control and Optimization*, vol. 43, no. 2, pp. 477–501, 2004.

[10] 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.

[11] P. Mertikopoulos, "Strange bedfellows: Riemann, Gibbs and vector Gaussian multiple access channels," in *NetGCoop '12: Proceedings of the 6th International Conference on Network Games, Control and Optimization*, 2012.

[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] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

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

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

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