Strange Bedfellows: Riemann, Gibbs and Vector Gaussian Multiple Access Channels

Panayotis Mertikopoulos
French National Center for Scientific Research (CNRS)
Laboratoire d'Informatique de Grenoble
E-mail: panayotis.mertikopoulos@imag.fr

Abstract—**Using tools and techniques from Riemannian geometry, we develop a novel distributed algorithm for optimizing the input signal distribution (and, in particular, its covariance matrix) in Gaussian multiple-input, multiple-output multiple access channels. To account for the problem's semidefiniteness constraints, we endow the space of positive-definitie matrices with a non-Euclidean spectral metric which becomes singular when the signal spectrum itself becomes singular. Quite remarkably, viewing the unit simplex as a subspace of the space of semidefinite matrices (corresponding to diagonal ones), we show that this metric generalizes the well-known Shahshahani metric on the simplex and extends the replicator dynamics of evolutionary game theory; in fact, gradient ascent trajectories defined with respect to this metric are shown to be equivalent to a Gibbs-based exponential learning process. In this way, we show that the resulting optimization algorithm converges to the optimum signal distribution exponentially fast: users attain an ε-neighborhood of the system's optimum configuration in time which is at most $\mathcal{O}(\log(1/\varepsilon))$ (and, in practice, within only a few iterations, even for large numbers of users).**

Index Terms—**Distributed optimization; semidefinite programming; Riemannian geometry; replicator dynamics; MIMO.**

I. INTRODUCTION

The widespread and widely successful use of multiple-input, multiple-output (MIMO) technologies in today's wireless networks (ranging from 3G LTE and 4G, to WiMAX and 802.11 WiFi) has brought into sharp focus the need for distributed optimization methods that allow wireless users to maximize their information throughput (especially in dense urban environments where centralized control is not practical and unregulated operation is the norm). In particular, since the radio spectrum (a scarce resource by itself) is shared by numerous users, the intended receiver of a signal must cope with unwanted interference from a large number of transmitters, a problem which may severely compromise the network's capacity if left unchecked. In view of the above, and given that the theoretical limits of MIMO systems are not known (even in basic models such as the interference channel), one usually starts with the mutual information for Guassian input and noise and aims to optimize the input spectrum of each user, considering all other users as (colored) interference.

A Gaussian channel model which has attracted significant interest in the literature is the MIMO multiple access channel (MAC) where a single multi-antenna receiver needs to simultaneously decode the incoming signals of several (multi-antenna) transmitters. Mathematically, this amounts to the (concave) problem of maximizing the users' sum rate function over a set of positive-semidefinite matrices representing the spread of the users' symbol distributions (their covariance to be exact) [1]. However, due to the implicitness of the problem's definiteness constraints, standard gradient ascent and interior point methods do not readily apply, so the problem is usually solved by water-filling methods which are known to converge to the system's optimum input spectrum [2–4].

In this paper, instead of relying on water-filling (which does not scale well with the number of users present in the system),[1] we introduce a distributed optimization algorithm based on Riemannian-geometric ideas. Specifically, we endow the cone of positive-definite matrices with a spectral geometry which makes it harder for users to ascend the system's sum rate function when the eigenvalues of the users' input covariance matrices are small. As a result, gradient ascent defined with respect to this metric is structurally confined to stay in the system's state space and the users' sum rate function increases along the way until eventually attaining capacity.

Remarkably, these Riemannian dynamics can be shown to be equivalent to a matrix-valued exponential learning scheme which was recently introduced in [7]. Indeed, by aggregating the differential of the users' sum rate function and then mapping this aggregate score back to the space of semidefinite matrices by means of a matrix-valued Gibbs distribution, users converge to an optimum input spectrum exponentially fast. More importantly, the speed of this convergence can be controlled by the temperature of the Gibbs distribution in question, and this parameter also determines the curvature of the Riemannian metric induced on the cone of semidefinite matrices. In this way, Riemannian and Gibbs-based learning are seen to be two sides of the same coin and they provide powerful tools for distributed optimization in Gaussian multiple access channels:[2] in practical implementations, the system achieves its sum capacity within only a few iterations, even for large numbers of users and/or antennas per user.

---

[1]Iterative and sequential water-filling methods require users to update their signal distributions in a round-robin fashion, a fact which greatly increases signaling overhead and convergence time in networks with large numbers of users. Simultaneous update (and, hence, faster) variants of water-filling algorithms do exist [5], but their convergence depends on the channels satisfying certain "mild interference" conditions which fail even when the antenna subchannels are assumed orthogonal (the parallel multiple access channel (PMAC) case) [6].

[2]In fact, our Riemann-Gibbs method can be used for a wide class of semidefinite problems; we only focus here on the MIMO MAC problem for concreteness and simplicity.

II. System Model

The system model that we will consider is a vector Gaussian multiple access channel where a finite number of wireless users $k \in \mathcal{K} \equiv \{1, \ldots, K\}$, each equipped with m_k ($k \in \mathcal{K}$) antennas, transmit their messages simultaneously to an n-antenna wireless receiver. This corresponds to the baseband signal model:

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{x}_k + \mathbf{z}, \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^n$ is the message 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 $n \times m_k$ channel matrix, and $\mathbf{z} \in \mathbb{C}^n$ is the channel noise (with covariance taken equal to $\mathbf{I}$ after suitable rescalings).

In this context, the transmit power of user k will be $P_k = \mathbb{E}\left[\|\mathbf{x}_k\|^2\right] = \mathrm{tr}(\mathbf{Q}_k)$, where the expectation is taken over the codebook of user k and $\mathbf{Q}_k = \mathbb{E}\left[\mathbf{x}_k \mathbf{x}_k^\dagger\right]$ denotes the corresponding signal covariance matrix. Thus, assuming single user decoding (SUD) at the receiver (instead of more costly interference cancellation techniques), the maximum information rate will be achieved for Gaussian codebooks and will be given by the familiar expression [1]:

$$\Phi(\mathbf{Q}) = \log \det \left(\mathbf{I} + \sum_k \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^\dagger\right), \tag{2}$$

Accordingly, if the users' power is constrained to some value $P_k = \mathrm{tr}(\mathbf{Q}_k)$,[3] we are led to the semidefinite problem:

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

where $\mathcal{X}_k$ is the set of feasible signal matrices of user k:

$$\mathcal{X}_k = \{\mathbf{Q}_k \in \mathbb{C}^{n \times m_k} : \mathbf{Q}_k \succcurlyeq 0, \mathrm{tr}(\mathbf{Q}_k) = P_k\}. \tag{3}$$

In what follows, our aim will be to provide a distributed method with which to solve (P) in a timely fashion.

III. Geometry, Dynamics and Optimization

Even though the maximization problem (P) is a concave one [1], the semidefiniteness constraints $\mathbf{Q}_k \succcurlyeq 0$ cannot be readily put in explicit functional form, so gradient ascent or interior point methods do not readily apply: for instance, by following the ordinary (Euclidean) gradient of Φ we will invariably end up violating the semidefiniteness constraints of (P). Our approach will thus be to change the geometry of the problem's state space $\mathcal{X} \equiv \prod_k \mathcal{X}_k$ in a way such that the modified (Riemannian) gradient ascent trajectories will always remain within $\mathcal{X}$ and will eventually converge to the maximum of Φ.

A. The Shahshahani metric and the replicator dynamics

To that end, recall first that a *Riemannian metric* on a smooth subset (or *submanifold*) M of $\mathbb{R}^m$ is a smooth assignment of an inner product $\langle \cdot, \cdot \rangle_x$ to the tangent space $T_x M$ of M at each $x \in M$ – see e.g. [8] for a masterful introduction to the subject. For instance, if $M = \mathbb{R}_+^m$ denotes the positive orthant of $\mathbb{R}^m$, then the *Shahshahani metric* [9] at the point $x \in M$ is defined as:

$$\langle \mathbf{z}, \mathbf{w} \rangle_x = \sum_{\alpha=1}^{m} \frac{z_\alpha w_\alpha}{x_\alpha}, \tag{4}$$

[3]Note that the inequality constraint $\mathrm{tr}(\mathbf{Q}_k) \leq P_k$ leads to the same problem.

where $\mathbf{z}, \mathbf{w}$ are (tangent) vectors in $\mathbb{R}^m \cong T_p M$.[4]

Importantly, a Riemannian metric $\langle \cdot, \cdot \rangle$ can be equivalently described as an isomorphism $\sharp \colon T_x^* M \to T_x M$, where $T_x^* M$ is the *cotangent space* of M at x, i.e. the space of linear functionals (or *covectors*) $\omega \colon T_x M \to \mathbb{R}$.[5] This index-raising isomorphism has the defining property that

$$\langle \omega^\sharp, \mathbf{z} \rangle = \omega(\mathbf{z}) \quad \text{for all } \mathbf{z} \in T_x M \text{ and } \omega \in T_x^* M, \tag{5}$$

so, in the case of the Shahshahani metric, we will have:

$$\omega_\alpha^\sharp = x_\alpha \omega_\alpha \quad \text{for all } \omega \in T_x^* M, \tag{6}$$

where ω_α and $\omega_\alpha^\sharp$ denote respectively the components of ω and $\omega^\sharp$ in the standard basis of $\mathbb{R}^m$.

Thanks to this isomorphism, the (Riemannian) *gradient* of a function $f \colon M \to \mathbb{R}$ is defined as:

$$\mathrm{grad}_x f = (\mathrm{d} f_x)^\sharp, \tag{7}$$

where now $\mathrm{d} f_x$ denotes the differential of f evaluated at x, i.e.: $\mathrm{d} f_x = \sum_{\alpha=1}^{m} \frac{\partial f}{\partial x_\alpha} \mathrm{d} x_\alpha$. In this way, the differential of a function (a covector by definition) is turned into a vector, which then allows us to write the gradient ascent dynamics:

$$\dot{x} = \mathrm{grad}_x f. \tag{8}$$

Accordingly, the expression (6) for the Shahshahani metric gives $\dot{x}_\alpha = x_\alpha \frac{\partial f}{\partial x_\alpha}$, and if we project $\mathrm{grad} f$ on the interior of the unit simplex $\Delta = \{x \in \mathbb{R}^m : x_\alpha \geq 0 \text{ and } \sum_\alpha x_\alpha = 1\}$, we readily obtain the well-known *replicator dynamics* [10, 11]:

$$\dot{x}_\alpha = x_\alpha \left(\frac{\partial f}{\partial x_\alpha} - \sum_{\beta=1}^{m} x_\beta \frac{\partial f}{\partial x_\beta}\right). \tag{9}$$

The replicator dynamics have remarkable optimization properties [12], some of which apply directly to (a restricted version of) the problem at hand. For instance, if the matrices $\mathbf{H}_k$ all admit a common singular decomposition and we restrict the matrices $\mathbf{Q}_k$ to be diagonal (the so-called parallel multiple access channel), it is shown in [13] that the replicator dynamics converge to the maximum of the sum rate function exponentially fast.[6]

B. The geometry of the cone of semidefinite matrices

In view of the above, our aim will be to properly extend the Shahshahani metric and the replicator dynamics to a semidefinite setting and take advantage of their optimization properties in order to solve (P). The first step in this task will be to derive a suitable metric for the convex cone of positive-definite matrices

$$\mathcal{C} = \{\mathbf{Q} \in \mathbb{C}^{m \times m} : \mathbf{Q} \succ 0\}, \tag{10}$$

and, as it turns out, it will be more convenient to define this metric in terms of the associated isomorphism $\sharp \colon T_\mathbf{Q}^* \mathcal{C} \to T_\mathbf{Q} \mathcal{C}$.

[4]Recall that M is an open subset of $\mathbb{R}^m$ so the tangent space at any point of M will be isomorphic to $\mathbb{R}^m$.

[5]In coordinates, one could think of vectors $\mathbf{z} \in T_x M$ as column vectors, and of covectors $\omega \in T_x^* M$ as row vectors, with $\omega(\mathbf{z})$ coinciding with the standard pairing between row and column vectors: $\omega(\mathbf{z}) = \omega \cdot \mathbf{z}$.

[6]Note that in the diagonal case, the problem's state space $\mathcal{X}$ becomes a product of simplices [13].

Of course, since $\mathcal{C}$ is an open subset of the (real) space of Hermitian matrices $\mathbb{H}_m \cong \mathbb{R}^{m^2}$ (implying that both $T_{\mathbf{Q}}\mathcal{C}$ and $T_{\mathbf{Q}}^*\mathcal{C}$ are canonically isomorphic themselves to $\mathbb{H}_m$), this boils down to constructing an automorphism of $\mathbb{H}_m$ which reduces to (6) for diagonal matrices. Thus, given a Hermitian matrix $\mathbf{\Omega} \in T_{\mathbf{Q}}^*\mathcal{C}$, a natural choice would be to set $\mathbf{\Omega}^\sharp = \mathbf{Q}\mathbf{\Omega}$; however, $\mathbf{\Omega}^\sharp$ defined in this way is not Hermitian (unless $\mathbf{\Omega}$ happens to commute with $\mathbf{Q}$); furthermore, if we symmetrize this choice by setting $\mathbf{\Omega}^\sharp = (\mathbf{\Omega}\mathbf{Q} + \mathbf{Q}\mathbf{\Omega})/2$, we lose the automorphism property: if $\mathbf{Q}$ anticommutes with some $\mathbf{\Omega} \neq 0$, then $\mathbf{\Omega}^\sharp = 0$, so $\sharp$ defined in this way cannot be an isomorphism.

To circumvent this difficulty, note that any expression of the form $\mathbf{Q}^s\mathbf{\Omega}\mathbf{Q}^{1-s}$ is linear in $\mathbf{\Omega}$ and reduces to (6) when $\mathbf{\Omega}$ and $\mathbf{Q}$ are simultaneously diagonalizable; furthermore, any superposition of such expressions will still be linear in $\mathbf{\Omega}$. Thus, instead of taking only the conjugate terms $\mathbf{Q}\mathbf{\Omega}$ and $\mathbf{\Omega}\mathbf{Q}$, we will take *all* powers $s \in [0, 1]$ and add them up to obtain:

$$\mathbf{\Omega}^\sharp = \int_0^1 \mathbf{Q}^s\mathbf{\Omega}\mathbf{Q}^{1-s}\,ds. \qquad (11)$$

With this definition, we then obtain:

Lemma 1. *The map $\sharp\colon \mathbf{\Omega} \in \mathbb{H}_m \mapsto \mathbf{\Omega}^\sharp$ as defined in (11) is an automorphism of $\mathbb{H}_m$ and reduces to (6) when $\mathbf{\Omega}$ and $\mathbf{Q}$ are simultaneously diagonalizable. Furthermore, in a basis where $\mathbf{Q}$ is diagonal, $\mathbf{Q} = \mathrm{diag}(q_1, \ldots, q_m)$, we will have:*

$$\mathbf{\Omega}_{\alpha\beta}^\sharp = \frac{q_\alpha - q_\beta}{\log q_\alpha - \log q_\beta}\mathbf{\Omega}_{\alpha\beta}, \quad (\alpha, \beta = 1, \ldots, m) \qquad (12)$$

with the convention that $(x - y)/(\log x - \log y) = x$ when $x = y$.

Sketch of proof: The change of variables $s \to 1 - s$ shows that $\mathbf{\Omega}^\sharp$ is Hermitian, so $\sharp$ is indeed an endomorphism of $\mathbb{H}$. Moreover, if we write everything in a basis where $\mathbf{Q}$ is diagonal, then we can carry out the matrix multiplications and the integration in (11) to obtain (12) – which obviously reduces to (6) when $\Omega_{\alpha\beta} = \omega_\alpha \delta_{\alpha\beta}$. By the form of (12) it is then easy to see that $\sharp$ is invertible and, hence, an automorphism of $\mathbb{H}_m$. ∎

Thanks to the above lemma (which extends the Shahshahani metric from $\mathbb{R}_+^m$ to $\mathcal{C}$), we may finally define a Riemannian gradient system for (P). Indeed, note first that the differential of Φ is simply be given by

$$\mathbf{V}_k = \partial\Phi/\partial\mathbf{Q}_k^* = \mathbf{H}_k^\dagger \mathbf{W}^{-1}\mathbf{H}_k, \qquad (13)$$

where $\mathbf{W} = \mathbf{I} + \sum_{\ell \in \mathcal{K}} \mathbf{H}_\ell \mathbf{Q}_\ell \mathbf{H}_\ell^\dagger$ is the aggregate signal-plus-noise covariance matrix at the receiver end. Thus, by using (11) to define the associated Riemannian gradient $\mathbf{V}_k^\sharp$ and projecting it on the problem's trace constraints $\mathrm{tr}(\mathbf{Q}_k) = P_k$, we readily obtain the *matrix-valued replicator dynamics*:

$$\dot{\mathbf{Q}}_k = \int_0^1 \mathbf{Q}_k^s\mathbf{V}_k\mathbf{Q}_k^{1-s}\,ds - P_k^{-1}\mathrm{tr}(\mathbf{Q}_k\mathbf{V}_k)\,\mathbf{I} \quad (k \in \mathcal{K}). \qquad (14)$$

In this way, we finally recover the powerful optimization properties of the replicator dynamics in the context of (P):

Theorem 1. *Let $\mathbf{Q}(t)$ be an interior solution trajectory of the matrix replicator dynamics (14). Then, the system's sum rate Φ*

increases along $\mathbf{Q}(t)$, and $\mathbf{Q}(t)$ converges to a solution $\mathbf{Q}^*$ of (P); in particular, for any sufficiently small $\varepsilon > 0$, $\mathbf{Q}(t)$ attains an ε-neighborhood of $\mathbf{Q}^*$ in time which is at most $\mathcal{O}(\log(1/\varepsilon))$.

C. Eigenvalues, eigenvectors and Gibbs learning

Of course, (14) is not very practical as an optimization scheme because it requires the evaluation of an integral at every update step. This limitation however is only an artifact of writing the matrix replicator dynamics (14) in a coordinate-independent way. The following proposition shows that things are considerably simpler if we write everything in a basis where $\mathbf{Q}$ is diagonal (i.e. examine the evolution of $\mathbf{Q}$ via the evolution of its eigenvalues and its eigenvectors):

Proposition 1. *Let $\mathbf{Q}(t)$ be an interior solution of the matrix replicator dynamics (14). If $\{q_{k\alpha}(t), \mathbf{u}_{k\alpha}(t)\}_{\alpha=1}^{m_k}$ is a smooth eigendecomposition of $\mathbf{Q}_k$ and $V_{\alpha\beta}^k \equiv \mathbf{u}_{k\alpha}^\dagger \mathbf{V}_k \mathbf{u}_{k\beta}$, then:*

$$\dot{q}_{k\alpha} = q_{k\alpha}\left(V_{\alpha\alpha}^k - P_k^{-1}\sum_{\beta=1}^{m_k} q_{k\beta}V_{\beta\beta}^k\right), \qquad (15a)$$

$$\dot{\mathbf{u}}_{k\alpha} = \sum_{\beta \neq \alpha} V_{\beta\alpha}^k \left(\log q_{k\alpha} - \log q_{k\beta}\right)^{-1}\mathbf{u}_{k\beta}. \qquad (15b)$$

Remarkably, these evolution equations were obtained in [7] as the offshoot of an exponential learning process based on the Gibbs distribution. In particular, if users try to optimize their spectrum choices $\mathbf{Q}_k$ by aggregating the differential of Φ over time and then turn these auxiliary aggregate "score" variables into input signal covariance matrices by using the Gibbs distribution, we obtain the exponential learning process:

$$\dot{\mathbf{Y}}_k = \mathbf{V}_k, \qquad (16a)$$

$$\mathbf{Q}_k = P_k\frac{\exp(\mathbf{Y}_k)}{\mathrm{tr}(\exp(\mathbf{Y}_k))}. \qquad (16b)$$

As a result of the Gibbs distribution (16b), $\mathbf{Q}_k$ will automatically satisfy the definiteness and trace constraints of (P); moreover, as was shown in [7], the eigenvalues and eigenvectors of $\mathbf{Q}$ evolve over time *exactly* as prescribed by Proposition 1. In other words, we thus see that *the matrix-valued Gibbs learning process (16) is equivalent to the Riemannian gradient ascent scheme (14), which in turn generalizes and extends the replicator dynamics of evolutionary game theory to a semidefinite programming context.*

Remark. Of course, one could adjust the temperature of the Gibbs distribution (16b) by using $\lambda\mathbf{Y}_k$ instead of $\mathbf{Y}_k$. This adjustment amounts to changing the time scale of (14) by a factor of λ with higher values of λ (low temperatures) corresponding to users learning very fast, essentially "freezing" to a best-response behavior. Interestingly, this temperature parameter also controls the curvature of the spectral geometry induced on $\mathcal{C}$: for larger λ, the geometry of $\mathcal{C}$ is commensurately flatter, so gradient ascent is faster as well.

IV. Algorithms and Numerical Simulations

In this section, our aim will be to derive a discrete-time algorithmic implementation of the matrix replicator dynamics (14) and then use it in numerical simulations of MIMO multiple access channels. At the user end, our assumptions will be that
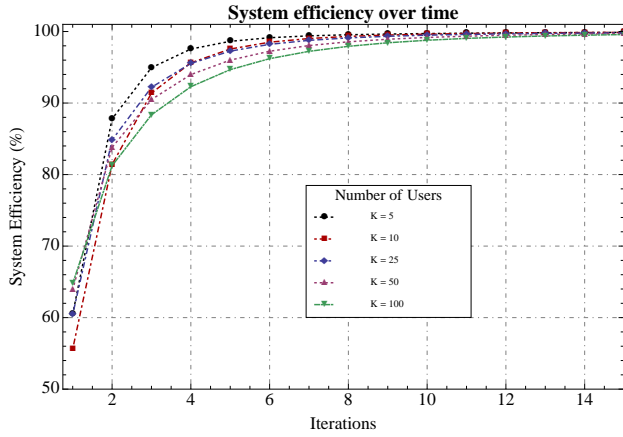
Fig. 1. Transmission rate maximization under the RGA scheme.

a) transmitters have accurate channel state information (i.e. each user knows his own channel matrix $\mathbf{H}_k$); and *b*) that the receiver measures the aggregate signal-plus-noise matrix $\mathbf{W}$ at his end and then broadcasts it (e.g. over a dedicated channel or by appending it to acknowledgment packets). In this way, (13) shows that each user can individually calculate the update step of the dynamics (15) in an efficient, distributed fashion, and *without* having to know the covariance matrices (or the channels) of other users.

Accordingly, under these assumptions (which are standard in the water-filling literature [2–4]), we obtain the Riemann-Gibbs ascent (RGA) algorithm:

Algorithm 1 Riemann-Gibbs ascent (RGA)

Require: Unitary matrices $\mathbf{U}_k$ of initial transmit directions $\mathbf{u}_{k\alpha} \in \mathbb{C}^{m_k}$ and power eigenvalues $q_{k\alpha} > 0$, $\sum_{\alpha=1}^{m_k} q_{k\alpha} = P_k$.
$t \leftarrow 0$;
repeat
$\quad t \leftarrow t + 1$;
$\quad$ **for all** $k \in \mathcal{K}$ **do**
$\quad\quad \{\mathbf{u}_{k\alpha}\} \leftarrow \left\{ \mathbf{u}_{k\alpha} + \delta(t) \sum_{\beta \neq \alpha} V_{\beta\alpha}^k (\log q_{k\alpha} - \log q_{k\beta})^{-1} \mathbf{u}_{k\beta} \right\}$;
$\quad\quad \{q_{k\alpha}\} \leftarrow \left\{ q_{k\alpha} + \delta(t) \, q_{k\alpha} \left(V_{\alpha\alpha}^k - P_k^{-1} \sum_{\beta=1}^{m_k} p_{k\beta} V_{\beta\beta}^k \right) \right\}$;
$\quad$ **end for**
until required accuracy is reached or transmission ends.

The only new element of the RGA algorithm with respect to the dynamics (15) is the variable step size $\delta(t)$ which is assumed to satisfy the "$L^2 - L^1$" summability conditions $\sum_t \delta(t) = \infty$, $\sum_t \delta(t)^2 < \infty$ for convergence purposes. This choice ensures that RGA will converge to the maximum of Φ but it also reduces the algorithm's convergence speed considerably; as we shall see however, we can take a constant step size $\delta(t) \equiv \delta$ without compromising the algorithm's convergence, all the while retaining its exponential convergence speed.

Indeed, to assess the performance of our RGA algorithm, we simulated in Fig. 1 a MIMO MAC system consisting of 5, 10, 25, 50 and 100 transmitters, each with a random number of antennas (uniformly drawn between 2 and 10), a receiver with $n = 5$ antennas, and random channel matrices $\mathbf{H}$. We then ran RGA with a constant step size and plotted over time the normalized efficiency ratio:

$$\text{eff}(t) = \frac{\Phi(t) - \Phi_{\min}}{\Phi_{\max} - \Phi_{\min}}, \qquad (17)$$

where $\Phi_{\min}$ and $\Phi_{\max}$ are the minimum and maximum values of Φ over $\mathcal{X}$ respectively, and $\Phi(t)$ is the users' sum rate at time t. Remarkably, we see in Fig. 1 that RGA scales very well with the number of users and achieves capacity within only a few iterations, even for $K = 100$ users.

V. Conclusions and Future Directions

In this paper, we introduced a novel distributed algorithm for solving concave semidefinite maximization problems in the context of multi-user MIMO networks. By using tools and ideas from Riemannian geometry, the semidefiniteness constraints of the users' signal covariance matrices are enforced naturally, and as these matrices are updated dynamically, the network's users achieve the system's capacity exponentially fast. Remarkably, our Riemannian algorithm is equivalent to a Gibbs-based exponential learning scheme whose temperature controls (and allows network users to tune) the convergence rate of the resulting algorithm.

References

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] P. Mertikopoulos, E. V. Belmega, and A. L. Moustakas, "Matrix exponential learning: Distributed optimization in MIMO systems," in *ISIT '12: Proceedings of the 2012 International Symposium on Information Theory*, 2012.

[8] J. M. Lee, *Introduction to Smooth Manifolds*, ser. Graduate Texts in Mathematics. New York, NY: Springer-Verlag, 2003, no. 218.

[9] S. M. Shahshahani, "A new mathematical framework for the study of linkage and selection," *Memoirs of the American Mathematical Society*, vol. 211, 1979.

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

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

[12] 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.

[13] 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.