Riemannian-Geometric Optimization Methods for MIMO Multiple Access Channels

Panayotis Mertikopoulos
Centre National de la Recherche Scientifique (CNRS)
Laboratoire d'Informatique de Grenoble
E-mail: panayotis.mertikopoulos@imag.fr

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

***Abstract*—Drawing ideas from Riemannian geometry, we develop a distributed optimization dynamical system for determining optimum input signal covariance matrices in MIMO multiple access channels. In this type of problems, standard (Euclidean) gradient ascent approaches fail because the problem's semidefiniteness constraints are generically violated along the gradient flow; however, by endowing the space of positive-definite matrices with a non-Euclidean geometry which becomes singular when the eigenvalues of the users' covariance matrices approach zero, we are able to derive a matrix-valued Riemannian gradient ascent scheme which converges to the system's optimum transmit spectrum. More to the point, we show that by tuning the geometry of the semidefinite cone, the algorithm's convergence speed changes significantly. As a result, for a specific choice of geometry (which extends the well-known replicator dynamics of evolutionary game theory to a matrix setting), our scheme converges within a few iterations and users are able to track the optimum signal profile even in the presence of rapidly changing channel conditions.**

***Index Terms*—Distributed optimization; non-Euclidean gradient flows; Riemannian geometry; multiple access channel; MIMO.**

I. Introduction

The massive deployment of multiple-input and multiple-output (MIMO) technologies in modern wireless communication networks (ranging from 3G LTE, 4G and HSPA+, to 802.11n WiFi and WiMAX) has thrown into sharp relief the need for distributed optimization methods that allow the users of MIMO-enabled system to attain its performance limits. In this often unregulated (and usually unlicensed) 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, a useful and widespread approximation is to start 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.

An important channel model of this kind is the MIMO multiple access channel (MAC) where a single receiver (conceivably representing a set of colocated receivers or even a set of non-colocated receivers, interconnected over a high-speed backbone network) is called to decode the simultaneous signals of several transmitters. From a theoretical point of view, the capacity of this channel is well-known [1], and the corresponding semidefinite optimization problem is usually solved by water-filling techniques [2], suitably adapted to mult-user environments [3, 4]. Unfortunately however, iterative and/or sequential water-filling techniques rely on users updating their covariance matrices in a round-robin fashion, so they converge very slowly when the number of users is large; on the other hand, the convergence of faster, simultaneous water-filling methods [5] depends on the channels satisfying certain "mild-interference" conditions that fail to hold even in orthogonal channels (e.g. in the parallel MAC setting which was considered in [6] as a reduced version of the full MIMO problem).

In view of the above, instead of taking a water-filling approach, we draw ideas from Riemannian geometry in order to derive a distributed Riemannian gradient optimization scheme for the MIMO sum capacity problem.[1] To that end, we endow the problem's state space (a subset of the positive-definite cone) with a spectral geometry which greatly increases the "cost" of ascending the system's sum rate function when the eigenvalues of the users' transmit spectrum become small. In this way, our Riemannian gradient ascent (RGA) scheme is structurally constrained to remain within the problem's state space, all the while increasing the users' sum rate and eventually achieving the system's sum capacity. More importantly, the speed of this convergence can be controlled by tuning the geometry of the semidefinite cone, thus allowing users to achieve the channel's sum capacity within a few iterations (even for large numbers of users and/or antennas per user); as a matter of fact, even in rapidly changing channel conditions (due to e.g. fading), the users are able to track the optimum transmit spectrum in a small fraction of the channel's coherence time.

II. System Model

Consider a vector Gaussian multiple access channel where a finite set of wireless users $k \in \mathcal{K} \equiv \{1, \dots, K\}$, each equipped with m_k antennas, transmit simultaneously to a wireless base receiver with n antennas. As is well-known, this system may 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 message reaching the receiver, $\mathbf{x}_k \in \mathbb{C}^{m_k}$ is the message transmitted by user $k \in \mathcal{K}$,

[1] In fact, the resulting optimization methods apply to a much wider class of concave semidefinite problems; we only focus here on the MIMO MAC case for the sake of concreteness and simplicity.

$\mathbf{H}_k \in \mathbb{C}^{n\times m_k}$ is the associated $n \times m_k$ (complex) channel matrix, and $\mathbf{z} \in \mathbb{C}^n$ is the noise in the channel, 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).

In this setting, the average transmit power of user k will be

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

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

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

Then, assuming successive interference cancellation (SIC) at the receiver, the maximum information transmission rate for a given profile $\mathbf{Q} = (\mathbf{Q}_1, \ldots, \mathbf{Q}_K)$ of covariance matrices is achieved for Gaussian codebooks [1] and is given by:

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

Hence, assuming users can only transmit with finite power $\operatorname{tr}(\mathbf{Q}_k) = P_k$, we are naturally led to the constrained *sum rate maximization problem*:

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

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

$$\mathcal{X}_k = \{\mathbf{Q}_k \in \mathbb{C}^{n\times m_k} : \mathbf{Q}_k \succcurlyeq 0, \operatorname{tr}(\mathbf{Q}_k) = P_k\}. \tag{5}$$

III. Riemannian Geometry and Semidefinite Optimization

A crucial challenge in solving (SRP) is that the positivity constraints $\mathbf{Q}_k \succcurlyeq 0$ are implicit, so standard (Lagrangian) steepest ascent methods do not readily apply: by following the (Euclidean) gradient of the sum rate function Φ, the system may (and generically will) violate the positivity constraints $\mathbf{Q}_k \succcurlyeq 0$. To counter this, our main tool will be a variant steepest ascent method where ascending the gradient of Φ becomes more and more "costly" as users approach the boundary of the state space, thus prohibiting the users' state variables (the signal covariance matrices $\mathbf{Q}_k$) from violating the constraints of (SRP). In more precise language, we will endow the state space $\mathcal{X}$ with a non-Euclidean Riemannian metric (scalar product) which blows up at the boundary $\operatorname{bd}(\mathcal{X})$ of $\mathcal{X}$; then, by ascending the gradient of Φ with respect to this new metric, the blow-up at the boundary will keep solution trajectories in $\mathcal{X}$, all the while ascending the sum rate function and eventually converging to a point which attains the sum capacity of the system.

The fundamental notion of (complex) Riemannian geometry is that of a *metric* [7], i.e. a smooth assignment of a Hermitian inner product to the tangent bundle of the space under consideration (in our MIMO framework, the space of Hermitian matrices and the positive-semidefinite cone respectively). In specifying such a "Riemannian metric" one then also prescribes the angles between tangent vectors, and thus the direction of steepest ascent of a function (i.e. its gradient); accordingly, our strategy will be to construct a well-behaved gradient scheme by endowing the semidefinite cone with a suitable metric.

In the case of the cone $\mathcal{C}$ of $m\times m$ positive-semidefinite matrices, a spectral metric which has attracted significant interest in the literature (see e.g. [8] and references thererin) is the pairing

$$[\![\mathbf{A}, \mathbf{B}]\!]_{\mathbf{Q}} = \operatorname{tr}\left(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}^{-1}\mathbf{B}\right), \tag{6}$$

where $\mathbf{A}, \mathbf{B} \in \mathbb{H}^m$ are $m \times m$ Hermitian matrices, "tangent" to $\mathcal{C}$ at $\mathbf{Q}$.[2] The *gradient* of a function $\Phi\colon \mathcal{C} \to \mathbb{R}$ with respect to $[\![\cdot,\cdot]\!]$ is then defined as the matrix $\mathbf{grad}\,\Phi$ with the property

$$[\![\mathbf{grad}\,\Phi, \mathbf{B}]\!]_{\mathbf{Q}} = \left.\frac{d}{dt}\right|_{t=0} \Phi(\mathbf{Q} + t\mathbf{B}) = \operatorname{tr}(\mathbf{D}\Phi\cdot\mathbf{B}) \quad \forall \mathbf{B} \in \mathbb{H}^m, \tag{7}$$

where $\mathbf{D}\Phi \equiv \frac{\partial\Phi}{\partial\mathbf{Q}^*}$ is the matrix derivative of Φ w.r.t. $\mathbf{Q}$. In particular, noting that (7) holds for all $\mathbf{B} \in \mathbb{H}^m$, then the definition (6) of $[\![\cdot,\cdot]\!]$ readily gives:

$$\mathbf{grad}\,\Phi = \mathbf{Q}\cdot\mathbf{D}\Phi\cdot\mathbf{Q}. \tag{8}$$

To tie things back to the sum-rate maximization problem (SRP), note first that some matrix calculus gives

$$\frac{\partial\Phi}{\partial\mathbf{Q}_k^*} \equiv \mathbf{V}_k = \mathbf{H}_k^\dagger\mathbf{W}^{-1}\mathbf{H}_k, \tag{9}$$

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. As a result, the matrix derivative $\mathbf{D}\Phi$ of Φ will be the direct sum $\mathbf{D}\Phi = \bigoplus_{k=1}^K \mathbf{V}_k$, denoted more simply by $\mathbf{V} = (\mathbf{V}_1, \ldots, \mathbf{V}_K)$. However, since the feasible region $\mathcal{X}$ of (SRP) is also subject to the trace constraints $\operatorname{tr}(\mathbf{Q}_k) = P_k$, we will also need to adjust (8) by projecting the Riemannian gradient (7) on the subspace defined by the equations $\operatorname{tr}(\mathbf{Q}_k) = P_k$. To that end, we have:

Proposition 1. *The gradient ascent dynamics for Φ with respect to the Riemannian metric (6) on $\mathcal{X}$ are:*

$$\dot{\mathbf{Q}}_k = \mathbf{Q}_k\mathbf{V}_k\mathbf{Q}_k - \frac{\operatorname{tr}(\mathbf{V}_k\mathbf{Q}_k^2)}{\operatorname{tr}(\mathbf{Q}_k^2)}\mathbf{Q}_k^2 \quad k \in \mathcal{K}. \tag{10}$$

Moreover, the gradient flow (10) remains in $\mathcal{X}$ for all $t \geq 0$.

Sketch of proof: For the first part of our claim, it suffices to show that $[\![\dot{\mathbf{Q}}_k, \mathbf{B}_k]\!] = \operatorname{tr}(\mathbf{V}_k\mathbf{B}_k)$ for any Hermitian matrix $\mathbf{B}_k \in \mathbb{H}^{m_k}$ with $\operatorname{tr}(\mathbf{B}_k) = 0$ (simply note that the tangent space to $\mathcal{X}_k$ consists of all *traceless* Hermitian matrices); this is easily verified by a direct calculation. For the second part of our claim, we then need to show the tangency requirement $\operatorname{tr}(\dot{\mathbf{Q}}_k) = 0$ and the positive-definiteness qualification $\mathbf{Q}_k \succcurlyeq 0$ for all $t \geq 0$; the first condition is again easily checked whereas the second follows from Proposition 2 below. ∎

In view of Proposition 1, it will also be important to determine the evolution of the eigenvalues and eigenvectors of the users' signal covariance matrices. Along these lines, we have:

Proposition 2. *Let $\mathbf{Q}(t)$ be an interior solution orbit of the matrix dynamics (10), and let $\{q_{k\alpha}(t), \mathbf{u}_{k\alpha}(t)\}$, $\alpha = 1, \ldots, m_k$, be an*

[2] Recall here that $\mathcal{C}$ is an open set in $\mathbb{H}^m$, so its tangent space is isomorphic to $\mathbb{H}^m$ (much as the tangent space to an open set of $\mathbb{R}^m$ is $\mathbb{R}^m$ itself.

eigen-decomposition of $\mathbf{Q}_k(t)$*,* $k \in \mathcal{K}$*. Then, if* $V^k_{\alpha\beta} \equiv \mathbf{u}^\dagger_{k\alpha}\mathbf{V}_k\mathbf{u}_{k\beta}$ *denotes the components of* $\mathbf{V}_k$ *in this basis, we will have:*

$$\dot{q}_{k\alpha} = q^2_{k\alpha}\left(V^k_{\alpha\alpha} - \left(\sum_{\gamma=1}^{m_k} q^2_{k\gamma}\right)^{-1} \sum_{\beta=1}^{m_k} q^2_{k\beta}V^k_{\beta\beta}\right), \tag{11a}$$

$$\dot{\mathbf{u}}_{k\alpha} = -\sum_{\beta\neq\alpha} \frac{q_{k\alpha}q_{k\beta}}{q_{k\beta} - q_{k\alpha}} V^k_{\beta\alpha}\,\mathbf{u}_{k\beta}. \tag{11b}$$

Sketch of proof: By writing $q_{k\alpha}\delta_{\alpha\beta} = \mathbf{u}^\dagger_{k\alpha}\mathbf{Q}\mathbf{u}_{k\beta}$ and differentiating, we obtain $\dot{q}_{k\alpha}\delta_{\alpha\beta} = (q_{k\alpha} - q_{k\beta})\mathbf{u}^\dagger_{k\alpha}\cdot\dot{\mathbf{u}}_{k\beta} + q_{k\alpha}q_{k\beta}V^k_{\alpha\beta} - \operatorname{tr}(\mathbf{Q}^2_k\mathbf{V}_k)/\operatorname{tr}(\mathbf{Q}^2_k)q^2_{k\alpha}\delta_{\alpha\beta}$, so our claim follows by rearranging. ■

Remark. To complete the proof of Proposition 1, note that (11a) implies that if an eigenvalue of $\mathbf{Q}_k$ becomes zero, then it stops evolving; thus, no eigenvalue of $\mathbf{Q}_k$ can become negative and $\mathbf{Q}_k(t)$ stays in $\mathcal{X}_k$ for all time.

The previous discussion is obviously specific to the Riemannian metric $[\![\cdot,\cdot]\!]$, so a natural question that emerges is whether we could consider a spectral metric other than (6). Armed with a fair amount of hindsight, we will thus also consider the following spectral variant of the *Shahshahani metric*:

$$\langle\mathbf{A},\mathbf{B}\rangle_{\mathbf{Q}} = \operatorname{tr}(\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{Q}^{-1/2}\mathbf{B}). \tag{12}$$

This metric is so named because it extends the well-known *Shahshahani metric* [9] which has important applications in evolutionary biology and game theory [10]. Indeed, in the special case where $\mathbf{A},\mathbf{B}$ and $\mathbf{Q}$ can be simultaneously diagonalized giving vectors $\mathbf{x},\mathbf{y}$ and $\mathbf{q} \in \mathbb{R}^m$, (12) gives $\langle\mathbf{x},\mathbf{y}\rangle = \sum_\beta x_\beta y_\beta/q_\beta$, viz. the Shahshahani metric on the positive orthant $\mathbb{R}^m_{++}$. However, owing to the $1/2$ exponents in (12), it is not immediately obvious that $\langle\cdot,\cdot\rangle$ is indeed a metric; this is the subject of the following lemma:

Lemma 1. *The spectral product (12) is a smooth Riemannian metric on* $\operatorname{int}(\mathcal{C})$*; in particular, for all positive-definite* $\mathbf{Q} \in \operatorname{int}(\mathcal{C})$*, and for all* $\mathbf{A},\mathbf{B} \in T_{\mathbf{Q}}\mathcal{C}$*, the pairing (12) satisfies the following properties:*

i. *Bilinearity:* $\langle\mathbf{A}, a\mathbf{B}_1 + b\mathbf{B}_2\rangle_{\mathbf{Q}} = a\langle\mathbf{A},\mathbf{B}_1\rangle_{\mathbf{Q}} + b\langle\mathbf{A},\mathbf{B}_2\rangle_{\mathbf{Q}}$ *for all* $a, b \in \mathbb{R}$ *(and similarly for the first argument).*
ii. *Symmetry:* $\langle\mathbf{A},\mathbf{B}\rangle_{\mathbf{Q}} = \langle\mathbf{B},\mathbf{A}\rangle_{\mathbf{Q}}$.
iii. *Positivity:* $\langle\mathbf{A},\mathbf{A}\rangle_{\mathbf{Q}} \geq 0$ *with equality if and only if* $\mathbf{A} = 0$.
iv. *Smoothness: the assignment* $\mathbf{Q} \mapsto \langle\cdot,\cdot\rangle_{\mathbf{Q}}$ *is smooth.*

Proof: Bilinearity and smoothness follow trivially from the definition of $\langle\cdot,\cdot\rangle$; as for symmetry, we have:

$$\begin{aligned}\langle\mathbf{A},\mathbf{B}\rangle_{\mathbf{Q}} &= \operatorname{tr}(\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{Q}^{-1/2}\mathbf{B})\\ &= \operatorname{tr}(\mathbf{B}\mathbf{Q}^{-1/2}\mathbf{A}\mathbf{Q}^{-1/2}) = \operatorname{tr}(\mathbf{Q}^{-1/2}\mathbf{B}\mathbf{Q}^{-1/2}\mathbf{A}) = \langle\mathbf{B},\mathbf{A}\rangle_{\mathbf{Q}}.\end{aligned} \tag{13}$$

Finally, as far as positive-definiteness is concerned, let $\mathbf{S} = \mathbf{Q}^{-1/4}\mathbf{A}\mathbf{Q}^{-1/4}$ and note that $\langle\mathbf{A},\mathbf{A}\rangle = \operatorname{tr}(\mathbf{S}\mathbf{S}^\dagger) \geq 0$, with equality if and only if $\mathbf{S} = 0$, i.e. if and only if $\mathbf{A} = 0$. ■

Then, as before, the gradient of Φ with respect to $\langle\cdot,\cdot\rangle$ will be

$$\mathbf{grad}(\Phi) = \mathbf{Q}^{1/2}\mathbf{V}\mathbf{Q}^{1/2}. \tag{14}$$

Hence, to obtain the gradient ascent dynamics for Φ on $\mathcal{X}$, we only need to project (14) on the problem's trace constraints:

Proposition 3. *The gradient ascent dynamics for* Φ *w.r.t. the spectral Shahshahani metric (12) on* $\mathcal{X}$ *are:*

$$\dot{\mathbf{Q}}_k = \mathbf{Q}^{1/2}_k\mathbf{V}_k\mathbf{Q}^{1/2}_k - P^{-1}_k\operatorname{tr}(\mathbf{Q}_k\mathbf{V}_k). \tag{15}$$

The state space $\mathcal{X}$ *of (SRP) remains invariant under the flow (15); in fact, with notation as in Proposition 2, we have:*

$$\dot{q}_{k\alpha} = q_{k\alpha}\left(V^k_{\alpha\alpha} - P^{-1}_k \sum_{\beta=1}^{m_k} q_{k\beta}V^k_{\beta\beta}\right), \tag{16a}$$

$$\dot{\mathbf{u}}_{k\alpha} = -\sum_{\beta\neq\alpha} \frac{\sqrt{q_{k\alpha}q_{k\beta}}}{q_{k\beta} - q_{k\alpha}} V^k_{\beta\alpha}\,\mathbf{u}_{k\beta}. \tag{16b}$$

Proof: Similar to the proofs of Propositions 1 and 2. ■

Remark 1. Importantly, the eigenvalue evolution equation (16) is equivalent to the well-known *replicator dynamics* of evolutionary game theory [10] which were used in [11] to solve (SRP) in the special case where the $\mathbf{Q}_k$ are all diagonal (the so-called parallel MAC case). In this way, (12) can be seen as a direct extension of the replicator dynamics to a matrix setting.

Remark 2. It is also important to note that the eigenvalue evolution equation (16) also appears in the matrix exponential learning approach of [12] which takes the coupled form:

$$\dot{\mathbf{Y}}_k = \mathbf{V}_k \tag{17a}$$

$$\mathbf{Q}_k = P_k\frac{\exp(\mathbf{Y}_k)}{\operatorname{tr}(\exp(\mathbf{Y}_k))}. \tag{17b}$$

However, even though the eigenvalues of $\mathbf{Q}(t)$ follow the same evolution law under (17) and (15), its eigenvectors follow a completely different law, so (17) and (15) evolve differently over time.

IV. Convergence Analysis

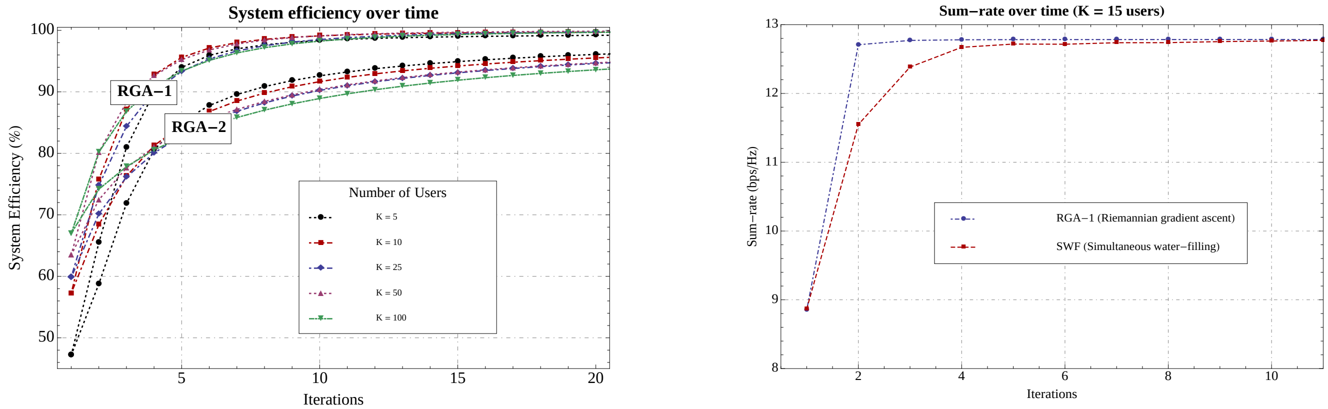
In view of the discussion of the previous section, we are now ready to state our main result:

Theorem 1. *The system's sum rate function* Φ *is weakly incrasing along the matrix dynamics (10)/(15); moreover, every interior trajectory of (10)/(15) converges to a solution of (SRP).*

Sketch of proof: A simple differentiation of Φ along (10) and (15) yields $\frac{d\Phi}{dt} = \sum_k \operatorname{tr}\left(\frac{\partial\Phi}{\partial\mathbf{Q}^*_k}\dot{\mathbf{Q}}_k\right) = \operatorname{tr}(\mathbf{Q}^{-r/2}\,\mathbf{grad}\,\Phi\cdot\mathbf{Q}^{-r/2}\,\mathbf{grad}\,\Phi) \geq 0$, (where we are taking $r = 2$ for (10) and $r = 1$ for (15)). This expression shows that Φ is weakly increasing along (15) and (10); furthermore, it can be shown that the concave objective Φ is also a Morse-Bott function (i.e. the kernel of its Hessian coincides with the tangent space of its solution set), so convergence to a solution of (SRP) follows from Proposition 3.9 in [13]. ■

Theorem 1 does not differentiate between the Riemannian metrics (6) and (12): gradient flow dynamics taken with respect to either metric converge to a solution of (SRP) from any interior initial power profile $\mathbf{Q}_k > 0$, $\operatorname{tr}(\mathbf{Q}_k) = P_k$. Nonetheless, the rate at which the gradient flow converges depends strongly on the metric being utilized.

For lack of space, we will only study here the behavior of these two dynamics in a simple linearized single-user diagonal system with one channel consistently better than the other,



(a) Maximization of transmission sum rate under RGA-1 and RGA-2.

(b) Riemannian gradient ascent and simultaneous water-filling.

Fig. 1. Convergence of the RGA algorithm and comparison to water-filling methods.

modeled by the differential matrix $\mathbf{V} = \operatorname{diag}(c, 0)$.[3] In this way, letting $\mathbf{Q} = \operatorname{diag}(x, 1-x)$ denote the corresponding power allocation matrix, the dynamics (10) and (15) respectively give

$$\dot{x} = cx(1-x), \tag{18a}$$

$$\dot{x} = c\frac{x^2(1-x)^2}{x^2 + (1-x)^2}. \tag{18b}$$

Both these equations can then be solved explicitly, yielding:

$$x = \left(1 + K_1 e^{-ct}\right)^{-1} \approx 1 - e^{-ct} \tag{19a}$$

$$x = \tfrac{1}{2} - \frac{1}{K_2 + ct} + \sqrt{\tfrac{1}{4} + \tfrac{1}{(K_2+ct)^2}} \approx 1 - \frac{1}{ct}, \tag{19b}$$

where K_1 and K_2 are constants depending only on initial conditions and $\approx$ denotes asymptotic equality as $t \to \infty$. Thus, even though both dynamics converge to 1 (the end-point of the gradient flow), they do so at vastly different rates: the replicator-like dynamics (15) converge at an exponential rate and reach an ε-neighborhood of the optimum point in time which is $\mathcal{O}(\log(1/\varepsilon))$, whereas the quadratic dynamics (10) only reach an ε-neighborhood in time $\mathcal{O}(1/\varepsilon)$.

In view of the above, the semidefinite optimization algorithm that we will consider is:

Algorithm 1 Riemannian gradient ascent (RGA)

Require: Metric exponent $r \in \{1, 2\}$, initial transmit directions $\mathbf{u}_{k\alpha} \in \mathbb{C}^{m_k}$ and eigenvalues $q_{k\alpha} > 0$, $\sum_{\alpha=1}^{m_k} q_{k\alpha} = P_k$.
  $t \leftarrow 0$;
  repeat
    $t \leftarrow t + 1$;
    for all $k \in \mathcal{K}$ **do**
      $\{\mathbf{u}_{k\alpha}\} \leftarrow \left\{\mathbf{u}_{k\alpha} - \delta(t) \sum_{\beta \neq \alpha} \frac{q_{k\alpha}^{r/2} q_{k\beta}^{r/2}}{q_{k\beta} - q_{k\alpha}} V_{\beta\alpha}^k \, \mathbf{u}_{k\beta}\right\}$;
      $\{q_{k\alpha}\} \leftarrow \left\{q_{k\alpha} + \delta(t) q_{k\alpha}^r \left(V_{\alpha\alpha}^k - \left(\sum_\gamma q_{k\gamma}^r\right)^{-1} \sum_{\beta=1}^{m_k} q_{k\beta}^r V_{\beta\beta}^k\right)\right\}$;
    end for
  until required accuracy is reached or transmission ends.

[3] By linearizing (10) and (15) near a stationary point, one can show that this simple example basically captures the general case as well.

Remark 1. Depending on the exponent $r = 1, 2$ of the Riemannian metric (6) or (12), we will refer to the above algorithm as RGA-1 or RGA-2 respectively. Obviously, RGA-1 refers to the Shahshahani gradient flow (15) while RGA-2 corresponds to the quadratic metric (6) and the associated dynamcis (10).

Remark 2. As in distributed water-filling algorithms [3–5], the only information requirements of the RGA algorithm is the channel matrix $\mathbf{H}_k$ at the transmitter end and the aggregate signal-plus-noise matrix $\mathbf{W} = \mathbf{I} + \sum_\ell \mathbf{H}_\ell \mathbf{Q}_\ell \mathbf{Q}_\ell^\dagger$ at the receiver end (which can then be broadcast to users, e.g. over a dedicated frequency band or by appending it to acknowledgment packets). That said, it is important to note that the update steps of RGA no longer contain a linear search for the water level, so its computational overhead is considerably lighter than that of water-filling algorithms.

Obviously, the RGA scheme above is just an Euler discretization of the gradient dynamics (10)/(15) with time-dependent steps $\delta(t)$ which will be assumed to satisfy the "$\ell^2 - \ell^1$" summability conditions of deterministic approximation, i.e. that $\sum_t \delta(t) = \infty$ and $\sum_t \delta(t)^2 < \infty$ (the gold standard being $\delta(t) = 1/t$). Under this assumption, we have:

Proposition 4. *For any step sequence $\delta(t)$ satisfying $\sum_t \delta(t) = \infty$ and $\sum_t \delta(t)^2 < \infty$, the algorithm RGA converges to a solution of the sum rate maximization problem (SRP).*

The proof is a standard application of deterministic approximation techniques to Theorem 1 (see e.g. [14]), so we omit it. Instead, it is more important to note here that a vanishing step size increases the convergence time of the algorithm in terms of actual iterations needed to converge to a given accuracy level. In practice however, the step size $\delta(t)$ can be taken constant without negatively impacting the quality of the algorithm's convergence, all the while increasing its convergence speed – cf. Section V below.

V. Numerical Simulations

To assess the convergence speed of the RGA scheme, we simulated in Fig. 1(a) a MIMO MAC system consisting of $K = 5, 10, 25, 50$ and 100 users, with a random number of transmit antennas (uniformly drawn between 2 and 10), a receiver with 5 antennas, and randomly drawn (but static once picked) channel matrices $\mathbf{H}$. We then ran RGA-1 and RGA-2 with a constant step from the same (random) initial conditions and we plotted over time the normalized efficiency ratio:

$$\mathrm{eff}(t) = \frac{\Phi(t) - \Phi_{\min}}{\Phi_{\max} - \Phi_{\min}}, \tag{20}$$

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.

We observe that RGA-1 achieves capacity within only a few iterations (even for large numbers of users and/or antennas per transmitter), and even though RGA-2 eventually achieves capacity as well, it does so at a much slower rate. From these simulations, we also see that a constant step size does not jeopardize the algorithm's convergence. In fact, by taking larger step sizes, the algorithm's convergence is sped up considerably: as we show in Fig. 1(b) where we compare the performance of our algorithm to simultaneous water-filling, RGA-1 with a larger step size achieves 99% of the channel's sum capacity in a *single* iteration.

Finally, to account for more realistic channel conditions, we also plotted the performance of RGA-1 in the presence of Rayleigh fading, following the well-known Jakes model [15]. Specifically, in Fig. 2, we consider a MIMO MAC system with 3 receive antennas, 3 users with 2 antennas each, transmitting at $\nu = 2\,\mathrm{GHz}$ and with average velocities of $v = 5\,\mathrm{km/h}$. We then ran RGA-1 with an update period of $\delta = 3\,\mathrm{ms}$, 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 $\mathbf{H}(t)$, and versus the "uniform" sum rate that users would have attained by spreading their power uniformly among their antennas. Despite the channels' fluctuations over time, RGA-1 tracks the channel's capacity remarkably well; moreover, the sum rate difference between the learned transmit spectrum and the uniform one can see that this tracking is not an artifact of all achievable sum rates falling within a very narrow band of the maximum one. In fact, by performing a correlation analysis of the maximum sum rate curve and the achieved one (inlay), we see that RGA-1 tracks the channel's sum capacity with a delay of about 6 ms, meaning that users achieve capacity within roughly 5% of the system's coherence time (108 ms for a fading velocity of $v = 5\,\mathrm{km/h}$).

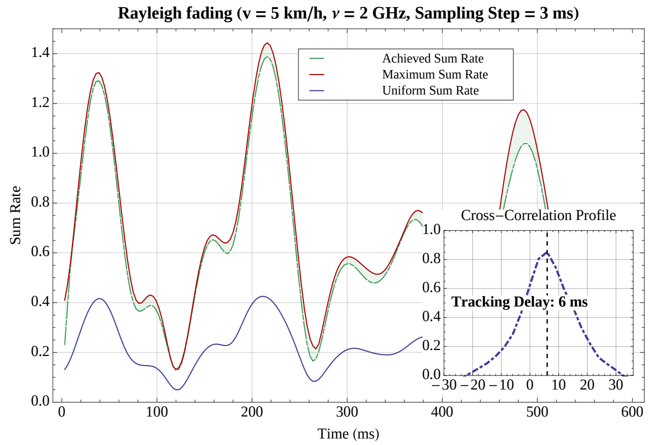


Fig. 2. Using RGA-1 to track the channel's sum capacity in the presence of Rayleigh fading conditions.

VI. Conclusions and Future Directions

In this paper, drawing tools and ideas from Riemannian geometry, we introduced two dynamic distributed algorithms for solving concave semidefinite maximization problems in the context of multi-user MIMO networks. The use of geometric concepts naturally enforces the positivity constraints of the users' signal covariance matrices and ensures that as these matrices are dynamically updated, the network's users will achieve the system's sum capacity. These methods can be easily generalized to MIMO optimization problems with other constraints (such as spectral mask constraints), and their rapid convergence ensures their robustness even in the presence of fading or measurement errors. Finally, we see that the choice of metric plays an important role in the algorithm's convergence speed, so optimizing the choice of geometry is a project of significant practical impact.

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] J. M. Lee, *Riemannian manifolds: an introduction to curvature*, ser. Graduate Texts in Mathematics. Springer, 1997, vol. 176.

[8] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

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

[11] 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.

[12] 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.

[13] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*. Springer-Verlag, 1996.

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

[15] 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.