Derivative-Free Optimization over Multi-User MIMO Networks[⋆]

Olivier Bilenne[1][0000−0003−1376−8281], Panayotis Mertikopoulos[1], and E. Veronica Belmega[2][0000−0003−4336−4704]

[1] Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, 38000 Grenoble, France
`olivier.bilenne@inria.fr` `panayotis.mertikopoulos@imag.fr`
[2] ETIS UMR 8051, CY University, ENSEA, CNRS, F-95000, Cergy, France
`belmega@ensea.fr`

Abstract. In wireless communication, the full potential of multiple-input multiple-output (MIMO) arrays can only be realized through optimization of their transmission parameters. Distributed solutions dedicated to that end include iterative optimization algorithms involving the computation of the gradient of a given objective function, and its dissemination among the network users. In the context of large-scale MIMO, however, computing and conveying large arrays of function derivatives across a network has a prohibitive cost to communication standards. In this paper we show that multi-user MIMO networks can be optimized without using any derivative information. With focus on the throughput maximization problem in a MIMO multiple access channel, we propose a "derivative-free" optimization methodology relying on very little feedback information: a single function query at each iteration. Our approach integrates two complementary ingredients: exponential learning (a derivative-based expression of the mirror descent algorithm with entropic regularization), and a single-function-query gradient estimation technique derived from a classic approach to derivative-free optimization.

Keywords: derivative-free optimization · zeroth-order optimization · exponential learning · MIMO systems · throughput maximization· SPSA.

1 Introduction

The appeal of multiple-input and multiple-output (MIMO) technologies in wireless communication is their ability to increase throughputs significantly and to improve the systems' robustness to ambient noise and channel fluctuations [1, 2]. On this account, large-scale deployment of *multiple-input and multiple-output* (MIMO) terminals is perceived as one of the key enabling technologies for next-generation wireless networks.

Releasing the full potential of large MIMO arrays requires, however, a principled approach to optimization, with the aim of minimizing computational overhead and related expenditures.

An essential aspect of the emblematic *throughput maximization* problem resides in the optimization of MIMO transmission parameters (such as the users' signal co-variance matrices) [3–7]. In multi-user networks, conventional optimization methods involve the use of water-filling (WF) techniques [4, 8, 9], which invariably rely on the availability of perfect channel state information at the transmitter (CSIT), and are vulnerable to observation noise, asynchronicities, and other operational impediments that arise in real-world networks.

More recently proposed in [10] as an alternative to water-filling, the *matrix exponential learning* (MXL) algorithm proceeds incrementally by combining (stochastic) gradient steps with a matrix exponential mapping that ensures feasibility of the users' signal covariance variables. In so doing, MXL guarantees fast convergence in cases where WF methods demonstrably fail. On the negative side, an important implementation bottleneck of MXL is the requirement to (*i*) invert a relatively large matrix at the receiver; and (*ii*) broadcast the resulting matrix to all connected users[3]. In consequence, the computation and communication overhead of MXL quickly becomes prohibitive in larger MIMO systems.

In this paper, we focus on the problem (stated in Section 2) of throughput maximization in a MIMO multiple access channel (MAC), with the objective to overcome the above limitations of the MXL by means of zeroth-order optimization, i.e., by making no gradient computations whatsoever. Following a classic approach from the *simultaneous perturbation stochastic approximation* (SPSA) framework [12, 13], we devise in Section 3 a "gradient-free" optimization algorithm by plugging into the chassis of the original MXL method a gradient estimator based no longer on first-order feedback but on function queries (a single one at each iteration). Our developments are followed by a discussion on the performances and potential of gradient-free matrix exponential learning (Section 4).

Notation We use bold capital letters for matrices, saving the letters k, l for user assignments and t, s for time indices, so that e.g., matrix $\mathbf{Q}_k$ relates to user k, $\mathbf{Q}_t$ to time t, and $\mathbf{Q}_{k,t}$ to user k at time t.

2   Problem statement

Consider a MIMO network where K users are transmitting simultaneously to a wireless receiver equipped with N antennas over a shared Gaussian vector MAC, modeled by

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

where $\mathbf{y} \in \mathbb{C}^N$ is the signal at the reception, M_k, $\mathbf{x}_k \in \mathbb{C}^{M_k}$ and $\mathbf{H}_k \in \mathbb{C}^{N \times M_k}$ respectively denote the number of antennas, the transmitted messages and the channel matrix of user k ($k = 1, \ldots, K$), and $\mathbf{z} \in \mathbb{C}$ models additive zero-mean Gaussian noise with unit

---

[3] In a MIMO array with $N = 128$ receive antennas, this would correspond to transmitting approximately 65 kB of data per frame, thus exceeding typical frame size limitations by a factor of 50× to 500× depending on the specific standard [11].

covariance. Without loss of generality, we assume every user to possess at least two antennas ($M_k \geq 2$). Let P_k be the maximum mean power consumption of user k due to transmissions, and let

$$\mathbf{Q}_k = \frac{1}{P_k} \mathbb{E}[\mathbf{x}_k \mathbf{x}_k^\dagger]$$

denote the normalized covariance matrix of $\mathbf{x}_k$. By definition, the matrix $\mathbf{Q}_k$ is Hermitian—we write $\mathbf{Q}_k \in \mathrm{Herm}\,(M_k)$—and positive semidefinite.

Our goal is to maximize, under the maximum available transmit power constraint $\mathrm{tr}(\mathbf{Q}_k) \leq 1$ for $k = 1, \ldots, K$, the achievable sum rate under successive interference cancellation (SIC),

$$R(\mathbf{Q}) = \log \det\left(\mathbf{I} + \sum_{k=1}^{K} P_k\, \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^\dagger \right) \tag{1}$$

where the aggregate form $\mathbf{Q} = (\mathbf{Q}_k, \ldots, \mathbf{Q}_K)$ contains all the unknowns of the problem. Since the maximum sum rate is achieved at a boundary point $\mathbf{Q}$ where $\mathrm{tr}(\mathbf{Q}_1) = \cdots = \mathrm{tr}(\mathbf{Q}_K) = 1$, the search domain of the problem is confined to the Cartesian product set $\mathcal{Q} = \mathcal{Q}_1 \times \cdots \times \mathcal{Q}_K$, where

$$\mathcal{Q}_k = \{\mathbf{Q}_k \in \mathrm{Herm}\,(M_k) : \mathrm{tr}(\mathbf{Q}_k) = 1, \mathbf{Q}_k \geq 0\}$$

is a compact subset of a d_k-dimensional real subspace, with $d_k = M_k^2 - 1 > 0$ for every user k.

The throughput maximization problem can be stated as the convex program:

$$\begin{aligned} &\text{maximize } R(\mathbf{Q}) \\ &\text{subject to } \mathbf{Q} \in \mathcal{Q}. \end{aligned} \tag{RM}$$

The structure of the feasible set $\mathcal{Q}$ makes the problem amenable to parallel optimization settings where (RM) is regarded as a collection of K sub-problems

$$\begin{aligned} &\text{maximize } R(\mathbf{Q}_k; \mathbf{Q}_{-k}) \\ &\text{subject to } \mathbf{Q}_k \in \mathcal{Q}_k \end{aligned} \tag{RM$_k$}$$

to be solved in parallel by the users. Equivalently, (RM$_k$) can be interpreted as maximizing the achievable transmission rate of user k when single-user decoding (SUD) is performed at the receiver,

$$R_k(\mathbf{Q}_k; \mathbf{Q}_{-k}) := R(\mathbf{Q}) - R(\mathbf{Q}_1, \ldots, \mathbf{Q}_{k-1}, 0, \mathbf{Q}_{k+1}, \ldots, \mathbf{Q}_K), \tag{2}$$

given the covariance matrices of the remaining users, thus regarding the interference due to the signals sent by other users as colored noise. Since the achievable sum rate (1) is a concave *potential function* for the game defined by (2), the solutions of (RM) are the solutions of the Nash equilibrium problem defined by (2), i.e. any solution $\mathbf{Q}^\star$ of (RM) satisfies, for $k = 1, \ldots, K$,

$$R_k(\mathbf{Q}_k^\star; \mathbf{Q}_{-k}^\star) \geq R_k(\mathbf{Q}_k; \mathbf{Q}_{-k}^\star) \quad \forall \mathbf{Q}_k \in \mathcal{Q}_k, \tag{NE}$$

and conversely. In other words, maximizing the achievable sum rate under SIC is equivalent to equilibrating the individual transmission rates (2) under SUD.

Many optimization methods rely on derivative information. Differentiation of the achievable sum rate (1) gives us the gradient $\nabla R = (\nabla_1 R, \ldots, \nabla_K R)$ where, for $k = 1, \ldots, K$,

$$\nabla_k R(\mathbf{Q}) = P_k \mathbf{H}_k{}^\dagger \left[\mathbf{I} + \sum_{l=1}^{K} P_l \, \mathbf{H}_l \mathbf{Q}_l \mathbf{H}_l^\dagger \right]^{-1} \mathbf{H}_k. \tag{3}$$

Making the derivatives $\nabla_k R(\mathbf{Q})$ available to the users implies the inversion of the $N \times N$ Hermitian matrix $\mathbf{I} + \sum_{l=1}^{K} P_l \mathbf{H}_l \mathbf{Q}_l \mathbf{H}_l^\dagger$ at the receiver, followed by the broadcast of the result towards the users, which then are able to compute (3) locally. On account that the communication overhead induced by the dissemination of the gradient may be prohibitive, we proceed under the assumption that the gradient is not accessible to the users, which instead are required to compute their own estimates of $\nabla_k R$, based no longer on derivative information but on mere measurements of $R(\mathbf{Q})$.

3   Derivative-free matrix exponential learning

3.1   The MXL algorithm

Among the existing (derivative-based) methods of solution for (RM) is the *matrix exponential learning* (MXL) [14], which in our developments will serve both as reference and as a starting point. We refer to [15, 16] for a characterization of the MXL algorithm as an instance of the mirror descent algorithm implemented with the *von Neumann relative entropy* for Bregman divergence. Given an initial point $\mathbf{Y}_0 = \mathbf{Y}_1 = (0, \ldots, 0)$ in the space of the gradients $\mathcal{Q}^* = \mathcal{Q}_1^* \times \cdots \times \mathcal{Q}_K^*$, where $\mathcal{Q}_k^* = \{\mathbf{Y}_k \in \mathrm{Herm}\,(M_k) : \mathrm{tr}(\mathbf{Y}_k) = 0\}$, the t-th step of the algorithm is defined for $t \geq 1$ by

$$\begin{aligned} \mathbf{Q}_t &= \mathbf{\Lambda}(\mathbf{Y}_t), \\ \mathbf{Y}_{t+1} &= \mathbf{Y}_t + \gamma_t \hat{\mathbf{V}}_t, \end{aligned} \tag{MXL}$$

where $\{\mathbf{Q}_t\}$ denotes the issued sequence of estimates for the optimal configuration, $\{\mathbf{Y}_t\}$ is a sequence generated in the space of the gradients, $\{\gamma_t\}$ is a sequence of positive step-sizes, $\hat{\mathbf{V}}_t = (\hat{\mathbf{V}}_{1,t}, \ldots, \hat{\mathbf{V}}_{K,t}) \in \mathcal{Q}^*$ is an estimate of the gradient $\nabla R(\mathbf{Q}_t)$, and we set $\mathbf{\Lambda} = (\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_K)$, where the exponential learning mapping $\mathbf{\Lambda}_k$ is defined by

$$\mathbf{\Lambda}_k(\mathbf{Y}_k) = \frac{\exp(\mathbf{Y}_k)}{\mathrm{tr}(\exp(\mathbf{Y}_k))},$$

in which exp denotes the (matrix) complex exponential function.

In contrast to the available implementations of MXL, which rely on full/noisy [14] or partial [17] gradient feedback for the computation of $\hat{\mathbf{V}}_t$, the gradient estimates $\hat{\mathbf{V}}_t$ in this work are derived without gradient information, as explained in Section 3.2.

3.2   Derivative-free MXL

Description of the the gradient-free MXL. Our developments build on an early approach to derivative-free optimization [12, 13] which, in time, has been seen as the cornerstone to the field of *simultaneous perturbation stochastic approximation* (SPSA).

After translation into our distributed, Hermitian setting, the SPSA approach can be described as follows.

In the absence of any gradient feedback, each user k infers an estimate $\hat{\mathbf{V}}_k \approx \nabla_k R$ of their individual gradient, derived from randomized queries of the sum rate R in the close neighborhood of the current iterate. For $k = 1, \ldots, K$, let $r_k > 0$ and $\mathbf{C}_k \in \mathcal{Q}_k$ such that the ball $\mathbf{C}_k + r_k \mathbf{B}_{d_k}$ is entirely contained by $\mathcal{Q}_k$. Concretely, each user k draws, randomly and independently, a matrix $\mathbf{Z}_k$ on the sphere $\mathbf{S}_{d_k-1} = \{\mathbf{Z}_k \in \mathcal{Q}_k^* : \|\mathbf{Z}_k\|_2 = 1\}$ living in the d_k-dimensional space $\mathcal{Q}_k^*$, and we let $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_K)$ aggregate the random matrices of all users. The gradient estimator for user $k = 1, \ldots, K$ is then defined as

$$\hat{\mathbf{V}}_k(\mathbf{Q}) = \frac{d_k}{\delta} R(\hat{\mathbf{Q}}) \, \mathbf{Z}_k, \tag{SPSA}$$

where $\hat{\mathbf{Q}} = (\hat{\mathbf{Q}}_1, \ldots, \hat{\mathbf{Q}}_K)$, and each test matrix

$$\hat{\mathbf{Q}}_k = \mathbf{Q}_k + \tfrac{\delta}{r_k}(\mathbf{C}_k - \mathbf{Q}_k) + \delta \mathbf{Z}_k \tag{4}$$

is derived from $\mathbf{Q}_k$ after deviation by random quantity $\delta \mathbf{Z}_k$, and prior shrinking of $\mathcal{Q}_k$ so as to keep the test configuration $\hat{\mathbf{Q}}$ inside the feasible set. The presence in (SPSA) of the factor $d_k = M_k^2 - 1$ can be explained as the ratio between the volumes of the sphere $\mathbf{S}_{d_k-1}$ (where $\mathbf{Z}_k$ is picked) and the containing ball $\mathbf{B}_{d_k} = \{\mathbf{Z}_k \in \mathcal{Q}_k^* : \|\mathbf{Z}_k\|_2 \leq 1\}$.

The distinguishing property of (SPSA) lies in that the the *bias* of the gradient estimator can be controlled by the parameter δ as this bias is uniformly bounded over $\mathcal{Q}$:

$$\|\mathbb{E}[\hat{\mathbf{V}}_k(\mathbf{Q}, \mathbf{Z}; \rho) - \nabla_k R(\mathbf{Q})]\|_* = O(\delta) \tag{5}$$

Besides, the *norm* of the gradient estimator (SPSA) satisfies

$$\|\hat{\mathbf{V}}_k(\mathbf{Q}, \mathbf{Z}; \rho)\|_* = O\left(\frac{1}{\delta}\right) \tag{6}$$

uniformly on $\mathcal{Q}$. Equations (5) and (6) thus unveil a tradeoff between the $O(\delta)$ bias of the estimator and its $O(\frac{1}{\delta})$ deviation from the true derivative. This bias–variance tradeoff induces in the present context strict restrictions on the choice of the query radius δ and of the step-size policy of the MXL algorithm, with consequences on the performance of the algorithm, as discussed towards the end of the section.

See Algorithm 1 for a pseudocode description of the gradient-free optimization algorithm obtained after combining MXL with (SPSA). Given a (typically non-increasing) query radius sequence $\{\delta_t\}$ and a step-size sequence $\{\gamma_t\}$, the task of user k at time step t consists of (i) sampling a random direction $\mathbf{Z}_{k,t} \in \mathcal{Q}_k^*$, (ii) implementing the test covariance matrix $\hat{\mathbf{Q}}_{k,t}$ obtained as in (4) by variation of the current covariance estimate $\mathbf{Q}_{k,t}$, (iii) receiving the value of the achievable total transmission rate $R(\hat{\mathbf{Q}}_t)$, (iv) inferring an estimate $\hat{\mathbf{V}}_{k,t}$ of the gradient along user direction k, and (v) updating $\mathbf{Y}_{k,t}$ and $\mathbf{Q}_{k,t}$ in accordance with (MXL).

Convergence of the gradient-free MXL. The convergence of the gradient-free version of MXL is guaranteed with probability 1 on condition that the implementation

---

Algorithm 1: Gradient-free MXL

    Parameters : $\{\gamma_t\}_{t=1}^{\infty}, \{\delta_t\}_{t=1}^{\infty}$

    Init.: $t \leftarrow 1$, $\mathbf{Y} \leftarrow 0$, $\forall k$: **transmit with** $\mathbf{Q}_k \leftarrow \frac{P_k}{M_k}\mathbf{I}_k$

1:   **Repeat until stopping criterion is reached**

2:       **For** $k \in \{1, \ldots, K\}$ **do in parallel**

          Sample $\mathbf{Z}_k$ **uniformly in** $\mathbf{S}_{d_k-1}$

          Transmit with $\hat{\mathbf{Q}}_k \leftarrow \mathbf{Q}_k + \frac{\delta_t}{r_k}(\mathbf{C}_k - \mathbf{Q}_k) + \delta_t\mathbf{Z}_k$

          Receive feedback $r \leftarrow R(\hat{\mathbf{Q}})$

          $\hat{\mathbf{V}}_k \leftarrow \left(\frac{M_k^2-1}{\delta_t}\right)r\,\mathbf{Z}_k$

          $\mathbf{Y}_k \leftarrow \mathbf{Y}_k + \gamma_t\hat{\mathbf{V}}_k$

          $\mathbf{Q}_k \leftarrow \mathbf{\Lambda}_k(\mathbf{Y}_k)$

3:       $t \leftarrow t + 1$

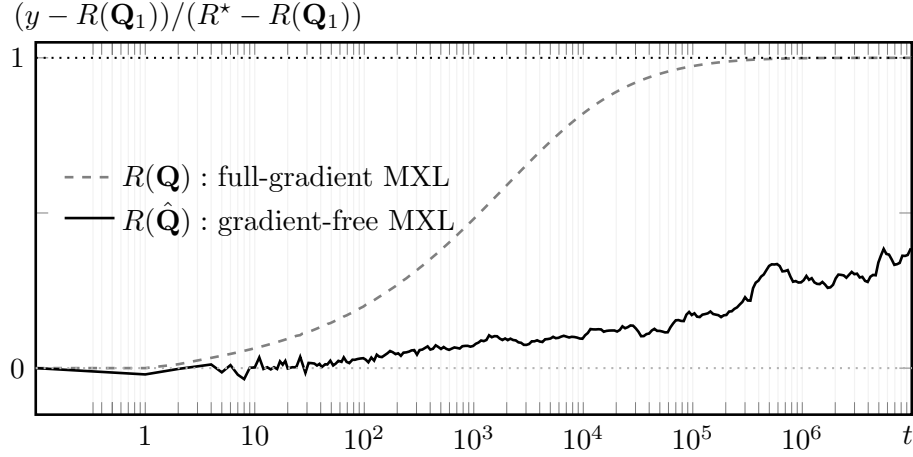$(y - R(\mathbf{Q}_1))/(R^{\star} - R(\mathbf{Q}_1))$

Fig. 1. Convergence of the gradient-free algorithm ($N = 16$, $K = 20$, $\mathbb{E}[M_k] = 3$): Algorithm 1 is run with policies $(\gamma_t, \delta_t) = (0.01\,t^{-3/4}, 0.1\,t^{-1/4})$, while MXL with full gradient feedback run with decreasing step size policy $\gamma_t = 0.01\,t^{-1/2}$.

parameters are chosen with care [16, 18]. Indeed, if Algorithm 1 is implemented with non-increasing step-size and query-radius policies satisfying the conditions

$$\text{(a) } \sum_{t=1}^{\infty}\gamma_t = \infty, \quad \text{(b) } \sum_{t=1}^{\infty}\frac{\gamma_t^2}{\delta_t^2} < \infty, \quad \text{(c) } \delta_t \downarrow 0, \quad \text{(d) } \delta_t < \min_k \frac{1}{\sqrt{M_k(M_k-1)}} \quad (\forall t),$$

(7)

then the sequences of estimates $\{\mathbf{Q}_t\}$ and of test configurations $\{\hat{\mathbf{Q}}_t\}$ converge almost surely towards the optimum $\mathbf{Q}^{\star}$.

Numerical simulations. Figure 1 reports experimental results for a network with 16 antennas at the receiver and 20 homogeneous users equipped with, on average, 3 an-

tennas. A comparison is made between the transmission rates iteratively realized by Algorithm 1 and those of the reference MXL algorithm with perfect gradient feedback. The gradient-free algorithm is run with decreasing step-size and query-radius sequences chosen in accordance with (7).

It can be seen on Figure 1 that the gradient-based algorithm finds optimal configuration within a handful of iterations. If the gradient-free algorithm also progresses towards the optimum, its convergence is less straightforward and much slower than with full gradient feedback. This tendency to slowness, which can be explained by the bias–variance tradeoff induced by the gradient estimator (SPSA), is only exacerbated in networks of larger sizes, where high problem dimensionality creates a bottleneck implying prohibitively slow convergence. In [16] it is shown that the convergence rate of Algorithm 1 is at best $O(1/\sqrt[4]{T})$ after T iterations, in contast to the considerably faster $O(1/\sqrt{T})$ rates that can be expected from the first-order methods.

4   Discussion and perspectives

Besides the very light nature of the feedback information it requires (a single query of the objective function per iteration), the distributed, zeroth-order (derivative-free) optimization methodology presented in this paper owes to the MXL algorithm the desirable feature that it is both easy to implement, and flexible in the sense that it can be run asynchronously for the users (cf. [16]). As seen in the previous section, its major drawback is slow convergence compared to gradient-based methods. The slowness issue is addressed in detail in our more recent work [16], where the formulation of the gradient estimator (SPSA) is revisited thoroughly in order to meet the $O(1/\sqrt{T})$ convergence rate of the fisrt-order methods. The interested reader is referred to the developments and discussions of [16] for an extensive analysis of the performances, possibilities, and guarantees of single-query zeroth-order optimization methods in the vein of Algorithm 1.

Bibliography

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, February 2014.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[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, "Optimal linear precoding strategies for wideband non-cooperative systems based on game theory – part I: Nash equilibria," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1230–1249, March 2008.

[6] ——, "Optimal linear precoding strategies for wideband non-cooperative systems based on game theory – part II: algorithms," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1250–1267, March 2008.

[7] E. V. Belmega, S. Lasaulce, M. Debbah, M. Jungers, and J. Dumont, "Power allocation games in wireless networks of multi-antenna terminals," *Telecommunication Systems*, vol. 47, no. 1-2, pp. 109–122, 2011.

[8] G. Scutari, D. P. Palomar, and S. Barbarossa, "Asynchronous iterative waterfilling for Gaussian frequency-selective interference channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 2868–2878, July 2008.

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

[10] P. Mertikopoulos and A. L. Moustakas, "Learning in an uncertain world: MIMO covariance matrix optimization with imperfect feedback," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 5–18, January 2016.

[11] R. Liao, B. Bellalta, M. Oliver, and Z. Niu, "MU-MIMO MAC protocols for wireless local area networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 162–183, January 2016.

[12] J. C. Spall, "A one-measurement form of simultaneous perturbation stochastic approximation," *Automatica*, vol. 33, no. 1, pp. 109–112, 1997.

[13] A. D. Flaxman, A. T. Kalai, and H. B. McMahan, "Online convex optimization in the bandit setting: gradient descent without a gradient," in *SODA '05: Proceedings of the 16th annual ACM-SIAM Symposium on Discrete Algorithms*, 2005, pp. 385–394.

[14] P. Mertikopoulos, E. V. Belmega, and A. L. Moustakas, "Matrix exponential learning: Distributed optimization in mimo systems," in *2012 IEEE International Symposium on Information Theory Proceedings*, July 2012, pp. 3028–3032.

[15] P. Mertikopoulos, E. V. Belmega, R. Negrel, and L. Sanguinetti, "Distributed stochastic optimization via matrix exponential learning," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2277–2290, May 2017.

[16] O. Bilenne, P. Mertikopoulos, and E. Belmega, "Efficient Gradient-Free Optimization for Distributed Multi-User MIMO Systems," Feb. 2020, in preparation.

[17] W. Li and M. Assaad, "Matrix exponential learning schemes with low informational exchange," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3140–3153, April 2019.

[18] M. Bravo, D. S. Leslie, and P. Mertikopoulos, "Bandit learning in concave N-person games," in *NIPS '18: Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018.